JMIR mHealth and uHealth

Impact Factor (2024): 5.4
Volume 10 (2022), Issue 8 ISSN 2291-5222 Editor in Chief: Lorraine Buis, PhD, MSI

Contents

Reviews

Accuracy and Precision of Consumer-Grade Wearable Activity Monitors for Assessing Time Spent in Sedentary Behavior in Children and Adolescents: Systematic Review (e37547)
Antonio Martinko, Josip Karuc, Petra Juri , Hrvoje Podnar, Maroje Sori 3

Wearables for Measuring the Physical Activity and Sedentary Behavior of Patients With Axial Spondyloarthritis: Systematic Review (e34734)
Julie Soulard, Thomas Carlin, Johannes Knitza, Nicolas Vuillerme. 22

Measures of Engagement With mHealth Interventions in Patients With Heart Failure: Scoping Review (e35657)
Ifeanyi Madujibeya, Terry Lennie, Adaeze Aroh, Misook Chung, Debra Moser. 37

mHealth Apps for Low Back Pain Self-management: Scoping Review (e39682)
Aki Rintala, Roy Rantalainen, Anu Kaksonen, Hannu Luomajoki, Kari Kauranen. 54

Original Papers

Quality of Life and Physical Activity in 629 Individuals With Sarcoidosis: Prospective, Cross-sectional Study Using Smartphones (Sarcoidosis App) (e38331)
Brian Chu, Daniel O'Connor, Marilyn Wan, Ian Barnett, Haochang Shou, Marc Judson, Misha Rosenbach. 66

Role of Social and App-Related Factors in Behavioral Engagement With mHealth for Improved Well-being Among Chronically Ill Patients: Scenario-Based Survey Study (e33772)
Freek Van Baelen, Melissa De Regge, Bart Larivière, Katrien Verleye, Sam Schelfout, Kristof Eeckloo. 78

The Use of Digital Health Tools for Health Promotion Among Women With and Without Chronic Diseases: Insights From the 2017-2020 Health Information National Trends Survey (e39520)
Kobi Ajayi, Elizabeth Wachira, Henry Onyeaka, Tyra Montour, Samson Olowolaju, Whitney Garney. 90

Health Care Workers' Need for Headspace: Findings From a Multisite Definitive Randomized Controlled Trial of an Unguided Digital Mindfulness-Based Self-help App to Reduce Healthcare Worker Stress (e31744)
Heather Taylor, Kate Cavanagh, Andy Field, Clara Strauss. 101

System Usability Scale Benchmarking for Digital Health Apps: Meta-analysis (e37290)
Maciej Hyzy, Raymond Bond, Maurice Mulvenna, Lu Bai, Alan Dix, Simon Leigh, Sophie Hunt. 124

Factors Associated With Using the COVID-19 Mobile Contact-Tracing App Among Individuals Diagnosed With SARS-CoV-2 in Amsterdam, the Netherlands: Observational Study (e31099)
Feiko Ritsema, Jizzo Bosdriesz, Tjalling Leenstra, Mariska Petrignani, Liza Coyer, Anja Schreijer, Yvonne van Duijnhoven, Janneke van de Wijgert, Maarten Schim van der Loeff, Amy Matser. 135

Predicting the Next-Day Perceived and Physiological Stress of Pregnant Women by Using Machine Learning and Explainability: Algorithm Development and Validation (e33850)
Ada Ng, Boyang Wei, Jayalakshmi Jain, Erin Ward, S Tandon, Judith Moskowitz, Sheila Krogh-Jespersen, Lauren Wakschlag, Nabil Alshurafa.
1 4 6

Mediators of Effects on Physical Activity and Sedentary Time in an Activity Tracker and Behavior Change Intervention for Adolescents: Secondary Analysis of a Cluster Randomized Controlled Trial (e35261)
Simone Verswijveren, Gavin Abbott, Samuel Lai, Jo Salmon, Anna Timperio, Helen Brown, Susie Macfarlane, Nicola Ridgers. 164

Corrigenda and Addenda

Correction: Technical Guidance for Clinicians Interested in Partnering With Engineers in Mobile Health Development and Evaluation (e41813)
Lochan Shah, William Yang, Ryan Demo, Matthias Lee, Daniel Weng, Rongzi Shan, Shannon Wongvibulsin, Erin Spaulding, Francoise Marvel, Seth Martin. 176

Review

Accuracy and Precision of Consumer-Grade Wearable Activity Monitors for Assessing Time Spent in Sedentary Behavior in Children and Adolescents: Systematic Review

Antonio Martinko[1], MSc; Josip Karuc[1,2], PhD; Petra Jurić[1], MSc; Hrvoje Podnar[1], PhD; Maroje Sorić[1,3], MD, PhD

[1]Faculty of Kinesiology, University of Zagreb, Zagreb, Croatia
[2]Proprio Centre, Physical Rehabilitation Centre, Zadar, Croatia
[3]Faculty of Sport, University of Ljubljana, Ljubljana, Slovenia

Corresponding Author:
Antonio Martinko, MSc
Faculty of Kinesiology
University of Zagreb
Horvaćanski zavoj 15
Zagreb, 10000
Croatia
Phone: 385 981302484
Email: antonio.martinko@kif.hr

Abstract

Background: A large number of wearable activity monitor models are released and used each year by consumers and researchers. As more studies are being carried out on children and adolescents in terms of sedentary behavior (SB) assessment, knowledge about accurate and precise monitoring devices becomes increasingly important.

Objective: The main aim of this systematic review was to investigate and communicate findings on the accuracy and precision of consumer-grade physical activity monitors in assessing the time spent in SB in children and adolescents.

Methods: Searches of PubMed (MEDLINE), Scopus, SPORTDiscus (full text), ProQuest, Open Access Theses and Dissertations, DART Europe E-theses Portal, and Networked Digital Library of Theses and Dissertations electronic databases were performed. All relevant studies that compared different types of consumer-grade monitors using a comparison method in the assessment of SB, published in European languages from 2015 onward were considered for inclusion. The risk of bias was estimated using Consensus-Based Standards for the Selection of Health Status Measurement Instruments. For enabling comparisons of accuracy measures within the studied outcome domain, measurement accuracy interpretation was based on group mean or percentage error values and 90% CI. Acceptable limits were predefined as –10% to +10% error in controlled and free-living settings. For determining the number of studies with group error percentages that fall within or outside one of the sides from previously defined acceptable limits, two 1-sided tests of equivalence were carried out, and the direction of measurement error was examined.

Results: A total of 8 studies complied with the predefined inclusion criteria, and 3 studies provided acceptable data for quantitative analyses. In terms of the presented accuracy comparisons, 14 were subsequently identified, with 6 of these comparisons being acceptable in terms of quantitative analysis. The results of the Cochran *Q* test indicated that the included studies did not share a common effect size (Q_5=82.86; *P*<.001). I^2, which represents the percentage of total variation across studies due to heterogeneity, amounted to 94%. The summary effect size based on the random effects model was not statistically significant (effect size=14.36, SE 12.04, 90% CI −5.45 to 34.17; *P*=.23). According to the equivalence test results, consumer-grade physical activity monitors did not generate equivalent estimates of SB in relation to the comparison methods. Majority of the studies (3/7, 43%) that reported the mean absolute percentage errors have reported values of <30%.

Conclusions: This is the first study that has attempted to synthesize available evidence on the accuracy and precision of consumer-grade physical activity monitors in measuring SB in children and adolescents. We found very few studies on the accuracy and almost no evidence on the precision of wearable activity monitors. The presented results highlight the large heterogeneity in this area of research.

Trial Registration: PROSPERO CRD42021251922; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=251922

(JMIR Mhealth Uhealth 2022;10(8):e37547) doi:10.2196/37547

KEYWORDS

accuracy; precision; sedentary behavior; children; adolescents; wearable activity monitor; eHealth; digital health; mobile health; mHealth; mobile phone

Introduction

Background

Wearable devices are part of a growing market and are trending in terms of monitoring physical activity (PA) and sleep. Widely attainable wearable activity monitors (WAMs) have a high demand, which is supported by projections of market size growth by the year 2028, with extrapolated values of US $138.7 billion being extrapolated [1]. In addition, the magnification of health problems related to sedentary lifestyles is expected to increase the demand for these types of products. In addition, the COVID-19 pandemic has added consciousness regarding an overall picture of fitness and health in the general public. Activity monitors function as a means of providing feedback to users, while also offering behavior change tools, tracking of progress, and data storage. Daily self-monitoring is a core component of WAMs, in addition to comparing results with those of other users, which could increase PA levels in the long term [2].

Consumer-based WAMs can be wrist-worn or attached to a piece of clothing on different parts of the body, such as the hip. Currently, WAMs generally use a triaxial accelerometer to capture bodily movement in 3 dimensions. The collected data are then analyzed by proprietary algorithms to estimate the daily number of steps, amount of expended energy, sleep quantity and quality, and time spent on activities of different intensities [3]. Although WAMs are directed toward and mostly used by consumers who are already conscious about their health and PA, these devices could also be used as measurement tools among researchers in the fields of health promotion and PA [2,4-6].

Research focused on technology (ie, accuracy and precision) is of great interest, whereas studies of WAMs in the context of treatment and in medical settings have also been increasing. A recent systematic review [7] that analyzed 463 studies demonstrated a significant growth rate in the annual number of publications that included WAMs between 2013 and 2017. Measurement accuracy is a vital consideration, as WAMs are frequently used as a tool in research and a way of advising health care decisions. Studies in this field of research rely on accurate and precise instruments with small errors to elucidate complex research questions in which measurement error limits statistical power [8]. Consumer-grade WAMs were deemed accurate when measuring heart rate and steps [3]. However, accuracy is susceptible to variation when different manufacturers and types of devices are considered, with lower accuracy being reported for sleep, distance covered, and time spent in different PA levels [5,6]. Regarding precision, it was reported that there is high precision among devices for steps, distance, expended energy, and sleep [5]. In contrast to the large number of already available and emerging WAMs, there is still limited evidence on the accuracy and precision of consumer-grade WAMs. Moreover, most of the limited evidence at present refers to the measurement of PA, whereas research on sedentary behavior (SB) is lacking. This also goes against the growing popularity of monitoring training, successive recovery, and components of individuals' anthropological status with this type of technology [9]. When discussing the interactions of children with WAMs, contrasting research findings have been found, where some studies suggest that WAMs could be used to increase PA levels, and others have reported that WAM use over prolonged periods declines over time among children and adolescents [7]. Studies on WAM feasibility in children have shown that design, feedback features, and comfort while wearing the device were the most important factors [10].

Previously registered trials and conducted studies have most commonly identified the number of steps taken as the outcome of interest, followed by time spent in activity, sleep, energy expenditure, and distance covered as some other outcomes [7,11]. As recommendations have been provided for the first time by the World Health Organization on the associations between SB and health outcomes [12], it seems that research on SB will gain greater interest in the future. SB relates to low-intensity activities (<1.5 metabolic equivalents of tasks) and includes several bodily positions, such as lying, sitting, and reclining. SB is accompanied by a set of adverse health outcomes, and this association is already apparent in childhood [13]. Therefore, accurate, precise, and low-cost methods for measuring SB in children are important SB. Consumer-based WAMs could be of assistance in terms of reducing the financial costs and time spent by professionals when providing support and guidance for behavior change in children and adolescents. Different issues may arise when measuring SB, such as the following: (1) WAM placement; (2) how nonwear time is defined, epochs, and cutoff points; (3) setting the criteria for SB bouts and breaks; and (4) a combination of posture and motion data [14]. Although several systematic reviews have attempted to synthesize evidence on the accuracy of WAMs in measuring PA [5,6,15,16], similar studies related to SB are not available.

Objectives

The main aim of this systematic review was to analyze the evidence available on the accuracy and precision of consumer-grade WAMs in assessing the time spent on any type of SB in children and adolescents.

Methods

Search Strategy

The search strategy followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [17]. The review protocol was registered with PROSPERO, an international prospective register of systematic reviews (CRD42021251922). Electronic databases PubMed

(MEDLINE), Scopus, and SPORTDiscus (full text) were searched to find all relevant studies; in addition, ProQuest, Open Access Theses and Dissertations, Dart Europe E-Theses Portal, and Networked Digital Library of Theses and Dissertations electronic databases were searched as alternative literature sources of possible gray literature [18]. For each electronic database, a modified search strategy concerning specific and controlled vocabulary was used, with a variation of the following terms: *(children, adolescent, teen, youth) AND (fitness tracker, physical fitness tracker, activity monitor, activity tracker, wearable device, wearable) AND (sedentary behavior, sedentary, sedentary lifestyle, physical inactivity, sedentary time, rest, sitting position, reclining)*. A filter covering studies published in European languages was applied, and a time frame was set for studies ranging from January 1, 2015, to the day of this systematic review's execution, April 15, 2021. The term European languages refers primarily to some of the most commonly spoken languages in Europe, namely English, German, French, Italian, and Spanish, as the authors can understand these languages; therefore, the search was not limited only to studies in English. In addition, reference lists of the included studies and secondary sources were examined to find additional studies that were acceptable for inclusion. The constructed search strategy is presented in Multimedia Appendix 1.

Study Selection and Eligibility Criteria

All papers retrieved from the electronic databases were gathered and organized into the Rayyan web application (Qatar Computing Research Institute, Hamad Bin Khalifa University) for systematic reviews [19]. Rayyan was used to screen for potential duplicates; a manual inspection was then performed to discover any additional duplicates. Titles and abstracts of the first 10% of the results were screened independently by 2 reviewers (AM and JK). Given that the interrater agreement was 100%, only one of the authors (AM) continued with the screening process for the remaining 90% of the results. In case of ambiguities, the authors resolved the situation through a discussion with a third reviewer (MS). After the initial screening, full texts of the selected studies were accessed and screened for eligibility by 2 independent reviewers (AM and JK). When needed, disagreements among reviewers were resolved through a discussion with the third reviewer (MS).

The search was limited to studies that involved participants aged <18 years, namely children and adolescents. Studies that included participants with a physical disability or any other condition precluding them from engaging in PA were excluded. Inclusion was possible for studies conducted in both controlled and free-living settings. Studies were considered for inclusion only when the accuracy of a consumer-grade PA monitor was examined in comparison with an appropriate research-grade device in relation to a specific study setting. Appropriate research-grade devices were predefined and included indirect calorimetry, direct observation, and accelerometers. The main outcome of the eligible studies was the duration of SB, which relates to activities that do not increase energy expenditure substantially above the resting level and includes activities performed in a sitting or lying down position. Studies that were not available in full text were excluded as were studies that reported energy expenditure as their only outcome. During the literature search stage, no restrictions were set in terms of study type, although only original scientific papers were considered for inclusion, and secondary sources were excluded after their reference sections were manually inspected. Studies comparing different types of consumer-grade monitors without including a research-grade comparison method were also excluded. Data from one of the studies [20] were sought from the study authors, but no response was received across several modes of communication. Accuracy metrics labeled as acceptable for quantitative analyses were as follows: mean absolute percentage error (MAPE), standardized regression coefficient, odds ratio, correlation statistics, average error, limits of agreement, area under the curve or % sensitivity, % specificity or % positive predictive value), % negative predictive value, and likelihood ratio.

Risk of Bias Assessment

All included studies were assessed for the risk of bias by AM. In case of uncertainty, a discussion with a second reviewer (MS) was required to reach a decision. For assessing the risk of bias, a Consensus-Based Standards for the Selection of Health Status Measurement Instruments (COSMIN) tool was used [21]. COSMIN is a checklist used to evaluate the methodological quality of included studies when conducting systematic reviews of measurement characteristics. Each aspect of methodological quality evaluation was appraised based on a proposed scoring system [22]; it could be either of excellent, good, fair, or poor quality. In line with previous studies [11], we used a modified checklist in which the assessment included 6 components relevant to our research aim. The design or methodology components focused on the following: (1) percentages of missing data, (2) missing data management, (3) adequate sample sizes, (4) acceptable criterion comparisons, (5) design or methodological flaws, and (6) reporting of acceptable accuracy metrics, the only analytical component.

Data Extraction and Coding

Study characteristics and outcomes were extracted by a single reviewer (PJ), whereas cross-checking of the table was performed by a second reviewer (AM) familiar with the details of the included studies. Potential conflicts were resolved through discussion with a senior reviewer (MS). The extracted data included the reference, study period, participants (number, age, sex, ethnicity, socioeconomic status, and inclusion and exclusion criteria), type of consumer-grade PA monitor, comparison method, context of SB (ie, setting and type of activity), cutoff points for SB, and reported accuracy metrics. Regarding accuracy metrics, MAPE, standardized regression coefficients, odds ratio, correlation statistics, average error, limits of agreement, area under the curve, % sensitivity, % specificity, % positive predictive value, % negative predictive value, and likelihood ratio were extracted if available. In cases where group percentage differences were not reported in the study, a group percentage error was calculated ($[\text{Consumer-grade}_{mean} - \text{Research-grade}_{mean}] / \text{Research-grade}_{mean} \times 100$). This was performed to acquire a common unit of measurement for enabling comparisons of accuracy measures within the studied outcome domain [6,11]. Except for % differences (ie, errors),

95% CIs were extracted if they were reported or calculated if appropriate data were accessible.

Statistical Analysis

All quantitative investigations were carried out in RStudio (version 4.1.2) [23], using the meta [24], metafor [25], TOSTER [26], gridExtra [27], dmetar [28], and ggplot2 [29] packages for producing the results and plots of the meta-analysis. A random effects model was used to pool the effect sizes and SEs of the included studies. Two 1-sided tests of equivalence were carried out to determine the number of studies with group error percentages that fall within or outside one of the sides from previously defined acceptable limits and the direction of measurement error was examined. Studies with large sample sizes conducting difference tests are more likely to find statistically significant differences, whereas studies with smaller sample sizes are less likely to do so; both cases lead to incorrect conclusions [8,26]. Although tests of mean difference are a common statistical approach in measurement agreement research, equivalence testing was developed to provide *evidence of equivalence* directly, in contrast with inferring *no evidence of differences* among different devices [8]. Two 1-sided tests of equivalence were conducted to compare the 90% CIs of the estimates from the consumer-grade PA monitors with the defined equivalence zone (EZ) extrapolated from the comparison method. Although no formal guidelines exist to define the best EZ, the interpretation of measurement accuracy in this study included predefined acceptable limits for measurement accuracy of –10% to +10% in controlled and free-living settings, in line with previous secondary publications [6,11] and based on a series of previous primary publications [30-32]. Testing for whether the 90% CI from the measurements of consumer-grade PA monitors falls within the determined EZ was conducted with a statistical significance set at .05.

Because of the variability in the consumer-grade and research-grade devices and the methods used, heterogeneity was suspected; therefore, the random effects model was chosen. Cochran Q test and I^2 test were used to assess heterogeneity and the degree of inconsistency, respectively, across studies [33,34]. Both tests were used, because significant heterogeneity among studies was poorly detected by the Cochran Q test when a small number of studies was included in the meta-analysis, as the power of the test is low under such conditions [33]. I^2 was used, as it represents the percentage of total variation across studies that is because of heterogeneity, ranging between 0% and 100%, where values of 25%, 50%, and 75% point to low, moderate, and high heterogeneity, respectively [33]. As noted earlier, some heterogeneity in the true effect sizes among studies was expected. Outlier and influence analyses were performed to investigate the causes of these problems. As several methods are present for determining outliers in meta-analyses [35], the dmetar package [28] in R contains a *find.outliers* function, which attempts to identify outlying studies included in the meta-analysis, after which these studies are removed, and the pooled effects are recalculated. Influential studies also have a substantial effect on the pooled effect or heterogeneity, and techniques used to identify these studies are based on the leave-one-out method [35]. In the leave-one-out method, recalculation of the results is performed as many times as there are included studies in the meta-analysis, leaving out one study each time [35]. Using the dmetar package [28] and an accompanying *InfluenceAnalysis* function, various influence diagnostics were calculated. A Baujat plot was constructed to illustrate studies that influence overall heterogeneity and the overall result, where those that fall to the top right quadrant have the most influence [36]. Two forest plots of the leave-one-out meta-analyses were also constructed: the first one being sorted by the pooled effect size and the second by the I^2 value. Effect estimates plotted against sample sizes used in the studies were visually inspected with funnel plots, where publication bias and other biases were identified if a skewed and asymmetrical plot was present [37].

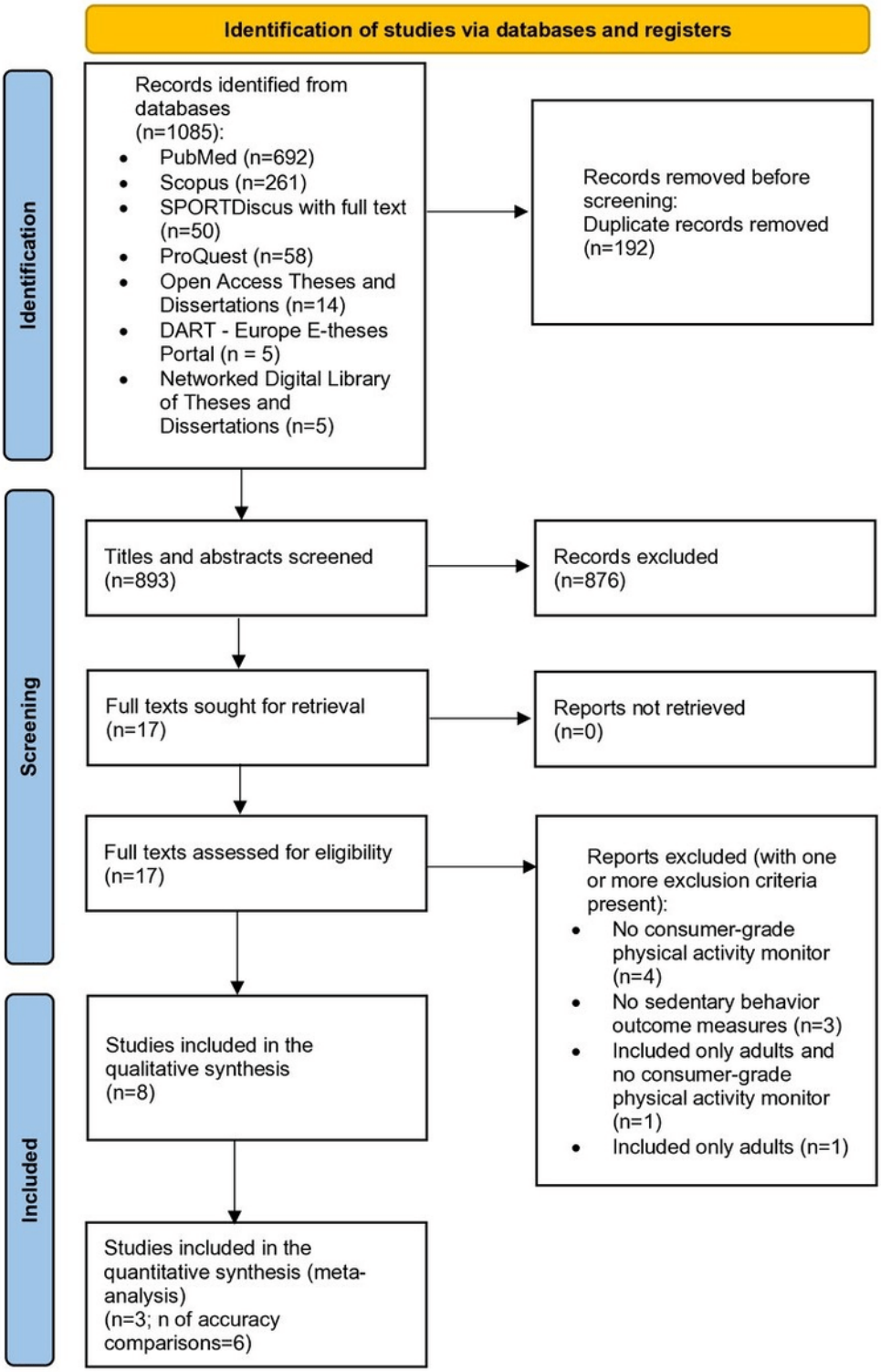
Alternatively, a narrative synthesis was performed for accuracy analyses that did not report data, allowing for the inspection of group percentage errors. Furthermore, the consistency of accuracy metrics available from those studies with the quantitative synthesis was narratively outlined, in addition to the direction of the measurement error.

Results

Study Selection

After applying our search strategy to designated electronic databases, 1085 studies were identified. Further, duplicates were removed, and 82.3% (893/1085) of titles and abstracts were carried over to screening for eligibility. The exclusion of 98.1% (876/1085) of studies left us with 1.9% (17/1085) of studies in which full-text screening was conducted. Following the full-text screening, 9 additional studies were excluded if at least one of the following reasons were present: no consumer-grade PA monitors were used (7/9, 78%), included only adults (2/9, 22%), and no SB outcome measures were assessed (2/9, 22%). Finally, 8 studies complied with all predefined inclusion criteria, and 3 (38%) of these studies provided acceptable data for inclusion in the quantitative analyses. Only studies in English were found and deemed eligible, even though the search was not limited only to this language. The steps taken to identify the studies included in this review are detailed in Figure 1.

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



Study Characteristics

All the information regarding study characteristics is summarized in Table 1. Of the included studies, 8 contained a total of 392 participants, whereas the mean number of participants was 49 (SD 41.36; range 10 to 144). Of the total number of participants, 195 were female (49.7%). Because one of the studies [38] included individuals aged 16 to 25 years with a median age of 19.3 (IQR 17-21) years, it was excluded from the participants' mean age calculation. Hence, the mean age of participants was 8.3 (SD 2.39) years, with the age ranging from 4.8 to 10.3 years. Participants in all studies were healthy, except for one study, which included youth with mental health problems [38] but that did not preclude these participants from engaging in PA. Of the 8 studies, 4 (50%) were conducted in free-living settings [20,38-40], 3 (38%) studies were conducted in controlled settings [41-43], and 1 (13%) study was conducted in both controlled and free-living settings [44]. Most studies

(6/8, 75%) used Fitbit devices (Fitbit Inc) as a consumer-grade WAM [20,38,39,41-43]. Of the 8 studies, the 2 (25%) remaining studies used the Polar active watch (Polar Electro Oy) [40] and Movband, Sqord, and Zamzee [44]. ActiGraph GT3X + (ActiGraph Inc) was most used as a comparison device in free-living settings [20,39,40], with both ActiGraph GT9X [40] and Actiwatch-64 (Philips Respironics) [38] being used in one study. In controlled settings, WAMs were compared with ActiGraph GT3X+ in 2 studies [42,44], whereas direct observation [41] and a portable indirect calorimeter (Cosmed K4B2; Cosmed Inc) [43] were used in one study each. Consumer-grade WAMs were worn on the nondominant wrist (6/8, 75%), dominant wrist (2/8, 25%), or the hip (2/8, 25%). Research-grade PA monitors were worn on the hip and attached to a belt on all the occasions. In studies conducted in free-living settings, children wore the WAM for 24 hours, whereas the duration of the monitoring period ranged from 1 to 7 consecutive days, although in one of the studies, children were observed across 5 days but only during an afterschool program that lasted for 80 minutes [40]. In studies conducted in controlled settings, sets of numerous unstructured or structured activities ranging in intensity are usually performed. The number and duration of these activities were similar, whereas the types of sedentary activity were also similar across studies and included sitting or lying while being quiet, watching television, listening to music, or playing video games. Across the included studies, some differences in the cutoff points used for identifying time spent in SB were present in both consumer-grade and research-grade PA monitors. For studies conducted in free-living settings, cutoff points for identifying SB were reported only in terms of research-grade PA monitors, and SB was usually equivalent to ≤25 counts per 15-second epoch or ≤100 counts per minute. A pair of studies conducted in free-living settings used 2 different cutoff points: SB <2.0 metabolic equivalent of tasks (METs) or <1.5 METs [40], and SB <37.5 counts or ≤25 counts per 15-second epoch. Most of the studies (3/4, 75%) conducted in controlled settings used similar cutoff points, that is, activities with MET values of 1.4 or <1.5 METs were identified as sedentary. Two of the studies did not report cutoff points for consumer- or research-grade PA monitors [38,44].

Table 1. Summary of study characteristics included in the systematic review (alphabetically by author and divided by study setting).

Author, year	Setting	Participants (female), n (%)	Age of participants (years), mean (SD)	Context of sedentary behavior (duration and type)	Type of device (body placement)		Cutoff point for sedentary behavior	Accuracy metric reported
					Consumer-grade	Research-grade		
Byun et al [20], 2018	Free-living	27 (11)	4.9 (1.0)	2 consecutive days (24-hour period)	Fitbit Flex (nondominant wrist)	ActiGraph GT3X+ (right hip)	Pate cutoff: <37.5 counts per 15 s; Evenson cutoff: ≤25 counts per 15 s	Pearson product moment correlation coefficients (*r*); MAPE[a]
Kim and Lochbaum [40], 2018	Free-living	51 (32)	10.30 (0.9)	Up to 5 consecutive days (afterschool program for 80 minutes in a predesignated classroom)	Polar active watch (nondominant wrist)	ActiGraph GT3X+ (waist); ActiGraph GT9X (nondominant wrist)	<2.0 METs[b]; <1.5 METs; Evenson cutoff: ≤50 counts per 30 s; Chandler cutoff:[c] <966 counts per 30 s	Pearson product moment correlation coefficients (*r*); MAPE; regression coefficients; mean ratios for equivalence tests; level of agreement
Mooses et al [39], 2018	Free-living	144 (72)	9-10 y	Only during school hours for 2 weeks (one in September and one in November 2016)	Fitbit zip (hip[d])	ActiGraph GT3x-BT (waist)	Evenson cutoff: ≤100 counts per min (1-min epochs)	Spearman correlation (*r*); limits of agreement[e]
Scott et al [38], 2019	Free-living	10 (6)	Median 19.3 (IQR 17-21)	7 consecutive days and nights (24-hour period)—the proportion of sedentary time	Fitbit[f] (nondominant wrist)	Actiwatch-64 (nondominant wrist)	—[g,h]	MAPE; level of agreement
Sirard et al [44], 2017	Controlled and free-living	Phase 2: 14 (7); phase 3: 16 (8)	Phase 2: 9.0 (2.0); phase 3: 8.6 (1.6)	Phase 2: 10 activities on 2 occasions (sedentary activity for 5 min)—sitting quietly; phase 3: 4 consecutive days (24-h period)	Movband (dominant wrist); Sqord (dominant wrist); Zamzee (right hip)	ActiGraph GT3X+ (right hip)	Evenson cutoff: ≤100 counts per min	Phase 2: spearman rho coefficients; phase 3: Spearman correlation (*r*)
Byun et al [41], 2018	Controlled	28 (13)	4.8 (1.0)	A total of 6 activities for 34 min (sedentary activity for 8 min with a 1 min rest between)—sedentary (watching television lying down for 4 min and watching television sitting on a couch for 4 min)	Fitbit Flex 1 (nondominant wrist)	Direct observation	<1.4 METs (Fitbit 1-min epochs; direct observation 5- to 15-s epochs)	Pearson correlation coefficients (*r*); MAPE; Cohen κ; sensitivity; specificity; ROC-AUC[i]
Godino et al [43], 2020	Controlled	59 (31)	9.9 (0.7)	14 activities for 2-3 h (each sedentary activity for 5 min)—sedentary (sitting quietly, listening to music, and playing games on iPad)	Fitbit Charge HR (nondominant wrist)	Cosmed K4B2 (fitted according to manufacturer recommendations)	<1.5 METs (1-min epochs)	MAPE; Cohen κ; sensitivity; specificity

Author, year	Setting	Participants (female), n (%)	Age of participants (years), mean (SD)	Context of sedentary behavior (duration and type)	Type of device (body placement)		Cutoff point for sedentary behavior	Accuracy metric reported
					Consumer-grade	Research-grade		
Kang et al [42], 2019	Controlled	43 (18)	9.7 (1.3)	12 activities for 48 min (each 3 min with a 1 min rest between)—sedentary (sitting quietly in a chair, playing a video game, and watching television)	Fitbit Charge HR (dominant and nondominant wrist)[j]	ActiGraph GT3X+ (dominant and nondominant wrist)[j]	<1.4 METs (1-min epochs)	Pearson correlation coefficients (*r*); MAPE; Cohen κ; sensitivity; specificity; ROC-AUC; ICC[k] (95% CI)

[a]MAPE: mean absolute percent error.

[b]MET: metabolic equivalent of task.

[c]Due to an error during production, Chandler cutoff points at the 30-second epoch length were incorrectly presented in the published paper (the corrected cutoff points are inserted in Table 1).

[d]The accelerometer and Fitbit Zip were attached on the hip with the same elastic belt and worn on the same side.

[e]Bland-Altman analysis with the calculation of bias between 2 devices (the mean of differences of the 2 devices).

[f]Model not reported.

[g]Not available.

[h]Sedentary behavior reported as a 0 to 1 value which represents the number of minutes sedentary divided by the morning time.

[i]ROC-AUC: area under the receiver operating curve.

[j]Random counterbalance of the wear position between the ActiGraph and Fitbit tracker on the wrist.

[k]ICC: intraclass correlation.

Risk of Bias

The risk of bias assessment results are presented in Table 2. On the individual COSMIN component level, all included studies were rated as either excellent or good in 3 components of the methodological quality evaluation, relating to the reporting of missing data, handling missing data, and use of an adequate criterion comparator (ie, device). In terms of acceptable accuracy metrics, 75% (6/8) of studies were rated as excellent (n=4, 67%) or good (n=2, 33%), and 25% (2/8) of studies were rated as poor because no percentage error was reported or a way to calculate it was present, although the studies reported other measures of accuracy. Instead of entirely excluding these studies from the review, the reported measures of accuracy, and their consistency with the examination of percentage measurement error are narratively outlined. When examining the components related to important methodological flaws in the design or execution of the studies, of the 8 studies, 6 (75%) studies were rated excellent (n=5, 83%) or good (n=1, 17%), and 2 (25%) were rated fair. In contrast to the scoring of most COSMIN components, most studies were rated as fair (1/8, 13%) or poor (4/8, 50%) in the adequate sample size component. Regarding the studies, of 8 studies, only 1 (13%) study was rated excellent and 2 (25%) studies were rated good in terms of sufficient sample size. No studies were excluded from the analysis owing to poor methodological quality, as only the adequate sample size component was unfavorable. Only one study had more than 100 participants (N=144) and was rated as excellent; therefore, no restrictions were set in terms of inclusion for the minimum number of participants needed in a study.

Table 2. Results of risk of bias assessment for studies included in the systematic review (N=8)[a].

Study details		Summary: excellent or good, n (%)	COSMIN[b] risk of bias assessment					
Author, year published	Study setting		Reporting missing data	Handling missing data	Adequate sample size	Acceptable comparison	Flaws in design or methods	Acceptable accuracy metrics
Kim and Lochbaum [40], 2018	F[c]	6 (100)	Good	Good	Good	Excellent	Excellent	Excellent
Mooses et al [39], 2018	F	6 (100)	Excellent	Good	Excellent	Excellent	Excellent	Excellent
Byun et al [20], 2018	F	5 (83)	Good	Good	Poor	Excellent	Excellent	Good
Scott et al [38], 2019	F	4 (67)	Excellent	Excellent	Poor	Good	Fair	Excellent
Godino et al [43], 2020	C[d]	6 (100)	Excellent	Excellent	Good	Excellent	Good	Excellent
Byun et al [41], 2018	C	5 (83)	Good	Good	Poor	Good	Excellent	Good
Kang et al [42], 2019	C	4 (67)	Good	Good	Fair	Excellent	Excellent	Poor
Sirard et al [44], 2017	C, F	3 (50)	Excellent	Excellent	Poor	Excellent	Fair	Poor

[a]The summary of excellent or good values of reporting missing data, handling missing data, adequate sample size, acceptable comparison, flaws in design or methods, and acceptable accuracy metrics are 8 (100%), 8 (100%), 3 (38%), 8 (100%), 6 (75%), and 6 (75%), respectively.

[b]COSMIN: Consensus-Based Standards for the Selection of Health Status Measurement Instruments.

[c]F: free-living.

[d]C: controlled.

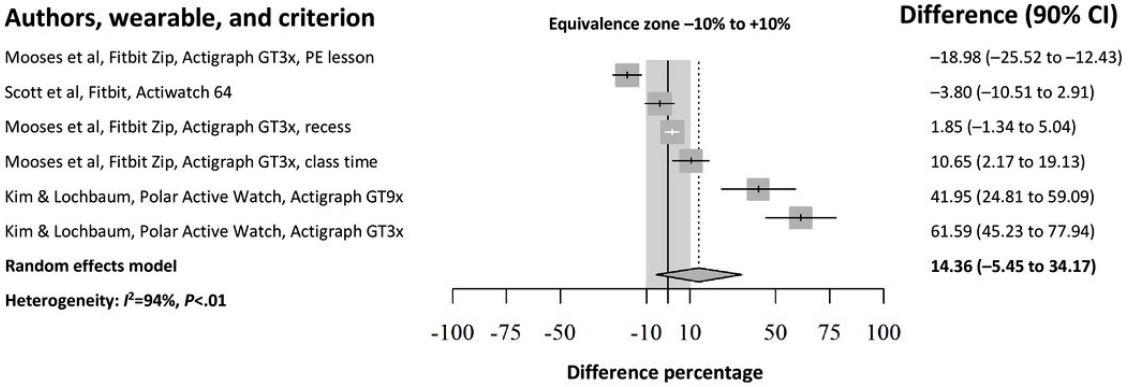
At the individual study level, the cutoff point for high study quality was set arbitrarily and was defined as scoring excellent or good on 88% (7/8) or 100% (8/8) of the components. In free-living settings, 3 studies were evaluated as being of high quality [20,39,40], whereas in controlled settings there were 2 studies of high quality [41,43]. All studies scored excellent or good on more than half of the COSMIN risk of bias components, although 3 of the studies could not be considered high quality [38,42,44]. The only study carried out in both free-living and controlled settings [44] had the lowest number of excellent or good scores (5/8, 62%) and also scored poorly on 2 components (ie, adequate sample size and acceptable accuracy metrics).

Accuracy of Time in SB Measurements

Quantitative Synthesis

In total, 8 studies containing 14 accuracy comparisons examined the accuracy of consumer-grade PA monitors in relation to a comparison method in assessing time spent in SB in children, although out of the total number, only 3 (38%) studies reported acceptable data for inclusion in the quantitative analyses. From the 3 studies included in the quantitative synthesis, 6 comparisons were examined. The random effects model results as seen in Figure 2. provided an estimated model coefficient (ie, the summary effect size) of 14.4% (SE 12%, 90 % CI −5.5% to 34.2%; *P*=.23).

Figure 2. Forest plot of consumer-grade physical activity monitors accuracy in relation to research-grade monitors for assessing sedentary behavior in children. Squares represent point estimates, and 90% CIs are indicated by lines. The pooled effect size and 90% CI of the random effects model are shown at the bottom, represented by a diamond. The dashed line represents the predefined equivalence zone of the comparison method (ie, –10% to +10%). PE: physical education. Mooses et al [40] (activity: physical education lesson); Scott et al [39]; Mooses et al [40] (activity: recess and activity: class time); Kim and Lochbaum [41] (research-grade physical activity monitor: ActiGraph GT9X and research-grade physical activity monitor: ActiGraph GT3X+).



The equivalence test results showed that consumer-grade PA monitors did not generate equivalent estimates of SB compared with research-grade monitors. The overall effect with the corresponding 90% CI was not completely within the predefined EZ of the comparison method (ie, –10% to +10%). It is also important to note the direction of the overall effect and the corresponding 90% CI with regard to the defined EZ of the comparison method, where it is evident that consumer-grade PA monitors overestimated SB compared with research-grade devices. Point estimates and corresponding 90% CIs of only 1 out of 6 accuracy analysis were located inside the predefined EZ [38], with a 90% CI of one additional accuracy analysis being borderline equivalent [39].

The results of the Cochran *Q* test indicated that heterogeneity among the population effect sizes estimated by the individual studies was present (Q_5=82.86; P<.001). Furthermore, the I^2 statistic [33] was 94% (95% CI 90.4%-96.2%), indicating very large heterogeneity. Although the Q-statistic and I^2 provide evidence regarding heterogeneity, there is no information on which studies may influence overall heterogeneity. The search for potential outlying accuracy analyses yielded 2 results [39,40]. After excluding the identified studies from the meta-analyses, I^2 decreased, although only marginally, from 94% to 84.5%, and the *Q* test of heterogeneity was still significant (P<.001). In addition, by removing these 2 studies, the pooled estimates were brought closer to the defined EZ from 14.4% (90% CI −5.5% to 34.2%) to 10.8% (90% CI −4.1% to 25.7%). Leave-one-out meta-analyses were also conducted with visualization of the results through forest plots that are presented in Multimedia Appendix 2. In the leave-one-out meta-analysis sorted by the pooled effect size, we found that the overall % difference was the largest when we removed 1 of the 2 outlying and influential studies [39] that had a very high contribution to the pooled effect size. In the second leave-one-out meta-analysis sorted by the values of I^2 (ie, heterogeneity), omitting one of the studies [40] led to the largest decrease in I^2. In conclusion, the results of the outlier and influence analyses indicate that the 2 studies [39,40] are likely influential outliers. Hence, a sensitivity analysis was conducted, in which these studies were excluded. The changes in the pooled effect size, I^2, and CIs associated with removing influential studies are shown in Table 3.

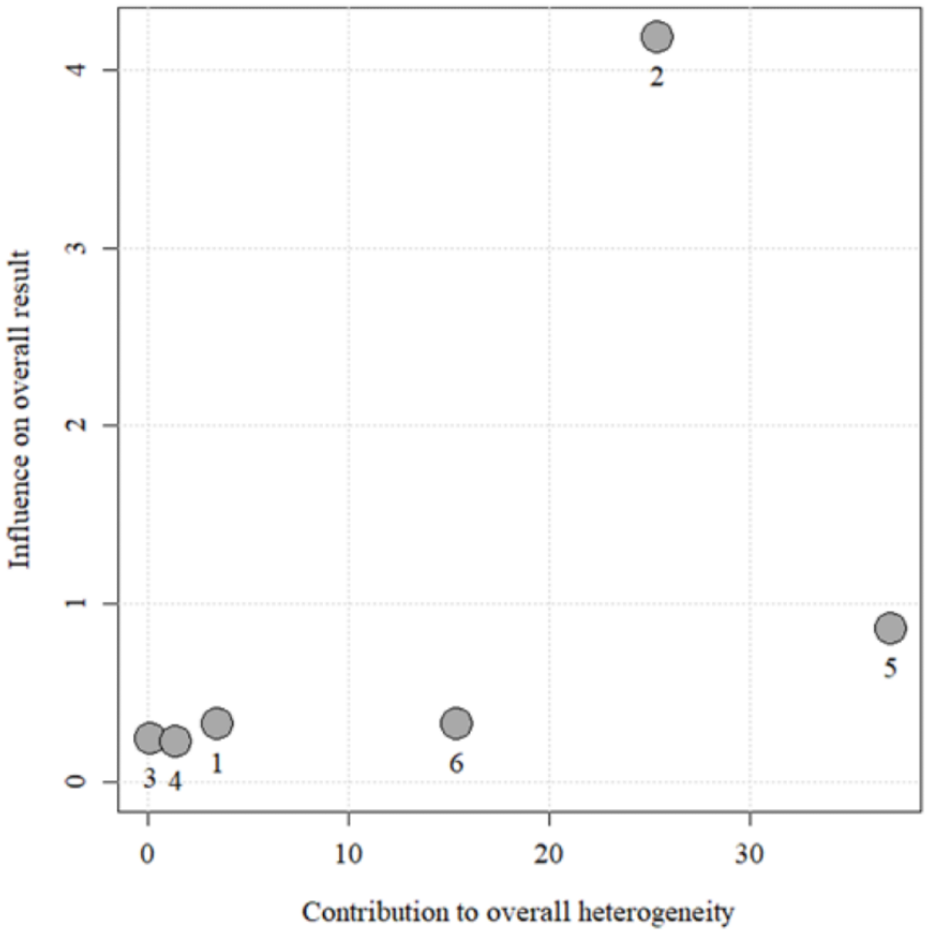
Table 3. Random effects model results before and after removing the outliers.

Analysis	Pooled effect size (%, 90% CI)	*P* value	I^2 (%, 95% CI)
Main analysis	14.4 (−5.5 to 34.2)	.23	94 (90.4 to 96.2)
Influential studies removed[a]	10.8 (−4.1 to 25.7)	.23	84.5 (66.4 to 92.8)

[a]Studies removed as outliers: Mooses et al [39] (activity: physical education lesson) and Kim and Lochbaum [40] (research-grade physical activity monitor: ActiGraph GT3X+).

As there is evidence of overall heterogeneity, the Baujat plot can display studies that contribute to overall heterogeneity and overall results [35]. A Baujat plot is shown in Figure 3. with the respective ID numbers used to differentiate the individual accuracy comparisons. Accuracy comparison [40] ID number 5 contributed the most to the overall result as well as the overall heterogeneity. Accuracy comparison [39] ID number 2 contributed the most to the overall heterogeneity and results, being closest to the upper right corner of the plot. A closer look at the characteristics of this accuracy comparison revealed that using different models and placements of consumer-grade and research-grade PA monitors could be potential moderating variables that may contribute to heterogeneity. In 2 identified outlying and influential accuracy analyses, one of the studies placed the consumer-grade PA monitor (ie, Fitbit Zip) at the hip [39], contrasting the placement in other included studies, whereas the other study used a Polar active watch as the consumer-grade PA monitor [40] also contrasting other studies.

Figure 3. Baujat plot showing the influence of individual studies on the overall heterogeneity and the overall result where studies falling closer to the top right quadrant have the most influence. 1: Mooses et al [40] (activity: class time); 2: Mooses et al [40] (activity: physical education lesson); 3: Mooses et al [40] (activity: recess); 4: Scott et al [39]; 5: Kim and Lochbaum [41] (research-grade physical activity monitor: ActiGraph GT3X+); 6: Kim and Lochbaum [41] (research-grade physical activity monitor: ActiGraph GT9X).



A funnel plot was constructed to assess publication bias and is provided in Multimedia Appendix 3. Because of the low power of asymmetry statistical tests when <10 studies were included [45], only a visual inspection of the funnel plot was carried out. After visual inspection, an asymmetry in the plot was noticed, indicating the possibility of publication bias, which should not be equated with it, as several conceivable causes are plausible [37,45].

Narrative Synthesis

Results from the included accuracy analyses that did not report data that would allow for the quantitative analysis of the accuracy of consumer-grade PA monitors in assessing the time spent in SB in children are narratively outlined in the following sections. In these studies, the results pertaining to other available accuracy metrics are summarized. This narrative synthesis encompasses 5 studies, in which 8 accuracy analyses were identified. Of the total 8 studies, 2 (40%) studies were conducted in free-living settings, of which a study was also carried out in controlled settings in one of the implementation phases. Byun et al [20] reported a mean difference of 42 to 71 minutes per day during 2 consecutive days among the devices when measuring SB, where the 90% CI for the mean estimates from the consumer-grade PA monitor (ie, Fitbit Flex) was within 15% of the mean estimates from ActiGraph GT3X+. In addition, a strong correlation of the time spent in SB between the 2 devices has been reported (r=0.87) [20]. Sirard et al [44] also used ActiGraph GT3X+ as a research-grade PA monitor in controlled and free-living settings for 4 consecutive days to assess the accuracy of several WAMs and reported a high correlation in SB time assessed in free-living conditions for Movband (r=0.76) and Sqord (r=0.86) PA monitors, whereas a moderate correlation was reported for Zamzee (r=0.59). Of note for this study, in phase 2, which was conducted in controlled settings, all the devices differentiated SB from light-intensity PA with similar accuracy as the research-grade PA monitor [44]. Moving on now to consider controlled settings, the sensitivity and specificity of SB detection for the Fitbit Flex device reported by Byun et al [41] amounted to 96.8% and 88.6%, respectively, with high SB classification accuracy (90.2%) and high area under the receiver operating characteristic curve values (0.92). In this study, the Fitbit Flex produced a negligible bias in SB estimation, with approximately 2 more minutes of SB recorded in relation to the criterion method (ie, direct observation) [41]. High sensitivity (84.8%) and specificity (83.1%) values have also been reported by Godino et al for Fitbit Charge HR for classifying SB or light PA [43]. A similar performance of Fitbit Charge HR was recorded in the study by Kang et al [42] with a sensitivity of 91.6% and specificity of 72.4%. Values were consistently high for studies conducted in

controlled settings, as also reported by Kang et al [42]. Herein, the classification accuracy (80.73%) and area under the receiver operating characteristic curve (0.82) values were also high for Fitbit Charge HR [42].

A common metric reported in most of the studies was MAPE, where considerable disagreement when measuring SB (ie, MAPE >60%) was present only in the study by Kim and Lochbaum [40]. Kim and Lochbaum [40] reported MAPEs of 121.68% (95% CI 84.87%-158.49%) and 122.73% (95% CI 53.9%-191.57%) for SB <2.0 METs but lower MAPEs of 69.92% (95% CI 63.39%-76.44%) and 79.84% (95% CI 55.21%-104.46%) for SB <1.5 METs when comparing Polar active watch with ActiGraph GT3X+ and ActiGraph GT9X, respectively. Hence, the results depended mostly on the defined SB cutoff points. Most studies that reported MAPEs reported values of <30% [20,38,41]. A total of 2 studies conducted in children of similar ages (4.9, SD 1.0 years and 4.8, SD 1.0 years) with the same consumer-grade PA monitor (Fitbit Flex) reported MAPEs of 9.1% and 13.0% based on different cutoff points [20] and 28.8% [41], respectively. Accordingly, differences appeared because the first study was conducted in free-living settings, comparing Fitbit Flex with a comparison device (ie, ActiGraph GT3X) [20], whereas the other study compared Fitbit in controlled settings with a criterion method (ie, direct observation) [41], where larger differences were expected. Undoubtedly, values depend mostly on the cutoff points used, settings or used devices. The study by Kim and Lochbaum [40] can be seen as an outlier in both quantitative synthesis and narrative synthesis. This disparity with other results could be due to the previously mentioned arguments revolving around differences in used devices, cutoff points, and settings. Not enough data were provided to try narratively synthesizing the direction of differences between consumer-grade and research-grade PA monitors. Regarding precision, no study has reported data on the precision of consumer-grade WAMs in assessing the time spent on any type of SB in children and adolescents. Only one of the included studies examined precision; however, SB was not considered an outcome in these analyses. In general, this study reported good precision for moderate to vigorous PA, energy expenditure, steps, and heart rate among devices carried on the wrists of both hands [42]. Furthermore, a study examined only interunit variability during orbital shaker testing, which is related to repeatability as one of the aspects of precision**.** Of the 3 devices used across a range of frequencies (1.3, 1.9, and 2.5 Hz), Movband showed the lowest interunit variability (coefficient of variation [CV] 0.62, 0.85, and 0.19, with respect to frequencies), whereas Sqord (CV 29.8, 3.85, and 1.93, respectively) and Zamzee produced worse results (CV 25.5, 12.1, and 9.75, respectively) [44]. These results could be relevant in situations when they are used in groups (eg, classrooms) where different children involved in the same activity may present different results in terms of the measured activity.

Discussion

Principal Findings

This review is one of the first studies that summarized data on the accuracy of consumer-grade PA monitors when measuring the time spent in any type of SB in children and adolescents. On the basis of the limited amount of data available for quantitative synthesis, it seems that consumer-grade PA monitors did not generate equivalent estimates of SB compared with research-grade monitors, with a tendency toward overestimation for these devices. In contrast, narrative synthesis suggested that at least some of these devices (ie, Fitbit) should be viewed as an accurate method of SB measurement in children and adolescents owing to the high levels of classification accuracy found in several individual studies.

The fact that WAMs were not found to be equivalent to research-grade monitors in this study should not be interpreted as having low accuracy. In all included studies conducted in free-living settings, accelerometers were used to determine the accuracy of consumer-grade PA monitors. Accelerometers cannot be regarded as the gold standard for measuring SB, although they produce results similar to a criterion method (ie, inclinometers). Accelerometers placed on the hip with their corresponding cutoff points overestimate the time spent in SB in comparison with a criterion method when young children, adolescents, and adults are considered because standing is also included as one of the inspected postures [14]. In general, criterion methods were only used in controlled settings in 25% (2/8) of the included studies. Controlled settings are appropriate for examinations of “genuine” accuracy although that does not necessarily translate to free-living settings, in which consumer-grade PA monitors are intended to be used. Comparing the accuracy of consumer-grade PA monitors in relation to criterion methods in free-living settings is difficult, where the use of these devices over several days when the participant is engaged in their everyday activities is not feasible. Therefore, it is uncertain whether the accuracy of consumer-grade PA monitors is poor when they are used to measure SB, especially in free-living settings. The specific aims of the study and the significance of accuracy and practicality should be considered when selecting a suitable device [46]. Using gold standards (eg, indirect calorimetry) could be needed in clinical studies; however, the cost and difficulties encountered with using them make them unsuitable for large samples located in free-living settings [47]. This problem is even more pronounced when working with children than with adults [47].

The random effects model results indicate that consumer-grade PA monitors overestimate the amount of time spent in SB, while removing 2 influential and outlying studies brought the estimates closer to the defined EZ. The placement of WAMs could be a potential moderating variable because Fitbit was placed in one of these studies at the hip [39], whereas some previous studies also reported overestimation of SB from hip-based accelerometers in adolescents and adults [14]. As overestimation of SB in our study was noticed for consumer-grade PA monitors mostly in relation to accelerometers, potentially even higher levels of overestimation would be present if the gold standard

was used for comparison. A research-grade device measuring the inclination of the thigh, such as activPal, is regarded as accurate for SB measurement among children in free-living settings [48], because it uses an inclinometer, a sensor capable of better horizontal (sitting or lying) and vertical (standing) position classification [49]. The use of these types of sensors in consumer-grade PA monitors could potentially have positive effects on SB measurement accuracy. Therefore, placing the PA monitor on the thigh might also be suggested, as changes in thigh positions proved to be the most accurate way of measuring SB [49]. As the thigh is at different inclinations when sitting and standing, the future might offer alternative solutions if the identification of different positions and inclinations of the wrist when shifting from sitting to standing and engaging in PA could be developed in consumer-grade PA monitors [14,50]. As it was shown that PA monitors wear time, over a longer period, declines, and comfort was defined as one of the most important factors [10], wearing the device on the wrist could increase their acceptance among children and adolescents. A visual inspection of the plots showed that the study by Kim and Lochbaum [40] contributed the most to the levels of heterogeneity and pooled results. The authors used a Polar active watch as the consumer-grade PA monitor, unlike other studies that used Fitbit devices [40]. Differences in measurements of SB are represented by values of MAPE >60%, in contrast to other studies, and it seems that the Polar active watch in this case [40] is not an acceptable device for SB measurement. Placing the research focus mostly on one device brand (ie, Fitbit) and a couple of models (eg, Charge, Zip, and Flex) of that brand produces limited knowledge regarding the accuracy of consumer-grade PA monitors. This is why this fact is pointed out as a potential confounder, as excluding the study by Kim and Lochbaum [40] would certainly provide better results in terms of accuracy. The results of our study cannot be generalized to all consumer-grade PA monitors, as only a few brands have been analyzed to date. The discontinuation of certain models is inevitable as the market and interest grow, as well as technological development. Even though we only included studies published since 2015, most of the devices used in these studies have been discontinued (ie, Fitbit Charge HR, Flex, and Zip and Polar active watch), although companies still provide consumer support [51]. Advanced algorithms and sensors, such as inclinometers and heart rate monitors, typically present in current WAM models could provide more accuracy when measuring SB, and a large part of the devices included in this review did not contain any of them. Used only in studies conducted in controlled settings, the Fitbit Charge HR, which also contains a heart rate monitor, did not prove to be superior in terms of accuracy when compared with other consumer-grade PA monitors. This might be because the used algorithms, as Fitbit Charge HR is an older model when compared with other included Fitbit devices, even though it has multiple built-in sensors.

When consumer-grade PA monitors are used by children and adolescents, their accuracy in detecting SB might be affected because of the greater amount of time spent in postures not typically observed in adults (eg, crawling, squatting, and kneeling). A previous study reported that as children spent more time in previously mentioned postures than adults, an overestimation of time spent in SB recorded by the activPal was found [14]. The various epoch lengths reported in the studies included in this review may have contributed to conflicting results when assessing the accuracy of consumer-grade PA monitors in free-living and controlled settings. Epoch lengths used in studies conducted in free-living settings have generally been shorter (ie, 15 and 30 seconds) than those used in studies with controlled settings (ie, 1 minute). It is up to discussion whether shorter epochs are better at assessing SB compared with longer epochs in children and adolescents. When PA is considered, shorter epochs seem to be better because of children's intermittent behavioral patterns [14,16]. However, in terms of SB, it is less likely that children will sit still for longer periods, which could partly explain the reported SB overestimation in free-living settings when shorter epochs were used. Applying longer epochs (ie, 1 minute) might result in underestimation of SB in children due to the sporadic nature of their movements, although no relevance of shorter SB epochs has been derived when it comes to impacts on the overall health [14].

Even if WAMs prove to be more accurate in assessing SB in the future, they may be limited by the fact that they do not recognize the context of SB. The context of SB is important, because higher durations or frequencies of screen time, television viewing, and video game use were previously mostly associated with a myriad of negative consequences (eg, body composition, cardiometabolic risk scores, physical fitness, and self-esteem) [52]. Although also defined as SB, more time spent reading and doing homework was associated with positive outcomes (eg, academic achievement) [52]. As limited data are available to discuss the precision of consumer-grade PA monitors, no specific discussion has focused on this issue. In future research, the precision of consumer-grade WAMs in assessing SB in children and adolescents should be considered. The characteristics of a good instrument emphasize both accuracy and precision, whereas the latter is neglected in this specific area of research.

High levels of heterogeneity were found in our study owing to differences in study protocols, type of wearable devices examined, comparison methods, sample sizes, and reported outcome measures, which complicated the analysis and comparisons among the results of the included studies. Although the risk of bias assessment showed high levels of methodological quality for all included studies and most acceptable accuracy metrics were reported, most included studies did not contain an adequate sample size. In line with our findings, it has recently been reported that studies evaluating data from wearable devices comprise different study designs with samples of varying characteristics and sizes, methodological approaches, devices used, and different cutoff points for activities across all intensities [9]. A recent review that included 23 validation studies of reported energy expenditure estimates from 58 devices comparing them to appropriate comparison devices suggested that most studies (87%) reported inappropriate accuracy indicators (eg, correlation coefficients) [53]. Sample sizes from the studies included in the review ranged from 13 to 60 participants (ie, 52% with sample sizes ranging from 20 to 30 participants). This agrees with our results that the sample sizes

in this area of research are not adequate. Only half (52%) or even fewer studies reported the recommended accuracy metrics (ie, MAPE and equivalence test results) needed to evaluate the actual individual error [53]. Equivalence tests and difference tests depend on arbitrary levels of significance and sample sizes; therefore, MAPE seems to be the most appropriate accuracy metric [53]. The quantitative synthesis of the data in this study was complicated by the fact that not every accuracy metric provides the same information, which is a major problem in this area of research [53]. In studies conducted in free-living settings, WAMs were worn during the study course, lasting for 1 day and up to 7 consecutive days. According to a study by Trost et al [49], a monitoring period of 7 days provided optimal approximations of daily moderate to vigorous PA among children and adolescents. The only study trying to determine how many days of monitoring are needed to provide precise estimates of SB for children was conducted with preschoolers [54]. Precise estimates of the total daily time spent in SB were possible after 6 to 9 consecutive days of monitoring [54]. Hence, it is questionable whether the most commonly used period of 7 days of monitoring would be acceptable in terms of SB analysis in children and adolescents.

Future Research

Regarding future research paths, the age and relevance of different consumer-grade devices and their models should be considered, as well as algorithms used, as they tend to constantly change with the growth of the accompanying market [1]. A large number of tested consumer-grade PA monitors are soon outdated or are no longer in use [55], adding to the complexity of this research area. In addition, Fitbit devices are the most commonly used as illustrated in this study and several other reviews [6,10,15,16,47], possibly because of their high market share and low cost. In contrast, no information on the accuracy of more expensive consumer-grade WAMs, such as smartwatches, in assessing SB is currently available. At the same time, several very low-cost WAMs are available on the market for prices as low as US $45.50 (eg, Mi Smart Band 6). These instruments provide an opportunity for mass PA promotion in children and adolescents, but their accuracy needs to be tested beforehand [55]. The transparency of the algorithms used by the devices and companies should be encouraged, because defining adequate wear time criteria and cutoff points for activities of different intensities is challenging at present. This is due to constant firmware updates, which are needed for further improvement of PA and SB measurement [11]. In addition to accuracy assessment, consumer-grade PA monitoring feasibility and acceptability research among children and adolescents is important. These types of studies are underrepresented in the literature, with results from a recent review showing that only approximately one-third of the studies (32%) investigated effectiveness, user engagement, and acceptability altogether [56]. In addition, descriptive statistics and visual analysis were performed in 60% of these studies when assessing effectiveness without using inferential statistics, and 18.9% of all studies had small sample sizes (ie, <13 participants) [56]. This could be of importance in terms of WAM acceptability among children, because the information from previous studies shows that one-third of consumer-grade PA monitor owners from the United States stopped using the device within 6 months of receiving the device, and just above 40% of them continued using it after 2 years have passed [14,50]. For children and adolescents, the definition of the epoch length that WAMs should use when measuring SB remains unresolved, and further research comparing the accuracy of consumer-grade with research-grade PA monitors conducted in free-living settings should be used to test the accuracy of different epochs [14]. Identification of contexts (ie, settings) in which examination of measurement properties has been previously conducted should be considered when choosing the appropriate device for examining SB in children and adolescents [15]. Hence, if the measurement properties of the selected tool are unknown in certain contexts, future research should also focus on examining the measurement properties of WAMs in these contexts to ensure certainty when these devices are applied outside the research settings [15]. Smartphones offer certain possibilities in this regard, as they could provide ways of context identification if data regarding screen time could be gathered and used in future research on SB. The rapid growth of the WAM market should be accompanied by additional validation studies, as the available evidence summarized in this study identified only a single study that has shown that consumer-grade PA monitors are comparable with research-grade devices in terms of SB measurement in children and adolescents. A caveat to consider is that this specific study included only 10 participants, and did not limit only to children and adolescents (median age 19.3, IQR 17-21 years) [38]. Also, all 10 participants reported depressive symptoms, 4 (40%) also reported anxiety symptoms, 3 (30%) hypomania symptoms and 1 (10%) had a history of hallucinations [38]. Therefore, conclusions regarding the accuracy of consumer-grade PA monitors for the entire childhood period cannot be drawn based solely on the results of this study.

Strengths and Limitations

The strength of this study relates to the fact that a broad search of electronic databases was performed, which included searching for gray literature and manual searching of the included studies reference lists and secondary sources. Another significant strength of this study is that it is the first to examine the accuracy of consumer-grade PA monitors in assessing SB, encompassing a quantitative synthesis of the available data as well as a narrative synthesis of studies not suitable for meta-analysis. Limitations relate to the fact that during the time needed to complete all stages of this review, new studies could have been published, as this area of research is very dynamic. The consumer-grade PA monitoring market is volatile, with new models being constantly brought to the market, and the technology is continuously improving. Another minor limitation could be that during the study selection phase, only 1 reviewer screened 90% of the studies, although an interrater agreement of 100% was reached after the first 10% of the abstracts and titles were screened independently by 2 reviewers. A limitation related to the small number of primary studies included in this review should also be noted. Not including smartphone apps in the review limits the generalizability of our findings, as they also provide data related to the time spent on activities of different intensities. Smartphone apps are already in wide use

among children; therefore, an examination of their accuracy in measuring SB should be performed in the future. Smart watches, which were not identified in any of the studies in this review, are also being accepted by children, although their price, battery life, and complex user interface represent certain disadvantages when used in this area of research with children [55]. None of the included studies used a smart watch to test the accuracy of SB measurement in children, and only smart bands were included. Studies lasting longer than 7 days were not included in this study as none have been identified, potentially serving as a limitation and a guide for future studies. Generally, at the individual component level, all included studies were either excellent or good in terms of missing data reporting, missing data handling, and use of an adequate comparison measure. However, most (5/8, 63%) of the studies consisted of small sample sizes (ie, <50), and the criterion method use was questionable, especially in controlled settings, as methods such as direct observation and indirect calorimetry were underrepresented.

Conclusions

To our knowledge, this is the first review to focus specifically on the accuracy and precision of consumer-grade PA monitors when measuring SB in children and adolescents, but we found a small number of available studies, especially those suitable for conducting a meta-analysis. In the quantitative synthesis, no equivalence in the average time spent in SB was found when consumer-grade PA monitors were compared with research-grade monitors. High levels of heterogeneity were noted in the results, although point estimates and corresponding 90% CIs of only one individual study were located inside the predefined EZ, with a 90% CI of an additional accuracy analysis being borderline equivalent. Moreover, heterogeneity was discernible in terms of different study designs with samples of varying characteristics and sizes, methodological approaches, devices used, and differences in the cutoff points used when defining SB. The narrative synthesis suggests that consumer-grade PA monitors could be considered a valid method of SB measurement in children and adolescents. The results of our study will inform researchers, clinicians, and consumers on the measurement accuracy of widely attainable PA monitors when measuring SB in children and adolescents. However, more evidence is needed to reach robust conclusions about the accuracy and precision in measuring SB of children and adolescents, even for the most prevalent devices currently available on the market.

Acknowledgments

The work of AM was funded by the Croatian Science Foundation under grant DOK-2020-01-3728. The work of JK was funded by the Croatian Science Foundation under grant DOK-2018-01-2328. The work of PJ was funded by the Croatian Science Foundation under grant DOK-2018-09-8532.

Authors' Contributions

AM and MS conceptualized this review. The methodology was determined by AM and MS. AM, JK, and MS performed the screening, and data extraction was done by PJ and checked by AM and HP. Relevant resources were determined by all authors. The preliminary draft was written by AM. All authors reviewed and edited the manuscript. Visualizations were performed by AM. The systematic review was conducted under the supervision of MS and HP. Finally, all authors read the final version of the manuscript and approved its submission.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy.
[DOCX File , 14 KB - mhealth_v10i8e37547_app1.docx]

Multimedia Appendix 2
Leave-one-out meta-analyses with accompanying forest plots.
[DOCX File , 46 KB - mhealth_v10i8e37547_app2.docx]

Multimedia Appendix 3
Funnel plot.
[DOCX File , 21 KB - mhealth_v10i8e37547_app3.docx]

References

1. Grand View Research. Fitness Tracker Market Size Worth $138.7 Billion By 2028 | CAGR: 18.9%: Grand View Research, Inc. CISION PR Newswire. 2021 Aug 24. URL: https://www.prnewswire.com/news-releases/fitness-tracker-market-size-worth-138-7-billion-by-2028--cagr-18-9-grand-view-research-inc-301361326.html [accessed 2021-11-30]

2. Brickwood KJ, Watson G, O'Brien J, Williams AD. Consumer-based wearable activity trackers increase physical activity participation: systematic review and meta-analysis. JMIR Mhealth Uhealth 2019 Apr 12;7(4):e11819 [FREE Full text] [doi: 10.2196/11819] [Medline: 30977740]
3. Wright SP, Hall Brown TS, Collier SR, Sandberg K. How consumer physical activity monitors could transform human physiology research. Am J Physiol Regul Integr Comp Physiol 2017 Mar 01;312(3):R358-R367 [FREE Full text] [doi: 10.1152/ajpregu.00349.2016] [Medline: 28052867]
4. Cheatham SW, Stull KR, Fantigrassi M, Motel I. The efficacy of wearable activity tracking technology as part of a weight loss program: a systematic review. J Sports Med Phys Fitness 2018 Apr;58(4):534-548. [doi: 10.23736/S0022-4707.17.07437-0] [Medline: 28488834]
5. Evenson KR, Goto MM, Furberg RD. Systematic review of the validity and reliability of consumer-wearable activity trackers. Int J Behav Nutr Phys Act 2015 Dec 18;12:159 [FREE Full text] [doi: 10.1186/s12966-015-0314-1] [Medline: 26684758]
6. Fuller D, Colwell E, Low J, Orychock K, Tobin MA, Simango B, et al. Reliability and validity of commercially available wearable devices for measuring steps, energy expenditure, and heart rate: systematic review. JMIR Mhealth Uhealth 2020 Sep 08;8(9):e18694 [FREE Full text] [doi: 10.2196/18694] [Medline: 32897239]
7. Shin G, Jarrahi M, Fei Y, Karami A, Gafinowitz N, Byun A, et al. Wearable activity trackers, accuracy, adoption, acceptance and health impact: a systematic literature review. J Biomed Inform 2019 May;93:103153 [FREE Full text] [doi: 10.1016/j.jbi.2019.103153] [Medline: 30910623]
8. Dixon PM, Saint-Maurice PF, Kim Y, Hibbing P, Bai Y, Welk GJ. A primer on the use of equivalence testing for evaluating measurement agreement. Med Sci Sports Exerc 2018 Apr;50(4):837-845 [FREE Full text] [doi: 10.1249/MSS.0000000000001481] [Medline: 29135817]
9. Düking P, Fuss FK, Holmberg HC, Sperlich B. Recommendations for assessment of the reliability, sensitivity, and validity of data provided by wearable sensors designed for monitoring physical activity. JMIR Mhealth Uhealth 2018 Apr 30;6(4):e102 [FREE Full text] [doi: 10.2196/mhealth.9341] [Medline: 29712629]
10. Ridgers ND, McNarry MA, Mackintosh KA. Feasibility and effectiveness of using wearable activity trackers in youth: a systematic review. JMIR Mhealth Uhealth 2016 Nov 23;4(4):e129 [FREE Full text] [doi: 10.2196/mhealth.6540] [Medline: 27881359]
11. Feehan LM, Geldman J, Sayre EC, Park C, Ezzat AM, Yoo JY, et al. Accuracy of Fitbit devices: systematic review and narrative syntheses of quantitative data. JMIR Mhealth Uhealth 2018 Aug 09;6(8):e10527 [FREE Full text] [doi: 10.2196/10527] [Medline: 30093371]
12. WHO 2020 guidelines on physical activity and sedentary behaviour. World Health Organization. Geneva, Switzerland: World Health Organization; 2020 Nov 25. URL: https://www.who.int/publications/i/item/9789240015128 [accessed 2021-12-05]
13. WHO guidelines on physical activity and sedentary behaviour: web annex: evidence profiles. World Health Organization. Geneva, Switzerland: World Health Organization; 2020. URL: https://apps.who.int/iris/handle/10665/336657 [accessed 2021-12-05]
14. Janssen X, Cliff DP. Issues related to measuring and interpreting objectively measured sedentary behavior data. Meas Phys Educ Exerc Sci 2015 Aug 19;19(3):116-124 [FREE Full text] [doi: 10.1080/1091367x.2015.1045908]
15. Phillips SM, Summerbell C, Hobbs M, Hesketh KR, Saxena S, Muir C, et al. A systematic review of the validity, reliability, and feasibility of measurement tools used to assess the physical activity and sedentary behaviour of pre-school aged children. Int J Behav Nutr Phys Act 2021 Nov 04;18(1):141 [FREE Full text] [doi: 10.1186/s12966-021-01132-9] [Medline: 34732219]
16. Lynch BA, Kaufman T, Rajjo TI, Mohammed K, Kumar S, Murad M, et al. Accuracy of accelerometers for measuring physical activity and levels of sedentary behavior in children: a systematic review. J Prim Care Community Health 2019;10:2150132719874252 [FREE Full text] [doi: 10.1177/2150132719874252] [Medline: 31509061]
17. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:n71 [FREE Full text] [doi: 10.1136/bmj.n71] [Medline: 33782057]
18. Phillips V, Barker E. Systematic reviews: structure, form and content. J Perioper Pract 2021 Sep;31(9):349-353 [FREE Full text] [doi: 10.1177/1750458921994693] [Medline: 34228554]
19. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a Web and mobile app for systematic reviews. Syst Rev 2016 Dec 05;5(1):210 [FREE Full text] [doi: 10.1186/s13643-016-0384-4] [Medline: 27919275]
20. Byun W, Kim Y, Brusseau T. The use of a Fitbit device for assessing physical activity and sedentary behavior in preschoolers. J Pediatr 2018 Aug;199:35-40. [doi: 10.1016/j.jpeds.2018.03.057] [Medline: 29754862]
21. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. Qual Life Res 2010 May;19(4):539-549 [FREE Full text] [doi: 10.1007/s11136-010-9606-8] [Medline: 20169472]

22. Terwee CB, Mokkink LB, Knol DL, Ostelo RW, Bouter LM, de Vet HC. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. Qual Life Res 2012 May;21(4):651-657 [FREE Full text] [doi: 10.1007/s11136-011-9960-1] [Medline: 21732199]
23. R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing. 2021. URL: https://www.r-project.org/ [accessed 2022-04-03]
24. Balduzzi S, Rücker G, Schwarzer G. How to perform a meta-analysis with R: a practical tutorial. Evid Based Ment Health 2019 Nov;22(4):153-160. [doi: 10.1136/ebmental-2019-300117] [Medline: 31563865]
25. Viechtbauer W. Conducting meta-analyses in R with the metafor Package. J Stat Softw 2010;36(3):1-48 [FREE Full text]
26. Lakens D. Equivalence tests: a practical primer for t tests, correlations, and meta-analyses. Soc Psychol Personal Sci 2017 May;8(4):355-362 [FREE Full text] [doi: 10.1177/1948550617697177] [Medline: 28736600]
27. gridExtra: Miscellaneous Functions for "Grid" Graphics – R package version 2.3. The Comprehensive R Archive Network. 2017 Sep 9. URL: https://cran.r-project.org/web/packages/gridExtra/index.html [accessed 2022-01-05]
28. Harrer M, Cuijpers P, Furukawa T, Ebert DD. dmetar: Companion R Package For The Guide 'Doing Meta-Analysis in R'. R package version 0.0.9000. 2019. URL: http://dmetar.protectlab.org/ [accessed 2022-01-05]
29. Wickham H. ggplot2: Elegant Graphics for Data Analysis. New York, NY, USA: Springer; 2009.
30. Bai Y, Tompkins C, Gell N, Dione D, Zhang T, Byun W. Comprehensive comparison of Apple Watch and Fitbit monitors in a free-living setting. PLoS One 2021 May 26;16(5):e0251975 [FREE Full text] [doi: 10.1371/journal.pone.0251975] [Medline: 34038458]
31. Bai Y, Hibbing P, Mantis C, Welk GJ. Comparative evaluation of heart rate-based monitors: Apple Watch vs Fitbit Charge HR. J Sports Sci 2018 Aug;36(15):1734-1741. [doi: 10.1080/02640414.2017.1412235] [Medline: 29210326]
32. Bai Y, Welk GJ, Nam YH, Lee JA, Lee JM, Kim Y, et al. Comparison of consumer and research monitors under semistructured settings. Med Sci Sports Exerc 2016 Jan;48(1):151-158. [doi: 10.1249/MSS.0000000000000727] [Medline: 26154336]
33. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ 2003 Sep 06;327(7414):557-560 [FREE Full text] [doi: 10.1136/bmj.327.7414.557] [Medline: 12958120]
34. Alba AC, Alexander PE, Chang J, MacIsaac J, DeFry S, Guyatt GH. High statistical heterogeneity is more frequent in meta-analysis of continuous than binary outcomes. J Clin Epidemiol 2016 Feb;70:129-135. [doi: 10.1016/j.jclinepi.2015.09.005] [Medline: 26386323]
35. Viechtbauer W, Cheung MW. Outlier and influence diagnostics for meta-analysis. Res Synth Methods 2010 Apr;1(2):112-125. [doi: 10.1002/jrsm.11] [Medline: 26061377]
36. Baujat B, Mahé C, Pignon JP, Hill C. A graphical method for exploring heterogeneity in meta-analyses: application to a meta-analysis of 65 trials. Stat Med 2002 Sep 30;21(18):2641-2652. [doi: 10.1002/sim.1221] [Medline: 12228882]
37. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. BMJ 1997 Sep 13;315(7109):629-634 [FREE Full text] [doi: 10.1136/bmj.315.7109.629] [Medline: 9310563]
38. Scott J, Grierson A, Gehue L, Kallestad H, MacMillan I, Hickie I. Can consumer grade activity devices replace research grade actiwatches in youth mental health settings? Sleep Biol Rhythms 2019 Jan 17;17(2):223-232 [FREE Full text] [doi: 10.1007/s41105-018-00204-x]
39. Mooses K, Oja M, Reisberg S, Vilo J, Kull M. Validating Fitbit Zip for monitoring physical activity of children in school: a cross-sectional study. BMC Public Health 2018 Jul 11;18(1):858 [FREE Full text] [doi: 10.1186/s12889-018-5752-7] [Medline: 29996797]
40. Kim Y, Lochbaum M. Comparison of polar active watch and waist- and wrist-worn ActiGraph accelerometers for measuring children's physical activity levels during unstructured afterschool programs. Int J Environ Res Public Health 2018 Oct 16;15(10):2268 [FREE Full text] [doi: 10.3390/ijerph15102268] [Medline: 30332785]
41. Byun W, Lee JM, Kim Y, Brusseau TA. Classification accuracy of a wearable activity tracker for assessing sedentary behavior and physical activity in 3-5-year-old children. Int J Environ Res Public Health 2018 Mar 26;15(4):594 [FREE Full text] [doi: 10.3390/ijerph15040594] [Medline: 29587432]
42. Kang S, Kim Y, Byun W, Suk J, Lee JM. Comparison of a wearable tracker with Actigraph for classifying physical activity intensity and heart rate in children. Int J Environ Res Public Health 2019 Jul 25;16(15):2663 [FREE Full text] [doi: 10.3390/ijerph16152663] [Medline: 31349667]
43. Godino JG, Wing D, de Zambotti M, Baker FC, Bagot K, Inkelis S, et al. Performance of a commercial multi-sensor wearable (Fitbit Charge HR) in measuring physical activity and sleep in healthy children. PLoS One 2020 Sep 4;15(9):e0237719 [FREE Full text] [doi: 10.1371/journal.pone.0237719] [Medline: 32886714]
44. Sirard JR, Masteller B, Freedson PS, Mendoza A, Hickey A. Youth oriented activity trackers: comprehensive laboratory- and field-based validation. J Med Internet Res 2017 Jul 19;19(7):e250 [FREE Full text] [doi: 10.2196/jmir.6360] [Medline: 28724509]
45. Sterne JA, Sutton AJ, Ioannidis JP, Terrin N, Jones DR, Lau J, et al. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. BMJ 2011 Jul 22;343:d4002. [doi: 10.1136/bmj.d4002] [Medline: 21784880]

46. Baranowski T, Simons-Morton BG. Dietary and physical activity assessment in school-aged children: measurement issues. J Sch Health 1991 May;61(5):195-197. [doi: 10.1111/j.1746-1561.1991.tb06009.x] [Medline: 1943040]
47. Welk GJ, Corbin CB, Dale D. Measurement issues in the assessment of physical activity in children. Res Q Exerc Sport 2000 Jun;71(2 Suppl):S59-S73. [Medline: 10925827]
48. Aunger J, Wagnild J. Objective and subjective measurement of sedentary behavior in human adults: a toolkit. Am J Hum Biol 2022 Jan;34(1):e23546. [doi: 10.1002/ajhb.23546] [Medline: 33277954]
49. Trost SG, Pate RR, Freedson PS, Sallis JF, Taylor WC. Using objective physical activity measures with youth: how many days of monitoring are needed? Med Sci Sports Exerc 2000 Feb;32(2):426-431. [doi: 10.1097/00005768-200002000-00025] [Medline: 10694127]
50. Kang M, Rowe DA. Issues and challenges in sedentary behavior measurement. Meas Phys Educ Exerc Sci 2015 Aug 19;19(3):105-115 [FREE Full text] [doi: 10.1080/1091367x.2015.1055566]
51. Fitbit kills Alta, Alta HR, Flex 2, and Zip. VentureBeat. 2019 Mar 6. URL: https://venturebeat.com/2019/03/06/fitbit-kills-alta-alta-hr-flex-2-and-zip/ [accessed 2021-11-14]
52. Carson V, Hunter S, Kuzik N, Gray CE, Poitras VJ, Chaput JP, et al. Systematic review of sedentary behaviour and health indicators in school-aged children and youth: an update. Appl Physiol Nutr Metab 2016 Jun;41(6 Suppl 3):S240-S265 [FREE Full text] [doi: 10.1139/apnm-2015-0630] [Medline: 27306432]
53. Welk GJ, Bai Y, Lee JM, Godino J, Saint-Maurice PF, Carr L. Standardizing analytic methods and reporting in activity monitor validation studies. Med Sci Sports Exerc 2019 Aug;51(8):1767-1780 [FREE Full text] [doi: 10.1249/MSS.0000000000001966] [Medline: 30913159]
54. Byun W, Beets MW, Pate RR. Sedentary behavior in preschoolers: how many days of accelerometer monitoring is needed? Int J Environ Res Public Health 2015 Oct 20;12(10):13148-13161 [FREE Full text] [doi: 10.3390/ijerph121013148] [Medline: 26492261]
55. Henriksen A, Haugen Mikalsen M, Woldaregay AZ, Muzny M, Hartvigsen G, Hopstock LA, et al. Using fitness trackers and smartwatches to measure physical activity in research: analysis of consumer wrist-worn wearables. J Med Internet Res 2018 Mar 22;20(3):e110 [FREE Full text] [doi: 10.2196/jmir.9157] [Medline: 29567635]
56. McCallum C, Rooksby J, Gray CM. Evaluating the impact of physical activity apps and wearables: interdisciplinary review. JMIR Mhealth Uhealth 2018 Mar 23;6(3):e58 [FREE Full text] [doi: 10.2196/mhealth.9054] [Medline: 29572200]

Abbreviations

COSMIN: Consensus-Based Standards for the Selection of Health Status Measurement Instruments
CV: coefficient of variation
EZ: equivalence zone
MAPE: mean absolute percentage error
MET: metabolic equivalent of task
PA: physical activity
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
SB: sedentary behavior
WAM: wearable activity monitor

Edited by L Buis; submitted 24.02.22; peer-reviewed by F Velayati, H Mehdizadeh; comments to author 02.06.22; revised version received 06.06.22; accepted 10.06.22; published 09.08.22.

Please cite as:
Martinko A, Karuc J, Jurić P, Podnar H, Sorić M
Accuracy and Precision of Consumer-Grade Wearable Activity Monitors for Assessing Time Spent in Sedentary Behavior in Children and Adolescents: Systematic Review
JMIR Mhealth Uhealth 2022;10(8):e37547
URL: https://mhealth.jmir.org/2022/8/e37547
doi:10.2196/37547
PMID:35943763

Review

Wearables for Measuring the Physical Activity and Sedentary Behavior of Patients With Axial Spondyloarthritis: Systematic Review

Julie Soulard[1,2,3], PT, PhD; Thomas Carlin[1,2], MSc; Johannes Knitza[1,2,4,5], MD; Nicolas Vuillerme[1,2,6], PhD

[1]Université Grenoble Alpes, AGEIS, La Tronche, France

[2]LabCom Telecom4Health, Orange Labs & Université Grenoble Alpes, CNRS, Inria, Grenoble INP-UGA, Grenoble, France

[3]Grenoble Alpes University Hospital, Grenoble, France

[4]Department of Internal Medicine 3–Rheumatology and Immunology, Friedrich-Alexander University, Erlangen-Nürnberg, and Universitätsklinikum Erlangen, Erlangen, Germany

[5]Deutsches Zentrum für Immuntherapie, Friedrich-Alexander University Erlangen-Nürnberg and Universitätsklinikum Erlangen, Erlangen, Germany

[6]Institut Universitaire de France, Paris, France

Corresponding Author:
Julie Soulard, PT, PhD
Université Grenoble Alpes, AGEIS
Batiment Jean Roget, Faculté de médecine
La Tronche, 38700
France
Phone: 33 476637104
Email: juliesoulard.physio@gmail.com

Abstract

Background: Axial spondyloarthritis (axSpA) is an inflammatory rheumatic disease associated with chronic back pain and restricted mobility and physical function. Increasing physical activity is a viable strategy for improving the health and quality of life of patients with axSpA. Thus, quantifying physical activity and sedentary behavior in this population is relevant to clinical outcomes and disease management. However, to the best of our knowledge, no systematic review to date has identified and synthesized the available evidence on the use of wearable devices to objectively measure the physical activity or sedentary behavior of patients with axSpA.

Objective: This study aimed to review the literature on the use of wearable activity trackers as outcome measures for physical activity and sedentary behavior in patients with axSpA.

Methods: PubMed, PEDro, and Cochrane electronic databases were searched in July 2021 for relevant original articles, with no limits on publication dates. Studies were included if they were original articles, targeted adults with a diagnosis of axSpA, and reported wearable device–measured physical activity or sedentary behavior among patients with axSpA. Data regarding the study's characteristics, the sample description, the methods used for measuring physical activity and sedentary behavior (eg, wearable devices, assessment methods, and outcomes), and the main results of the physical activity and sedentary behavior assessments were extracted.

Results: A total of 31 studies were initially identified; 13 (13/31, 42%) met the inclusion criteria, including 819 patients with axSpA. All the studies used accelerometer-based wearable devices to assess physical activity. Of the 13 studies, 4 (4/31, 31%) studies also reported outcomes related to sedentary behavior. Wearable devices were secured on the wrists (3/13 studies, 23%), lower back (3/13, 23%), right hip (3/13, 23%), waist (2/13, 15%), anterior thigh (1/13, 8%), or right arm (1/13, 8%). The methods for reporting physical activity and sedentary behavior were heterogeneous. Approximately 77% (10/13) of studies had a monitoring period of 1 week, including weekend days.

Conclusions: To date, few studies have used wearable devices to quantify the physical activity and sedentary behavior of patients with axSpA. The methodologies and results were heterogeneous, and none of these studies assessed the psychometric properties of these wearables in this specific population. Further investigation in this direction is needed before using wearable device–measured physical activity and sedentary behavior as outcome measures in intervention studies in patients with axSpA.

Trial Registration: PROSPERO CRD42020182398; https://tinyurl.com/ec22jzkt

International Registered Report Identifier (IRRID): RR2-10.2196/23359

(JMIR Mhealth Uhealth 2022;10(8):e34734) doi:10.2196/34734

KEYWORDS

axial spondyloarthritis; rheumatology; physical activity; sedentary behavior; objective measures; wearable; mobile health; mHealth; eHealth; systematic review; mobile phone

Introduction

Axial spondyloarthritis (axSpA) is a chronic inflammatory rheumatic disease that can cause inflammatory back pain, structural damage, and disability [1]. New therapeutic agents allow for effective therapy [2]. Physical activity should be an integral part of standard care, according to current guidelines [3], because of its multiple health benefits, including pain reduction [4,5], increased mobility [4-6], physical function [5-8], and cardiorespiratory fitness [9], which ultimately reduces disease activity [4-6,8,10]. Furthermore, increasing physical activity represents a viable strategy for improving quality of life [7,11] and reducing the psychological comorbidities of patients with axSpA [7]. Physical activity is a predictor of mortality and cardiovascular events in the general population [12]. In addition, high-intensity exercises (12 weeks of endurance and strength exercises) have been reported to significantly improve disease activity and reduce cardiovascular risks in patients with axSpA [13]. However, a good therapeutic response depends on short symptom duration and close disease monitoring [14]. A recent review reported that most measures of physical activity used in patients with axSpA were "subjective and limited by patient recall, reporting bias, and relatively short study intervals" [15].

Wearable technology comprises "a device fitted to the participant's body which detects and collects data" [16]. These wearables can include accelerometers, pedometers, or inertial measurement units, which are small and transportable. They can be advantageously used to monitor physical activity data under real-world conditions in various chronic populations [16-18], including patients with axSpA [11,19-28]. By allowing longitudinal physical activity monitoring, these devices can remotely monitor the disease and evolution of health status in patients with axSpA [20,29]. Indeed, greater disease activity is associated with lower levels of physical activity in axSpA and could help detect flares [20,29].

In addition to measuring physical activity, in a complementary manner, wearables can measure the time spent sitting, also called sedentary behavior. Indeed, it is of particular interest in patients with axSpA as the more time patients with axSpA spend sitting, the greater the association with disease activity [27], decreased physical function [27], and decreased quality of life [30].

Thus, physical activity and sedentary behavior assessments using wearable devices represent an attractive and feasible health monitoring option for patients with axSpA. Interestingly, a multicentric prospective observational study, which involved 157 patients with chronic inflammatory rheumatic diseases, found good acceptability of wearing activity trackers for physical activity assessment in this population [31].

However, to the best of our knowledge, no study has assessed the use of wearables to objectively monitor physical activity and sedentary behavior in patients with axSpA. We designed the present review to identify and synthesize the currently available evidence on the use of wearable activity trackers as outcome measures for physical activity and sedentary behavior in patients with axSpA [32]. We aimed to answer the following research question: which wearable devices, assessment methods, and associated outcomes are commonly used to quantify physical activity or sedentary behavior among patients with axSpA?

Methods

This systematic review's protocol was developed based on the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) statement. It was registered in PROSPERO (International Prospective Register of Systematic Reviews; CRD42020182398) and was published in November 2021 [32].

Inclusion Criteria

Studies were included if they (1) were original articles published in English-language peer-reviewed journals, (2) targeted adults (aged ≥18 years) with a diagnosis of axSpA, and (3) reported wearable device–measured physical activity or sedentary behavior among patients with axSpA.

Studies were excluded if they (1) were case reports, abstracts, editorials, conference abstracts, letters to the editor, reviews, or meta-analyses or (2) did not use wearable devices to quantify the physical activity or sedentary behavior of patients with axSpA.

Data Sources and Search Strategy

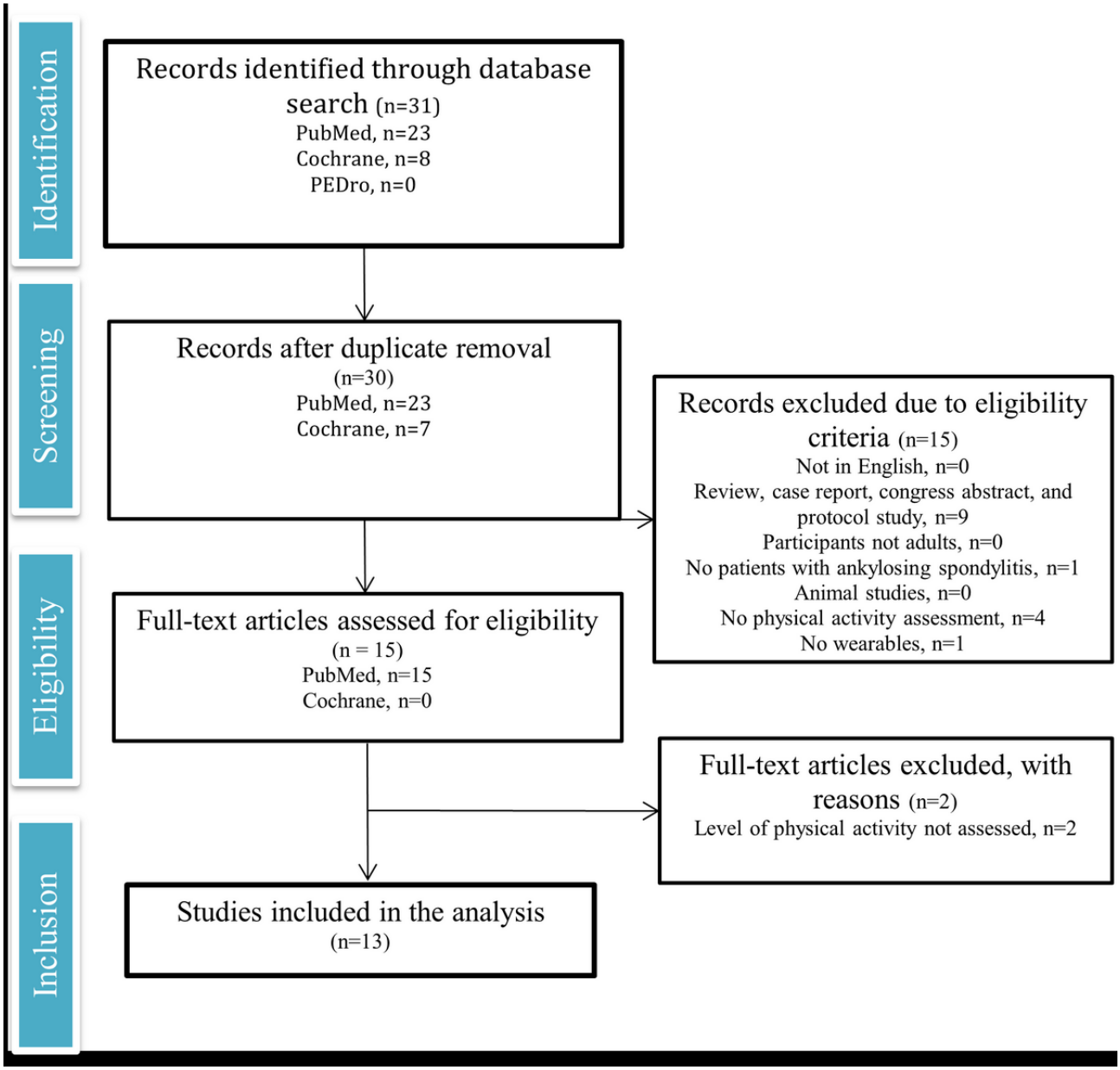
In July 2021, we conducted searches with no date restrictions in 3 electronic bibliographic databases (PubMed, PEDro, and Cochrane). The Boolean operators *AND* and *OR* were used to combine keywords relevant to the population, wearable devices, and the outcomes of physical activity or sedentary behavior and were searched in all fields. The detailed search strategy is presented in the review protocol recently published in *JMIR Research Protocols* [32].

Study Selection

A total of 2 independent reviewers (TC and JS) screened the titles, abstracts, and keywords of all the studies found in the search to identify potentially relevant articles. Duplicates were manually removed. The selected full-length text articles were then screened for eligibility according to the criteria abovementioned. In cases of disagreement, the reviewers reached a consensus through discussion. If their disagreement persisted, a third reviewer (NV) was asked to make the final

decision. In accordance with the PRISMA guidelines [33], a flow chart was constructed to summarize each step of the selection process with its corresponding number of citations (Figure 1).

Figure 1. PRISMA (Preferred Reporting Item for Systematic Reviews and Meta-Analyses) flowchart of the selection process.



Risk of Bias in Individual Studies

As indicated in the published review protocol [32], as the purpose of this review was not to assess the clinical effects of interventions, we did not perform a risk of bias assessment [34-36]. Indeed, as mentioned in the *Introduction* section, this review was designed to identify and synthesize the available evidence on the use of wearable devices to quantify physical activity or sedentary behavior among patients with axSpA.

Data Extraction

Data extraction was performed independently by 2 reviewers (TC and JS) who were not blinded to the authors or journals. Information was extracted on (1) the study's characteristics, (2) the sample description, (3) the methods used for measuring physical activity and sedentary behavior (eg, wearable devices, assessment methods, and outcomes), and (4) the main results of the physical activity and sedentary behavior assessments.

Data Synthesis and Analysis

Owing to the significant heterogeneity of data types, we decided to perform only a narrative synthesis [37,38]. As per the data extraction strategy, the tables and figures found in this review only summarize the available information on wearable devices used to objectively assess the physical activity and sedentary behavior of patients with axSpA.

Results

Study Selection

The study selection process is illustrated in Figure 1. A preliminary search of the 3 electronic bibliographic databases

identified 31 citations. A duplicate was removed, leaving 97% (30/31) of unique records for preliminary screening, focusing on the title, abstract, and keywords. Finally, of the 30 articles, 13 (43%) met our eligibility criteria and were included in this review.

Study Characteristics

The characteristics of the studies included (N=13) are outlined in Table 1.

Table 1. Study characteristics (N=13).

First author	Year of publication	Country	Study design	Outcomes of interest
Arends et al [11]	2013	The Netherlands	Observational validation study	Physical activity
Bayraktar et al [39]	2021	Turkey	Cross-sectional study	Physical activity
Carbo et al [40]	2021	The Netherlands	Part of a prospective, longitudinal, observational cohort study	Physical activity
Coulter et al [30]	2019	United Kingdom	Prospective cohort study	Physical activity and sedentary behavior
Gossec et al [20]	2019	France	Prospective, multicenter, longitudinal, observational study	Physical activity and sedentary behavior
Jacquemin et al [29]	2017	France	Prospective, multicenter, longitudinal, observational study	Physical activity
Jacquemin et al [31]	2018	France	Prospective, multicenter, longitudinal, observational study	Physical activity
O'Dwyer et al [27]	2015	Ireland	Cross-sectional controlled study	Physical activity and sedentary behavior
Plasqui et al [28]	2012	The Netherlands	Observational case-control study	Physical activity
Swinnen et al [21]	2014	Belgium	Observational cross-sectional controlled study	Physical activity and sedentary behavior
van Genderen et al [41]	2015	The Netherlands	Multicenter cross-sectional study	Physical activity
van Genderen et al [42]	2014	The Netherlands	Cross-sectional case-control study	Physical activity
Yuksel et al [43]	2021	Turkey	Observational cross-sectional controlled study	Physical activity

Most studies were conducted in Europe (11/13, 85%), namely, the Netherlands (5/13, 38%) [11,28,40-42], France (3/13, 23%) [20,29,31], the United Kingdom (1/13, 8%) [30], Ireland (1/13, 8%) [27], and Belgium (1/13, 8%) [21]. Approximately 15% (2/13) of studies were conducted in Turkey [39,43]. Designs of the 13 studies included 1 (8%) observational validation study [11], 1 (8%) reproducibility study [40], 4 (31%) longitudinal studies [20,29-31], and 7 (54%) cross-sectional studies [21,27,28,39,41-43]. None of the studies reported interventions for the levels of physical activity or lifestyle. All studies (13/13, 100%) [11,20,21,27-31,39-43] focused on assessing physical activity and reported at least one corresponding outcome, whereas some studies (4/13, 31%) also reported outcomes related to sedentary behavior in patients with axSpA [20,21,27,30].

Sample Characteristics

Table 2 presents participants' descriptive characteristics. The 13 studies included covered a total of 819 patients with axSpA, of whom 490 (59.8%) were male. The mean sample size was 63 (SD 33.1), ranging from 24 [42] to 135 [41] participants with axSpA. The mean patient age was 44.43 (SD 4.4) years. Disease diagnoses were based on the Assessment of Spondyloarthritis international Society recommendations (6/13, 46%) [20,29-31,39,40], the modified New York criteria (4/13, 31%) [27,28,41,42], or both (2/13, 15%) [11,40]. Approximately 8% (1/13) of studies used the European Spondyloarthropathy Study Group recommendations for disease diagnosis [21]. Mean or median disease duration ranged from 4 [39] to 20.5 years [42].

All the studies used the Bath Ankylosing Spondylitis Disease Activity Index (13/13, 100%) [11,20,21,27-31,39-43], with mean or median scores ranging from 3.1 [20] to 4.5 [30]. Approximately 85% (11/13) studies used the Bath Ankylosing Spondylitis Functional Index [11,21,27,28,30,31,39-43], with mean or median scores ranging from 1.7 [31] to 4.4 [30]. Approximately 31% (4/13) of studies used the Bath Ankylosing Spondylitis Metrology Index [21,30,39,43], with mean scores ranging from 1.8 [39,43] to 3.05 [21].

Approximately 54% (7/13) studies included a healthy control group [21,27,28,31,41-43], including 281 healthy participants, of whom 169 (60.1%) were male. The mean sample size was 40.14 (SD 27.2), ranging from 24 [42] to 99 [41] control participants. The mean participant age across the healthy control groups was 42.56 (SD 16.8) years.

Table 2. Characteristics of patients with axSpA[a].

First author	Participants, N	Males, n (%)	Age (years), mean (SD) or median (IQR)[b]	BMI, mean (SD) or median (IQR)[b]	Duration (years), mean (SD) or median (IQR)[b]	Criteria for diagnosis	BASFI[c] score, mean (SD) or median (IQR)[b]	BASDAI[d] score, mean (SD) or median (IQR)[b]	BASMI[e] score, mean (SD) or median (IQR)[b]
Arends et al [11]	115	71 (62)	44.6 (12.1)	26.4 (4.4)	10.0 (0-42)[b]	ASAS[f]+NY[g]	3.8 (2.4)	3.7 (0-9)[b]	—[h]
Bayraktar et al [39]	58	32 (55)	39.0 (30.0-46.0)[b]	26.1 (23.7-28.7)[b]	4 (3-10)	ASAS	2.2 (0.5-3.8)[b]	2.8 (1.4-4.7)[b]	1.8 (1.1-2.95)[b]
Carbo et al [40]	45	23 (51)	50.7 (11.6)	—	27.0 (18-36)[b]	ASAS+NY	3.3 (1.4-5.7)[b]	3.4 (2.0-5.7)[b]	—
Coulter et al [30]	45	23 (46)	49.0 (11.7)	27.4 (5.6)	15.6 (11.2)	ASAS	4.4 (2.6)	4.5 (2.3)	3.6 (1.8)
Gossec et al [20]	73	41 (56.2)	41.2 (10.3)	24.6 (4.6)	10.8 (9.1)	ASAS	—	3.1 (2.0)	—
Jacquemin et al [29]	79	44 (55.7)	41.4 (10.2)	25.0 (4.6)	10.4 (8.9)	ASAS	—	3.3 (2.1)	—
Jacquemin et al [31]	74	43 (58.1)	41.3 (10.4)	25.3 (4.6)	10.4 (9.1)	ASAS	1.7 (1.8)	3.2 (2.1)	—
O'Dwyer et al [27]	39	32 (82.1)	40.0 (9.0)	28.6 (6.8)	6.0 (10.0)	NY	2.9 (3.8)	3.6 (2.2)	—
Plasqui et al [28]	25	15 (60)	48.0 (11.0)	26.2 (5.0)	19.0 (12.0)	NY	4.0 (2.2)	4.3 (2.2)	—
Swinnen et al [21]	40	24 (60)	44.38 (11.3)	26.3 (5.1)	11.4 (9.5)	ESSG[i]	3.52 (2.5)	3.7 (2.6)	3.1 (1.2)
van Genderen et al [42]	24	14 (58.3)	48.0 (11.0)	26.0 (4.6)[b]	20.5 (22.0)[b]	NY	3.8 (2.1)[b]	4.0 (3.7)[b]	—
van Genderen et al [41]	135	81 (80)	51.0 (13.0)	26 (4.3)	16.5 (12.1)	NY	4.1 (2.6)	4.3 (2.2)	—
Yuksel et al [43]									
AS[j]	34	47 (70.1)	41.0 (31-46)[b]	26.1 (22.9-29.6)[b]	8.0 (4-13)[b]	ASAS	2.4 (0.5-3.9)[b]	3.4 (1.5-5.8)[b]	2.1 (1.5-3.9)[b]
Nr-SpA[k]	33	47 (70.1)	37.0 (32-40)[e]	26.3 (25.4-28.7)[e]	4.0 (2-9)[e]	ASAS	1.2 (0.6-2.9)[e]	2.4 (1.4-5.4)[e]	1.5 (1.1-2.0)[e]

[a]axSpA: axial spondyloarthritis.

[b]Outcomes are reported with median (IQR) values.

[c]BASFI: Bath Ankylosing Spondylitis Functional Index.

[d]BASDAI: Bath Ankylosing Spondylitis Disease Activity Index.

[e]BASMI: Bath Ankylosing Spondylitis Metrology Index.

[f]ASAS: Assessment of Spondyloarthritis international Society.

[g]NY: modified New York criteria.

[h]Not available

[i]ESSG: European Spondyloarthropathy Study Group.

[j]AS: ankylosing spondylitis.

[k]nr-SpA: nonradiologic form of axial spondyloarthritis.

Methods of Measuring Physical Activity and Sedentary Behavior

Table 3 presents the methods used to objectively assess physical activity or sedentary behavior among patients with axSpA. Information regarding wearable devices (eg, device name, manufacturer, and sensor), assessment methods (device location, length of monitoring, requisite conditions for valid monitoring, visualization of physical activity by the participants, and instructions to the participants on physical activity), and outcomes (physical activity and sedentary behavior) are reported.

Table 3. Wearable device and monitoring characteristics.

First author	Device name	Manufacturer (country)	Sensor	Device location	Length of monitoring	Requisite conditions for valid monitoring	Visualization of physical activity by the participants	Instructions to the participants on physical activity	Outcomes reported
Arends et al [11]	The ActiGraph: GT1M	ActiGraph (United States)	Uniaxial accelerometer	Right hip	7 consecutive days	Minimum wear time of 10 hours per day and 5 days with both weekend day	No information	No information	Kilocounts per day and mean wear time
Bayraktar et al [39]	GT3X	ActiGraph (United States)	Triaxial accelerometer	Waist	7 consecutive days	Patients not wearing accelerometer as instructed (<10 hours per day wear; <7 days total wear) were removed from the analysis	No information	No information	Total activity duration and activity duration intensity (light or moderate or vigorous); METs[a] for total and for each physical activity intensities
Carbo et al [40]	GT3X	ActiGraph (United States)	Triaxial accelerometer	Right hip	7 consecutive days	Data excluded if accelerometer worn <10 hours per day, for <5 days or for <2 weekend days	No information	No information	Total activity kilocounts and activity duration intensity (light, moderate, and vigorous) in minutes per week
Coulter et al [30]	The activPAL3	PAL Technologies Ltd (Scotland)	Triaxial accelerometer	Anterior thigh of the dominant leg	7 consecutive days	Minimum wear time of 24 hours for a valid day	No information	No information	Daily standing, walking, sedentary time, and steps per day
Gossec et al [20]	Withings Activité Pop	Withings (France)	—[b]	Wrist	90 consecutive days	—	"patients could visualize their physical activity on their smartphones."	"No instruction about physical activity was given to the participants"	Steps per minute
Jacquemin et a. [29]	Withing Activité Pop	Withings (France)	—	Wrist	90 consecutive days	—	"patients could visualize their physical activity on their smartphones."	"No instruction about physical activity was given to the participants"	Steps per day, total activity duration, and activity duration in moderate to vigorous intensity
Jacquemin et al [31]	Withing Activité Pop	Withings (France)	—	Wrist	90 consecutive days	—	No information	"No intervention was specifically performed to increase physical activity, and no instruction about physical activity was given to the participants."	Steps per day, morning step count, total activity duration, and activity duration in moderate to vigorous intensity

First author	Device name	Manufacturer (country)	Sensor	Device location	Length of monitoring	Requisite conditions for valid monitoring	Visualization of physical activity by the participants	Instructions to the participants on physical activity	Outcomes reported
O'Dwyer et al [27]	RT3	Stayhealthy Inc (United States)	Triaxial accelerometer	Right hip	7 consecutive days	Minimum wear time of 10 hours per day, including at least one weekend day	No information	No information	Counts per day
Plasqui et al [28]	Tracmor	Philips Research (the Netherlands)	3 uniaxial piezoelectric accelerometers	Lower back	7 consecutive days	—	No information	No information	Kilocounts per day
Swinnen et al [21]	SenseWear Pro 3 Armband	Bodymedia Inc (United States)	Biaxial accelerometer	Right triceps muscle	5 consecutive days	Minimum wear time of 1296 minutes, corresponding to 90% of a 24-hour period, including both weekend days	No information	No information	Weekly average of kilocounts per day
van Genderen et al [42]	Tracmor	Philips Research (the Netherlands)	3 uniaxial piezoelectric accelerometers	Lower back	7 consecutive days	Minimum wear time of 10 hours for a valid day	No information	No information	Kilocounts per day and mean wear time
van Genderen et al [41]	GT3X	ActiGraph (United States)	Triaxial accelerometer	Lower back	7 consecutive days	Minimum wear time of 10 hours for a valid day	No information	No information	Vector magnitude counts, counts per day, and counts per minute
Yuksel et al [43]	GT3X	ActiGraph (United States)	Triaxial accelerometer	Waist	7 consecutive days	Not specified; however, participants "instructed to wear the device for at least 10 h/day except for water-related activities such as showering or swimming"	No information	No information	Activity duration (light, moderate, vigorous) in minutes per week

[a]MET: metabolic equivalent of task.

[b]Not available

All studies (13/13, 100%) [11,20,21,27-31,39-43] reported wearable device–measured physical activity outcomes, whereas some (4/13, 40%) studies reported wearable device–measured sedentary behavior outcomes among patients with axSpA [20,21,27,30]. Only 15% (2/13) of studies provided information on the visualization of physical activity levels by the participants [20,29], and 23% (3/13) of studies provided instructions to the participants on physical activity [20,29,31].

Types of Sensors

All 13 studies used accelerometer-based wearable devices, with 6 (46%) using triaxial accelerometers [27,30,39-41,43], 2 (15%) using 3 uniaxial piezoelectric accelerometers [28,42], 1 (8%) using a biaxial accelerometer [21], and 1 (8%) using a uniaxial accelerometer [11]. Approximately 23% (3/13) of studies did not mention the type of accelerometer [20,29,31]. The brands used were ActiGraph (5/13, 38%) [11,39-41,43], Withings (3/13, 23%) [20,29,31], Philips Research (2/13, 15%) [28,42], PAL Technologies (1/13, 8%) [30], Stayhealthy (1/13, 8%) [27], and Bodymedia (1/13, 8%) [21].

The Withings device was the Withing Activité Pop, an accelerometer-based activity tracker worn on the wrist. ActiGraph devices included the GT1M, a uniaxial accelerometer fixed on the participant's right hip, and the GT3X, a triaxial accelerometer placed on the lower back using a belt. The Tracmor sensor is a combination of 3 uniaxial piezoelectric accelerometers that are fixed to the lower back. PAL Technologies' activPAL3 is a triaxial accelerometer fixed to

the anterior thigh of a participant's dominant leg. Stayhealthy's RT3 is a triaxial accelerometer worn on the right hip. Bodymedia's SenseWear Pro 3 Armband is a biaxial accelerometer worn on the back of the right triceps muscle.

Wearable devices were secured on the wrist (3/13, 23%) [20,29,31], lower back (3/13, 23%) [28,41,42], right hip (3/13, 23%) [11,27,40], waist (2/13, 15%) [39,43], anterior thigh of the dominant leg (1/13, 8%) [30], and right arm (1/13, 8%) [21].

Monitoring Protocol

Approximately 69% (9/13) of studies used 1-week monitoring (7 consecutive days) [11,27,28,30,39-43], and 8% (1/13) used a 5-day period (including both weekend days) [21]. Approximately 23% (3/13) of studies used 3-month monitoring and follow-up [20,29,31]. None of the studies assessed only days of the week. Monitoring was considered complete when wearable devices were worn on both weekend days in 38% (5/13) of studies [11,21,39,40,43] or when 1 weekend day was included in the follow-up period in 8% (1/13) of studies [27]. Approximately 54% (7/13) of studies imposed wearing the tracker for at least 10 hours per day [11,27,39-43], 10% (1/13) imposed wearing trackers for at least 1296 minutes (corresponding to 90% of 24 hours) [21], and 10% (1/13) imposed a minimum wear time of 24 hours per day [30]. Approximately 23% (3/13) of studies reported wear time [11,40,42] and 23% (3/13) others reported activity duration or time spent on specific activities (walking and standing) [20,29,31].

Physical Activity Outcomes

Twelve objective measures of physical activity were used to assess patients with axSpA in the 13 studies: 4 (31%) studies reported steps per day [20,29-31], 6 (46%) reported activity counts [11,27,28,40-42], 4 (46%) reported total activity duration [29,31,39,41], 2 (15%) reported energy expenditure [21,39], 2 (15%) reported levels of physical activity [21,28], and 4 (31%) reported average wear time in hours per day [11,21,40,42]. Approximately 69% (9/13) of studies also reported activity intensity [21,27,29-31,39-41,43] using six different expressions of measurement: light, moderate, vigorous, and very vigorous levels of activity levels [21,27,39-41,43]; minutes spent in moderate to vigorous activity [21,27,29-31]; and mean walking-event cadence [30].

The number of steps per day was reported as the average daily step count [20,29-31]. Activity counts were reported as the average of daily activity counts [11,27,28,41,42]. Total activity duration was reported as the sum of minutes involving at least 20 steps recorded in a week [29] or as active minutes in a day, derived from either step count (minutes with at least 20 steps recorded) [31] or activity count [41]. Energy expenditure was expressed as the metabolic equivalent of task (MET) hours per day [21] or MET minutes per week [39]. METs were used to report the overall and objective levels of physical activity in 15% (2/13) of studies [21,28]. Activity intensity was obtained from activity counts [21,27,30,41] or number of steps [29,31]. Established cutoff points were used to convert raw data from daily activity counts [27,41] or MET values [21,30] into each activity intensity. The average time spent doing light, moderate, and vigorous activities was reported in minutes per day [21,27,41] or hours per day [27]. One of the studies reported the score for very vigorous activity and expressed it in minutes per day [21]. The time spent performing light to vigorous activities was expressed in minutes per day [21,27,31] or minutes per week [29,30,39,40,43]. Another study reported cadence using steps per minute [30].

Sedentary Behavior Outcomes

A total of 5 measures were used to assess sedentary behavior in patients with axSpA. These included (1) the number of sitting events (1/13, 8%) [30], (2) total sitting time (1/13, 8%) [30], (3) the number of bouts of prolonged sitting time (>30 minutes; 1/13, 8%) [30], (4) the total duration of this prolonged sitting time (1/13, 8%) [30], and (5) duration of sedentary behavior (3/13, 23%) [21,27,41].

Cutoff values based on daily activity counts [27,41] or MET values [21,30] were used to derive sedentary time from the raw data. Coulter et al [30] also reported sitting events per day, duration of sitting time in hours per day, and the number of periods of prolonged sitting time (>30 minutes).

Discussion

To the best of our knowledge, this is the first systematic review to identify and synthesize available evidence on the use of wearable activity trackers as outcome measures for physical activity and sedentary behavior in patients with axSpA. For the sake of clarity, we discuss our findings through three main themes: (1) the wearable devices themselves, (2) reference outcomes for physical activity and sedentary behavior assessment using wearable devices, and (3) monitoring protocols and assessment methods.

Wearable Devices

Our findings showed the broad use of wearable devices, mostly incorporating triaxial accelerometers [27,28,30,39-43], and less use of simple devices such as pedometers [20,29,31] or uniaxial or biaxial accelerometers [11,21]. Among the studies included in our synthesis of directly comparable data (Table 2), accelerometers were the most frequently used direct measuring devices.

To implement these wearables in clinical practice, measurements should be both feasible (ie, used by patients and health professionals) and accurate (ie, validity and reliability) [44-46].

Monitoring of health and physical activity seems feasible in patients with axSpA. Indeed, in a recent study by Jacquemin et al [31], 157 patients reported that wearing a wristwatch-type device for 3 months was acceptable, with a mean acceptability score of 8 out of 10 [31]. However, the interpretation of the data provided by these devices requires digital health skills that not every patient with axSpA may have. The implementation of wearables in clinical practice also necessitates the formation and training of health professionals supporting patients with the use of wearable activity trackers. None of the included studies have addressed this issue.

Depending on their purpose (ie, with specific conditions of use), the validity of activity trackers can vary significantly, making

them more or less suitable for research purposes. Some activity trackers are specifically designed for research purposes (*research activity trackers*), with relatively short-term use, fewer needs for the interface with the users, and more precise and detailed parameters. This is the case for the ActiGraph [47,48], Philips Research [48], and ActivPAL [48,49] sensors, which have been widely validated against doubly labeled water in healthy control populations and presented a high degree of accuracy [47-49]. Other activity trackers (such as Fitbit or Withings devices) are primarily designed for consumers to monitor and improve their physical activity levels, are easy to wear, and are adapted for long-term use. At this point, it is important to note that previous studies have reported that the validity of these devices is lower than that of research activity trackers [50-53], the estimation of energy expenditure was outside the acceptable accuracy [54-57], and the availability of raw data is not always warranted.

Interestingly, we found no published studies assessing the psychometric properties of wearable devices to monitor physical activity and sedentary behavior in our specific population, neither in free-living conditions nor in a more standardized environment, such as a laboratory. Therefore, it would be appropriate to conduct studies addressing the metrological properties of these devices in this population. Importantly, patients with axSpA seem to present with motor and gait specificities [36,58] that could affect the validity of wearables designed to monitor walking activity.

The positioning of the wearable devices on the body should also be considered. The 3 main locations used were the wrist [20,29,31], lower back [28,41,42], and hip [11,27,40]. In the general population, previous studies have reported that where devices are placed on the body has a significant impact on the accuracy of the number of steps counted during various walking activities [44,46,59,60]. The major trend reported in these studies was the outperformance of hip-worn devices. For example, in the laboratory-based validation protocol described by Kooiman et al [46], a hip-worn activity tracker (Fitbit Zip, Fitbit Inc) had the highest validity and reliability in counting healthy participants' steps. Hip-mounted devices were also the best for counting steps at the preferred walking speeds of healthy individuals, with a lower absolute mean relative error [60]. However, when walking speeds decreased, the wrist-worn devices in the same study provided more accurate step count estimations than the hip-worn devices [60].

Previous studies have shown that physical activity and sedentary behavior can be modified in patients with axSpA compared with healthy controls. Indeed, if no significant differences were found between patients with axSpA and healthy controls in light physical activity [21,27,41,43], counts per day [27,28,41,42], or duration of sedentary behavior [21,27,41], 31% (4/13) of studies found that patients with axSpA performed significantly less vigorous activity [21,27,41,43]. Results regarding durations of moderate or moderate to vigorous activity were inconsistent (ie, some studies found significantly less moderate physical activity [41,43] or moderate to vigorous activity [21] in patients with axSpA, whereas others did not [21,27]).

Moreover, considering the symptoms caused by spondyloarthritis, such as limitations in the sagittal range of motion of the lower limbs during gait [61,62] and lower gait speed [63], previous results in the literature regarding wearable devices in axSpA may be questionable. Future studies on how the location of wearable devices affects the overall accuracy of measurements of physical activity and sedentary behavior among this specific population are needed, particularly regarding the gait specificities of patients with axSpA [36,58].

Furthermore, only 15% (2/13) of studies mentioned that participants could visualize physical activity [20,29], and 23% (3/13) of studies mentioned that they provided no instructions on physical activity [20,29,31]. Wearing a wearable tracker and visualizing physical activity can increase the physical activity of participants [64]. Thus, future studies on wearables and physical activity should include information on visualization and instructions on physical activity.

Reference Outcomes for Physical Activity and Sedentary Behavior Assessment Using Wearable Devices

The studies included in this systematic review reported two main categories of outcomes related to physical activity: first, the number of daily steps taken [20,29-31], and second, data based on the daily activity counts recorded by accelerometers [11,21,27,28,39-43]. These 2 types of data allowed us to estimate the time spent doing activities of different intensities (ie, light, moderate, or vigorous). In other words, using threshold values, it is possible to estimate the intensity of an activity based on the number of steps or activity counts.

In contrast, even if recording the number of steps per day (4/13, 31% studies) [20,29-31] requires fewer raw data than recording the activity count, parameters related to the activity count were used slightly more for tracking individual behavior (9/13, 69% studies) [21,27,29-31,39-41,43]. Indeed, monitoring techniques based on activity count data require devices with greater memory and storage capacity; however, recent technological advances have made data compression and storage problems almost irrelevant. There is extensive literature related to activity count cutoff points [65-71] but not to step cutoff points, which could explain why activity counts were more common in the present review. Further investigation in a laboratory-type setting is needed to draw firm conclusions on the pros and cons of each measurement method in the axSpA population. It would also be appropriate to determine specific cutoff points for this population for both step and activity counts.

Using thresholds allowed us to categorize low, moderate, vigorous, and very vigorous activities. However, in line with the recommended levels of physical activity already stated for the general population [72], and especially in this population [73], we believe that it would be preferable to group individuals' moderate-intensity and vigorous-intensity activities into one moderate to vigorous activity category [74], which would also facilitate future comparisons with the rest of the literature.

Another way of expressing physical activity levels is by using METs. METs are a method of expressing an activity's energy costs, and they refer to energy expenditure rather than an

activity's intensity. Using METs would seem to be more relevant and more likely to accurately report an individual's true level of physical activity or energy expenditure [75]. Nevertheless, the lack of available data did not allow us to validate this outcome's use among pathological populations; therefore, using kilocalories was a preferable way of expressing energy expenditure. An increasing number of studies used activity counts to report physical activity levels in the axSpA patient population [11,27,28,41,42]; however, it would seem advisable to assess this outcome under standardized laboratory and open-field conditions.

The number of daily steps remains the outcome of choice for monitoring an individual's activity level [44,76-78]. This enables the particularly straightforward detection of decreases in ambulatory activities that prevent the onset of runaway evolution in the disease or marked disease flares [29]. As one of the studies already addressed this question by linking daily numbers of steps over 3 months to acute disease flare-ups [29], we believe that data over longer periods could also be of interest. Just as Tudor-Locke et al [79] reported changes in daily activity patterns over the course of a year, we think it would be interesting to track the daily number of steps taken by patients with axSpA over 1 year.

The literature concerning assessments of sedentary behavior is much more scattered, and more studies are needed to confirm the trends reported to date. As with the intensity of physical activity, sedentary behavior can be defined using certain thresholds. Some researchers have used activity count [27,41,42], whereas others have estimated sedentary behavior using METs provided by the manufacturers' algorithms [30].

At present, only one method of monitoring sedentary behavior exists for patients with axSpA, although it has different outcomes. This method uses wearable trackers and the acceleration data obtained from them. The outcomes are overall reports of sedentary time based on thresholds and detailed reports of sitting times and the number of sitting events.

To increase the monitoring precision and for comparative purposes, we suggest that all studies clearly mention the duration of carrying the wearable devices.

Monitoring Protocols and Assessment Methods

The studies included in this systematic review reported two ways of monitoring sedentary behavior and physical activity levels of patients with axSpA. We found studies with short follow-up periods of 5 to 7 days [11,21,27,28,30,39-43] and others with follow-up periods of >90 days [20,29,31], depending on each study's objectives. Importantly, sufficient daily wear time and a sufficient number of follow-up days, including specific weekend days, had to be ensured for that follow-up to be valid. Interestingly, all studies included weekends in their monitoring. For example, Gossec et al [20] identified the critical time intervals for classifying activities and tracking accuracy. These authors also reported that the significantly different activities performed on Saturday afternoons were associated with the detection of axSpA flare-ups and changes in flare-up state [20]. When examining the general population data [79], the days of the week included in the follow-up may have influenced the monitoring results. Furthermore, owing to the technical properties of wearable devices, some activities, such as swimming, could not be monitored.

The length of monitoring (ie, 1 week or several weeks) and days included in the monitoring (ie, weekday, workday, day off, and weekend) should be harmonized as they influence physical activity performance [20,80,81]. If reliable results can often be found with 1 week of monitoring, longer monitoring can help health professionals capture days of the week in which the participant is always inactive to further adapt the intervention program [80].

Limitations and Perspectives of the Review

Wearable trackers are promising as they have the potential to better monitor the physical activity and sedentary behavior evolution of patients with axSpA and study its impact on the disease. Mobile health, including trackers, permits health monitoring when outside health structures and could limit the number of visits to the hospital or clinic [82]. The literature on wearable trackers is rapidly building, and it is possible that some studies were published between the search and publication of this review. Moreover, most systematic reviews were limited by the small number of studies included. To avoid this, the search strategy included all fields and was not limited to the titles and abstracts. We used a thorough systematic and transparent methodology to conduct this review [32]. Despite this, only a few studies were included in the review, and we encountered some challenges when comparing across studies because of varying methods and reported results.

Furthermore, although the present review did not focus on the role of wearables as interventions to improve physical activity and sedentary behavior, this area of research could represent a relevant future research direction. Indeed, trackers can also be used as an intervention to motivate users to increase physical activity and decrease sedentary behavior [17,64,83-85] and could further prevent inactivity- or sedentary-related diseases (eg, cardiovascular diseases) [12].

Conclusions

This review identified and synthesized currently available evidence on the use of wearable activity trackers as outcome measures for physical activity and sedentary behavior in patients with axSpA.

We have underlined some trends regarding (1) the types of wearable devices used, (2) the outcomes reported, and (3) the follow-up protocols used. To date, few studies have used wearable devices to quantify physical activity among patients with axSpA, and the methods used have been heterogeneous. To fill this gap in knowledge and the literature, we suggest that future research focus on testing the feasibility and accuracy of physical activity and sedentary behavior assessments in patients with axSpA. The best locations to position the sensors should also be considered. This should occur in both the short-term, controlled, and supervised conditions of a laboratory environment and the long-term, varied, and everyday conditions of normal living environments.

Acknowledgments

This work was partly supported by the French National Research Agency within the framework of the Investissements d'avenir program (ANR-10-AIRT-05 and ANR-15-IDEX-02). The sponsors had no involvement in the study design, collection, analysis, and interpretation of data or in writing the manuscript. This study also forms part of a broader translational and interdisciplinary GaitAlps research program (NV).

Authors' Contributions

NV designed the systematic review protocol. JS and TC searched for and selected the articles included in this review and extracted data from the articles. JS and TC drafted the first version of the manuscript. JK and NV critically revised the article for important intellectual content. All authors read and approved the final version and agreed to be accountable for all aspects related to the accuracy or integrity of the work.

Conflicts of Interest

None declared.

References

1. Sieper J, Rudwaleit M, Baraliakos X, Brandt J, Braun J, Burgos-Vargas R, et al. The Assessment of SpondyloArthritis international Society (ASAS) handbook: a guide to assess spondyloarthritis. Ann Rheum Dis 2009 Jul 11;68 Suppl 2(Suppl 2):ii1-i44. [doi: 10.1136/ard.2008.104018] [Medline: 19433414]
2. Dubash S, McGonagle D, Marzo-Ortega H. New advances in the understanding and treatment of axial spondyloarthritis: from chance to choice. Ther Adv Chronic Dis 2018 Mar;9(3):77-87 [FREE Full text] [doi: 10.1177/2040622317743486] [Medline: 29511503]
3. Kiltz U, Landewé RB, van der Heijde D, Rudwaleit M, Weisman MH, Akkoc N, et al. Development of ASAS quality standards to improve the quality of health and care services for patients with axial spondyloarthritis. Ann Rheum Dis 2020 Feb 11;79(2):193-201 [FREE Full text] [doi: 10.1136/annrheumdis-2019-216034] [Medline: 31604704]
4. Zão A, Cantista P. The role of land and aquatic exercise in ankylosing spondylitis: a systematic review. Rheumatol Int 2017 Dec 5;37(12):1979-1990. [doi: 10.1007/s00296-017-3829-8] [Medline: 28983663]
5. Millner JR, Barron JS, Beinke KM, Butterworth RH, Chasle BE, Dutton LJ, et al. Exercise for ankylosing spondylitis: an evidence-based consensus statement. Semin Arthritis Rheum 2016 Mar;45(4):411-427 [FREE Full text] [doi: 10.1016/j.semarthrit.2015.08.003] [Medline: 26493464]
6. Liang H, Xu L, Tian X, Wang S, Liu X, Dai Y, et al. The comparative efficacy of supervised- versus home-based exercise programs in patients with ankylosing spondylitis: a meta-analysis. Medicine (Baltimore) 2020 Mar;99(8):e19229 [FREE Full text] [doi: 10.1097/MD.0000000000019229] [Medline: 32080122]
7. Liang H, Zhang H, Ji H, Wang C. Effects of home-based exercise intervention on health-related quality of life for patients with ankylosing spondylitis: a meta-analysis. Clin Rheumatol 2015 Oct 15;34(10):1737-1744. [doi: 10.1007/s10067-015-2913-2] [Medline: 25771852]
8. Pécourneau V, Degboé Y, Barnetche T, Cantagrel A, Constantin A, Ruyssen-Witrand A. Effectiveness of exercise programs in ankylosing spondylitis: a meta-analysis of randomized controlled trials. Arch Phys Med Rehabil 2018 Feb;99(2):383-9.e1. [doi: 10.1016/j.apmr.2017.07.015] [Medline: 28860095]
9. Saracoglu I, Kurt G, Okur EO, Afsar E, Seyyar GK, Calik BB, et al. The effectiveness of specific exercise types on cardiopulmonary functions in patients with ankylosing spondylitis: a systematic review. Rheumatol Int 2017 Mar 11;37(3):409-421. [doi: 10.1007/s00296-016-3603-3] [Medline: 27837263]
10. Verhoeven F, Guillot X, Prati C, Mougin F, Tordi N, Demougeot C, et al. Aerobic exercise for axial spondyloarthritis - its effects on disease activity and function as compared to standard physiotherapy: a systematic review and meta-analysis. Int J Rheum Dis 2019 Mar;22(2):234-241. [doi: 10.1111/1756-185X.13385] [Medline: 30187695]
11. Arends S, Hofman M, Kamsma YP, van der Veer E, Houtman PM, Kallenberg CG, et al. Daily physical activity in ankylosing spondylitis: validity and reliability of the IPAQ and SQUASH and the relation with clinical assessments. Arthritis Res Ther 2013 Aug 23;15(4):R99 [FREE Full text] [doi: 10.1186/ar4279] [Medline: 23971767]
12. Lear SA, Hu W, Rangarajan S, Gasevic D, Leong D, Iqbal R, et al. The effect of physical activity on mortality and cardiovascular disease in 130 000 people from 17 high-income, middle-income, and low-income countries: the PURE study. Lancet 2017 Dec 16;390(10113):2643-2654. [doi: 10.1016/S0140-6736(17)31634-3] [Medline: 28943267]
13. Sveaas SH, Berg IJ, Provan SA, Semb AG, Hagen KB, Vøllestad N, et al. Efficacy of high intensity exercise on disease activity and cardiovascular risk in active axial spondyloarthritis: a randomized controlled pilot study. PLoS One 2014;9(9):e108688 [FREE Full text] [doi: 10.1371/journal.pone.0108688] [Medline: 25268365]
14. Poddubnyy D, Sieper J. Treatment of axial spondyloarthritis: what does the future hold? Curr Rheumatol Rep 2020 Jul 20;22(9):47 [FREE Full text] [doi: 10.1007/s11926-020-00924-5] [Medline: 32691259]

15. Deodhar A, Gensler LS, Magrey M, Walsh JA, Winseck A, Grant D, et al. Assessing physical activity and sleep in axial spondyloarthritis: measuring the gap. Rheumatol Ther 2019 Dec 31;6(4):487-501 [FREE Full text] [doi: 10.1007/s40744-019-00176-5] [Medline: 31673975]
16. Pericleous P, van Staa TP. The use of wearable technology to monitor physical activity in patients with COPD: a literature review. Int J Chron Obstruct Pulmon Dis 2019;14:1317-1322 [FREE Full text] [doi: 10.2147/COPD.S193037] [Medline: 31354259]
17. Phillips SM, Cadmus-Bertram L, Rosenberg D, Buman MP, Lynch BM. Wearable technology and physical activity in chronic disease: opportunities and challenges. Am J Prev Med 2018 Jan;54(1):144-150 [FREE Full text] [doi: 10.1016/j.amepre.2017.08.015] [Medline: 29122356]
18. Hammond-Haley M, Allen C, Han J, Patterson T, Marber M, Redwood S. Utility of wearable physical activity monitors in cardiovascular disease: a systematic review of 11464 patients and recommendations for optimal use. Eur Heart J Digit Health 2021 Jun;2(2):231-243. [doi: 10.1093/ehjdh/ztab035]
19. Brophy S, Cooksey R, Davies H, Dennis MS, Zhou S, Siebert S. The effect of physical activity and motivation on function in ankylosing spondylitis: a cohort study. Semin Arthritis Rheum 2013 Jul;42(6):619-626 [FREE Full text] [doi: 10.1016/j.semarthrit.2012.09.007] [Medline: 23351615]
20. Gossec L, Guyard F, Leroy D, Lafargue T, Seiler M, Jacquemin C, et al. Detection of flares by decrease in physical activity, collected using wearable activity trackers in rheumatoid arthritis or axial spondyloarthritis: an application of machine learning analyses in rheumatology. Arthritis Care Res (Hoboken) 2019 Oct;71(10):1336-1343. [doi: 10.1002/acr.23768] [Medline: 30242992]
21. Swinnen TW, Scheers T, Lefevre J, Dankaerts W, Westhovens R, de Vlam K. Physical activity assessment in patients with axial spondyloarthritis compared to healthy controls: a technology-based approach. PLoS One 2014 Feb 28;9(2):e85309 [FREE Full text] [doi: 10.1371/journal.pone.0085309] [Medline: 24586239]
22. Fongen C, Halvorsen S, Dagfinrud H. High disease activity is related to low levels of physical activity in patients with ankylosing spondylitis. Clin Rheumatol 2013 Dec 31;32(12):1719-1725. [doi: 10.1007/s10067-013-2320-5] [Medline: 23900575]
23. Haglund E, Bergman S, Petersson IF, Jacobsson LT, Strömbeck B, Bremander A. Differences in physical activity patterns in patients with spondylarthritis. Arthritis Care Res (Hoboken) 2012 Dec 28;64(12):1886-1894 [FREE Full text] [doi: 10.1002/acr.21780] [Medline: 22745014]
24. Manning VL, Hurley MV, Scott DL, Bearne LM. Are patients meeting the updated physical activity guidelines? Physical activity participation, recommendation, and preferences among inner-city adults with rheumatic diseases. J Clin Rheumatol 2012;18(8):399-404. [doi: 10.1097/rhu.0b013e3182779cb6]
25. Fabre S, Molto A, Dadoun S, Rein C, Hudry C, Kreis S, et al. Physical activity in patients with axial spondyloarthritis: a cross-sectional study of 203 patients. Rheumatol Int 2016 Dec 24;36(12):1711-1718. [doi: 10.1007/s00296-016-3565-5] [Medline: 27665288]
26. Niedermann K, Sidelnikov E, Muggli C, Dagfinrud H, Hermann M, Tamborrini G, et al. Effect of cardiovascular training on fitness and perceived disease activity in people with ankylosing spondylitis. Arthritis Care Res (Hoboken) 2013 Dec 01;65(11):1844-1852 [FREE Full text] [doi: 10.1002/acr.22062] [Medline: 23836515]
27. O'Dwyer T, O'Shea F, Wilson F. Decreased physical activity and cardiorespiratory fitness in adults with ankylosing spondylitis: a cross-sectional controlled study. Rheumatol Int 2015 Dec 9;35(11):1863-1872. [doi: 10.1007/s00296-015-3339-5] [Medline: 26254884]
28. Plasqui G, Boonen A, Geusens P, Kroot EJ, Starmans M, van der Linden S. Physical activity and body composition in patients with ankylosing spondylitis. Arthritis Care Res (Hoboken) 2012 Jan 28;64(1):101-107 [FREE Full text] [doi: 10.1002/acr.20566] [Medline: 22213726]
29. Jacquemin C, Molto A, Servy H, Sellam J, Foltz V, Gandjbakhch F, et al. Flares assessed weekly in patients with rheumatoid arthritis or axial spondyloarthritis and relationship with physical activity measured using a connected activity tracker: a 3-month study. RMD Open 2017 Jun 29;3(1):e000434 [FREE Full text] [doi: 10.1136/rmdopen-2017-000434] [Medline: 28879046]
30. Coulter EH, McDonald MT, Cameron S, Siebert S, Paul L. Physical activity and sedentary behaviour and their associations with clinical measures in axial spondyloarthritis. Rheumatol Int 2020 Mar 17;40(3):375-381 [FREE Full text] [doi: 10.1007/s00296-019-04494-3] [Medline: 31848736]
31. Jacquemin C, Servy H, Molto A, Sellam J, Foltz V, Gandjbakhch F, et al. Physical activity assessment using an activity tracker in patients with rheumatoid arthritis and axial spondyloarthritis: prospective observational study. JMIR Mhealth Uhealth 2018 Jan 02;6(1):e1 [FREE Full text] [doi: 10.2196/mhealth.7948] [Medline: 29295810]
32. Carlin T, Soulard J, Aubourg T, Knitza J, Vuillerme N. Objective measurements of physical activity and sedentary behavior using wearable devices in patients with axial spondyloarthritis: protocol for a systematic review. JMIR Res Protoc 2021 Dec 25;10(11):e23359 [FREE Full text] [doi: 10.2196/23359] [Medline: 34842559]
33. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]

34. Mustapa A, Justine M, Mohd Mustafah N, Jamil N, Manaf H. Postural control and gait performance in the diabetic peripheral neuropathy: a systematic review. Biomed Res Int 2016;2016:9305025 [FREE Full text] [doi: 10.1155/2016/9305025] [Medline: 27525281]
35. Parreira RB, Grecco LA, Oliveira CS. Postural control in blind individuals: a systematic review. Gait Posture 2017 Sep;57:161-167. [doi: 10.1016/j.gaitpost.2017.06.008] [Medline: 28641161]
36. Soulard J, Vaillant J, Agier C, Vuillerme N. Gait characteristics in patients with ankylosing spondylitis: a systematic review. Clin Exp Rheumatol 2021;39(1):173-186 [FREE Full text] [Medline: 33025884]
37. St Fleur RG, St George SM, Leite R, Kobayashi M, Agosto Y, Jake-Schoffman DE. Use of fitbit devices in physical activity intervention studies across the life course: narrative review. JMIR Mhealth Uhealth 2021 May 28;9(5):e23411 [FREE Full text] [doi: 10.2196/23411] [Medline: 34047705]
38. Wurz A, Brunet J. The effects of physical activity on health and quality of life in adolescent cancer survivors: a systematic review. JMIR Cancer 2016 May 24;2(1):e6 [FREE Full text] [doi: 10.2196/cancer.5431] [Medline: 28410184]
39. Bayraktar D, Yuksel Karsli T, Ozer Kaya D, Sarac DC, Gucenmez S, Gercik O, et al. Is the international physical activity questionnaire (IPAQ) a valid assessment tool for measuring physical activity of patients with axial spondyloartritis? Musculoskelet Sci Pract 2021 Oct;55:102418. [doi: 10.1016/j.msksp.2021.102418] [Medline: 34171605]
40. Carbo MJ, Paap D, Maas F, Baron AJ, van Overbeeke LC, Siderius M, et al. The mSQUASH; a valid, reliable and responsive questionnaire for daily physical activity in patients with axial spondyloarthritis. Semin Arthritis Rheum 2021 Aug;51(4):719-727 [FREE Full text] [doi: 10.1016/j.semarthrit.2021.05.004] [Medline: 34144381]
41. van Genderen S, Boonen A, van der Heijde D, Heuft L, Luime J, Spoorenberg A, et al. Accelerometer quantification of physical activity and activity patterns in patients with ankylosing spondylitis and population controls. J Rheumatol 2015 Dec 01;42(12):2369-2375. [doi: 10.3899/jrheum.150015] [Medline: 26523021]
42. van Genderen S, van den Borne C, Geusens P, van der Linden S, Boonen A, Plasqui G. Physical functioning in patients with ankylosing spondylitis. J Clin Rheumatol 2014;20(3):133-137. [doi: 10.1097/rhu.0000000000000086]
43. Yuksel Karsli T, Bayraktar D, Ozer Kaya D, Oz HE, Gucenmez S, Gercik O, et al. Comparison of physical activity levels among different sub-types of axial spondyloarthritis patients and healthy controls. Mod Rheumatol 2021 Dec;31(6):1202-1207 [FREE Full text] [doi: 10.1080/14397595.2021.1891676] [Medline: 33627021]
44. Evenson KR, Goto MM, Furberg RD. Systematic review of the validity and reliability of consumer-wearable activity trackers. Int J Behav Nutr Phys Act 2015 Dec 18;12:159 [FREE Full text] [doi: 10.1186/s12966-015-0314-1] [Medline: 26684758]
45. Straiton N, Alharbi M, Bauman A, Neubeck L, Gullick J, Bhindi R, et al. The validity and reliability of consumer-grade activity trackers in older, community-dwelling adults: a systematic review. Maturitas 2018 Jul;112:85-93. [doi: 10.1016/j.maturitas.2018.03.016] [Medline: 29704922]
46. Kooiman TJ, Dontje ML, Sprenger SR, Krijnen WP, van der Schans CP, de Groot M. Reliability and validity of ten consumer activity trackers. BMC Sports Sci Med Rehabil 2015;7:24 [FREE Full text] [doi: 10.1186/s13102-015-0018-5] [Medline: 26464801]
47. Plasqui G, Westerterp KR. Physical activity assessment with accelerometers: an evaluation against doubly labeled water. Obesity (Silver Spring) 2007 Oct;15(10):2371-2379 [FREE Full text] [doi: 10.1038/oby.2007.281] [Medline: 17925461]
48. Plasqui G, Bonomi AG, Westerterp KR. Daily physical activity assessment with accelerometers: new insights and validation studies. Obes Rev 2013 Jul;14(6):451-462. [doi: 10.1111/obr.12021] [Medline: 23398786]
49. Skipworth RJ, Stene GB, Dahele M, Hendry PO, Small AC, Blum D, European Palliative Care Research Collaborative (EPCRC). Patient-focused endpoints in advanced cancer: criterion-based validation of accelerometer-based activity monitoring. Clin Nutr 2011 Dec;30(6):812-821. [doi: 10.1016/j.clnu.2011.05.010] [Medline: 21733604]
50. Redenius N, Kim Y, Byun W. Concurrent validity of the Fitbit for assessing sedentary behavior and moderate-to-vigorous physical activity. BMC Med Res Methodol 2019 Feb 07;19(1):29 [FREE Full text] [doi: 10.1186/s12874-019-0668-1] [Medline: 30732582]
51. Brewer W, Swanson BT, Ortiz A. Validity of Fitbit's active minutes as compared with a research-grade accelerometer and self-reported measures. BMJ Open Sport Exerc Med 2017;3(1):e000254 [FREE Full text] [doi: 10.1136/bmjsem-2017-000254] [Medline: 29018543]
52. Reid RE, Insogna JA, Carver TE, Comptour AM, Bewski NA, Sciortino C, et al. Validity and reliability of Fitbit activity monitors compared to ActiGraph GT3X+ with female adults in a free-living environment. J Sci Med Sport 2017 Jul;20(6):578-582. [doi: 10.1016/j.jsams.2016.10.015] [Medline: 27887786]
53. Roberts-Lewis SF, White CM, Ashworth M, Rose MR. Validity of Fitbit activity monitoring for adults with progressive muscle diseases. Disabil Rehabil 2021 Oct 31:1-11. [doi: 10.1080/09638288.2021.1995057] [Medline: 34719329]
54. Fuller D, Colwell E, Low J, Orychock K, Tobin MA, Simango B, et al. Reliability and validity of commercially available wearable devices for measuring steps, energy expenditure, and heart rate: systematic review. JMIR Mhealth Uhealth 2020 Sep 08;8(9):e18694 [FREE Full text] [doi: 10.2196/18694] [Medline: 32897239]
55. Xie J, Wen D, Liang L, Jia Y, Gao L, Lei J. Evaluating the validity of current mainstream wearable devices in fitness tracking under various physical activities: comparative study. JMIR Mhealth Uhealth 2018 May 12;6(4):e94 [FREE Full text] [doi: 10.2196/mhealth.9754] [Medline: 29650506]

56. Siddall AG, Powell SD, Needham-Beck SC, Edwards VC, Thompson JES, Kefyalew SS, et al. Validity of energy expenditure estimation methods during 10 days of military training. Scand J Med Sci Sports 2019 Oct;29(9):1313-1321. [doi: 10.1111/sms.13488] [Medline: 31136027]
57. Leung W, Case L, Sung M, Jung J. A meta-analysis of Fitbit devices: same company, different models, different validity evidence. J Med Eng Technol 2022 Mar;46(2):102-115. [doi: 10.1080/03091902.2021.2006350] [Medline: 34881682]
58. Soulard J, Vaillant J, Vuillerme N. Gait in patients with axial spondyloarthritis: a systematic review of the literature. Curr Rheumatol Rev 2022 Sep 21;18(2):117-123. [doi: 10.2174/1573397117666210921114949] [Medline: 34548003]
59. Sears T, Avalos E, Lawson S, Eschbach C, Bunn J. Wrist-worn physical activity trackers tend to underestimate steps during walking. Int J Exerc Sci 2017;10:764-773.
60. Ehrler F, Weber C, Lovis C. Influence of pedometer position on pedometer accuracy at various walking speeds: a comparative study. J Med Internet Res 2016 Oct 06;18(10):e268 [FREE Full text] [doi: 10.2196/jmir.5916] [Medline: 27713114]
61. Zebouni L, Helliwell PS, Howe A, Wright V. Gait analysis in ankylosing spondylitis. Ann Rheum Dis 1992 Jul 01;51(7):898-899 [FREE Full text] [doi: 10.1136/ard.51.7.898] [Medline: 1632666]
62. Del Din S, Carraro E, Sawacha Z, Guiotto A, Bonaldo L, Masiero S, et al. Impaired gait in ankylosing spondylitis. Med Biol Eng Comput 2011 Jul;49(7):801-809. [doi: 10.1007/s11517-010-0731-x] [Medline: 21229328]
63. Soulard J, Vaillant J, Baillet A, Gaudin P, Vuillerme N. Gait and axial spondyloarthritis: comparative gait analysis study using foot-worn inertial sensors. JMIR Mhealth Uhealth 2021 Nov 09;9(11):e27087 [FREE Full text] [doi: 10.2196/27087] [Medline: 34751663]
64. Brickwood K, Watson G, O'Brien J, Williams AD. Consumer-based wearable activity trackers increase physical activity participation: systematic review and meta-analysis. JMIR Mhealth Uhealth 2019 Apr 12;7(4):e11819 [FREE Full text] [doi: 10.2196/11819] [Medline: 30977740]
65. Tudor-Locke C, Camhi S, Troiano R. A catalog of rules, variables, and definitions applied to accelerometer data in the National Health and Nutrition Examination Survey, 2003-2006. Prev Chronic Dis 2012 Jun;9:E113 [FREE Full text] [doi: 10.5888/pcd9.110332] [Medline: 22698174]
66. Troiano RP, Berrigan D, Dodd KW, Mâsse LC, Tilert T, McDowell M. Physical activity in the United States measured by accelerometer. Med Sci Sports Exerc 2008 Jan;40(1):181-188. [doi: 10.1249/mss.0b013e31815a51b3] [Medline: 18091006]
67. Matthews CE, Chen KY, Freedson PS, Buchowski MS, Beech BM, Pate RR, et al. Amount of time spent in sedentary behaviors in the United States, 2003-2004. Am J Epidemiol 2008 May 01;167(7):875-881 [FREE Full text] [doi: 10.1093/aje/kwm390] [Medline: 18303006]
68. Aguilar-Farías N, Brown WJ, Peeters GM. ActiGraph GT3X+ cut-points for identifying sedentary behaviour in older adults in free-living environments. J Sci Med Sport 2014 May;17(3):293-299. [doi: 10.1016/j.jsams.2013.07.002] [Medline: 23932934]
69. Sasaki JE, John D, Freedson PS. Validation and comparison of ActiGraph activity monitors. J Sci Med Sport 2011 Oct;14(5):411-416. [doi: 10.1016/j.jsams.2011.04.003] [Medline: 21616714]
70. Freedson PS, Melanson E, Sirard J. Calibration of the computer science and applications, Inc. accelerometer. Med Sci Sports Exerc 1998 May;30(5):777-781. [doi: 10.1097/00005768-199805000-00021] [Medline: 9588623]
71. Copeland J, Esliger D. Accelerometer assessment of physical activity in active, healthy older adults. J Aging Phys Act 2009 Jan;17(1):17-30. [doi: 10.1123/japa.17.1.17] [Medline: 19299836]
72. Global Recommendations on Physical Activity for Health. Genève: World Health Organization; Jan 1, 2010.
73. Rausch Osthoff A, Niedermann K, Braun J, Adams J, Brodin N, Dagfinrud H, et al. 2018 EULAR recommendations for physical activity in people with inflammatory arthritis and osteoarthritis. Ann Rheum Dis 2018 Sep;77(9):1251-1260. [doi: 10.1136/annrheumdis-2018-213585] [Medline: 29997112]
74. Lai T, Liao Y, Lin C, Huang W, Hsueh M, Chan D. Moderate-to-vigorous physical activity duration is more important than timing for physical function in older adults. Sci Rep 2020 Dec 07;10(1):21344 [FREE Full text] [doi: 10.1038/s41598-020-78072-0] [Medline: 33288797]
75. Casey B, Coote S, Donnelly A. Objective physical activity measurement in people with multiple sclerosis: a review of the literature. Disabil Rehabil Assist Technol 2018 Feb 13;13(2):124-131. [doi: 10.1080/17483107.2017.1297859] [Medline: 28285547]
76. Bassett DR, Toth LP, LaMunion SR, Crouter SE. Step counting: a review of measurement considerations and health-related applications. Sports Med 2017 Jul 22;47(7):1303-1315 [FREE Full text] [doi: 10.1007/s40279-016-0663-1] [Medline: 28005190]
77. Saint-Maurice PF, Troiano RP, Bassett DR, Graubard BI, Carlson SA, Shiroma EJ, et al. Association of daily step count and step intensity with mortality among US adults. JAMA 2020 Mar 24;323(12):1151-1160 [FREE Full text] [doi: 10.1001/jama.2020.1382] [Medline: 32207799]
78. Tudor-Locke CE, Myers AM. Methodological considerations for researchers and practitioners using pedometers to measure physical (ambulatory) activity. Res Q Exerc Sport 2001 Mar;72(1):1-12. [doi: 10.1080/02701367.2001.10608926] [Medline: 11253314]
79. Tudor-Locke C, Williams JE, Reis JP, Pluto D. Utility of pedometers for assessing physical activity: convergent validity. Sports Med 2002;32(12):795-808. [doi: 10.2165/00007256-200232120-00004] [Medline: 12238942]

80. Berlin J, Storti K, Brach J. Using activity monitors to measure physical activity in free-living conditions. Phys Ther 2006 Aug;86(8):1137-1145. [Medline: 16879047]
81. Pasanen S, Halonen JI, Pulakka A, Kestens Y, Thierry B, Brondeel R, et al. Contexts of sedentary time and physical activity among ageing workers and recent retirees: cross-sectional GPS and accelerometer study. BMJ Open 2021 May 18;11(5):e042600 [FREE Full text] [doi: 10.1136/bmjopen-2020-042600] [Medline: 34006539]
82. Seppen BF, den Boer P, Wiegel J, Ter Wee MM, van der Leeden M, de Vries R, et al. Asynchronous mHealth interventions in rheumatoid arthritis: systematic scoping review. JMIR Mhealth Uhealth 2020 Nov 05;8(11):e19260 [FREE Full text] [doi: 10.2196/19260] [Medline: 33151161]
83. Böhm B, Karwiese SD, Böhm H, Oberhoffer R. Effects of mobile health including wearable activity trackers to increase physical activity outcomes among healthy children and adolescents: systematic review. JMIR Mhealth Uhealth 2019 Apr 30;7(4):e8298 [FREE Full text] [doi: 10.2196/mhealth.8298] [Medline: 31038460]
84. Laranjo L, Ding D, Heleno B, Kocaballi B, Quiroz JC, Tong HL, et al. Do smartphone applications and activity trackers increase physical activity in adults? Systematic review, meta-analysis and metaregression. Br J Sports Med 2021 May;55(8):422-432. [doi: 10.1136/bjsports-2020-102892] [Medline: 33355160]
85. Tang MS, Moore K, McGavigan A, Clark RA, Ganesan AN. Effectiveness of wearable trackers on physical activity in healthy adults: systematic review and meta-analysis of randomized controlled trials. JMIR Mhealth Uhealth 2020 Jul 22;8(7):e15576 [FREE Full text] [doi: 10.2196/15576] [Medline: 32706685]

Abbreviations

axSpA: axial spondyloarthritis
MET: metabolic equivalents task
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PROSPERO: International Prospective Register of Systematic Reviews

Edited by L Buis; submitted 05.11.21; peer-reviewed by C Romanzini, T Vliet Vlieland, RG St Fleur; comments to author 18.01.22; revised version received 02.02.22; accepted 21.03.22; published 22.08.22.

Please cite as:
Soulard J, Carlin T, Knitza J, Vuillerme N
Wearables for Measuring the Physical Activity and Sedentary Behavior of Patients With Axial Spondyloarthritis: Systematic Review
JMIR Mhealth Uhealth 2022;10(8):e34734
URL: https://mhealth.jmir.org/2022/8/e34734
doi:10.2196/34734
PMID:35994315

Review

Measures of Engagement With mHealth Interventions in Patients With Heart Failure: Scoping Review

Ifeanyi Madujibeya[1], BSN; Terry Lennie[1], PhD; Adaeze Aroh[2], BSc, MSc, MPH, DPH; Misook L Chung[1], PhD; Debra Moser[1], PhD

[1]College of Nursing, University of Kentucky, Lexington, KY, United States

[2]Department of Public Health, College of Health Professions, Slippery Rock University, Slippery Rock, PA, United States

Corresponding Author:
Ifeanyi Madujibeya, BSN
College of Nursing
University of Kentucky
751 Rose Street
Lexington, KY, 40536
United States
Phone: 1 859 334 0561
Email: ima232@uky.edu

Abstract

Background: Despite the potential of mobile health (mHealth) interventions to facilitate the early detection of signs of heart failure (HF) decompensation and provide personalized management of symptoms, the outcomes of such interventions in patients with HF have been inconsistent. As engagement with mHealth is required for interventions to be effective, poor patient engagement with mHealth interventions may be associated with mixed evidence. It is crucial to understand how engagement with mHealth interventions is measured in patients with HF, and the effects of engagement on HF outcomes.

Objective: In this review, we aimed to describe measures of patient engagement with mHealth interventions and the effects of engagement on HF outcomes.

Methods: We conducted a systematic literature search in 7 databases for relevant studies published in the English language from 2009 to September 2021 and reported the descriptive characteristics of the studies. We used content analysis to identify themes that described patient engagement with mHealth interventions in the qualitative studies included in the review.

Results: We synthesized 32 studies that operationalized engagement with mHealth interventions in 4771 patients with HF (3239/4771, 67.88%, male), ranging from a sample of 7 to 1571 (median 53.3) patients, followed for a median duration of 90 (IQR 45-180) days. Patient engagement with mHealth interventions was measured only quantitatively based on system usage data in 72% (23/32) of the studies, only qualitatively based on data from semistructured interviews and focus groups in 6% (2/32) of studies, and by a combination of both quantitative and qualitative data in 22% (7/32) of studies. System usage data were evaluated using 6 metrics of engagement: number of physiological parameters transmitted (19/30, 63% studies), number of HF questionnaires completed (2/30, 7% studies), number of log-ins (4/30, 13% studies), number of SMS text message responses (1/30, 3% studies), time spent (5/30, 17% studies), and the number of features accessed and screen viewed (4/30, 13% studies). There was a lack of consistency in how the system usage metrics were reported across studies. In total, 80% of the studies reported only descriptive characteristics of system usage data. The emotional, cognitive, and behavioral domains of patient engagement were identified through qualitative studies. Patient engagement levels ranged from 45% to 100% and decreased over time. The effects of engagement on HF knowledge, self-care, exercise adherence, and HF hospitalization were inconclusive.

Conclusions: The measures of patient engagement with mHealth interventions in patients with HF are underreported and lack consistency. The application of inferential analytical methods to engagement data is extremely limited. There is a need for a working group on mHealth that may consolidate the previous operational definitions of patient engagement into an optimal and standardized measure.

(JMIR Mhealth Uhealth 2022;10(8):e35657) doi:10.2196/35657

KEYWORDS

heart failure; mobile health interventions; mHealth interventions; patient engagement; system usage data; heart failure outcomes; mobile phone

Introduction

Background

Heart failure (HF) is a progressive chronic health condition characterized by the inability of the heart muscle to pump sufficient blood to meet the metabolic demands of the body [1]. HF is characterized by a high incidence of acute exacerbations, leading to poor health-related quality of life, and high hospitalization and mortality rates [2]. An estimated 6.2 million adults aged 20 years and older have HF in the United States [2]. This prevalence rate is projected to increase by 46% by 2030 [2].

Mobile devices are increasingly leveraged in mobile health (mHealth) interventions to provide comprehensive and personalized care that may decrease the incidence of HF exacerbations, improve health-related quality of life, and decrease HF hospitalization and mortality rates. mHealth is the use of mobile devices, such as smartphones, wearable sensors, PDAs, tablet computers, and mobile telemonitoring devices to deliver care, maintain health, and manage chronic conditions [3-5]. The outcomes of mHealth interventions for patients with HF have been inconsistent. Previous meta-analyses [6,7] and a systematic review [8] of mHealth-based interventions have shown mixed evidence on the effectiveness of these interventions in improving outcomes in patients with HF. Considering that engagement with mHealth interventions is a prerequisite for the effectiveness of the interventions [9], poor patient engagement with the interventions might be associated with mixed results [10-14]. Hence, it is crucial to measure engagement with mHealth interventions in patients with HF.

Conceptualization of Patient Engagement With mHealth

On the basis of an expert consensus, Yardley et al [9] conceptualized engagement with mHealth interventions as a dynamic process involving microengagement and macroengagement. Microengagement is the moment-to-moment use of mHealth interventions or systems and the subjective experience that is derived from using the systems. Macroengagement is the degree of health-related behavior change that is mediated by the use of mHealth interventions [9]. Perski et al [15] extended the framework proposed by Yardley et al [9] by describing subjective user experience as attention, interest, and affect [10,15]. Hence, patient engagement with mHealth was operationalized in previous studies as the intensity, duration, and frequency of mHealth system use [10,15-18], as well as the subjective experiences of the users, defined as attention, interest, and affect [15].

Short et al [11] advanced previous work [9,15] by identifying 8 subthemes that may be used in qualitative research to describe subjective user experience (Multimedia Appendix 1). Accordingly, we conceptualized engagement with mHealth interventions as a dynamic and multidimensional construct that consists of behavioral, cognitive, and emotional domains. The behavioral domain is measured using system usage data, which are quantitative data generated by the physical interaction of a user with mHealth systems [9,11]. Cognitive (pertains to what a patient thinks or knows) and emotional (what a patient feels) domains describe subjective user experiences of using mHealth [11,15,19].

Gap in Evidence

There is a dearth of information on how engagement with mHealth interventions has been conceptualized and measured in patients with HF. Recent scoping reviews [10,20] of evidence on measures of engagement in mHealth interventions for the management of chronic conditions included 51 articles in which patient engagement measures were reported. Only 3 articles were reviewed related to patients with HF. However, 3 previous reviews of mHealth applications for HF self-care [21-23] identified 70 available mHealth applications for the management of HF. Patient engagement with these applications has not been reported. Thus, previous scoping reviews [10,20] might not be a full representation of current mHealth engagement research in patients with HF.

In addition, previous scoping reviews [10,20] focused only on system usage data, which is an objective measure of usage logs generated during a user's interaction with mHealth systems. System usage data may not capture subjective user experiences, which are an essential aspect of patient engagement with mHealth interventions [9,11,17]. Thus, a review that includes both objective and subjective measures of patient engagement with mHealth is warranted. This review aimed to synthesize current evidence on measures of engagement of patients with HF with mHealth interventions and examine the effects of patient engagement with mHealth interventions on HF outcomes.

Specifically, we addressed the following questions: (1) How was engagement with mHealth interventions operationalized quantitatively and qualitatively in patients with HF? (2) How was engagement with mHealth interventions in patients with HF analyzed and reported in previous studies in patients with HF? (3) What were the patterns of engagement over time? (4) What factors predicted patterns of engagement over time? (5) What was the relationship between engagement and HF outcomes?

Methods

Methodological Framework

As a result of the novelty and heterogeneity of mHealth interventions in patients with HF [6,8], we used a scoping review to synthesize current evidence on engagement with mHealth interventions in patients with HF [24-28]. The review followed the checklist of the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) [25]. The review was guided by the 6-step methodological framework for scoping reviews by Arksey and

O'Malley [28-30] except for the optional consultation phase (step 6) [26]. The 5 steps used in this review are as follows: (1) formulate a research question; (2) search the literature to identify relevant studies; (3) select the relevant studies based on predefined eligibility criteria; (4) chart the data to identify key information; and (5) organize, summarize, and report the findings [27-30].

We conducted a comprehensive search of the literature in 7 databases (CINAHL, Cochrane Central Register of Controlled Trials [CENTRAL], PubMed, Scopus, PsycINFO, MEDLINE, and Ovid) for relevant literature published in the English language from 2009 to September 2021. The search was conducted with the help of an experienced medical librarian. A combination of keywords was used to search the databases (Multimedia Appendix 2).

Eligibility Criteria

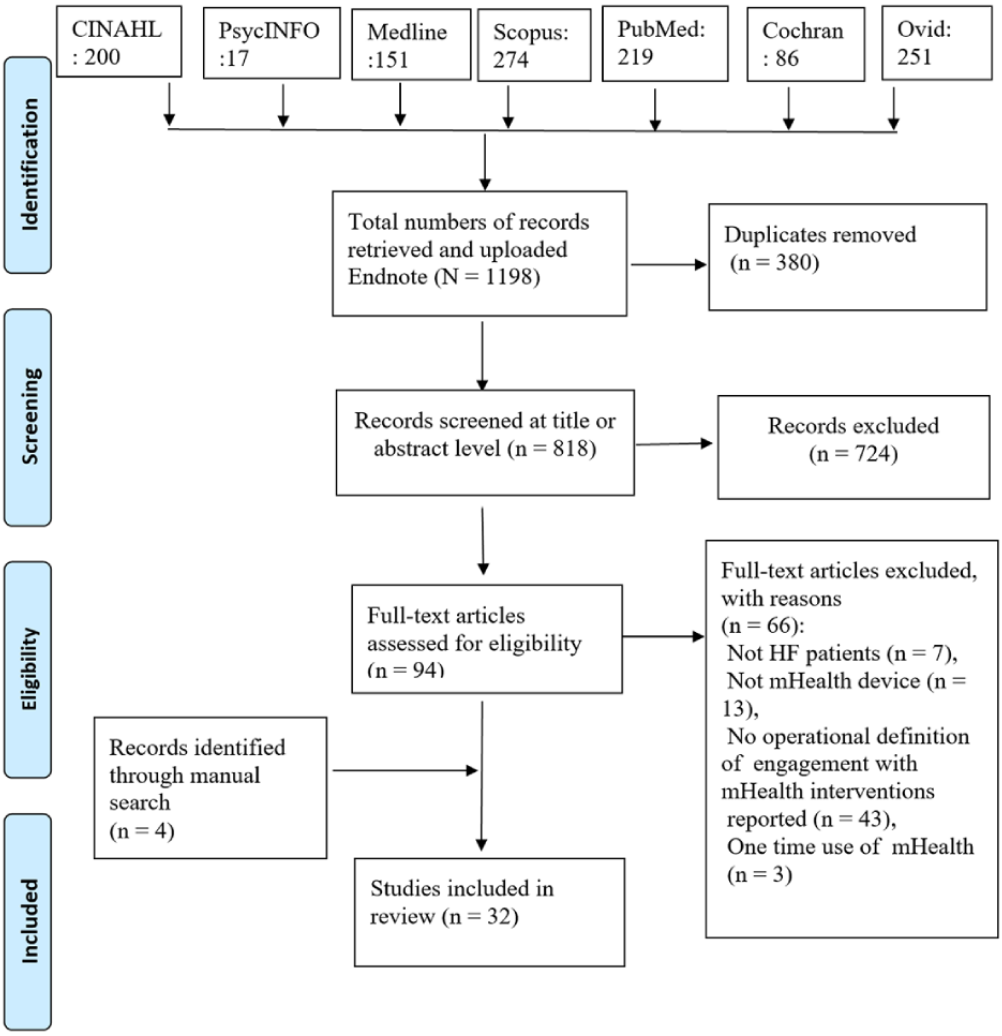
The following were the inclusion criteria: (1) studies that included adult patients with HF, aged ≥18 years, in New York Heart Association class 1 to 4, of any sex, ethnicity, and nationality, and published in the English language between 2009 and September 2021; (2) studies that operationalized engagement with mHealth interventions or usage of mHealth systems; and (3) studies that included results of patient engagement with mHealth interventions or effects of engagement with mHealth interventions on patient outcomes. Usability and feasibility studies in which patients explored mHealth application features only once were excluded because one-time usage is insufficient to establish patient engagement with the intervention [10]. Landline telephone–based interventions were also excluded because landline telephones are not considered mobile devices.

Data Extraction and Analysis

The initial database search yielded 1198 articles. The articles were uploaded to the Endnote software (version 20) for analysis. The selection process is illustrated in the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram (Figure 1). Two reviewers (first and third authors) independently selected 32 studies from the 1198 that met the inclusion criteria. The study and intervention characteristics were coded using a data extraction form based on related constructs from the CONSORT‐EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth) checklist (V.1.6.1) [31]. The coded characteristics are presented in Textbox 1. The descriptive characteristics of the studies were reported. System usage data reported in the studies were categorized using the frequency, intensity, time, and type (FITT) principle to provide more insight into the usage data. The FITT principle has been previously described by Short et al [11] and applied in analyzing the system usage data [32]. Frequency describes how often a patient completes a required task. The intensity or depth is the proportion of an assigned task completed by a patient. Time measures the duration for completing a task, and type is attributed to the type of intervention [11].

All studies in which qualitative methods were used to measure patient engagement were uploaded to the qualitative data analysis software Atlas.ti (version 8). The 3 phases of deductive content analysis outlined by Elo et al [33] were used to analyze the qualitative data. In the first phase, line-by-line coding was performed by grouping the data into clusters of information and assigning labels to the clusters. In the second phase, the list of codes was combined into potential subthemes and themes in accordance with the 8 main constructs used by Short et al [11] to describe the emotional, cognitive, and behavioral domains of engagement. Although the constructs overlapped, Short et al [11] provided a concise description of each construct (Multimedia Appendix 2). In the third phase, the potential themes and subthemes were refined to ensure that the data within each theme were distinctive. Two authors (IM and AA) independently conducted the initial analysis, which was reviewed by all the coauthors. Any disagreement during the analytical process was discussed until a consensus was reached.

Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses flow diagram.



Textbox 1. Codes extracted from included studies.

Study characteristics

- General information about the studies including the first author's last name, year, country of publication, duration of follow-up, patient characteristics, and the purpose of the study

Mobile health (mHealth) characteristics

- mHealth devices: mobile phone, PDA devices, sensor, and mobile telemonitor systems
- Measured physiological parameters: blood pressure, heart rate, weight, oxygen saturation, and electrocardiogram transmitted by patients
- Data transmission: mode of transmitting data from peripheral devices, such as weighing scale to the mHealth device
- Transmission frequency: how often patients transmit physiological parameters to providers or central monitoring centers
- The interactive user interface: interface for patient's interaction with mHealth systems

Engagement measures

- Operationalization of engagement: how engagement was measured
- Objective measures: objective measures of engagement, such as quantitative measures of system usage
- Subjective measures: measures of engagement using self-reported questionnaires or through a qualitative method, such as interviews
- Data collection method: methods for collecting engagement information
- Analytical methods: methods used for analyzing engagement data
- Reported engagement: the results of engagement with mHealth
- Effect of engagement: the reported effects of engagement on patient-reported outcomes.
- The strengths and limitations of studies

Results

Study and Patient Characteristics

Of the 32 studies, 16 (50%) [14,34-48] were conducted in the United States. The remaining studies were conducted in Germany (3/32, 9%) [49-51], Canada (2/32, 6%) [14,52], Belgium (2/32, 6%) [53,54], Italy (2/32, 6%) [55,56], the United Kingdom (1/32, 3%) [57], Austria (1/32, 3%) [58], Sweden (1/32, 3%) [59], Poland (1/32, 3%) [60], Singapore (1/32, 3%) [61], China (1/32, 3%) [62], and Australia (1/32, 3%) [63]. The duration of the studies ranged from 1 to 26 months, with a median of 3 months. The sample sizes ranged from 7 to 1571, with a median of 53.3 patients. Most of the patients (3239/4771, 67.9%) were male.

Intervention Characteristics

The key characteristics of the interventions are presented in (Multimedia Appendix 3). In approximately 50% (16/32) of the studies [14,35,38,40,43,45,47-49,52-55,57,58,60], patients used smartphones, 28% (9/32) [36,37,39,42,46,59,61-63] used tablet computers, 6.25% (2/32) [50,56] used PDAs, and 16% (5/32) used either portable telemonitoring devices [34,44] or a combination of smartphones, smart watches, and tablet computers [45,51,64] as integral components of mHealth systems for the management of HF symptoms or for the provision of HF-related self-care education.

In 84% (27/32) of the studies, physiological parameters (weight, blood pressure, heart rate, or electrocardiogram), patient-reported HF symptoms, and self-care activities that were transmitted to either secured servers or telemonitoring centers were used to provide personalized HF remote monitoring and management. In 9% (3/32) of the studies, the mHealth intervention focused solely on providing HF-related self-care education through daily HF quizzes [34], game application [37], or daily SMS text messages [35]. In the remaining 6% (2/32) of the studies [60,64] investigators used mHealth systems to implement home-based cardiac rehabilitation or to target exercise adherence via videoconferencing. The investigators in 91% (29/32) of the studies incorporated the user interface of the mHealth devices to provide interactive HF education, graphic display of monitored parameters, activity reminders, or interaction with a web-based assistant.

Operational Definitions of Patient Engagement With mHealth

In addition, the operational definitions of patient engagement with mHealth interventions are summarized in Multimedia Appendix 4. Patient engagement with mHealth interventions was measured solely based on system usage data in 72% (23/32) of the studies. Among the remaining studies, 6% (2/32) used only qualitative methods to determine engagement (focus groups and semistructured interviews) [34,39]; 19% (6/32) [14,40,48,54,57,62] used both system usage data and qualitative methods; and 3% (1/32) [43] planned to use system usage data, qualitative methods (think aloud), and user engagement questionnaires.

Analytical Methods Applied to System Usage Data

As shown in Multimedia Appendix 5, in 94% (30/32) of the reviewed studies, patient engagement with mHealth interventions was evaluated using six main system usage data: (1) number of physiological parameters measured and transmitted (19/30, 63%), (2) number of HF symptom

questionnaires completed (2/30, 7%), (3) number of log-ins (4/30, 13%), (4) number of SMS text message responses (1/30, 3%), (5) time spent (5/30, 17%), and (6) number of features accessed or screens viewed (4/30, 13%). Descriptive statistics (mean, range, median, and percentage) were used to summarize patient engagement in 80% (24/30) of the reviewed studies that analyzed system usage data. The remaining 20% (6/30) of the studies applied both descriptive and inferential statistics to system usage data. The analytical methods and studies that used them are presented in Multimedia Appendix 4.

Operational Definitions of System Usage Data

Operational definitions of system usage data and reported outcomes are presented in Multimedia Appendices 4 and 6. The operational definitions differed across the 30 studies that reported the metric. In 47% (14/30) of the 30 studies, system usage data were operationalized as the proportion of patients who used an mHealth system to complete 70% [49,50], 80% [52], 85% [61], or 100% [36,38,40,42,48,60-64] of the required tasks as expected during the duration of intervention or system use. The engagement levels reported in the 14 studies ranged from 45% [63] to 100% [60]. In 23% (7/30) of the 30 studies [14,40,41,46,53,57,59], system usage data were measured as the proportion of days during which patients completed assigned tasks or used mHealth, as expected, during the total number of days equipped with the system. Median engagement rates of 88% and 96% were reported in 2 studies [41,59], while 1 study [40] reported a mean engagement of 18.2%. The remaining 4 studies [14,46,53,57] reported engagement outcomes as a percentage, ranging from 73.6% [46] to 88% [57].

In 20% (6/30) of the 30 studies [35,44,52,54,55,58], system usage data were operationalized as the number of assigned tasks completed per patient per number of days equipped with a mobile device or intervention. In 1 of the 6 studies [35], investigators reported a mean engagement of 5.7%, with a range of 0 to 27, while in the remaining studies, engagement was reported as an overall rate, ranging from 53.3% to 95%. In 7% (2/30) of the studies [37,51], investigators measured system usage data as the number of times an mHealth system was used per patient per duration of intervention. The reported engagement ranged from 9.7 hours in 28 days [37], to 11.3 hours in 60 days [51]. Other investigators [45,47] measured system usage data as the ratio of the number of hours a patient had heart rate readings to the total hours in the study. For example, Sohn et al [45] reported an engagement rate of 79.1%.

The categorization of system usage data based on the FITT principle is presented in Table 1. The intensity category was the most predominant (22/30, 73.3%) among the reviewed studies, followed by frequency (12/30, 40%), and time (8/30, 26.7%). Only 2 studies [40,48] reported the frequency, intensity, and time spent.

Table 1. Categorization of system usage data based on the frequency, intensity, time, and type (FITT) principle.

Study	mHealth[a] device	Frequency	Intensity	Time spent	Type of intervention
Apergi et al [46]	Tablet	N/A[b]	✓	N/A	Telemonitoring
Athilingam et al [38]	Smartphone	N/A	✓	✓	Telemonitoring, HF[c] education, and physical activity
Bartlett et al [57]	Smartphone	✓	N/A	✓	Telemonitoring, physical activity, and HF education
Buck et al [39]	Tablet	N/A	N/A	N/A	Telemonitoring and physical activity
Chow et al [61]	Tablet	N/A	✓	N/A	Telemonitoring and HF education
Dang et al [40]	Smartphone	✓	✓	✓	Telemonitoring
Deka et al [64]	Smartwatch	N/A	✓	N/A	Physical activity
Dendale et al [53]	Smartphone	N/A	✓	N/A	Telemonitoring
Ding et al [63]	Tablet	✓	N/A	N/A	Telemonitoring
Guo et al [62]	Tablet	✓	N/A	N/A	Telemonitoring
Hägglund et al [59]	Tablet	✓	N/A	N/A	Telemonitoring and HF education
Hayes et al [44]	Tablet, WTD[d]	✓	✓	N/A	Telemonitoring and HF education
Kitsiou et al [47]	Smartphone, smartwatch	N/A	✓	✓	Telemonitoring and physical activity
Koehler et al [49]	WTD, smartphone, or tablet,	N/A	✓	N/A	Telemonitoring and HF[a] education
Koehler et al [50]	PDA	N/A	✓	N/A	Telemonitoring
Lloyd et al [42]	Tablet	✓	✓	N/A	Self-care and physical activity
Louise et al [34]	WTD	N/A	N/A	N/A	HF education
Nundy et al [35]	Smartphone	✓	N/A	N/A	HF education
Piotrowicz et al [60]	Smartphone	N/A	✓	N/A	Cardiac rehab and HF education
Pedone et al [55]	Smartphone	N/A	✓	N/A	Telemonitoring
Radhakrishnan et al [37]	Tablet	N/A	✓	✓	HF education via gaming
Rosen et al [41]	Tablet	N/A	✓	N/A	Telemonitoring and HF education
Scherr et al [58]	Smartphone	✓	N/A	N/A	Telemonitoring
Seto et al [52]	Smartphone	N/A	✓	N/A	Telemonitoring
Smeets et al [54]	Smartphone	N/A	✓	N/A	Telemonitoring and HF education
Sohn et al [45]	Smartwatch, smartphone	N/A	✓	✓	Telemonitoring and physical activity
Villani et al [56]	PDA	N/A	✓	N/A	Telemonitoring
Ware et al [14]	Smartphones	N/A	✓	N/A	Telemonitoring
Wei et al [48]	Smartphones	✓	✓	✓	Telemonitoring and HF education
Werhahn et al [51]	Smartphones, tablet smartwatch	✓	N/A	✓	Telemonitoring and physical activity
Zan et al [36]	Tablet, web portal	N/A	✓	N/A	Telemonitoring
Zhang et al [43]	Smartphones with virtual reality–based self-care assistance	N/A	N/A	N/A	Telemonitoring and physical activity

[a]mHealth: mobile health.

[b]N/A: not applicable; represents qualitative studies or studies that did not report elements of the FITT principle.

[c]HF: heart failure.

[d]WTD: wireless telemonitoring device.

Longitudinal Patterns of Patient Engagement With mHealth Interventions

The investigators in one of the 8 studies [41] that reported longitudinal patterns of patient engagement with mHealth interventions concluded that patient engagement did not change over time. However, the investigators did not state how the effect of time on patient engagement patterns was examined. In the remaining 7 studies, the investigators used descriptive statistics (plots of engagement over time) [36,40,42,44,46,52,62] or a longitudinal analysis [14] to examine the effects of time on engagement patterns. All the investigators reported that patient engagement decreased over time.

Predictors of Patient Engagement With mHealth

Four groups [14,41,45,46] examined the effects of age on patient engagement with mHealth interventions, and the findings were inconclusive. Apergi et al [46] reported a positive association between age and patient engagement, whereas Sohen et al [45] and Rosen et al [41] reported a nonsignificant association between age and patient engagement. The investigators in 2 studies [14,41] examined the effects of sex and HF severity (measured by New York Heart Association class) on patient engagement. They reported a nonsignificant association among gender, HF severity, and patient engagement with mHealth interventions.

Qualitative Measures of Patient Engagement

The emotional, cognitive, and behavioral domains of patient engagement with mHealth interventions, and the constructs used to describe them in qualitative research are summarized in Table 2. In 8 [14,34,39,40,48,54,57,62] out of the 9 studies that included qualitative measures, open-ended questionnaires, focus groups, and semistructured interviews were used to describe patients' experience of using mHealth devices. The experiences were categorized under the behavioral, cognitive, and emotional domains of patient engagement. Intervention usage, which is a construct of the behavioral domain that describes a user's patterns of interaction with mHealth interventions or systems [11,19], was the most reported subcategory (7/8, 88%) in the studies. For example, in the postintervention interviews with patients who participated in a tablet-delivered self-care intervention (Penn State Heart Assistant), patients stated that they recorded their blood pressure and weight every morning and exercised daily whenever the mHealth system was functional [39].

Three [14,40,62] out of the 8 studies used affect to describe the emotional domain of patient engagement. For example, in a mobile phone–based telemonitoring intervention, patients stated that they felt guilty when they missed measuring the required daily physiological parameters [14]. In 11.1% (1/9) [43] of the studies that included qualitative measures, investigators planned to use think aloud to capture the patient's cognitive process while patients were performing tasks on mHealth applications.

Table 2. Qualitative constructs used to describe the emotional, cognitive, and behavioral engagement.

Study	Subcategories	Quotes
Barlett et al [57]	• Intervention usage[a] • Immersion[b]	• "The interview data report higher engagement with the walking than was recorded in the step count in the mobile device." (Intervention usage) • "I cannot use the system every day, I will use it as it fit my lifestyle." (Immersion)
Buck et al [39]	• Intervention usage[a]	• "I still record my blood pressure, weight, and exercise every day. So, instead of a paper, I would put it on my iPad." (Intervention usage)
Dang et al [40]	• Affect[c] • Intervention usage[a]	• "All participants said that the program made them feel more secure about their health and that they would stay enrolled." (Affect) • "Since participants received daily reminders to weigh themselves, it had become a habit." (Intervention usage)
Guo et al [62]	• Interest[c] • Affect[c]	• Participants were more interested in smart health tracking devices, which could help them keep track of health conditions anywhere, (interest) so that they felt more secure and involved in their care (affect)
Laframboise et al [34]	• Intervention usage[a] • Interest[b]	• "Many participants perceived the daily interaction with the Health Buddy (mobile device) as social contact and something they looked forward to, as well as something to do daily." (Interest) • "The Health Buddy was kind of like a good friend. It gave me something to do every day." (Intervention usage and interest)
Smeets et al [54]	• Intervention usage[a]	• "50% of patients were eager to continue using the CardioCoach follow-up tool after the study ended." (Intervention usage)
Ware et al [14]	• Intervention usage[a] • Affect[c] • Interest[b]	• "Taking my readings is what I do first thing in the morning before I get the phone call with the annoying ringing" (Intervention usage, affect) • "Feel kind of guilty because I haven't got it [Taking daily readings] done." (Affect)
Wei et al [48]	• Intervention usage[a]	• "One participant reported synching issues between the scale and the app." (Intervention usage)

[a]Behavioral domain.

[b]Cognitive domain.

[c]Emotional domain.

Effects of Patient Engagement With mHealth Interventions on HF Outcomes

Few researchers reported the effect of patient engagement with mHealth interventions on HF outcomes (HF knowledge, self-care, weight loss, and exercise engagement) using both quantitative and qualitative methods. Patient engagement with mHealth interventions was positively correlated with an improvement in HF knowledge. Three studies aimed to improve HF knowledge using daily HF quizzes [57], mobile game applications [37], or watching HF educational videos on smartphone interfaces [48]. There was a significant positive correlation between patient engagement and improvement in HF knowledge in all 3 studies.

Radhakrishnan et al [37] reported a positive correlation between the average game-playing time and HF-related self-care. In contrast, Sohn et al [45] showed a negative correlation between patient engagement and self-care confidence. In the 3 studies in which semistructured interviews were used [14,34,40], patients stated that engagement in telemonitoring was associated with improvement in their HF self-care [14,34] and self-care confidence [40]. However, only 33% (8/24) of the patients interviewed in one study [14] agreed that engagement with the intervention improved their self-care confidence. Thus, based on qualitative data, the effect of patient engagement on self-care is inconclusive.

Only one investigative team [42] examined the effects of patient engagement with interventions on weight loss and exercise. The investigators reported positive associations between patient engagement, weight, and exercise engagement. On the other hand, only Haynes et al [44] examined the effect of patient engagement on hospitalization because of HF. They reported that every 1-day increase in patient engagement was associated with a 19% decrease in HF hospitalization [44].

Discussion

Principal Findings

We used a scoping review to present an overview of how engagement with mHealth interventions was operationalized among patients with HF. Across the studies, patient engagement with the interventions was evaluated using both quantitative measures based on system usage data and qualitative measures based on semistructured interviews and focus groups. System

usage data were evaluated as physiological parameters transmitted to telemonitoring centers, number of HF questionnaires completed, number of log-ins, number of SMS text message responses, time spent engaging with interventional features, features accessed, or screen viewed. The measures of system usage data were underreported and lacked consistency. The application of inferential analytical methods to the data is extremely limited.

Evaluation of System Usage Data

In most studies in our review (23/32, 72%), only system usage data were measured to quantify engagement with mHealth. The predominant focus on system usage data in the reviewed studies was expected, considering that these metrics are the most reported measures of patient engagement with mHealth interventions [10,11,15]. mHealth devices can automatically track the user's patterns of interaction with mHealth interventions and generate quantitative data that reflect the patterns of the interaction. The ready availability of the data may have contributed to its popularity among investigators. However, this method alone misses important components of engagement.

Reporting all 4 main elements of the FITT principle is essential to capture all the aspects of system usage data [11,65-67]. However, only 2 studies in our review reported all 4 components of the FITT intervention. In 47% (14/30) of the studies that evaluated system usage data, investigators reported only intensity. The emphasis on intensity was consistent with previous studies [10,15] that categorized system usage metrics as amount, breadth, duration, and depth. Pham et al [10] reported that the majority (31/41, 76%) of the studies in their review measured the depth of engagement category, which is the same as the intensity [11]. It is likely that the investigators were not examining the frequency and time components of the FITT principle or were underreporting them. This could obscure the differences in patient engagement profiles when patients showed similar intensity levels, but differed in either frequency or time spent in mHealth interventions. Examining all components of FITT is essential in gaining more insight into patient engagement behaviors than measuring only one component. Such insight could guide actions and policies to promote engagement behaviors that are congruent with interventional outcomes [11].

Longitudinal Patterns of Patient Engagement

Cheikh-Moussa et al [20] concluded in their review of 10 articles that patients with cardiometabolic conditions' engagement with mHealth interventions decreased over time. The findings are consistent with the results from 8 articles in our review that showed that patient engagement with mHealth interventions decreased over time. However, our findings should be interpreted with caution. The investigators in 7 of the 8 studies used only simple plots (descriptive statistics) to examine the relationship between patient engagement and time. Similarly, researchers in 2 studies [41,45] out of the 4 that examined the effect of age on engagement limited their analysis to descriptive statistics. Thus, the application of inferential statistics in evaluating system usage data is extremely limited, making it challenging to draw definitive conclusions on the longitudinal patterns of patient engagement and the predictors of patient engagement with mHealth interventions.

Subjective Measures of Engagement

Intervention usage was the most identified qualitative measure of patient engagement, indicating that most investigators focused on usage (behavioral domain). These findings appear consistent with a previous qualitative review of 11 studies that evaluated patient engagement with eHealth [19]. The investigators highlighted the behavioral and cognitive domains of engagement as the most assessed aspects of patient engagement [19]. However, the emotional domain is equally important in understanding the complexity of patient engagement with mHealth interventions. For example, the experience of technical challenges with mHealth interventions could trigger negative emotions in patients, such as emotional exhaustion and sadness. Patients may be inclined to regulate these emotions by decreasing the extent of their interaction with the intervention. Hence, the interplay between the emotional and behavioral domains of engagement within the context of technical problems could influence patterns of patient engagement with mHealth [68]. Thus, assessing the 3 domains of patient engagement may be pivotal in understanding the complexity of patient engagement with mHealth interventions.

The qualitative assessment of intervention usage may be combined with system usage data to provide more insight into the patterns of patient engagement with mHealth interventions. For example, in 2 studies, the SMART Personalized Self-management System for HF intervention [57] and phone-based telemonitoring intervention for patients with HF [14], the investigators deduced from interview reports that system usage data captured by the mHealth system did not reflect the actual patient engagement. The patients reported a higher degree of engagement, but it was not captured by the mHealth systems because of technical problems such as poor connectivity between peripheral devices and mobile phones, server downtime, and system malfunction [57]. Thus, the use of both qualitative and quantitative approaches to measure patient engagement with mHealth is recommended.

Focusing only on the qualitative method may present an inaccurate representation of patient engagement, considering that the findings of qualitative methods are subject to social desirability and recall bias. For example, in the Health Buddy intervention [34], the interview was conducted approximately 2 years after the intervention was completed. However, the patients may not recall their experiences of using the intervention. Thus, both system usage data and qualitative methods have limitations that may hamper the accurate capture of patient engagement data. However, both methods may complement each other when combined.

Effects of Engagement on HF Outcomes

We determined that the effects of patient engagement on HF outcomes were inconclusive owing to the lack of rigorous analytical methods in the reviewed studies. For example, in 3 studies [37,48,57] that examined the relationship between patient engagement and HF knowledge, only correlation analyses were used. Correlation analysis can be used to summarize sample

characteristics, but an inferential analytical approach is essential for making an inference about a population from a sample. The effects of patient engagement on weight loss, exercise, and HF hospitalizations were examined in only 1 study. Although the findings were promising, there is insufficient evidence to conclude that patient engagement with mHealth is associated with improvements in HF outcomes [11].

Study Implications

The CONSORT-EHEALTH checklist for reporting eHealth and mHealth interventions highly recommends reporting operational definitions of patient engagement [31]. The findings from our study and previous reviews [10,20,69,70] indicate the lack of a standard approach for measuring patient engagement with mHealth interventions. Across studies, different cutoff points were used to indicate effective patient engagement, without any supporting evidence for choosing the cutoff points. To ensure the comparison of findings across studies, addressing the inconsistency in measures of patient engagement should be a key research priority.

International working groups on mHealth have been previously used to develop strategies and policies to support the global implementation of effective mHealth initiatives [71], and unify previous conceptual definitions of patient engagement into an integrative definition of patient engagement [72,73]. Thus, a working group on mHealth could be established to consolidate previous operational definitions of patient engagement into a standardized measure and determine a cutoff point for effective engagement that could be applied across studies. Moreover, when possible, validated self-reported questionnaires of patient engagement with mHealth, such as the Digital Behavior Change Interventions scale [17] and User Engagement Scale [74,75], may be integrated into mHealth interventions in patients with HF to enable comparison of findings across studies [9].

In 80% (24/30) of studies in which system usage data were analyzed, only descriptive statistics were reported as engagement outcomes. Although patient engagement is conceptualized as a dynamic process that changes over time [9,15,44,76], only 3 studies [14,42,44] in our review applied a longitudinal analytical method to analyze system usage data. The application of longitudinal methods in examining system usage may offer an understanding of how patient engagement with mHealth interventions changes within a person over time, and the effects of the interventions on HF outcomes. Thus, future longitudinal studies with methodological rigor are essential to understand the relationship between patient engagement and HF outcomes and the predictors that influence engagement.

Contemporary mobile devices are embedded with third-party analytical applications, such as Google Analytics [77], Amazon Mobile Analytics [78], Android's UsageStatsManager [79], and Apple's Use Screen Time [80]. These applications can capture real-time patterns of patient engagement with mHealth interventions. Surprisingly, only 2 studies in our review used third-party analytical tools to capture patient engagement data. A previous review attributed the minimal usage of analytical applications to investigators' lack of knowledge of how to extract engagement data from the application [10]. Hence, future investigators should consider collaborating with software developers to design effective approaches for using analytical applications to understand patients' patterns of engagement with mHealth interventions.

Strengths and Limitations of the Study

To the best of our knowledge, this is the first scoping review to focus on engagement with mHealth interventions in patients with HF. Unlike previous studies that focused only on quantitative measures [10,69], our review included both objective and subjective measures to capture the wide range of methods that have been used to measure engagement in mHealth interventions among patients with HF.

Our study had some limitations. There was a paucity of studies that examined the relationship between patient engagement with mHealth intervention and HF outcomes, making it challenging to draw conclusions on the effect of the engagement on HF outcomes. The limited number of studies may be related to the small number of articles (N=32) included in our review. The focus of the review on only patients with HF may account for the small number of studies, as mHealth interventions in patients with HF is still at an early stage [8]. Therefore, we conducted a comprehensive literature search with the help of a medical librarian to ensure that all relevant studies were included in the review.

The use of a standardized method to appraise the quality and methodological rigor of the included studies is optional in a scoping review and may be required when the purpose of a review is to appraise the quality of the existing evidence [25]. Considering that the main objective of the present review was to examine the operational definitions of patient engagement with mHealth interventions, a critical appraisal of the existing evidence was not conducted.

In addition, the lack of consistency in the operational definitions of patient engagement in the reviewed studies made it challenging to compare the engagement levels reported across studies. Thus, only the descriptive characteristics of the engagement outcomes are presented in our findings.

Conclusions

This review indicates that engagement with mHealth interventions in patients with HF has been measured using both quantitative and qualitative approaches. There was a lack of consistency in how the quantitative data were measured across the reviewed studies, making comparisons across studies difficult. The effect of mHealth interventions on HF-related outcomes was inconclusive, possibly related to the investigators' use of different and incomplete measures of engagement. More research focusing on developing optimal and standardized measures of patient engagement that may be applied across different study designs is warranted. This will facilitate a deeper understanding of patterns of patient engagement with mHealth interventions that may explain variations in intervention outcomes as well as inform future research and policies regarding mHealth interventions.

Acknowledgments

The authors wish to thank the faculty of the University of Kentucky Medical Library for their assistance in conducting the literature search, especially the nursing liaison librarian, Stephanie Henderson, for her dedication in designing the database search strategies used in the review. This study was not funded.

Conflicts of Interest

The authors disclose that the preliminary findings of this research were previously presented by IM in a meeting (with a published abstract); however, the findings reported in this manuscript are updated.

Multimedia Appendix 1
Conceptual definitions of behavioral, cognitive, and emotional domains.
[DOCX File , 15 KB - mhealth_v10i8e35657_app1.docx]

Multimedia Appendix 2
Databases search strategies and Boolean operators.
[DOCX File , 13 KB - mhealth_v10i8e35657_app2.docx]

Multimedia Appendix 3
The characteristics of mobile health interventions.
[DOCX File , 21 KB - mhealth_v10i8e35657_app3.docx]

Multimedia Appendix 4
The operational definitions of patient engagement with mobile health interventions.
[DOCX File , 21 KB - mhealth_v10i8e35657_app4.docx]

Multimedia Appendix 5
A summary of descriptive characteristics of mobile health (mHealth) intervention engagement metrics and reported outcomes of patient engagement with mHealth interventions.
[DOCX File , 18 KB - mhealth_v10i8e35657_app5.docx]

Multimedia Appendix 6
Reported outcomes of patient engagement with mobile health interventions.
[DOCX File , 22 KB - mhealth_v10i8e35657_app6.docx]

References

1. Puggia I, Rowland TJ, Miyamoto SD, Sinagra G, Mestroni L. Chapter 1 - Molecular and cellular mechanisms in heart failure. In: Jefferies JL, Chang AC, Rossano JW, Shaddy RE, Towbin JA, editors. Heart Failure in the Child and Young Adult: From Bench to Bedside. Boston, MA, USA: Academic Press; 2018:3-19.
2. Virani SS, Alonso A, Benjamin EJ, Bittencourt MS, Callaway CW, Carson AP, American Heart Association Council on Epidemiology and Prevention Statistics Committee and Stroke Statistics Subcommittee. Heart Disease and Stroke Statistics-2020 update: a report from the American Heart Association. Circulation 2020 Mar 03;141(9):e139-e596 [FREE Full text] [doi: 10.1161/CIR.0000000000000757] [Medline: 31992061]
3. Kumar S, Nilsen WJ, Abernethy A, Atienza A, Patrick K, Pavel M, et al. Mobile health technology evaluation: the mHealth evidence workshop. Am J Prev Med 2013 Aug;45(2):228-236 [FREE Full text] [doi: 10.1016/j.amepre.2013.03.017] [Medline: 23867031]
4. Burke LE, Ma J, Azar KM, Bennett GG, Peterson ED, Zheng Y, American Heart Association Publications Committee of the Council on Epidemiology and Prevention, Behavior Change Committee of the Council on Cardiometabolic Health, Council on Cardiovascular and Stroke Nursing, Council on Functional Genomics and Translational Biology, Council on Quality of Care and Outcomes Research, and Stroke Council. Current science on consumer use of mobile health for cardiovascular disease prevention: a scientific statement from the American Heart Association. Circulation 2015 Sep 22;132(12):1157-1213 [FREE Full text] [doi: 10.1161/CIR.0000000000000232] [Medline: 26271892]
5. Free C, Phillips G, Felix L, Galli L, Patel V, Edwards P. The effectiveness of M-health technologies for improving health and health services: a systematic review protocol. BMC Res Notes 2010 Oct 06;3:250 [FREE Full text] [doi: 10.1186/1756-0500-3-250] [Medline: 20925916]

6. Carbo A, Gupta M, Tamariz L, Palacio A, Levis S, Nemeth Z, et al. Mobile technologies for managing heart failure: a systematic review and meta-analysis. Telemed J E Health 2018 Apr 02;24(12):958-968. [doi: 10.1089/tmj.2017.0269] [Medline: 29608430]
7. Son YJ, Lee Y, Lee HJ. Effectiveness of mobile phone-based interventions for improving health outcomes in patients with chronic heart failure: a systematic review and meta-analysis. Int J Environ Res Public Health 2020 Mar 07;17(5):1749 [FREE Full text] [doi: 10.3390/ijerph17051749] [Medline: 32156074]
8. Cajita MI, Gleason KT, Han HR. A systematic review of mHealth-based heart failure interventions. J Cardiovasc Nurs 2016;31(3):E10-E22 [FREE Full text] [doi: 10.1097/JCN.0000000000000305] [Medline: 26544175]
9. Yardley L, Spring BJ, Riper H, Morrison LG, Crane DH, Curtis K, et al. Understanding and promoting effective engagement with digital behavior change interventions. Am J Prev Med 2016 Nov;51(5):833-842. [doi: 10.1016/j.amepre.2016.06.015] [Medline: 27745683]
10. Pham Q, Graham G, Carrion C, Morita PP, Seto E, Stinson JN, et al. A library of analytic indicators to evaluate effective engagement with consumer mHealth apps for chronic conditions: scoping review. JMIR Mhealth Uhealth 2019 Jan 18;7(1):e11941 [FREE Full text] [doi: 10.2196/11941] [Medline: 30664463]
11. Short CE, DeSmet A, Woods C, Williams SL, Maher C, Middelweerd A, et al. Measuring engagement in eHealth and mHealth behavior change interventions: viewpoint of methodologies. J Med Internet Res 2018 Nov 16;20(11):e292 [FREE Full text] [doi: 10.2196/jmir.9397] [Medline: 30446482]
12. Holmen H, Torbjørnsen A, Wahl AK, Jenum AK, Småstuen MC, Arsand E, et al. A mobile health intervention for self-management and lifestyle change for persons with type 2 diabetes, part 2: one-year results from the Norwegian randomized controlled trial RENEWING HEALTH. JMIR Mhealth Uhealth 2014 Dec 11;2(4):e57 [FREE Full text] [doi: 10.2196/mhealth.3882] [Medline: 25499872]
13. Bradway M, Pfuhl G, Joakimsen R, Ribu L, Grøttland A, Årsand E. Analysing mHealth usage logs in RCTs: explaining participants' interactions with type 2 diabetes self-management tools. PLoS One 2018 Aug 30;13(8):e0203202 [FREE Full text] [doi: 10.1371/journal.pone.0203202] [Medline: 30161248]
14. Ware P, Dorai M, Ross HJ, Cafazzo JA, Laporte A, Boodoo C, et al. Patient adherence to a mobile phone-based heart failure telemonitoring program: a longitudinal mixed-methods study. JMIR Mhealth Uhealth 2019 Feb 26;7(2):e13259 [FREE Full text] [doi: 10.2196/13259] [Medline: 30806625]
15. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2017 Jun;7(2):254-267 [FREE Full text] [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
16. O'Brien HL, Toms EG. What is user engagement? A conceptual framework for defining user engagement with technology. J Am Soc Inf Sci 2008 Apr;59(6):938-955. [doi: 10.1002/asi.20801]
17. Perski O, Blandford A, Garnett C, Crane D, West R, Michie S. A self-report measure of engagement with digital behavior change interventions (DBCIs): development and psychometric evaluation of the "DBCI Engagement Scale". Transl Behav Med 2020 Feb 03;10(1):267-277 [FREE Full text] [doi: 10.1093/tbm/ibz039] [Medline: 30927357]
18. Eysenbach G. The law of attrition. J Med Internet Res 2005 Mar 31;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
19. Barello S, Triberti S, Graffigna G, Libreri C, Serino S, Hibbard J, et al. eHealth for patient engagement: a systematic review. Front Psychol 2016 Jan 8;6:2013 [FREE Full text] [doi: 10.3389/fpsyg.2015.02013] [Medline: 26779108]
20. Cheikh-Moussa K, Mira JJ, Orozco-Beltran D. Improving engagement among patients with chronic cardiometabolic conditions using mHealth: critical review of reviews. JMIR Mhealth Uhealth 2020 Apr 08;8(4):e15446 [FREE Full text] [doi: 10.2196/15446] [Medline: 32267239]
21. Athilingam P, Jenkins B. Mobile phone apps to support heart failure self-care management: integrative review. JMIR Cardio 2018 May 02;2(1):e10057 [FREE Full text] [doi: 10.2196/10057] [Medline: 31758762]
22. Masterson Creber RM, Maurer MS, Reading M, Hiraldo G, Hickey KT, Iribarren S. Review and analysis of existing mobile phone apps to support heart failure symptom monitoring and self-care management using the Mobile Application Rating Scale (MARS). JMIR Mhealth Uhealth 2016 Jun 14;4(2):e74 [FREE Full text] [doi: 10.2196/mhealth.5882] [Medline: 27302310]
23. Mortara A, Vaira L, Palmieri V, Iacoviello M, Battistoni I, Iacovoni A, et al. Would you prescribe mobile health apps for heart failure self-care? An integrated review of commercially available mobile technology for heart failure patients. Card Fail Rev 2020 Mar;6:e13 [FREE Full text] [doi: 10.15420/cfr.2019.11] [Medline: 32537246]
24. Armstrong R, Hall BJ, Doyle J, Waters E. Cochrane Update. 'Scoping the scope' of a cochrane review. J Public Health (Oxf) 2011 Mar;33(1):147-150. [doi: 10.1093/pubmed/fdr015] [Medline: 21345890]
25. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 02;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
26. Fan M, Tscheng D, Hamilton M, Hyland B, Reding R, Trbovich P. Diversion of controlled drugs in hospitals: a scoping review of contributors and safeguards. J Hosp Med 2019 Jul;14(7):419-428. [doi: 10.12788/jhm.3228] [Medline: 31251158]

27. Sucharew H, Macaluso M. Progress notes: methods for research evidence synthesis: the scoping review approach. J Hosp Med 2019 Jul 01;14(7):416-418. [doi: 10.12788/jhm.3248] [Medline: 31251164]
28. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]
29. Levac D, Colquhoun H, O'Brien KK. Scoping studies: advancing the methodology. Implement Sci 2010 Sep 20;5:69 [FREE Full text] [doi: 10.1186/1748-5908-5-69] [Medline: 20854677]
30. Peters M, Godfrey C, McInerney P, Soares B, Parker D. Joanna Briggs Institute Reviewer's Manual. Joanna Briggs Institute. 2017. URL: https://wiki.joannabriggs.org/spaces/flyingpdf/flyingpdf.action?key=MANUAL [accessed 2019-12-23]
31. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011 Dec 31;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]
32. Adu MD, Malabu UH, Malau-Aduli AE, Drovandi A, Malau-Aduli BS. User retention and engagement with a mobile app intervention to support self-management in Australians with type 1 or type 2 diabetes (My Care Hub): mixed methods study. JMIR Mhealth Uhealth 2020 Jun 11;8(6):e17802 [FREE Full text] [doi: 10.2196/17802] [Medline: 32525491]
33. Elo S, Kyngäs H. The qualitative content analysis process. J Adv Nurs 2008 Apr;62(1):107-115. [doi: 10.1111/j.1365-2648.2007.04569.x] [Medline: 18352969]
34. LaFramboise LM, Woster J, Yager A, Yates BC. A technological life buoy: patient perceptions of the Health Buddy. J Cardiovasc Nurs 2009;24(3):216-224. [doi: 10.1097/JCN.0b013e318199a60f] [Medline: 19390339]
35. Nundy S, Razi RR, Dick JJ, Smith B, Mayo A, O'Connor A, et al. A text messaging intervention to improve heart failure self-management after hospital discharge in a largely African-American population: before-after study. J Med Internet Res 2013 Mar 11;15(3):e53 [FREE Full text] [doi: 10.2196/jmir.2317] [Medline: 23478028]
36. Zan S, Agboola S, Moore SA, Parks KA, Kvedar JC, Jethwani K. Patient engagement with a mobile Web-based telemonitoring system for heart failure self-management: a pilot study. JMIR Mhealth Uhealth 2015 Apr 01;3(2):e33 [FREE Full text] [doi: 10.2196/mhealth.3789] [Medline: 25842282]
37. Radhakrishnan K, Toprac P, O'Hair M, Bias R, Kim MT, Bradley P, et al. Interactive digital e-Health game for heart failure self-management: a feasibility study. Games Health J 2016 Dec;5(6):366-374 [FREE Full text] [doi: 10.1089/g4h.2016.0038] [Medline: 27976955]
38. Athilingam P, Jenkins B, Johansson M, Labrador M. A mobile health intervention to improve self-care in patients with heart failure: pilot randomized control trial. JMIR Cardio 2017 Aug 11;1(2):e3 [FREE Full text] [doi: 10.2196/cardio.7848] [Medline: 31758759]
39. Buck H, Pinter A, Poole E, Boehmer J, Foy A, Black S, et al. Evaluating the older adult experience of a Web-based, tablet-delivered heart failure self-care program using gerontechnology principles. Geriatr Nurs 2017;38(6):537-541. [doi: 10.1016/j.gerinurse.2017.04.001] [Medline: 28554497]
40. Dang S, Karanam C, Gómez-Orozco C, Gómez-Marín O. Mobile phone intervention for heart failure in a minority urban county hospital population: usability and patient perspectives. Telemed J E Health 2017 Jul;23(7):544-554. [doi: 10.1089/tmj.2016.0224] [Medline: 28051761]
41. Rosen D, McCall JD, Primack BA. Telehealth protocol to prevent readmission among high-risk patients with congestive heart failure. Am J Med 2017 Nov;130(11):1326-1330. [doi: 10.1016/j.amjmed.2017.07.007] [Medline: 28756266]
42. Lloyd T, Buck H, Foy A, Black S, Pinter A, Pogash R, et al. The Penn State Heart Assistant: a pilot study of a Web-based intervention to improve self-care of heart failure patients. Health Informatics J 2019 Jun;25(2):292-303 [FREE Full text] [doi: 10.1177/1460458217704247] [Medline: 28504048]
43. Zhang L, Babu SV, Jindal M, Williams JE, Gimbel RW. A patient-centered mobile phone app (iHeartU) with a virtual human assistant for self-management of heart failure: protocol for a usability assessment study. JMIR Res Protoc 2019 May 23;8(5):e13502 [FREE Full text] [doi: 10.2196/13502] [Medline: 31124472]
44. Haynes SC, Tancredi DJ, Tong K, Hoch JS, Ong MK, Ganiats TG, Better Effectiveness After Transition–Heart Failure (BEAT-HF) Research Group. Association of adherence to weight telemonitoring with health care use and death: a secondary analysis of a randomized clinical trial. JAMA Netw Open 2020 Jul 01;3(7):e2010174 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.10174] [Medline: 32648924]
45. Sohn A, Speier W, Lan E, Aoki K, Fonarow GC, Ong MK, et al. Integrating remote monitoring into heart failure patients' care regimen: a pilot study. PLoS One 2020 Nov 19;15(11):e0242210 [FREE Full text] [doi: 10.1371/journal.pone.0242210] [Medline: 33211733]
46. Apergi LA, Bjarnadottir MV, Baras JS, Golden BL, Anderson KM, Chou J, et al. Voice interface technology adoption by patients with heart failure: pilot comparison study. JMIR Mhealth Uhealth 2021 Apr 01;9(4):e24646 [FREE Full text] [doi: 10.2196/24646] [Medline: 33792556]
47. Kitsiou S, Gerber BS, Kansal MM, Buchholz SW, Chen J, Ruppar T, et al. Patient-centered mobile health technology intervention to improve self-care in patients with chronic heart failure: protocol for a feasibility randomized controlled trial. Contemp Clin Trials 2021 Jul;106:106433 [FREE Full text] [doi: 10.1016/j.cct.2021.106433] [Medline: 33991686]

48. Wei KS, Ibrahim NE, Kumar AA, Jena S, Chew V, Depa M, et al. Habits heart app for patient engagement in heart failure management: pilot feasibility randomized trial. JMIR Mhealth Uhealth 2021 Jan 20;9(1):e19465 [FREE Full text] [doi: 10.2196/19465] [Medline: 33470941]
49. Koehler F, Koehler K, Deckwart O, Prescher S, Wegscheider K, Kirwan BA, et al. Efficacy of telemedical interventional management in patients with heart failure (TIM-HF2): a randomised, controlled, parallel-group, unmasked trial. Lancet 2018 Sep 22;392(10152):1047-1057. [doi: 10.1016/S0140-6736(18)31880-4] [Medline: 30153985]
50. Koehler F, Winkler S, Schieber M, Sechtem U, Stangl K, Böhm M, Telemedical Interventional Monitoring in Heart Failure Investigators. Impact of remote telemedical management on mortality and hospitalizations in ambulatory patients with chronic heart failure: the telemedical interventional monitoring in heart failure study. Circulation 2011 May 03;123(17):1873-1880. [doi: 10.1161/CIRCULATIONAHA.111.018473] [Medline: 21444883]
51. Werhahn SM, Dathe H, Rottmann T, Franke T, Vahdat D, Hasenfuß G, et al. Designing meaningful outcome parameters using mobile technology: a new mobile application for telemonitoring of patients with heart failure. ESC Heart Fail 2019 Jun;6(3):516-525 [FREE Full text] [doi: 10.1002/ehf2.12425] [Medline: 30868756]
52. Seto E, Leonard KJ, Cafazzo JA, Barnsley J, Masino C, Ross HJ. Mobile phone-based telemonitoring for heart failure management: a randomized controlled trial. J Med Internet Res 2012 Feb 16;14(1):e31 [FREE Full text] [doi: 10.2196/jmir.1909] [Medline: 22356799]
53. Dendale P, De Keulenaer G, Troisfontaines P, Weytjens C, Mullens W, Elegeert I, et al. Effect of a telemonitoring-facilitated collaboration between general practitioner and heart failure clinic on mortality and rehospitalization rates in severe heart failure: the TEMA-HF 1 (TElemonitoring in the MAnagement of Heart Failure) study. Eur J Heart Fail 2012 Mar;14(3):333-340 [FREE Full text] [doi: 10.1093/eurjhf/hfr144] [Medline: 22045925]
54. Smeets CJ, Storms V, Vandervoort PM, Dreesen P, Vranken J, Houbrechts M, et al. A novel intelligent two-way communication system for remote heart failure medication uptitration (the CardioCoach Study): randomized controlled feasibility trial. JMIR Cardio 2018 Apr 04;2(1):e8 [FREE Full text] [doi: 10.2196/cardio.9153] [Medline: 31758773]
55. Pedone C, Rossi FF, Cecere A, Costanzo L, Antonelli Incalzi R. Efficacy of a physician-led multiparametric telemonitoring system in very old adults with heart failure. J Am Geriatr Soc 2015 Jun;63(6):1175-1180. [doi: 10.1111/jgs.13432] [Medline: 26031737]
56. Villani A, Malfatto G, Compare A, Della Rosa F, Bellardita L, Branzi G, et al. Clinical and psychological telemonitoring and telecare of high risk heart failure patients. J Telemed Telecare 2014 Dec;20(8):468-475. [doi: 10.1177/1357633X14555644] [Medline: 25339632]
57. Bartlett YK, Haywood A, Bentley CL, Parker J, Hawley MS, Mountain GA, et al. The SMART personalised self-management system for congestive heart failure: results of a realist evaluation. BMC Med Inform Decis Mak 2014 Nov 25;14:109 [FREE Full text] [doi: 10.1186/s12911-014-0109-3] [Medline: 25421307]
58. Scherr D, Kastner P, Kollmann A, Hallas A, Auer J, Krappinger H, MOBITEL Investigators. Effect of home-based telemonitoring using mobile phone technology on the outcome of heart failure patients after an episode of acute decompensation: randomized controlled trial. J Med Internet Res 2009 Aug 17;11(3):e34 [FREE Full text] [doi: 10.2196/jmir.1252] [Medline: 19687005]
59. Hägglund E, Lyngå P, Frie F, Ullman B, Persson H, Melin M, et al. Patient-centred home-based management of heart failure. Findings from a randomised clinical trial evaluating a tablet computer for self-care, quality of life and effects on knowledge. Scand Cardiovasc J 2015 Aug;49(4):193-199. [doi: 10.3109/14017431.2015.1035319] [Medline: 25968968]
60. Piotrowicz E, Dobraszkiewicz-Wasilewska B, Grzebielichowski A, Jasionowska A, Bednarczyk-Banaszak M, Gwilkowska J, et al. New model of home telemonitored cardiac rehabilitation based on nordic walking training in patients with heart failure: safety, effectiveness and compliance. Circulation 2012 Apr 15;125(19):e61 [FREE Full text] [doi: 10.1161/CIR.0b013e31824fcd6b]
61. Leng Chow W, Aung CY, Tong SC, Goh GS, Lee S, MacDonald MR, et al. Effectiveness of telemonitoring-enhanced support over structured telephone support in reducing heart failure-related healthcare utilization in a multi-ethnic Asian setting. J Telemed Telecare 2020 Jul;26(6):332-340. [doi: 10.1177/1357633X18825164] [Medline: 30782070]
62. Guo X, Gu X, Jiang J, Li H, Duan R, Zhang Y, et al. A hospital-community-family-based telehealth program for patients with chronic heart failure: single-arm, prospective feasibility study. JMIR Mhealth Uhealth 2019 Dec 13;7(12):e13229 [FREE Full text] [doi: 10.2196/13229] [Medline: 31833835]
63. Ding H, Jayasena R, Chen SH, Maiorana A, Dowling A, Layland J, et al. The effects of telemonitoring on patient compliance with self-management recommendations and outcomes of the innovative telemonitoring enhanced care program for chronic heart failure: randomized controlled trial. J Med Internet Res 2020 Jul 08;22(7):e17559 [FREE Full text] [doi: 10.2196/17559] [Medline: 32673222]
64. Deka P, Pozehl B, Williams MA, Norman JF, Khazanchi D, Pathak D. MOVE-HF: an Internet-based pilot study to improve adherence to exercise in patients with heart failure. Eur J Cardiovasc Nurs 2019 Feb;18(2):122-131. [doi: 10.1177/1474515118796613] [Medline: 30129790]
65. Burnet K, Kelsch E, Zieff G, Moore J, Stoner L. How fitting is F.I.T.T.?: a perspective on a transition from the sole use of frequency, intensity, time, and type in exercise prescription. Physiol Behav 2019 Feb 01;199:33-34. [doi: 10.1016/j.physbeh.2018.11.007] [Medline: 30414400]

66. DeSimone GT. The Tortoise factor — get FITT. ACSM's Health Fit J 2019 Sep 15;23(2):3-4. [doi: 10.1249/fit.0000000000000456]
67. Dipla K. The FITT principle in individuals with type 2 diabetes: from cellular adaptations to individualized exercise prescription. J Adv Med Med Res 2017 Jul 15;22(11):1-18. [doi: 10.9734/jammr/2017/34927]
68. Flaherty SJ, McCarthy M, Collins AM, McCafferty C, McAuliffe FM. Exploring engagement with health apps: the emerging importance of situational involvement and individual characteristics. Eur J Market 2021 May 06;55(13):122-147. [doi: 10.1108/ejm-06-2019-0531]
69. Sieverink F, Kelders SM, van Gemert-Pijnen JE. Clarifying the concept of adherence to eHealth technology: systematic review on when usage becomes adherence. J Med Internet Res 2017 Dec 06;19(12):e402 [FREE Full text] [doi: 10.2196/jmir.8578] [Medline: 29212630]
70. Nelson LA, Coston TD, Cherrington AL, Osborn CY. Patterns of user engagement with mobile- and Web-delivered self-care interventions for adults with T2DM: a review of the literature. Curr Diab Rep 2016 Jul;16(7):66 [FREE Full text] [doi: 10.1007/s11892-016-0755-1] [Medline: 27255269]
71. K MP, Ba H, Bouverot A, Childs R, Gnoumou S, Kapila N, et al. m-Powering Development Initiative Report of the Working Group on m-Health. International Telecommunication Union. 2014. URL: https://www.itu.int/en/ITU-D/Initiatives/m-Powering/Documents/mHealth_Report_of_the_Working_Group.pdf [accessed 2022-04-22]
72. Michie S, Yardley L, West R, Patrick K, Greaves F. Developing and evaluating digital interventions to promote behavior change in health and health care: recommendations resulting from an international workshop. J Med Internet Res 2017 Jun 29;19(6):e232 [FREE Full text] [doi: 10.2196/jmir.7126] [Medline: 28663162]
73. Edney S, Ryan JC, Olds T, Monroe C, Fraysse F, Vandelanotte C, et al. User engagement and attrition in an app-based physical activity intervention: secondary analysis of a randomized controlled trial. J Med Internet Res 2019 Nov 27;21(11):e14645 [FREE Full text] [doi: 10.2196/14645] [Medline: 31774402]
74. O'Brien HL, Cairns P, Hall M. A practical approach to measuring user engagement with the refined user engagement scale (UES) and new UES short form. Int J Human Comput Stud 2018 Apr;112:28-39 [FREE Full text] [doi: 10.1016/j.ijhcs.2018.01.004]
75. O'Brien HL, Toms EG. Examining the generalizability of the User Engagement Scale (UES) in exploratory search. Inf Process Manag 2013 Sep;49(5):1092-1107. [doi: 10.1016/j.ipm.2012.08.005]
76. Taki S, Lymer S, Russell CG, Campbell K, Laws R, Ong KL, et al. Assessing user engagement of an mHealth intervention: development and implementation of the growing healthy app engagement index. JMIR Mhealth Uhealth 2017 Jun 29;5(6):e89 [FREE Full text] [doi: 10.2196/mhealth.7236] [Medline: 28663164]
77. Google Analytics. Google. URL: https://analytics.google.com/analytics/web/?authuser=0#/report-home/a54654253w87630799p90987413 [accessed 2022-07-15]
78. Tracking App Usage Data with Amazon Mobile Analytics. Amazon Web Services. URL: https://docs.aws.amazon.com/mobile/sdkforxamarin/developerguide/getting-started-analytics.html [accessed 2022-07-15]
79. UsageStatsManager. Android. URL: https://developer.android.com/reference/android/app/usage/UsageStatsManager [accessed 2022-07-15]
80. Use Screen Time on your iPhone, iPad, or iPod touch. Apple. URL: https://support.apple.com/en-us/HT208982 [accessed 2022-07-15]

Abbreviations

CONSORT-EHEALTH: Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth
FITT: frequency, intensity, time, and type
HF: heart failure
mHealth: mobile health
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews

Edited by L Buis; submitted 13.12.21; peer-reviewed by F Van Baelen, H Li, J Vranken; comments to author 22.01.22; revised version received 04.05.22; accepted 10.06.22; published 22.08.22.

Please cite as:
Madujibeya I, Lennie T, Aroh A, Chung ML, Moser D
Measures of Engagement With mHealth Interventions in Patients With Heart Failure: Scoping Review
JMIR Mhealth Uhealth 2022;10(8):e35657
URL: https://mhealth.jmir.org/2022/8/e35657
doi:10.2196/35657
PMID:35994345

Review

mHealth Apps for Low Back Pain Self-management: Scoping Review

Aki Rintala[1], PhD; Roy Rantalainen[1], BSc; Anu Kaksonen[1], MSc; Hannu Luomajoki[2], PhD; Kari Kauranen[3], PhD

[1]Faculty of Social Services and Health Care, LAB University of Applied Sciences, Lahti, Finland

[2]Department of Health, Zürich University of Applied Sciences ZHAW, Institute for Physiotherapy, Winterthur, Switzerland

[3]Faculty of Social Services and Health Care, LAB University of Applied Sciences, Lappeenranta, Finland

Corresponding Author:
Aki Rintala, PhD
Faculty of Social Services and Health Care
LAB University of Applied Sciences
Mukkulankatu 19
Lahti, 15120
Finland
Phone: 358 447085254
Email: aki.rintala@lab.fi

Abstract

Background: The role of self-management in health promotion, as well as prevention and rehabilitation, is increasing through the use of mobile health (mHealth) apps. Such mHealth apps are also increasingly being used for self-management of low back pain (LBP), but their effectiveness has not been sufficiently explored.

Objective: The aim of this scoping review was to provide an overview of the literature on self-management mHealth apps and their effects on the levels of pain and disability in people with LBP.

Methods: We applied the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) methodology, including a priori research questions. A literature search was conducted in 2 databases (PubMed and PEDro) for studies published between January 1, 2015, and June 17, 2021. Interventional, cohort, or case series studies with an interventional period were included if the mHealth app included built-in self-management content, the app was used for self-management for people with LBP, and the study reported outcomes regarding pain and disability in people with LBP.

Results: In total, 7 studies were selected for the review with overall 2307 persons with LBP, of whom 1328 (57.56%) were women. Among the studies (5/7, 71%) that reported the type of pain, 85% (390/459) of the participants were experiencing chronic LBP. A total of 5 different mHealth apps were identified, of which 4 contributed to a statistically significant reduction in LBP and clinically meaningful changes. Of the 7 studies, 4 (57%) used 4 different assessments for disability, of which 3 (75%) showed statistically significant improvements in the level of functional ability of participants in the experimental groups using an mHealth app with built-in self-management content for LBP.

Conclusions: This scoping review supports the conclusion that people with LBP may benefit from mHealth apps that provide self-management content. However, the generalizability of the findings is limited because of heterogeneity in the pain characterization of the included participants and the intervention durations. More high-quality studies with longer follow-up periods to investigate personalized mHealth approaches are recommended for LBP self-management.

(JMIR Mhealth Uhealth 2022;10(8):e39682) doi:10.2196/39682

KEYWORDS

low back pain; mobile health; mHealth; app; disability; self-management; mobile phone

Introduction

Background

Low back pain (LBP) is one of the greatest concerns in health care worldwide, and it is one of the major factors in a decline in overall function [1,2]. Almost 80% of the world's population will encounter LBP at some period in their lives, and approximately 50% will experience multiple pain periods during their lifetime [1]. In 2019, the number of prevalent LBP cases was shown to increase with age, with LBP peaking at age 45

to 54 years for both genders [3]. The origin of LBP is still unclear in the literature, but many factors contribute to its existence, such as an individual's genome, obesity, smoking, sedentary behavior, physical labor, work posture, and excessive sitting, as well as psychological factors such as stress [3,4].

Self-management is an important treatment strategy for health promotion. Self-management is defined as any treatment method that improves or maintains health, prevents disease, and supports management of disease or disability [5,6]. The active participation of patients in their care of the symptoms or illness aiming to prevent the progression of medical conditions plays a key role in self-management [5]. Self-management content is usually focused on ergonomics, weight management, behavioral changes, and physical activity [6]. Self-management also plays a role in LBP rehabilitation, and it has been found to have a small-to-moderate effect on decreasing the levels of pain and disability in people with LBP [7].

The rapid advances in new technology have also led to a merger of new therapy approaches in rehabilitation and self-management. Mobile health (mHealth) is defined as a health and well-being mobile service that enables 2-way health-related information delivery and communication [8]. A search for mobile apps in the Apple App Store and Google Play Store at the end of 2021 revealed that >5 million different mobile apps were available, of which >100,000 apps were related to mHealth content. Furthermore, there were >500 million users worldwide using services related to mHealth apps already in 2011 [9]. These statistics show that mHealth apps are already an important part of people's everyday life and will play an increasing role in the management of their medical care because the number of mHealth apps and their use is expected to increase in the near future.

Self-management interventions have been studied with supporting evidence as part of LBP treatment [7,10]. However, studies investigating the effects of mHealth apps with built-in self-management content are still lacking. A previous review focusing on eHealth (web-based and mobile-based) self-management programs for LBP found preliminary evidence from the subgroup meta-analysis consisting of 3 studies that supported the claim that mHealth programs may have a role in decreasing the levels of pain and disability [10]. However, the meta-analysis included only 3 randomized controlled trials (RCTs), which included mHealth apps without built-in self-management content. Only 1 systematic review has identified the commercial use of mHealth apps related to self-management for people with LBP, and it provided an overview of existing mHealth apps and their content [1]. Machado et al [1] found 61 commercially available mHealth apps targeting LBP in the Australian iTunes Store and Google Play Store. The content of the mHealth apps for LBP management was designed to provide a wide variety of exercises related to strengthening, stretching, range of motion, motor control, Pilates or the McKenzie method, yoga, tai chi, and mindfulness, or a combination of these [1]. However, it is still unclear whether such apps can be used in clinical settings.

Objectives

Health care professionals have called for efficient tools that can be used to motivate and engage their patients in managing their LBP [1,11]. An updated overview is lacking for studies on mHealth apps with built-in self-management content and their effects on the levels of LBP and disability. New technology approaches such as mHealth apps may support traditional pain management and care and, in turn, improve patients' abilities to self-manage their LBP in the home environment setting. The objective of this scoping review was to map the number of current mHealth apps used in research settings and to identify their effects on the levels of pain and disability.

Methods

Search Strategy

We conducted a literature search for studies published between January 1, 2015, and June 17, 2021, in the PubMed and PEDro databases. The literature search was also expanded to manual search, using the same search terms, in Google Scholar and reference lists of the retrieved articles. The search strategy focused on health care interventions in the form of mHealth apps and, therefore, did not include the development, construction, or evaluation of the technology itself. The search strategy contained general terms related to mHealth technology (ie, "smartphone app," "technology," "health app," and "mobile health"), LBP (ie, "low back pain," "low back ache," "back pain," "lumbago," "acute lower back pain," and "chronic lower back pain"), and self-management (ie, "self-management" and "self-care"). As the databases differed in terms of technical search options, in the PEDro database, we used the following simple search terms: "back pain mobile self," "back pain mhealth," and "back pain smartphone."

The literature search was conducted by 2 reviewers (RR and AR) who screened and assessed the published articles independently. The screening was conducted with a predefined strategy, which involved first screening the titles and abstracts, followed by an assessment of the included studies based on a full-text reading by the 2 reviewers of the research team (RR and AR). A list of the studies to be included was agreed upon after resolving disagreements on eligibility through discussions. In case of disagreement, a third reviewer (AK) evaluated the studies. After including relevant studies, both reviewers also reviewed the references of the included studies for additional relevant publications.

Inclusion and Exclusion Criteria

The literature search was limited to peer-reviewed articles published between January 1, 2015, and June 17, 2021. Studies were included in this review if they aimed to explore the use of mHealth apps for self-management in people with LBP. The definition of self-management and self-care was derived from the World Health Organization guideline on self-care interventions for health and well-being [12]: "the ability of individuals, families and communities to promote health, prevent disease, maintain health and to cope with illness and disability with or without the support of a health worker." The scope was limited to health promotion; disease prevention and control

relates to providing care to dependent persons in a rehabilitation setting.

We included any type of interventional study, cohort study, or possible case series, including an interventional period for >1 participant. Studies also needed to be published in English. Studies were excluded if the mHealth apps did not include self-management content (as described previously), did not include an interventional study period, or if participants had been diagnosed with pain other than LBP. In addition, we excluded study protocols, opinion articles, and studies other than interventional, cohort, or case series (ie, reviews, case studies, and qualitative studies).

Data Extraction

We extracted the predefined data from the included studies after the full-text screening. Data charting was conducted first by 1 member of the research team and then reviewed by another member of the research team. Predefined data included study details (authors, year, study method, objectives, country, measurement of pain, and number of participants), mHealth details (name and content of the mHealth app), and personal (gender and age) and clinical (type of pain and duration of pain) characteristics. Type of pain was defined as acute (<6 weeks), subacute (6-12 weeks), or chronic (>12 weeks). If needed, research team members contacted the corresponding authors of the included studies for additional inquiry if the aforementioned data were not reported adequately in the original article.

Quality Assessment

Two reviewers assessed the methodological quality of the selected studies using JBI's critical appraisal tool [13,14]. In case of disagreement, a third reviewer evaluated the methodological quality of the selected study. Depending on the type of study methods, the number of methodological questions ranged from 11 to 13 items; for example, the maximum score for RCTs is 13 points, for case-cohort studies 11 points, and for non-RCTs 9 points. Each item was rated *yes*, *no*, *unclear*, or *not applicable*. Total and mean scores of items rated *yes* (1 point) were computed for each included study.

Data Synthesis

Extraction data were analyzed descriptively, and if applicable, a frequency analysis was conducted for retrieved data related to study characteristics (number of participants [n]), personal characteristics (age [mean, range] and gender [n, %]), and clinical characteristics (type of pain [n, %] and pain duration [mean, range]). The effects of mHealth apps on the levels of pain and disability were calculated with a vote-counting analysis.

Rationale

We conducted a scoping review following the JBI's Manual for Evidence Synthesis and the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) [15,16]. For developing the research questions, we used JBI's population, concept, and context framework to formulate the primary review questions. The primary research questions in this review are as follows:

1. How many self-management mHealth apps have been used in research settings for people with LBP?
2. What type of self-management content do mHealth apps have for people with LBP?
3. What are the effects of using self-management mHealth apps on the levels of pain and disability in people with LBP?

Results

Overview

Of the initial 87 studies identified in the literature search, 4 (5%) duplicates were removed. Of the 83 remaining studies, 73 (88%) that did not meet the inclusion criteria after title and abstract screening were excluded. The screening of the included 10 full-text studies revealed 7 (70%) that met the inclusion criteria [17-23]. Of these 7 studies, 4 (57%) were RCTs, 2 (29%) were cohort studies, and 1 (14%) was a non-RCT (Table 1). A flowchart of the study selection is presented in Figure 1. Of the 7 studies, 4 (57%) were conducted in Europe [19,20,22,23], and 1 (14%) each in the United States, Jordan, and India [17,18,21].

Table 1. Study details and results of studies using a mobile health (mHealth) app for low back pain (LBP).

Study, country	Content of the study	Participants; woman, n (%), age	Type of pain: n (%); duration; medication	Name of the mHealth app	Content of mHealth app	Outcome type of the levels of pain and disability	Effects of the mHealth app on the levels of pain and disability	Vote-counting, pain;disability
RCTs[a]								
Irvine et al [21], 2015, United States	Investigating the efficacy of the mHealth app to guide user implementation of personalized strategies for LBP management and prevention (16 weeks); group 1: mHealth app; group 2: web-based email support for the LBP program; group 3: no LBP program (control)	Workers; group 1: 199 (58), age not provided; group 2: 199 (59), age not provided; group 3: 199 (63), age not provided	Not reported	FitBack	Personalized content depending on whether the person on a daily average basis is sitting, standing, driving, or lifting; general well-being; mindfulness exercises; strength exercises; stretching exercises; diary	How bad is your LBP? (6-point Likert scale); how often have you experienced LBP? (6-point Likert scale); when you experienced LBP, on average how intense was the pain? (7-point Likert scale); when you experienced LBP, on average how long did it usually last? (5-point Likert scale)	Back pain measures: group 1 level of pain decreased by 0.4 points at 16 weeks, in group 2 by 0.3 points, and in group 3 by 0.1 points; group 1 vs group 2=not statistically significant difference after 16 weeks (*P*=.17), at 8 weeks not tested; group 1 vs group 3=statistically significant differences after 16 weeks (*P*=.002), at 8 weeks not tested	+;—[b]
Chhabra et al [18], 2018, India	Investigating the effect of using an mHealth app on pain and function in patients with chronic LBP (12 weeks); group 1: mHealth app; group 2: conventional group receiving a written prescription from the physician and a list of prescribed medicines and dosages	People with chronic LBP; group 1: 45 (not reported), 41; group 2: 48 (not reported), 41	Type of pain: chronic (>12 weeks): 93 (100); duration: group 1: 23 months; group 2: 28 months; medication: not reported	Snapcare	Personalized set of exercises based on the health status of the user using gamification: physical activity goals (eg, daily steps); home therapeutic exercises; possibility of progression based on the use of the app; focus on increasing daily life activities and increasing basic routines as independently as possible	NRS[c] (0-10); MODI[d] (0-50); CSS[e] (0-25)	NRS: pain decreased by 4.0 points in group 1 and by 3.4 points in group 2 at 12 weeks; no statistically significant group differences (*P*=.23); MODI: functional ability improved by 31.9 points in group 1 and by 11.5 points in group 2 at 12 weeks; difference was statistically significant (*P*<.001); CSS improved by –9.0 points in group 1 at 12 weeks compared with baseline (*P*<.001); group differences were not tested	0;+

Study, country	Content of the study	Participants; woman, n (%), age	Type of pain: n (%); duration; medication	Name of the mHealth app	Content of mHealth app	Outcome type of the levels of pain and disability	Effects of the mHealth app on the levels of pain and disability	Vote-counting, pain;disability
Toelle et al [23], 2019, Germany	Investigating the clinical effects of a multidisciplinary mHealth app for LBP (12 weeks); group 1: mHealth app; group 2: 6 physiotherapy sessions and web-based education	Adults with LBP; group 1: 48 (73), 41; group 2: 46 (67), 43	Type of pain: chronic (>12 weeks): 94 (100); duration: group 1: 7.2 months; group 2: 6.7 months; medication: MQS[f], group 1: 2.4; group 2: 2.8	Kaia app	Therapeutic exercises; mindfulness exercises; education regarding LBP; possibility of progression	NRS (0-11); HFAQ[g]	NRS: pain decreased by 2.4 points in group 1 and by 2.0 points in group 2 at 12 weeks; group difference was statistically significant in favor of the mHealth app group (*P*=.02); group differences were not statistically significant at baseline nor after 6 weeks (*P*>.05); within groups, both groups showed a significant decrease in pain symptoms over time (baseline vs 6 weeks and 6 weeks vs 12 weeks, all *P*<.01); HFAQ: no statistically significant differences between the groups (*P*>.05)	+;0
Almhdawi et al [17], 2020, Jordan	Investigating the efficacy of a newly developed evidence-based mHealth app for LBP management (6 weeks); group 1: mHealth app; group 2: placebo app containing nutritional facts with no LBP management	Workers; group 1: 21 (34), 41; group 2: 20 (20), 42	Type of pain: chronic (>12 weeks): 41 (100); VAS[h]>3.0; duration: pain chronicity>3 months; medication: not reported	Relieve My Back	Education regarding LBP (general advice and instruction); home therapeutic exercises for lower back and abdominal muscles; stretching exercises for lower back and abdominal muscles	VAS (0-10); ODI[i] (0-100)	VAS: pain decreased by –3.5 in group 1 and by –0.1 in group 2 at 6 weeks; group difference was statistically significant in favor of the mHealth app group (*P*<.001); ODI: functional ability improved by 11.5 points in group 1 and by 0.6 points in group 2 at 6 weeks; the difference was statistically significant in favor of the mHealth app group (*P*=.002)	+;+
Cohort studies								
Huber et al [20], 2017, Germany	Investigating short-term changes effected by an mHealth app for the treatment of LBP (12 weeks)	Users of the Kaia app with a history of medical treatment of back pain and no history of specific back pain; 180 (58), 34	Type of pain: acute (<6 weeks): 25 (14); subacute (6-12 weeks): 23 (13); chronic (>12 weeks): 132 (73); duration: not reported; medication: not reported	Kaia app	Therapeutic exercises; mindfulness exercises; education regarding LBP; possibility of progression	NRS (0-10)	NRS score decreased at 4 weeks by 1.3 points (*P*<.001), at 8 weeks by 1.5 points (*P*<.001), and at 12 weeks by 2.0 points (*P*=.21), with no difference between pain types (*P*>.30)	—

Study, country	Content of the study	Participants; woman, n (%), age	Type of pain: n (%); duration; medication	Name of the mHealth app	Content of mHealth app	Outcome type of the levels of pain and disability	Effects of the mHealth app on the levels of pain and disability	Vote-counting, pain;disability
Clement et al [19], 2018, Germany	Investigating the effect on user retention and clinical outcomes of the Kaia app during development between 2 groups (24 weeks); users were grouped depending on the available version at the time of the sign-up; group 1: older version (0.x) of Kaia app; group 2: new version (1.x)	Users of the Kaia app with a history of medical treatment of back pain and no history of specific back pain; group 1: 196 (58), age not provided; group 2: 1055 (49), age not reported	Not reported	Kaia app	Pain self-management app containing several domains with the possibility of personalization; therapeutic exercises; mindfulness exercises; education regarding LBP; pain diary and self-test; chat; feedback system available for training and pain levels; possibility of progression	NRS (0-10)	Levels of pain decreased in both groups after 24 weeks: group 1 by 0.9 points and group 2 by 1.2 points with no difference between the groups (P=.29); within the group, the decrease in pain was statistically significant in group 1 (P=.008)	—
Non-RCT								
Sandal et al [22], 2020, Denmark and Norway	Investigating the basis for recruitment and screening procedures to explore the associations between the inclusion process and questionnaires and app installation and to examine the changes in clinical outcomes (6 weeks)	People with LBP within the past 8 weeks; 51 (58), 46	Type of pain: acute (<6 weeks): 11 (22); subacute (6-12 weeks): 10 (20); chronic (>12 weeks): 30 (58); duration: not reported; medication: infrequent use: 51 (58)	selfBACK	Weekly personalized self-management plans: physical activity (number of steps per day); strength and mobility exercises; mindfulness exercises; education regarding LBP; goal setting	NRS (0-10) average past week; NRS (0-10) worst past week; pain-related disability (RMDQ[j])	NRS average past week: pain decreased by 1.0 point (95% CI –1.6 to –0.4); NRS worst past week: pain decreased by 1.0 point (95% CI –1.6 to –0.4); RMDQ: functional ability improved by 1.8 points (95% CI –2.9 to –0.7)	—

[a]RCT: randomized controlled trial.

[b]Not available.

[c]NRS: numeric rating scale (0-10 with higher scores indicating worse pain).

[d]MODI: Modified Oswestry Disability Index.

[e]CSS: current symptom score.

[f]MQS: Medication Quantification Scale.

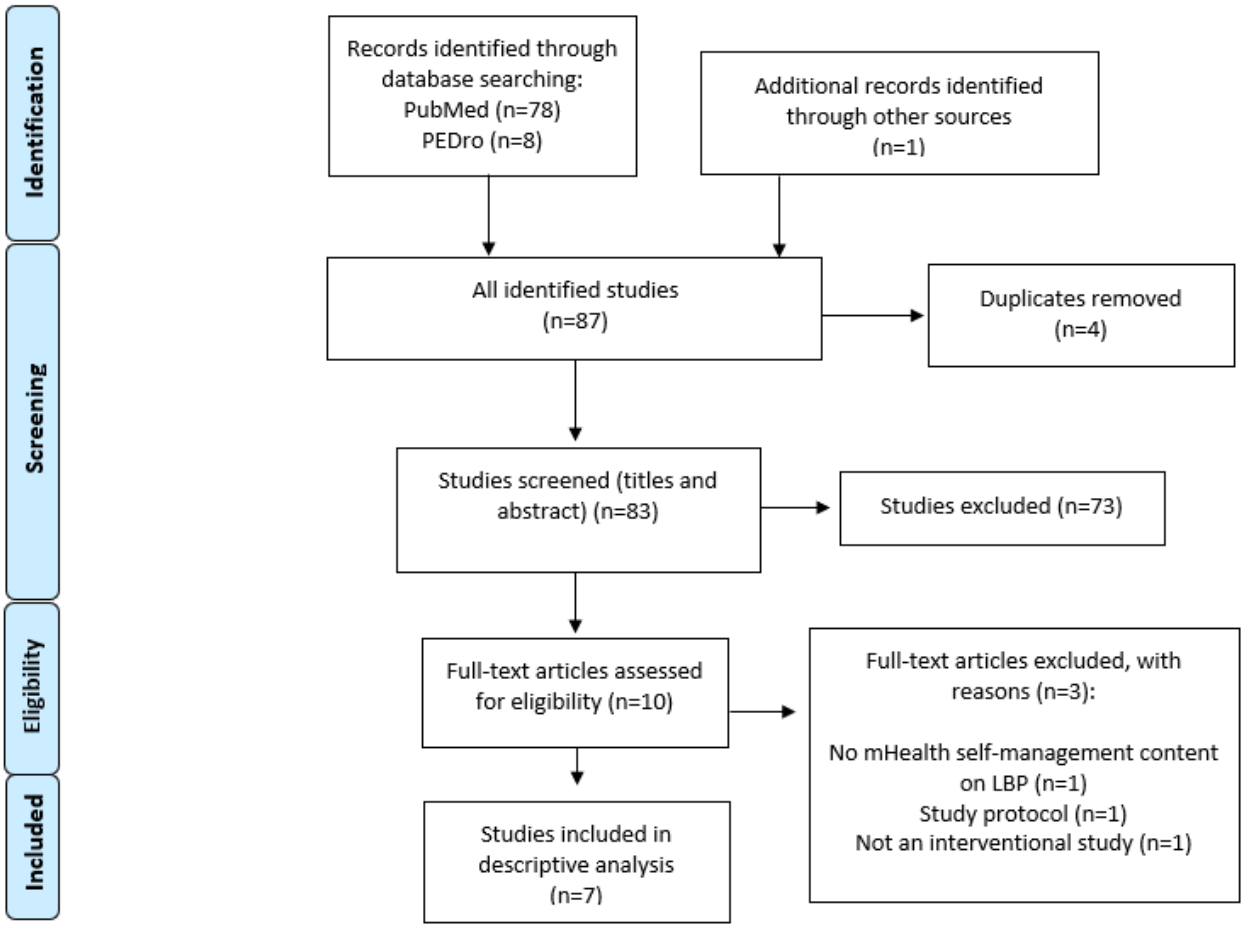
[g]HFAQ: Hannover Functional Ability Questionnaire.

[h]VAS: visual analog scale.

[i]ODI: Oswestry Disability Index.

[j]RMDQ: Roland-Morris Disability Questionnaire.

Figure 1. Flow chart of study selection. LBP: low back pain; mHealth: mobile health.



Description of the Participants

The selected studies included 2307 people with LBP, of whom 825 (35.76%) were included in RCTs, 1431 (62.03%) in cohort studies, and 51 (2.21%) in a non-RCT. Of the 825 participants with LBP in RCTs, 313 (37.9%) were included in the experimental group and 512 (62.1%) in the control group. In the total sample, the mean age of the participants was 40.7 (SD 4.2; range 40-46) years, and 57.56% (1328/2307) were women. Of the 7 studies, 2 (29%) did not report age [19,21], and 1 (14%) did not report gender distribution [18].

Of the 7 studies, 5 (71%) included 390 participants with chronic LBP [17,18,20,22,23]; of these 5 studies, 2 (40%) also included people with acute (n=36) and subacute (n=33) LBP [20,22]. In total, 14% (2/7) of the studies did not report the type of pain; both these studies together accounted for the highest number of participants (n=1848) [19,21]. The exact duration of pain at baseline was only reported in 14% (2/7) of the studies (Table 1).

The main inclusion criteria for eligibility to participate in the included studies varied. The inclusion criteria were as follows: age (>18 years), ability to use a smartphone, and experiencing any level of LBP in the past days or months [17-23]. Other criteria included a minimum pain score of 4 on the visual analog scale (VAS) or ≥4 to ≥5 points on the numeric rating scale (NRS) scale [17,18,23], sufficient level of self-reported physical fitness [18-20,22], a declaration of medical treatment for back pain [18-20], or a declaration of no medical treatment for back pain [21]. In total, 57% (4/7) of the studies included participants with nonspecific LBP as a definition of perceived pain in the lower back region that was not attributable to a recognizable specific pathology (eg, infection, tumor, osteoporosis, lumbar spine fracture, structural deformity, inflammatory disorder, cauda equina syndrome, ankylosing spondylitis, spondylitis, spondylolisthesis, spondyloarthritis, spinal stenosis, or spinal disk herniation) [19,20,22,23].

Methodological Quality

The overall methodological quality of the studies is described in Multimedia Appendix 1 [17-23]. The methodological quality of the RCTs (n=4) varied from fair to good. A general drawback was the blinding of outcome assessors and reporting of the outcome assessment reliability procedures. This was expected because the blinding procedure is difficult to achieve (outcome assessors as well as participants) in these types of interventional studies. There were similar issues with the cohort studies (n=2) because they did not clearly report confounding factors or strategies thereof; furthermore, strategies for completion were not always reported.

Self-management Content of mHealth Apps

This review identified 5 different mHealth apps containing self-management content for people with LBP (Table 1). Content varied based on the app and the level of reporting in each study, but commonalities were personalization, increasing as well as

monitoring daily life activities and physical activity, targeted home exercises (strengthening and stretching), mindfulness training, and education regarding LBP. Of the 7 studies, 4 (57%) reported the possibility of personalizing the mHealth app content in the apps (Kaia app, Snapcare, FitBack, and SelfBACK) [18,19,21,22], and 3 (43%) studies reported a possibility for participants to build a training progression in 2 apps (Kaia app and Snapcare) [18,20,23]. More specified description of the app versions, use, training progression, and personalization of the content was not reported. We have listed the main content for each app in the following paragraphs.

Kaia app was used in 43% (3/7) of the studies [19,20,23]. The app was designed to include three domains: (1) back pain–specific education, (2) physiotherapy (a pool of 145 exercises adapted to the user's fitness level), and (3) mindfulness exercises. The app also contained units dedicated to breathing techniques, body scan, and progressive muscle relaxation. Users had the possibility to optimize the content on a daily basis depending on their status with regard to knowledge, practice, and progress. Educational content was focused on general pain–related and back pain–specific education (overall, there were 30 different educational units).

The remaining (4/5, 80%) apps—SelfBACK [22], FitBack [21], Relieve My Back [17], and Snapcare [18]—were used in individual studies.

SelfBACK contained weekly user-tailored general physical activity, strengthening and flexibility exercises, and patient education [22]. Other minor units in the app were a goal-setting tool, audio mindfulness exercises, pain-relieving exercises, and general information about LBP.

FitBack included a self-tailored cognitive behavioral approach where the main focus was to monitor the levels of pain and activity and to provide LBP-related in-app text and video messages [21]. The content was administered based on the user's job type (sitting most of the day, standing most of the day, driving most of the day, or lifting most of the day). Other units in the app were unlimited access to pain management education, instructional videos on strengthening and stretching exercises tailored based on the job type, and live web-based streaming instructions on ergonomics and exercises.

Relieve My Back contained general advice and instructions to conduct exercises at work and home. Office-based exercises were focused on stretching, whereas home-based exercises in the evening included strengthening exercises, both focusing on the lower back and abdominal muscles [17]. The app also provided prompts to remind the users to take a walk break, check their posture, and perform the exercises.

Snapcare included 2 main units: monitoring the levels of daily activities and monitoring the user's symptomatic profile [18]. Daily activity goals, including back and aerobic exercises, were developed based on the user's health status, activities of daily living, and daily activity progress.

mHealth Interventions

Interventions in RCTs

The training periods ranged from 6 to 16 weeks (Table 1). Participants used solely a smartphone-based mHealth solution in their interventions. Regarding the environmental settings of the mHealth interventions, all interventions were applied in an at-home environment. mHealth interventions were compared with either web-based email support for LBP and no rehabilitation [21], conventional training [18], physiotherapy and web-based education [23], or a placebo app containing nutritional facts without LBP self-management [17].

Interventions in the Cohort Studies and Non-RCT

The cohort studies [19,20] and non-RCT [22] followed participants' use of mHealth apps for 24 weeks, 12 weeks, and 6 weeks, respectively (Table 1). Of these 3 studies, 2 (67%) focused on user retention and clinical outcomes of pain [19,22], and 1 (33%) investigated short-term changes in LBP using an mHealth app [20].

Assessments of Pain and Disability

The clinical outcomes of pain are described in Table 1. Of the 7 studies, 6 (86%) used either the VAS or the NRS to measure self-reported LBP. Of these 6 studies, 1 (17%) used 4 items, including 5- and 7-point Likert scales, to determine the current level, frequency, intensity, and duration of LBP [21], and 4 (67%) determined the level of disability related to LBP using the Modified Oswestry Disability Index, Hannover Functional Ability Questionnaire, Oswestry Disability Index, or Roland-Morris Disability Questionnaire [17,18,22,23].

Effects of Using mHealth Apps for LBP Self-management on the Level of Pain

RCTs

Of the 4 RCTs, 3 (75%) reported statistically significant changes in decreases in the level of LBP in favor of the mHealth app group compared with the physiotherapy and web-based education [23], web-based email support group or no training [21], or placebo (nutritional) [17] groups (Table 1), whereas 1 (25%) did not show statistically significant changes in the level of pain between the groups when mHealth app interventions were compared with conventional training [18].

Cohort Studies and Non-RCT

All (3/3, 100%) of the studies reported a statistically significant decrease in LBP in participants using mHealth apps for 6 weeks [22], 4 and 8 weeks [20], and 24 weeks [19]. Of the 3 studies, 1 (33%) found a decrease in the level of pain at 12 weeks, but this was not statistically significant when observing the main effect of time for the pain ratings [20].

Effects of Using mHealth Apps for LBP Self-management on the Level of Disability

The level of disability was assessed in 57% (4/7) of the included studies using 4 different disability assessments (Table 1). Of the 3 RCTs, 2 (67%) reported a statistically significant change in improving functional ability compared with conventional training [18] or placebo (nutritional) group [17], whereas 1

(33%) did not report statistically significant differences in functional ability between mHealth app intervention and traditional physiotherapy [23]. The non-RCT showed a statistically significant improvement in functional ability for participants in the mHealth intervention group over 6 weeks [22].

Discussion

Principal Findings

This scoping review found 5 mHealth apps for LBP self-management that were used in research settings (n=7 studies) to investigate their effects on the levels of LBP and disability. The majority of the studies reported promising evidence of the effects of the mHealth apps on decreasing the levels of pain (6/7, 86%) and disability (3/4, 75%) when the focus of the studies was on self-managing LBP. However, heterogeneity was observed across the studies regarding the mHealth apps, the type and duration of pain across participants, and the comparison groups, all of which diminish the possibility of a robust conclusion in this review. Despite these heterogeneity aspects, some general conclusions can be drawn.

When we view our findings regarding the content of the mHealth apps, our analyses were similar to those presented in 2 previous studies [1,10]. Most of the content included therapeutic exercises focusing on strength, mobility, and mindfulness. Our review identified only 5 mHealth apps that were used in a research setting, whereas the systematic review by Machado et al [1] provided a general overview of existing commercial apps that included 61 different apps. The review by Du et al [10] used 3 mHealth studies in the subgroup meta-analyses. The reason for narrowing our focus and including only studies involving mHealth self-management apps in our review was to ascertain the current state of these apps to provide preliminary scientific support to self-manage the levels of LBP and disability when using mHealth apps with self-management content. We also excluded studies if the mHealth apps did not include built-in, self-management content, which makes the overview of this scoping review more targeted to such mHealth apps.

Our review showed supporting evidence that mHealth apps targeting self-management may have their place as an additional tool in LBP self-management in a home environment setting. This was also supported by a previous meta-analysis of 3 studies [10]. However, Du et al [10] also included in their meta-analysis a study on an app that was targeted to only report daily data without specific built-in self-management content in the app itself [24]. For providing such services in clinical or home environment settings, it must be taken into account how clinically meaningful the results are for the level of pain. The included studies assessed the level of pain mostly using an 11-point Likert scale assessment (eg, the VAS and the NRS) that is commonly used in clinical practice because of its ease of use as well as evidence of the validity and reliability of its measures [25]. Another review also pointed out that when comparing the measurement properties of the VAS and the NRS, no evidence was provided to indicate that one was superior to the other in the measurement of LBP [26].

For the included RCTs in our review that reported statistically significant differences in favor of the mHealth group in the level of pain measured with the NRS, the changed values varied from –2.0 to –4.0 points, which can be considered a minimal clinically important change according to a previous study reporting a minimum threshold of –2.0 points or a percentage value of –33%, each of which was associated with better improvement in chronic musculoskeletal pain intensity [27]. For the VAS, an included RCT reported a decreased value of 3.5 points [17], which also can be considered within the threshold (30 mm) of a minimal clinically important difference score that was reported in a previous study investigating the levels of minimal clinically important difference scores on the VAS to measure pain [28]. Given that all included studies reported the levels of pain to be above the minimal clinical threshold, we may carefully conclude that mHealth interventions targeted at self-management may achieve a clinically meaningful change in the level of pain within intervention periods lasting from 6 to 16 weeks. Although this is a promising finding, more studies are required to investigate whether such clinically meaningful change is detectable and sustainable over a much longer period of time.

Another aspect of investigating the use of an mHealth self-management app for LBP was to identify its effects on the level of disability. In our review, this was measured in clinical trials (3 RCTs and 1 non-RCT) showing that, of the 4 studies, 3 (75%) did show a statistically significant change in improving functional ability. It seems that using mHealth self-management apps in a home setting may improve the functioning of patients with LBP. This could be a game changer, especially given the fact that the functional ability of patients with LBP is usually affected by anxiety and fear [29]. That said, mHealth self-management apps could provide help for these patients to decrease the worries related to LBP, in addition to providing clinical care. However, more research is needed to investigate the relationship between mHealth app content and the level of disability to confirm these early findings using more sophisticated analyses (eg, meta-analysis and meta-regression).

Achieving optimal management of LBP also requires the patient to play an active role and participate in the treatment. This was highlighted in another review that pointed out several aspects with regard to the patients wishing for more patient-centered care mapping the desirable characteristics of health care professionals, patients' information needs, aspects of care, and barriers to care [30]. From these key elements, the mHealth approach could facilitate some factors related to care, where Chou et al [30] reported that participants wished for more holistic, personalized, emotionally supportive, and encouraging health care as well as the need for continuity of care. mHealth self-management apps could support this when providing an extension of care alongside clinical care. In addition, participants wished to have more information available related to their diagnosis and cause of pain [30]. This was also part of the content of the mHealth apps included in our review.

When we explore the use of mHealth apps in LBP self-management, we should also think critically about the patient for whom this may be more feasible. Although our review consisted of studies involving >2000 participants, almost half (3/7, 43%) of the studies reported very poorly the duration

of pain at baseline in people with LBP. In addition, among the studies (5/7, 71%) that reported the type of pain, 85% (390/459) of the participants were experiencing chronic LBP. Therefore, the majority (5/7, 71%) of the included studies that reported the type of pain included participants with chronic pain; even so, it is still too early to conclude whether mHealth apps are beneficial for a certain type of LBP when they are targeted at self-management of the symptoms. It seems that mHealth apps may be an alternative method alongside individual treatment strategies in coping and dealing with pain. However, a question mark remains over the timing and use of mHealth apps in LBP to maximize support for patients.

The methodological quality of the included trials varied from fair to good. Overall, none of the included studies showed a poor methodological quality, which can be considered a promising finding. The included RCTs had mainly inadequate reporting related to treatment allocation, blinding of participants, and blinding of outcome assessors. Given the types of interventions, the difficulty of blinding participants or outcome assessors can be considered understandable. However, the reliability of the selected outcome assessments was only adequately reported in 50% (2/4) of the RCTs. The RCTs had sample sizes ranging from 20 to 199 participants in the experimental groups, with 75% (3/4) of the studies including relatively low sample sizes (<50 participants), which may lower statistical power and hinder the vote-counting analysis of this scoping review. With regard to the other included studies, mainly the cohort studies, the primary issues concerned insufficient reporting of possible confounding factors. Finally, the methodological quality assessment did not assess the existence of possible participation in other therapies (cointervention bias), which, if not reported, can be considered a confounding factor with regard to drawing conclusions about the effects of mHealth apps on our outcomes of interest. Furthermore, this should be reported more clearly and taken into account when assessing the effects of mHealth on the levels of pain and disability in people with LBP.

This review includes some limitations. First, a selection bias cannot be ruled out during the literature screening procedures of this review. Studies were excluded if they did not explicitly report an mHealth app–based intervention in the title or abstract. Second, we only included studies that were published after January 1, 2015. It is possible that older studies have been published that should have been included in this review. However, this decision was made based on a previous review by Du et al [10] and also based on our presearch to identify proper keywords for our search strategy. Third, the generalizability of the results is limited because a few studies that included a high sample size did not report the duration of pain or the type of pain (acute, subacute, or chronic). Another aspect that limits the generalizability was the lack of reporting to understand patients' acceptance of using an mHealth app for LBP self-management, as well as the intensity and frequency of use.

Future Study Recommendations

More large-scale RCTs investigating the effects of mHealth apps in LBP self-management are needed with a comparison of similar treatments. In our review, all (4/4, 100%) included RCTs were relatively heterogeneous, precluding a comparison of treatments. In addition, the duration of the included interventions in the RCTs ranged from 6 weeks to 16 weeks and in the cohort studies from 12 to 24 weeks. We cannot yet draw conclusions regarding long-term effects of using mHealth apps for LBP self-management, and the feasibility of the apps for a targeted type of pain is still not fully explored. Therefore, we require longer follow-up periods (>16 weeks) to investigate the effects as well as clinically meaningful change over time. Another important clinical aspect for future studies is to measure the role of mHealth apps in behavioral changes in LBP because mHealth apps may provide additional support to patients to overcome barriers related to LBP and provide further support in home environment settings in addition to clinical care.

Clinical Implications

Current research supports the use of mHealth as an additional tool alongside traditional care. Such mHealth apps may provide additional support for clinical care targeted to provide support in home environment settings for LBP. However, this review was limited to the information provided in each study for the content of the apps. It is possible that the app versions and the content of each app have been developed further. In addition, the use of mHealth apps for a longer period of time may require additional costs, and the apps may not be publicly available worldwide, which may narrow the targeted need for such mHealth apps. Future studies should also report more specific details bearing in mind the clinical use of the apps.

Conclusions

Promising results were found for mHealth self-management apps on decreasing the levels of pain and disability in people with LBP. However, more high-quality RCTs with longer study periods are needed to provide further evidence on whether mHealth apps have longer-term effects on LBP self-management in home environment settings.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The methodological quality of the included studies that used a mobile health app designed for low back pain.
[DOCX File , 20 KB - mhealth_v10i8e39682_app1.docx]

References

1. Machado GC, Pinheiro MB, Lee H, Ahmed OH, Hendrick P, Williams C, et al. Smartphone apps for the self-management of low back pain: a systematic review. Best Pract Res Clin Rheumatol 2016 Dec;30(6):1098-1109. [doi: 10.1016/j.berh.2017.04.002] [Medline: 29103552]
2. Hartvigsen J, Hancock MJ, Kongsted A, Louw Q, Ferreira ML, Genevay S, Lancet Low Back Pain Series Working Group. What low back pain is and why we need to pay attention. Lancet 2018 Jun 09;391(10137):2356-2367. [doi: 10.1016/S0140-6736(18)30480-X] [Medline: 29573870]
3. Chen S, Chen M, Wu X, Lin S, Tao C, Cao H, et al. Global, regional and national burden of low back pain 1990-2019: a systematic analysis of the Global Burden of Disease study 2019. J Orthop Translat 2022 Jan;32:49-58 [FREE Full text] [doi: 10.1016/j.jot.2021.07.005] [Medline: 34934626]
4. Alhowimel AS, Alotaibi MA, Alenazi AM, Alqahtani BA, Alshehri MA, Alamam D, et al. Psychosocial predictors of pain and disability outcomes in people with chronic low back pain treated conservatively by guideline-based intervention: a systematic review. J Multidiscip Healthc 2021 Dec;Volume 14:3549-3559. [doi: 10.2147/jmdh.s343494]
5. Novak M, Costantini L, Schneider S, Beanlands H. Approaches to self-management in chronic illness. Semin Dial 2013;26(2):188-194. [doi: 10.1111/sdi.12080] [Medline: 23520989]
6. Jonkman NH, Schuurmans MJ, Jaarsma T, Shortridge-Baggett LM, Hoes AW, Trappenburg JC. Self-management interventions: proposal and validation of a new operational definition. J Clin Epidemiol 2016 Dec;80:34-42. [doi: 10.1016/j.jclinepi.2016.08.001] [Medline: 27531245]
7. Oliveira VC, Ferreira PH, Maher CG, Pinto RZ, Refshauge KM, Ferreira ML. Effectiveness of self-management of low back pain: systematic review with meta-analysis. Arthritis Care Res (Hoboken) 2012 Nov 27;64(11):1739-1748 [FREE Full text] [doi: 10.1002/acr.21737] [Medline: 22623349]
8. MHealth New Horizons for Health Through Mobile Technologies. Geneva: World Health Organization; 2011.
9. Mosa AS, Yoo I, Sheets L. A systematic review of healthcare applications for smartphones. BMC Med Inform Decis Mak 2012 Jul 10;12:67 [FREE Full text] [doi: 10.1186/1472-6947-12-67] [Medline: 22781312]
10. Du S, Liu W, Cai S, Hu Y, Dong J. The efficacy of e-health in the self-management of chronic low back pain: a meta analysis. Int J Nurs Stud 2020 Jun;106:103507. [doi: 10.1016/j.ijnurstu.2019.103507] [Medline: 32320936]
11. Franchini M, Salvatori M, Denoth F, Molinaro S, Pieroni S. Participation in low back pain management: it is time for the to-be scenarios in digital public health. Int J Environ Res Public Health 2022 Jun 25;19(13):7805 [FREE Full text] [doi: 10.3390/ijerph19137805] [Medline: 35805463]
12. World Health Organization. WHO Guideline on Self-care Interventions for Health and Well-being. Geneva: World Health Organization; 2021. URL: https://www.who.int/publications/i/item/9789240052192
13. Porritt K, Gomersall J, Lockwood C. JBI's systematic reviews: study selection and critical appraisal. Am J Nurs 2014 Jun;114(6):47-52. [doi: 10.1097/01.NAJ.0000450430.97383.64] [Medline: 24869584]
14. Buccheri RK, Sharifi C. Critical appraisal tools and reporting guidelines for evidence-based practice. Worldviews Evid Based Nurs 2017 Dec 12;14(6):463-472. [doi: 10.1111/wvn.12258] [Medline: 28898556]
15. Peters MD, Marnie C, Colquhoun H, Garritty CM, Hempel S, Horsley T, et al. Scoping reviews: reinforcing and advancing the methodology and application. Syst Rev 2021 Oct 08;10(1):263 [FREE Full text] [doi: 10.1186/s13643-021-01821-3] [Medline: 34625095]
16. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Sep 04;169(7):467. [doi: 10.7326/M18-0850]
17. Almhdawi KA, Obeidat DS, Kanaan SF, Oteir AO, Mansour ZM, Alrabbaei H. Efficacy of an innovative smartphone application for office workers with chronic non-specific low back pain: a pilot randomized controlled trial. Clin Rehabil 2020 Oct 30;34(10):1282-1291. [doi: 10.1177/0269215520937757] [Medline: 32602362]
18. Chhabra HS, Sharma S, Verma S. Smartphone app in self-management of chronic low back pain: a randomized controlled trial. Eur Spine J 2018 Nov 15;27(11):2862-2874. [doi: 10.1007/s00586-018-5788-5] [Medline: 30324496]
19. Clement I, Lorenz A, Ulm B, Plidschun A, Huber S. Implementing systematically collected user feedback to increase user retention in a mobile app for self-management of low back pain: retrospective cohort study. JMIR Mhealth Uhealth 2018 Jun 06;6(6):e10422 [FREE Full text] [doi: 10.2196/10422] [Medline: 29875088]
20. Huber S, Priebe JA, Baumann K, Plidschun A, Schiessl C, Tölle TR. Treatment of low back pain with a digital multidisciplinary pain treatment app: short-term results. JMIR Rehabil Assist Technol 2017 Dec 04;4(2):e11 [FREE Full text] [doi: 10.2196/rehab.9032] [Medline: 29203460]
21. Irvine AB, Russell H, Manocchia M, Mino DE, Cox Glassen T, Morgan R, et al. Mobile-web app to self-manage low back pain: randomized controlled trial. J Med Internet Res 2015 Jan 02;17(1):e1 [FREE Full text] [doi: 10.2196/jmir.3130] [Medline: 25565416]
22. Sandal LF, Øverås CK, Nordstoga AL, Wood K, Bach K, Hartvigsen J, et al. A digital decision support system (selfBACK) for improved self-management of low back pain: a pilot study with 6-week follow-up. Pilot Feasibility Stud 2020 May 23;6(1):72 [FREE Full text] [doi: 10.1186/s40814-020-00604-2] [Medline: 32489674]
23. Toelle TR, Utpadel-Fischler DA, Haas K, Priebe JA. App-based multidisciplinary back pain treatment versus combined physiotherapy plus online education: a randomized controlled trial. NPJ Digit Med 2019;2:34 [FREE Full text] [doi: 10.1038/s41746-019-0109-x] [Medline: 31304380]

24. Yang J, Wei Q, Ge Y, Meng L, Zhao M. Smartphone-based remote self-management of chronic low back pain: a preliminary study. J Healthc Eng 2019;2019:4632946 [FREE Full text] [doi: 10.1155/2019/4632946] [Medline: 30881606]
25. Rabea Begum M, Hossain M. Validity and reliability of Visual Analogue Scale (VAS) for pain measurement. J Med Case Reports Rev 2019;2:394-402 [FREE Full text]
26. Chiarotto A, Maxwell LJ, Ostelo RW, Boers M, Tugwell P, Terwee CB. Measurement properties of visual analogue scale, numeric rating scale, and pain severity subscale of the brief pain inventory in patients with low back pain: a systematic review. J Pain 2019 Mar;20(3):245-263. [doi: 10.1016/j.jpain.2018.07.009] [Medline: 30099210]
27. Salaffi F, Stancati A, Silvestri CA, Ciapetti A, Grassi W. Minimal clinically important changes in chronic musculoskeletal pain intensity measured on a numerical rating scale. Eur J Pain 2004 Aug;8(4):283-291. [doi: 10.1016/j.ejpain.2003.09.004] [Medline: 15207508]
28. Lee JS. Clinically important change in the visual analog scale after adequate pain control. Acad Emergency Med 2003 Oct 01;10(10):1128-1130. [doi: 10.1197/s1069-6563(03)00372-5]
29. Kawi J. Chronic low back pain patients' perceptions on self-management, self-management support, and functional ability. Pain Manag Nurs 2014 Mar;15(1):258-264. [doi: 10.1016/j.pmn.2012.09.003] [Medline: 23232149]
30. Chou L, Ranger TA, Peiris W, Cicuttini FM, Urquhart DM, Sullivan K, et al. Patients' perceived needs of health care providers for low back pain management: a systematic scoping review. Spine J 2018 Apr;18(4):691-711. [doi: 10.1016/j.spinee.2018.01.006] [Medline: 29373836]

Abbreviations

LBP: low back pain
mHealth: mobile health
NRS: numeric rating scale
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews
RCT: randomized controlled trial
VAS: visual analog scale

Edited by L Buis; submitted 20.05.22; peer-reviewed by HL Tam, R Jacobs; comments to author 02.06.22; revised version received 04.07.22; accepted 27.07.22; published 26.08.22.

Please cite as:
Rintala A, Rantalainen R, Kaksonen A, Luomajoki H, Kauranen K
mHealth Apps for Low Back Pain Self-management: Scoping Review
JMIR Mhealth Uhealth 2022;10(8):e39682
URL: https://mhealth.jmir.org/2022/8/e39682
doi:10.2196/39682
PMID:36018713

Original Paper

Quality of Life and Physical Activity in 629 Individuals With Sarcoidosis: Prospective, Cross-sectional Study Using Smartphones (Sarcoidosis App)

Brian Chu[1*], BS; Daniel M O'Connor[2*], MD; Marilyn Wan[1], MD; Ian Barnett[3], PhD; Haochang Shou[3], PhD; Marc Judson[4], MD; Misha Rosenbach[1], MD

[1]Department of Dermatology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, United States

[2]Department of Dermatology, Harvard Medical School, Boston, MA, United States

[3]Department of Biostatistics, Epidemiology, and Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, United States

[4]Division of Pulmonary and Critical Care Medicine, Albany Medical College, Albany, NY, United States

[*]these authors contributed equally

Corresponding Author:
Misha Rosenbach, MD
Department of Dermatology
Perelman School of Medicine
University of Pennsylvania
3400 Spruce Street
2nd Floor, Maloney Building
Philadelphia, PA, 19104
United States
Phone: 1 2156627883
Email: misha.rosenbach@uphs.upenn.edu

Abstract

Background: Large gaps exist in understanding the symptomatic and functional impact of sarcoidosis, a rare multisystem granulomatous disease affecting fewer than 200,000 individuals in the United States. Smartphones could be used for prospective research, especially for rare diseases where organizing large cohorts can be challenging, given their near ubiquitous ownership and ability to track objective and subjective data with increasingly sophisticated technology.

Objective: We aimed to investigate whether smartphones could assess the quality of life (QoL) and physical activity of a large cohort of individuals with sarcoidosis.

Methods: We developed a mobile app (Sarcoidosis App) for a prospective, cross-sectional study on individuals with sarcoidosis. The Sarcoidosis App was made available on both Apple and Android smartphones. Individuals with sarcoidosis were recruited, consented, and enrolled entirely within the app. Surveys on sarcoidosis history, medical history, and medications were administered. Patients completed modules from the Sarcoidosis Assessment Tool, a validated patient-reported outcomes assessment of physical activity, fatigue, pain, skin symptoms, sleep, and lungs symptoms. Physical activity measured by smartphones was tracked as available.

Results: From April 2018 to May 2020, the App was downloaded 2558 times, and 629 individuals enrolled (404, 64.2% female; mean age 51 years; 513, 81.6% White; 86, 13.7% Black). Two-thirds of participants had a college or graduate degree, and more than half of them reported an income greater than US $60,000. Both QoL related to physical activity (P<.001, ρ=0.250) and fatigue (P<.01, ρ=–0.203) correlated with actual smartphone-tracked physical activity. Overall, 19.0% (98/517) of participants missed at least 1 week of school or work in an observed month owing to sarcoidosis, and 44.4% (279/629) reported that finances "greatly" or "severely" affected by sarcoidosis. Furthermore, 71.2% (437/614) of participants reported taking medications for sarcoidosis, with the most common being prednisone, methotrexate, hydroxychloroquine, and infliximab. Moreover, 46.4% (244/526) reported medication side effects, most commonly due to prednisone.

Conclusions: We demonstrate that smartphones can prospectively recruit, consent, and study physical activity, QoL, and medication usage in a large sarcoidosis cohort, using both passively collected objective data and qualitative surveys that did not require any in-person encounters. Our study's limitations include the study population being weighted toward more educated and

wealthier individuals, suggesting that recruitment was not representative of the full spectrum of patients with sarcoidosis in the United States. Our study provides a model for future smartphone-enabled clinical research for rare diseases and highlights key technical challenges that future research teams interested in smartphone-based research for rare diseases should anticipate.

(JMIR Mhealth Uhealth 2022;10(8):e38331) doi:10.2196/38331

KEYWORDS

sarcoidosis; smartphone; quality of life; mobile app; mobile health; mHealth; digital health; rare disease; physical activity; exercise; fitness; development; tracking; recruit; enroll

Introduction

Sarcoidosis is a multisystem granulomatous disease that most commonly affects the lungs, skin, eyes, and lymph nodes. In the United States, over 185,000 patients with sarcoidosis seek medical care annually, and 25,000 new cases of sarcoidosis are diagnosed [1]. While many cases appear mild, the disease can cause substantial functional morbidity including exertional dyspnea, generalized pain, and decreased physical activity [2]. Fatigue is common among patients with sarcoidosis and can be debilitating. Given the variable clinical presentation of sarcoidosis, large gaps remain in understanding the daily impact of sarcoidosis.

Sarcoidosis is a rare disease, which is defined as affecting fewer than 200,000 individuals. Clinical research of rare diseases is challenging—nearly one-third of rare disease clinical trials are discontinued with insufficient patient accrual being the most common reason [3]. Large cohorts are difficult to recruit and may not be representative of the patient population owing to overrepresentation of patients with access to academic medical centers. However, more than 85% of Americans currently own a smartphone, and manufacturers are increasingly including sophisticated health-tracking technology [4]. For example, Apple ResearchKit allows investigators to not only measure the number of daily walking steps and distance traveled, but also estimate cardiac fitness levels and walking stability [5]. In addition, mobile apps allow patients to complete validated survey instruments on smartphones. Therefore, smartphones are a compelling channel to conduct clinical research on rare disease populations.

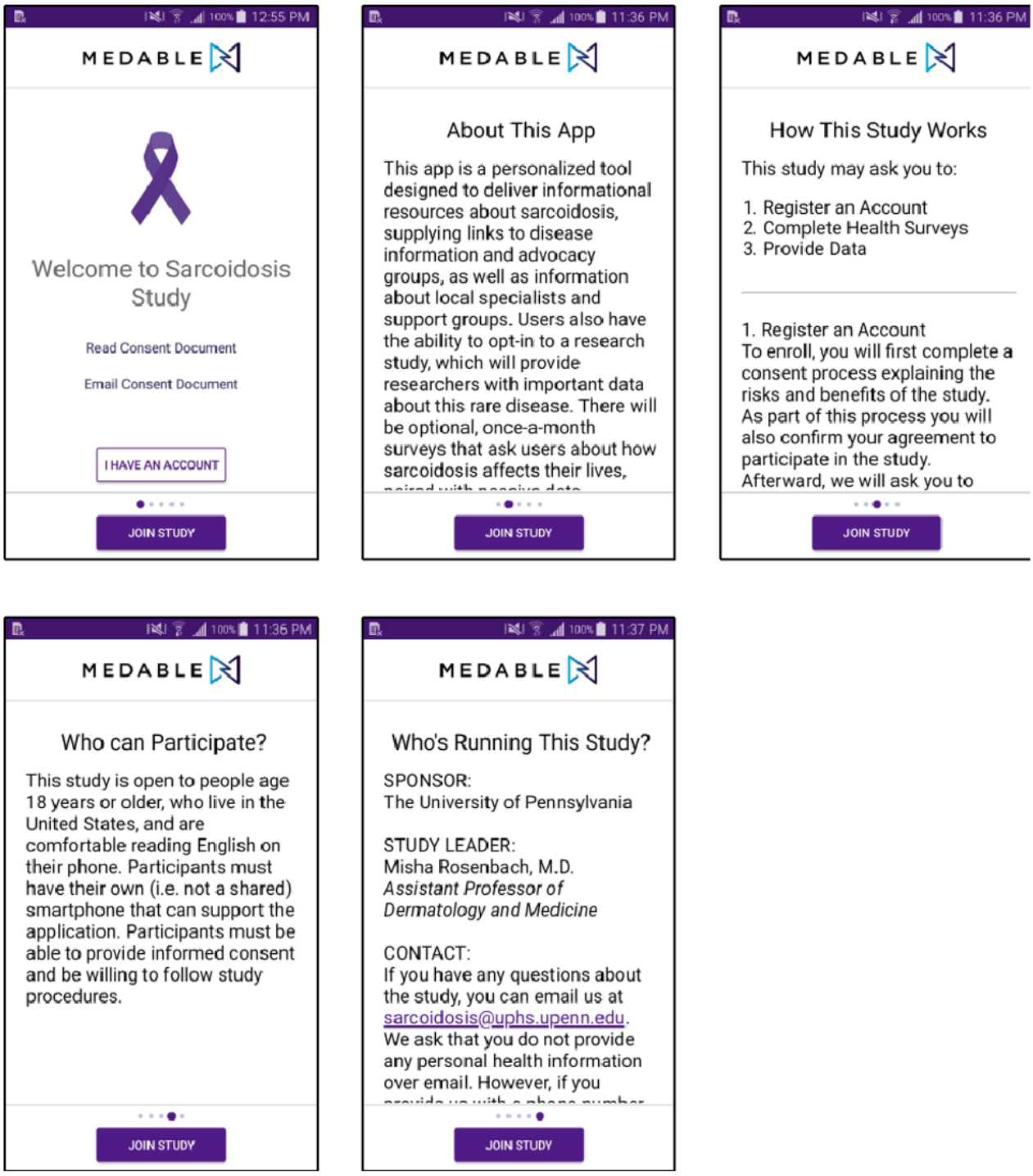
The goal of our study was to develop and launch a smartphone app to assess patient-reported quality of life (QoL) in a large population of patients with sarcoidosis and to characterize patients' symptomatology and functional status. We designed this app to describe physical activity and correlate physical activity data with self-reported characteristics through QoL surveys, medication use, and adverse effects and comorbidities. Here, we describe the development and launch of the app and present results from the surveys and physical activity data collected from patients.

Methods

Smartphone App Development and Launch

The Sarcoidosis App [6], developed by authors DMO and MR, is a smartphone app that measures physical activity levels and records patient-reported responses to questions (Figure 1) [7]. The Sarcoidosis App was initially designed using the open-source Apple ResearchKit framework. The app was then ported onto the Medable trial platform [8], to allow for distribution on both Android and Apple operating systems. Data were automatically encrypted, deidentified, and uploaded directly to secure servers, adhering to guidelines specified by the Health Insurance Portability and Accountability Act. Study data were not shared with any organization including Apple, Alphabet, Medable, or with nonstudy personnel. The Sarcoidosis App was made available on the Google Play and the Apple App Store in April 2018. Data were collected through May 2020, though the length of time spent using the App varied by participant. The Foundation for Sarcoidosis Research shared an announcement about the Sarcoidosis App's launch with their patient email list.

Figure 1. Screenshots of new participant experience in the Sarcoidosis App (Android version).



Patient Recruitment, Consent, and Enrollment

Patients with sarcoidosis were recruited via outpatient dermatology clinic visits at the University of Pennsylvania Health System, the Foundation for Sarcoidosis Research mailing list, and targeted social media advertisements. After prospective participants downloaded the app, they were presented with an inclusion and exclusion criteria questionnaire to provide multiple means of participant recruitment. Participants were eligible for study enrollment if they were aged greater than 18 years, lived in the United States, and self-reported a diagnosis of sarcoidosis. Participants were asked for permission to enable the app to read HealthKit or Google Fit data. To ensure participants understood the risks, benefits, and options of study participation, they were required to pass a quiz concerning these issues before digitally signing the informed consent document (Multimedia Appendix 1).

Patient-Reported Outcome Measures: Survey Design and Data Collection

The Sarcoidosis App administered baseline surveys of (1) sarcoidosis history, (2) pertinent medical history, (3) sarcoidosis treatment medications, and (4) items from modules of the Sarcoidosis Assessment Tool (SAT). The SAT is a validated

patient-reported assessment that comprises select generic measures from the Patient-Reported Outcomes Measurement Information System, as well as several sarcoidosis-specific item banks, including physical functioning, satisfaction with roles and activities, fatigue, pain interference, sleep disturbance, lung concerns, skin concerns, and skin stigma and embarrassment [7]. After the intake process, several surveys were administered. Participants could skip questions that they did not wish to answer and could be completed at any time. The app also asked participants for permission to import certain smartphone physical activity data, including daily counts of steps, distance walked or run, flights of stairs climbed, and exercise time.

Statistical Analysis

Descriptive statistics were calculated for demographic information, patient-reported outcomes, and physical activity data. Missing data were excluded from analyses except where described. Spearman correlation coefficients were calculated for correlating responses to the SAT modules for physical activity, lungs, and fatigue with average daily steps and traveled distance from smartphone-recorded data, as much of the survey and physical activity measures were not normally distributed. In addition, the use of ranks through Spearman correlation diminishes the influence of outliers in some of the physical activity measures. All statistics were calculated using Stata (version 16.1; StataCorp) and R (version 3.6.1; The R Foundation for Statistical Computing). The survey data and physical activity data used in this study is available upon reasonable request to BC and MR.

Ethical Considerations

This study was approved by the institutional review board of the University of Pennsylvania (824080). This paper adheres to STROBE (Strengthening the Reporting of Observational studies in Epidemiology) reporting guidelines. Participants who met eligibility criteria proceeded to provide electronic informed consent.

Results

Study Enrollment and Background Information

From April 2018 to May 2020, the app was downloaded 2558 times: 1603 from the Apple App Store and 955 from Google Play. A quarter (629/2558) of downloads converted to study participation, with 629 unique participants completing at least one component of the background survey concerning basic demographic data (Table 1). Of them, 64.2% (n=404) of participants were female, and 81.6% (n=513) were White. Two-thirds (n=416, 66.1%) of participants had a college or graduate degree, and more than half reported full-time employment (n=316, 50.2%) and annual incomes greater than US $60,000 (n=360, 57.2%). In terms of background information on sarcoidosis disease and QoL, 60.1% (n=378) of participants reported that their sarcoidosis was diagnosed within the past 5 years. Subjectively, 19.9% (n=125) of participants reported “poor” health and 39.7% (n=250) reported “fair” health. Overall, 44.4% (n=279) of participants reported that their family’s finances were “greatly affected” or “severely affected” by sarcoidosis.

Table 1. Baseline characteristics of participants in the Sarcoidosis App study (N=629).

Characteristic	Value
Gender	
Female	404 (64.2)
Male	199 (31.6)
Missing	26 (4.1)
Mean age (years), mean (median; SD)	51.0 (50; 10.95)
Race, n (%)	
White	513 (81.6)
Black or African American	86 (13.7)
Other/Unknown	27 (4.3)
Missing	3 (0.5)
Ethnicity, n (%)	
Hispanic/Latino	32 (5.1)
Non-Hispanic	587 (93.3)
Missing	10 (1.6)
BMI[a], mean (SD)	32.12 (7.75)
Years since diagnosis, n (%)	
<1	127 (20.2)
1-5	251 (39.9)
5-20	197 (31.3)
>20	52 (8.3)
Missing	2 (0.3)
Education, n (%)	
High school	8 (1.3)
General Educational Development	205 (32.6)
College	266 (42.3)
Graduate	150 (23.8)
Employment status, n (%)	
Student	6 (1.0)
Part-time	44 (7.0)
Full-time	316 (50.2)
Unemployed	61 (9.7)
Disabled	114 (18.1)
Retired	82 (13.0)
Missing	6 (1.0)
Income (US $), n (%)	
<15,000	49 (7.8)
15,000-30,000	58 (9.2)
30,000-60,000	139 (22.1)
60,000-100,000	166 (26.4)
>100,000	194 (30.8)
Missing	23 (3.7)
Financial impact, n (%)	

Characteristic	Value
No financial impact	124 (19.7)
Slightly affected	223 (35.5)
Greatly affected	178 (28.3)
Severely affected	101 (16.1)
Missing	3 (0.5)

[a]BMI was obtained from 605 participants.

SAT Results

In total, 597 unique participants filled out at least one of the SAT modules. The mean scores of each SAT module and their SDs are reported in Table 2. Owing to technical issues with the app, there were fewer reported outcomes for the skin symptoms module of the SAT.

Table 2. Sarcoidosis Assessment Test (SAT) survey results at baseline. A score of 50 represents the mean score of the original calibration sample of the SAT.

SAT module	Participants, n	Module score, mean (SD)
Activity [+][a]	552	39.90 (7.97)
Fatigue [–][b]	564	62.67 (9.32)
Lungs [–]	572	45.87 (8.46)
Pain [–]	544	60.48 (10.63)
Skin symptoms [–]	208	57.80 (6.99)
Sleep [–]	567	58.38 (9.51)
Stigma/embarrassment/skin impact [–]	535	49.73 (8.65)

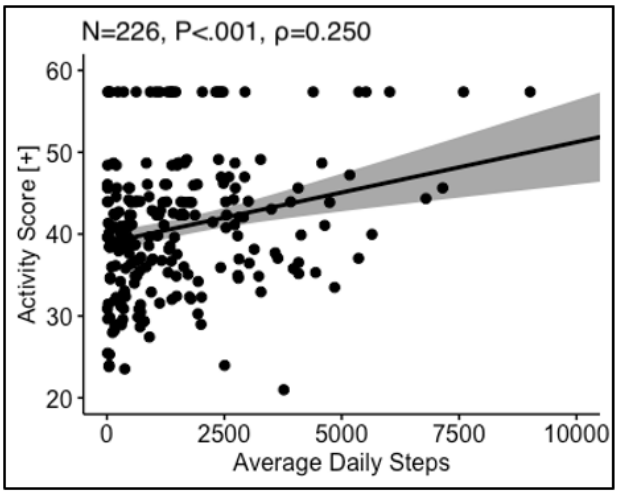
[a]A higher score representing a higher quality of life.

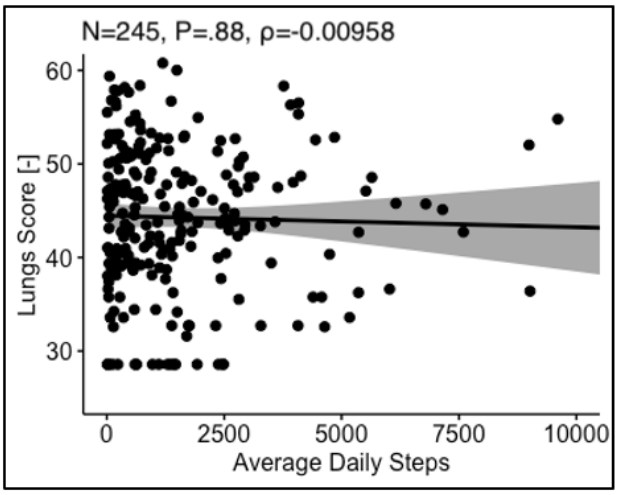
[b]A higher score representing a lower quality of life.

Correlational Data

Where data were available for both SAT survey responses and device-measured physical activity, correlational analyses were performed. SAT physical activity scores positively correlated with average daily steps (n=226), and SAT fatigue scores negatively correlated with average daily steps (n=245) (Figure 2). SAT lung symptoms scores did not correlate with average daily steps (n=238). These trends were replicated when comparing SAT survey responses to device-tracked daily distance moved (Figure 3).

Figure 2. Correlation between Sarcoidosis Assessment Test surveys of physical activity, lung symptoms, and fatigue with device-reported physical activity data. [+] indicates that a higher score represents a higher quality of life. [-] indicates that a higher score represents a lower quality of life. P and ρ are the *P* value and Spearman correlation coefficient, respectively.





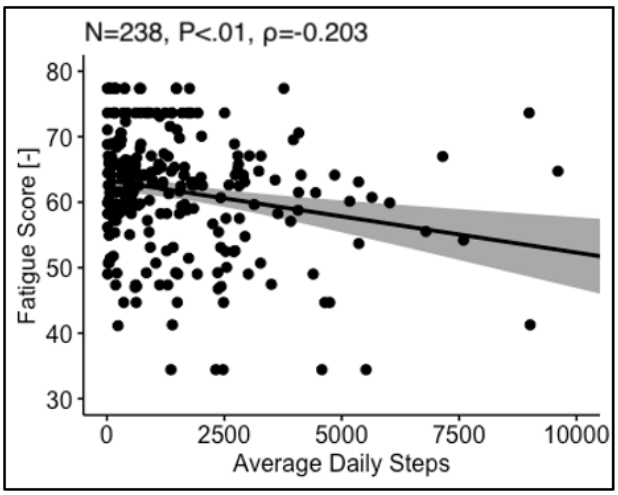
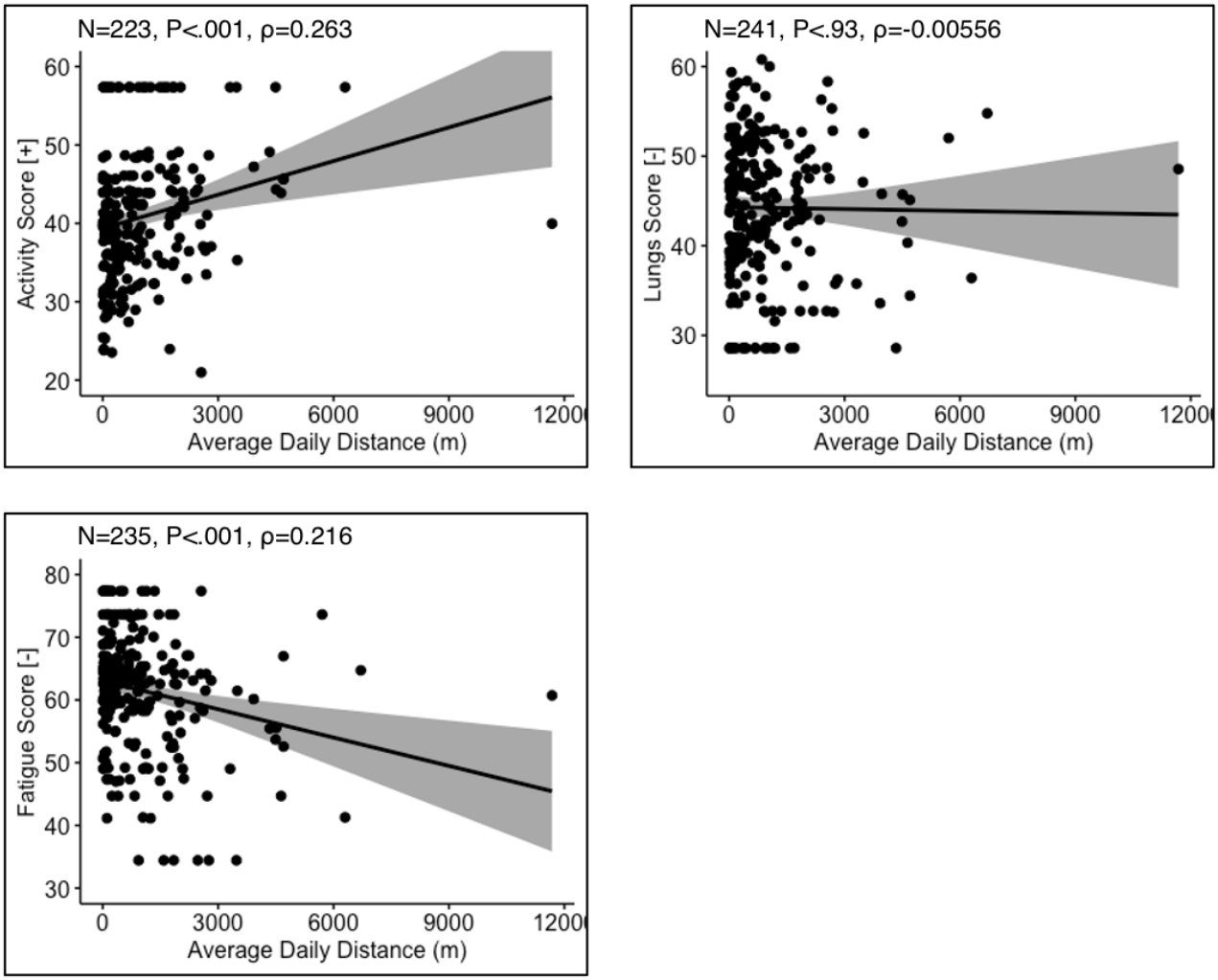


Figure 3. Correlation between Sarcoidosis Assessment Test surveys of physical activity, lung symptoms, and fatigue with device-reported average daily distance traveled. [+] indicates that a higher score represents a higher quality of life. [-] indicates that a higher score represents a lower quality of life. P and ρ are the *P* value and Spearman correlation coefficient, respectively.



Sarcoidosis Medical Resource Usage

Medical resource usage was assessed among 517 unique participants who completed at least one question of an initial survey (Table 3). Overall, 59.8% (n=309) of participants reported at least one regularly scheduled clinic visit in the previous month, while 23.6% (n=122) reported one or more unscheduled clinic visits for sarcoidosis. Furthermore, 11.0% (n=57) of participants reported an emergency room visit for sarcoidosis in the previous month, and 7.4% (n=38) reported a hospitalization related to sarcoidosis. Moreover, 12.4% (n=64) of participants reported missing the entire previous month of school or work because of sarcoidosis, and another 29.4% (n=152) reported missing at least one day of school or work in the previous month.

Table 3. Initial response to survey of medical resource usage (N=517).

Variables	Value, n (%)
How many regularly scheduled clinic visits for your sarcoidosis have you had in the last month?	
None	206 (39.8)
1	153 (29.6)
More than 1	156 (30.1)
In the past month, how many unscheduled clinic visits have you had for your sarcoidosis?	
None	392 (75.8)
1	70 (13.5)
More than 1	52 (10.1)
In the past month, how many visits to the ER[a] have you had for your sarcoidosis?	
None	442 (85.5)
1	57 (11.0)
More than 1	18 (3.5)
In the past month, how many times have you been hospitalized for your sarcoidosis?	
None	478 (94.5)
1	31 (6.0)
More than 1	7 (1.4)
How many days did you miss school/work in the past month because of your sarcoidosis?	
None	279 (54.0)
1 day	45 (8.7)
2-6 days	73 (14.1)
7-14 days	22 (4.3)
>Half month	12 (2.3)
Entire month	64 (12.4)

[a]ER: emergency room.

Medication Use and Side Effects

In total, 614 unique participants completed at least one question from the baseline medication survey (Table 4). Overall, 71.2% (437/614) of participants reported using medications to treat their sarcoidosis; of them, 58.8% (257/437) were being treated with prednisone and 81.0% (354/437) were being treated with medications other than prednisone for sarcoidosis. The most common specified medications other than prednisone were methotrexate, hydroxychloroquine, and infliximab.

Participants were also asked to complete a survey concerning medication adverse effects within the prior month (Table 5). Overall, 64.4% (244/379) of participants reported major (requiring changes in medications) or minor side effects from their sarcoidosis medications, with another 20.6% (78/379) of participants reporting possible side effects. The most common medication causing side effects was prednisone. In a follow-up question, participants were most often recommended to continue medications at the same dose when side effects were discussed with physicians.

Table 4. Responses to survey of baseline medication usage (N=614).

Variables	Participants, n/N (%)
Are you currently being treated with medications for your sarcoidosis? (Yes)	437/614 (71.2)
Are you taking prednisone for sarcoidosis?	257/437 (41.9)
Daily prednisone dose (mg)	
1-10	131/257 (50.9)
11-20	65/257 (25.3)
21-60	57/257 (22.2)
>61	4/257 (1.6)
Are you taking any medications other than prednisone for sarcoidosis?	354/437 (57.7)
What other medications are you taking?	
Hydroxychloroquine	71/354 (20.1)
Methotrexate	134/354 (37.9)
Chloroquine	4/354 (1.1)
Azathioprine	29/354 (8.2)
Leflunomide	6/354 (1.7)
Mycophenolate mofetil	33/354 (9.3)
Infliximab	51/354 (14.4)
Adalimumab	28/354 (7.9)
Other medications	182/354 (51.4)

Table 5. Medication adverse effects survey (N=526).

Variables	Participants, n/N (%)
Have you taken any medications for sarcoidosis in the past month?	379/526 (72.1)
Did you have any side effects from sarcoidosis medications in past month?	
Yes, major side effects requiring change in medications	104/379 (27.4)
Yes, minor side effects	140/379 (36.9)
Possibly/unsure	78/379 (20.6)
No	70/379 (18.5)

Discussion

In this study, we demonstrate the novel use of smartphones to prospectively recruit, consent, and study physical activity and QoL in a large cohort of individuals with sarcoidosis, using both objective health tracking data and qualitative survey responses. We were able to demonstrate a strong correlation between the assessment of physical state and the activity level of participants with sarcoidosis, measured by smartphone apps. Specifically, participants who were more active, as measured by daily steps and distance traveled as tracked by their smartphones, also had physical activity and fatigue scores, representing a smaller impact of their disease on these domains on the SAT—a previously-defined patient-reported QoL metric [7]. Nearly half of the participants missed at least a day of school or work monthly, and nearly one-fifth missed at least a week, reflecting the poor QoL related to physical activity and fatigue reported on the SAT, and demonstrating the profound impact of sarcoidosis on patients. The challenge of pharmaceutical management of sarcoidosis was highlighted by the finding that 6 in 10 participants reported medication side effects, of whom 43% required changes in their medications. Furthermore, 4 in 5 participants reported taking medications for sarcoidosis other than prednisone. Together, these data provide a detailed view into how individuals live with sarcoidosis and demonstrate that smartphones are a compelling method of prospective research for rare diseases, where such wide-scale data collection would otherwise be unfeasible.

The findings of this study suggest that smartphone technology may have advantages in the conduct of prospective clinical research in sarcoidosis and other rare diseases, though robust human and technical resources are critical. First, these data suggest smartphones have potential to enroll patients with sarcoidosis in clinical trials and reliably assess them without requiring a traditional in-person clinic visit. In this way, the clinical research study can be brought directly to participants, bypassing financial and geographic barriers of many socially disadvantaged patients with sarcoidosis, who are often

unrepresented in clinical trials. However, these methods cannot replace studies that require laboratory testing or imaging. Second, as these assessments can be made in real time without investigator prompting, this method may avoid significant recall bias and investigator influence respectively. Third, the capability of integrating objective health tracking data with subjective survey data offers a multidimensional assessment of patients with sarcoidosis. In this way, studies using smartphones could provide alternative channels of demonstrating construct validity of patient-reported outcomes.

The strengths of our study were tempered by the technical challenges of developing and maintaining a mobile app. Clinical research teams without strong technical experience will encounter many obstacles in designing and launching mobile apps, and partnerships with technology companies are essential to the success of these projects. However, lack of cross-disciplinary understanding presents substantial challenges to meaningful collaboration with external stakeholders; clinical research teams may not even be able to envision what is technically possible, and developers may lack perspective on how these products are actually delivered to patients [10]. Even with industry partnership, research teams require members skilled in computer science, database management, and data analysis to process the large amount of complex data generated by apps and sensors. For instance, our study lacked longitudinal data collection owing to the challenge of long-term maintenance of the app without dedicated program management or use of participant-engagement rewards systems. Another limitation is that physical activity data could only be tracked if participants were consistently carrying their smartphones, which could not be enforced remotely. As a result, for some participants, physical activity data were sporadically recorded and may not be representative of their actual activity.

There were several limitations regarding the background of participants. First, individuals self-reported a diagnosis of sarcoidosis, which introduces the possibility of participants without a true diagnosis, even though participants were recruited directly from sarcoidosis clinics and advocacy groups for patients with sarcoidosis. Another limitation of our study is that our study population was weighted toward more educated and wealthier individuals. It is possible that the rate of response and familiarity with smartphone apps in our cohort was not representative of the full population of patients with sarcoidosis in the United States, particularly older individuals and those with low technology literacy. Future mobile app studies of sarcoidosis and other rare diseases should prioritize recruiting from a diverse set of sarcoidosis clinics that would provide a more representative sample, in addition to patient advocacy groups. Given that sarcoidosis results in disparate outcomes by race, sex, and socioeconomic class [11] and an annual health care cost of US $20,000 [12], more work is necessary to realize the benefits of the ubiquitous smartphone ownership across all socioeconomic groups.

Future apps could also integrate environmental data, such as location, weather, and air quality to provide additional dimensions of analysis. Wearable devices, such as smartwatches, can also provide valuable data, though ownership is not as prevalent. Beyond observational research, smartphones also present opportunities for digital therapeutics, which are evidence-based interventions driven by software. For example, one group has demonstrated that a smartphone-based stress management tool significantly reduced stress and fatigue in patients with sarcoidosis compared to control patients [13]. In an era of rapid adoption of telehealth driven by the COVID-19 pandemic, clinicians and patients may be more accepting of such tools [14]. The future management of rare chronic diseases such as sarcoidosis may evolve toward using patient-owned devices to actively monitor symptomatology and medication side effects in real time outside of medical centers, allowing rapid treatment adjustment. At the same time, they can also serve as trusted patient education and community platforms, which are highly desired by patients.

Acknowledgments

We acknowledge Kathleen Chen, the University of Pennsylvania Patient Centered Outcomes Research pilot grant, Siegel Family philanthropic support, the Foundation for Sarcoidosis Research, Medable (Stephanie Adamkiewicz), and the Department of Dermatology Clinical Studies Unit.

Authors' Contributions

DMO and MR developed the Sarcoidosis App with advice from MJ. BC, DMO, and MW performed statistical analyses. HS and IB provided statistical advice. BC and DMO acted as cofirst authors for the manuscript. All authors contributed to the drafting, editing, and submission of the manuscript, led by BC, DMO, MR, and MJ. BC and MR take responsibility for the content of the manuscript, including the data and analysis.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Consent quiz to ensure participants understood the risks, benefits, and options of study participation.
[PDF File (Adobe PDF File), 3487 KB - mhealth_v10i8e38331_app1.pdf]

References

1. Baughman RP, Field S, Costabel U, Crystal RG, Culver DA, Drent M, et al. Sarcoidosis in America. Analysis Based on Health Care Use. Annals ATS 2016 Aug;13(8):1244-1252. [doi: 10.1513/annalsats.201511-760oc]
2. Drent M, Crouser ED, Grunewald J. Challenges of Sarcoidosis and Its Management. N Engl J Med 2021 Sep 09;385(11):1018-1032. [doi: 10.1056/nejmra2101555]
3. Mitani AA, Haneuse S. Small Data Challenges of Studying Rare Diseases. JAMA Netw Open 2020 Mar 02;3(3):e201965 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.1965] [Medline: 32202640]
4. Mobile Fact Sheet. Pew Research Center. 2021 Apr 07. URL: https://www.pewresearch.org/internet/fact-sheet/mobile/ [accessed 2021-10-21]
5. ResearchKit and CareKit: Empowering medical researchers, doctors, and you. Apple. URL: https://www.apple.com/lae/researchkit/ [accessed 2021-10-21]
6. The Sarcoidosis App. URL: http://sarcoidosisapp.com/ [accessed 2022-07-20]
7. Judson MA, Mack M, Beaumont JL, Watt R, Barnathan ES, Victorson DE. Validation and Important Differences for the Sarcoidosis Assessment Tool. A New Patient-reported Outcome Measure. Am J Respir Crit Care Med 2015 Apr 01;191(7):786-795. [doi: 10.1164/rccm.201410-1785oc]
8. Medable. URL: https://www.medable.com/ [accessed 2022-07-20]
10. Marcu G, Ondersma SJ, Spiller AN, Broderick BM, Kadri R, Buis LR. Barriers and Considerations in the Design and Implementation of Digital Behavioral Interventions: Qualitative Analysis. J Med Internet Res 2022 Mar 30;24(3):e34301 [FREE Full text] [doi: 10.2196/34301] [Medline: 35353043]
11. Gerke AK, Judson MA, Cozier YC, Culver DA, Koth LL. Disease Burden and Variability in Sarcoidosis. Annals ATS 2017 Dec;14(Supplement_6):S421-S428. [doi: 10.1513/annalsats.201707-564ot]
12. Rice JB, White A, Lopez A, Conway A, Wagh A, Nelson WW, et al. Economic burden of sarcoidosis in a commercially-insured population in the United States. J Med Econ 2017 Oct 21;20(10):1048-1055 [FREE Full text] [doi: 10.1080/13696998.2017.1351371] [Medline: 28678623]
13. James W, Chandler J, Brunner-Jackson B, Sox L, Swem M, Neely A, et al. Acceptability and Efficacy of a mHealth App for Sarcoidosis Associated Fatigue. 2019 Presented at: American Thoracic Society 2019 International Conference; May 17-22, 2019; Dallas, TX. [doi: 10.1164/ajrccm-conference.2019.199.1_meetingabstracts.a3059]
14. Weiner JP, Bandeian S, Hatef E, Lans D, Liu A, Lemke KW. In-Person and Telehealth Ambulatory Contacts and Costs in a Large US Insured Cohort Before and During the COVID-19 Pandemic. JAMA Netw Open 2021 Mar 01;4(3):e212618 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.2618] [Medline: 33755167]

Abbreviations

QoL: quality of life
SAT: Sarcoidosis Assessment Test
STROBE: Strengthening the Reporting of Observational studies in Epidemiology

Edited by L Buis; submitted 29.03.22; peer-reviewed by R Dellavalle, D Sherifi; comments to author 09.05.22; revised version received 12.05.22; accepted 13.05.22; published 10.08.22.

Please cite as:
Chu B, O'Connor DM, Wan M, Barnett I, Shou H, Judson M, Rosenbach M
Quality of Life and Physical Activity in 629 Individuals With Sarcoidosis: Prospective, Cross-sectional Study Using Smartphones (Sarcoidosis App)
JMIR Mhealth Uhealth 2022;10(8):e38331
URL: https://mhealth.jmir.org/2022/8/e38331
doi:10.2196/38331
PMID:35947439

Original Paper

Role of Social and App-Related Factors in Behavioral Engagement With mHealth for Improved Well-being Among Chronically Ill Patients: Scenario-Based Survey Study

Freek Van Baelen[1*], MA; Melissa De Regge[2,3*], PhD; Bart Larivière[4,5], PhD; Katrien Verleye[3], PhD; Sam Schelfout[6], MD; Kristof Eeckloo[2,7], PhD

[1]School of Business and Management, University College Ghent, Ghent, Belgium

[2]Strategic Policy Cell, Ghent University Hospital, Ghent, Belgium

[3]Department of Marketing, Innovation and Organisation, Faculty of Economics and Business Administration, Ghent University, Ghent, Belgium

[4]Department of Marketing, Faculty of Economics and Business, Catholic University of Leuven, Leuven, Belgium

[5]Center for Service Intelligence, Ghent University, Ghent, Belgium

[6]Multidisciplinary Pain Center, Ghent University Hospital, Ghent, Belgium

[7]Department of Public Health and Primary Care, Faculty of Medicine and Health Sciences, Ghent University, Ghent, Belgium

[*]these authors contributed equally

Corresponding Author:
Melissa De Regge, PhD
Strategic Policy Cell
Ghent University Hospital
10, Corneel Heymanslaan
Ghent, 9000
Belgium
Phone: 32 92643493
Email: melissa.deregge@uzgent.be

Abstract

Background: The last decade has seen a considerable increase in the number of mobile health (mHealth) apps in everyday life. These mHealth apps have the potential to significantly improve the well-being of chronically ill patients. However, behavioral engagement with mHealth apps remains low.

Objective: The aim of this study was to describe the behavioral engagement of chronically ill patients with mHealth apps by investigating (1) how it is affected by social factors (ie, physician recommendation) and app-related factors (ie, app integration) and (2) how it affects patient well-being. This study also considers the moderating effect of attachment to traditional health care and the mobile app experience among patients.

Methods: We carried out a scenario-based survey study of chronically ill patients (N=521). A Bayesian structural equation modeling with mediation and moderation analysis was conducted in MPlus.

Results: Both physician recommendations for mHealth app use and app integration have positive effects on the behavioral engagement of chronically ill patients with mHealth apps. Higher behavioral engagement positively affects the hedonic well-being (extent of pleasure) and the eudaemonic well-being (extent of self-efficacy) of chronically ill patients. Mobile app experience, however, positively moderates the relationship between app integration and behavioral engagement, whereas patient attachment to traditional care does not moderate the relationship between physician recommendation and behavioral engagement. Taken together, the proportion of variance explained (R^2) equals 21% for behavioral engagement and 52.8% and 62.2% for hedonic and eudaemonic well-being, respectively, thereby providing support for the strong influence of app integration and physician recommendation via the mediation of the patients' behavioral engagement on both patients' hedonic and eudaemonic well-being.

Conclusions: Physician recommendation and app integration enable behavioral engagement and promote well-being among chronically ill patients. It is thus important to take social and app-related factors into consideration during and after the development of mHealth apps.

(JMIR Mhealth Uhealth 2022;10(8):e33772) doi:10.2196/33772

KEYWORDS

mHealth app; engagement; social influence; app integration; well-being; Belgium; mHealth; behavioral; behavioral engagement; mobile health; mobile health apps; mobile phone

Introduction

Background

With the growth in smartphone use and increasing demands from patients for immediate access to web-based services, mobile health (mHealth) apps that allow patients to actively manage their own health through mobile and wireless technologies are on the rise [1-3]. Internationally, the popularity of mHealth apps to support the achievement of health objectives is increasing [4], especially for chronically ill patients [1]. For this group of patients, research suggests that mHealth apps lead to increased confidence in disease management [5,6], improved therapy compliance, better health care outcomes [7], and even reduced costs [6].

Despite the proven impact of mHealth apps on patient well-being [8], patients do not always show high levels of behavioral engagement with them [9-11]. Here, behavioral engagement refers to the adoption and continued usage of mHealth apps by chronically ill adults. User data from popular app stores even show that most mHealth apps are only used a few times before being abandoned [12]. Less than a third of chronically ill patients aged 50 years and older currently use an mHealth app. More than a third of patients have used mHealth apps in the past but have then stopped using them [13]. In an attempt to better understand why behavioral engagement with mHealth apps is low, we explore the drivers of behavioral engagement with mHealth apps.

In line with the technology acceptance model and the unified theory of acceptance and use of technology, several researchers have, in recent years, pointed out the importance of social-related and technology-related factors when explaining behavioral engagement with health care technologies [14,15]. In light of the behavioral engagement with mHealth apps, there has been considerable research on effort and performance expectancy [16]. However, less research has been dedicated to the role of social and app-related drivers of behavioral engagement with mHealth apps. Pham et al [17] also call for more research on the relationship between mHealth engagement and well-being for chronically ill patients.

Against this background, this research characterizes social enablers (ie, physician recommendation) and app-related enablers (ie, app integration) of behavioral engagement with mHealth apps among chronically ill patients, thereby also considering the impact of behavioral engagement with mHealth apps for patient well-being (that is, hedonic well-being defined as the extent of pleasure and eudaemonic well-being defined as the extent of self-efficacy [18]).

Conceptual Framework and Hypotheses

Social Enablers of Patients' Behavioral Engagement With mHealth Apps

With regard to the social enablers of behavioral engagement with technologies, it is well established that people can affect each other [19]. Specifically, several researchers have shown that behavioral engagement with technologies is—as suggested by the technology acceptance model—a function of social influence [14,20]. Social influence refers to any "change in an individual's thoughts, feelings, attitudes, or behaviors that results from interaction with another individual or a group" [21]. A key question revolves around which individuals or groups can change an individual's thoughts, feelings, attitudes, or behaviors in the context of mHealth apps.

Cajita et al [22] have shown that physicians have a significant influence on mHealth app usage among older adults with heart failure. Likewise, Apolinário-Hagen et al [23] demonstrated that physicians significantly affect behavioral engagement with mHealth apps among people with multiple sclerosis. Specifically, patients with different health conditions may interpret the efforts of health care professionals to use mHealth as an incentive to use mHealth themselves [9,24]. Alternatively, patients may show more behavioral engagement with mHealth apps when they receive recommendations for the use of mHealth apps from health care professionals [10], including physicians [25]. As chronically ill patients often have longstanding relationships with their physician and since they tend to follow their physicians' instructions, we suggest that physician recommendation—that is, the extent to which physicians recommend the use of mHealth apps—can play an important role when engaging these patients with mHealth apps. We hypothesize as follows:

H1a: Physician recommendation positively affects behavioral engagement with mHealth apps among chronically ill patients.

As suggested by the diffusion of innovation theory [26], the impact of physician recommendation on behavioral engagement with mHealth apps among chronically ill patients also relates to patients' own perceptions of the relative advantages of these health care technologies in relation to the idea it supersedes (here, traditional care). If patients are attached to traditional care, the relative advantages of mHealth apps may be less for them. As relative advantage is one of the strongest predictors of the emerging use of technological innovation [27], patients who are more attached to traditional care are less likely to show behavioral engagement with the use of mHealth apps recommended by their physicians. We thus hypothesize as follows:

H1b: The positive impact of physician recommendation on behavioral engagement with mHealth apps among chronically ill patients will decrease when they are more attached to traditional care.

App-Related Enablers of Patients' Behavioral Engagement With mHealth Apps

To ensure that health care technologies such as mHealth apps are relevant to patients and health care professionals, several researchers have called for these users to be involved in the app development process [28,29]. In this regard, patients and physicians have emphasized that health care technologies need to enable data exchange with other systems or applications [30-32]. Indeed, health care technologies that lack interoperability (the ability to exchange data with other systems or applications) have been described as information silos [33,34]; the same holds for mobile apps, including mHealth apps, which are not compatible with other systems such as electronic patient records [25,35].

Empirical evidence also suggests that mHealth apps with low levels of interoperability may deteriorate health care outcomes [34,35]. In contrast, allowing mobile apps to exchange data with each other and other digital systems may help to avoid duplication of medical care, increase patient safety, improve the continuity of care, and reduce administrative burdens [33,36]. Moreover, mHealth apps with high levels of interoperability contribute to increased functionality and better experiences for patients [37] while allowing patients to access, store, or make certain information digitally available, thereby making them, to a greater extent, into managers of their own health [31,33,36]. As chronically ill patients often encounter multiple health care providers, we contend that app integration, that is, the extent to which mHealth apps are interoperable is even more important [15,33]. We therefore hypothesize as follows:

H2a: App integration positively affects behavioral engagement with mHealth apps among chronically ill patients.

If chronically ill patients have more experience with mobile apps, they are more likely to have tried apps with high levels of interoperability and hence to have experienced how app integration can benefit them. The diffusion of innovation theory [26] confirms that innovations that users can experiment with are (in line with the idea of trialability) more likely to be embraced. Building upon the trialability idea, we contend that the positive effect of app integration on behavioral engagement with mHealth apps is strengthened when patients have more experience with mobile apps. We thus hypothesize as follows:

H2b: The positive effect of app integration on behavioral engagement with mHealth apps among chronically ill patients increases when such patients have more mobile app experience.

Patient Behavioral Engagement With mHealth Apps for Improved Well-being

The behavioral engagement of patients with health care technologies has been associated with improved well-being [8,38,39]. As widely acknowledged in the well-being literature [18], well-being incorporates hedonic well-being, with its focus on pleasure attainment, and eudaemonic well-being, with its focus on self-realization, that is, the degree to which a person is fully functioning. Research suggests that health care technologies like mHealth apps can contribute to improved hedonic and eudaemonic well-being by providing pleasant experiences to patients and by helping the patients to reach their goals [40,41]. We therefore hypothesize as follows:

H3a: Behavioral engagement with mHealth apps positively affects the hedonic well-being of chronically ill patients.

H3b: Behavioral engagement with mHealth apps positively affects the eudaemonic well-being of chronically ill patients.

Methods

Research Design and Procedure

In this study, we rely upon a scenario-based survey study, which is very common in business research [42] and in technology acceptance studies [43]; an increasing use of scenario-based surveys are also being used in health care [44]. Scenario-based survey studies have the advantage of eliminating the difficulties associated with observation or enactment of events in real life, such as in this study, with undesirable outcomes, and with not reaching a sufficiently large sample size, as can happen when forcing patients to use a nonintegrated app [42]. Compared to recall-based surveys, scenario-based surveys also have the advantage of reducing biases from memory retrieval [45].

This scenario-based survey study involves a between-participant 2×2 design and introduces participants to a scenario. In all scenarios, the patient receives a pamphlet with information on a fictional mHealth app, but the scenarios differed in terms of the recommendation by the physician to use the mHealth app (ie, strong vs weak recommendation to use the mHealth app) and app integration (integrated vs nonintegrated mHealth app). This 2×2 design has 4 different possible scenarios, and each participant was randomly assigned to 1 of these 4 scenarios. After reading the scenario, the participant filled out a questionnaire. The scenarios are detailed in Multimedia Appendix 1.

Sampling

G*Power 3.1.9 (Heinrich Heine Universität) was used to calculate the required sample size for detecting a medium effect (Cohen d=0.5) in an independent sample t-test (2-tailed). With 80% power at an α level of .05, a total sample size of 204 participants (51 per group) was needed to test the hypotheses. To achieve the required sample size, respondents with chronic conditions were recruited by more than 60 organizations representing the interests of chronically ill people in the Flemish region of Belgium through sharing the survey in their e-newsletter, website, or Facebook page. Eligible respondents (1) had been diagnosed by their physician with a chronic disease and (2) were aged between 18 and 65 years. These age boundaries were set because the empirical literature identified strong differences in the adoption of technology among young people, adults, and older adults [46]. In total, 722 respondents completed the questionnaire. After quality checks (including age and condition checks and a control question), 521 respondents were retained.

Ethical Considerations

The Ghent University Hospital review board approved the study protocol (2019/1975-670202042704), and participants were asked for consent.

Measures

We conducted a web-based survey from March to May 2019. Study data were collected and managed using REDCap electronic data capture tools hosted at Ghent University Hospital [47,48]. The survey involved 5 different constructs, including skip patterns. All constructs were measured using previously validated multi-item scales with proven validity and reliability (See Multimedia Appendix 2 [49-53]). The original scales were translated into Dutch using the forward and backward translation technique. Although validated by previous research, the measurement instrument was further tested to ensure reliability within the study context. Cronbach α values of the validated constructs ranged from .748 to .952 and showed that the reliability requirements were met. Reponses were provided using a 7-point Likert scale, with anchors ranging from 1 ("strongly disagree") to 7 ("strongly agree"). Finally, the survey included questions about age, gender, and the duration of the chronic condition, as it is common to include these demographics in research relating to chronic conditions [54].

Analytical Approach

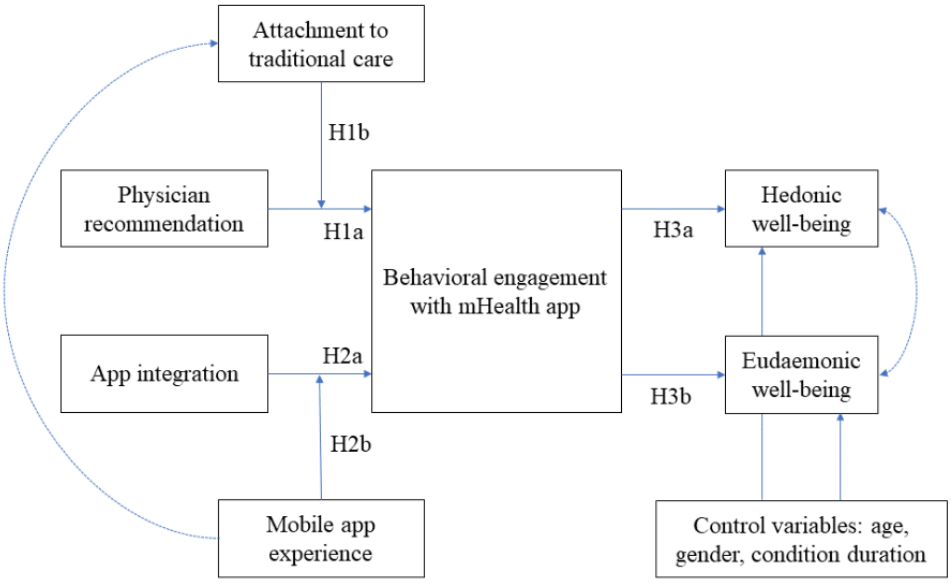
We assessed the experimental interventions by comparing the mean score on a single item measuring physician recommendation ("my physician recommends me to use this app") and a single item measuring app integration ("this is an integrated app") between the different scenarios. The mean differences for both interventions were significant. The mean score on a 7-point Likert scale for physician recommendation was 4.24 in the weak physician recommendation scenario versus 5.47 in the strong physician recommendation scenario (P=.02). The mean score for app integration was 3.67 in the scenario with a nonintegrated app versus 5.87 in the scenario with an integrated app (P=.004).

To simultaneously test all hypotheses (including drivers, consequences, and moderators), we used a mediation approach [55] with Bayesian estimation [56]. As suggested by Iacobucci [57] and Yuan and MacKinnon [56], the following 3 equations were jointly estimated using structural equation modeling in order to test our proposed conceptual model:

in which the $BehavioralEngagement_i$ denotes the individual i's (i=1 to 521) behavioral engagement with mHealth apps and the $WellBeing_{di}$ denotes the 2 (d=1 to 2) well-being dimensions: hedonic well-being (d=1) and eudaemonic well-being (d=2). β_1 denotes the effect of the influence of the physician ($PhysicianRecommendation_i$) on behavioral engagement in order to test H1a, whereas β_2 denotes the effect of app integration ($AppIntegration_i$) on behavioral engagement in order to test H2a. β_{10d}represents the effect of behavioral engagement on both hedonic well-being and eudaemonic well-being and respectively enables testing of H3a and H3b. β_3 denotes the moderating effect of patient attachment to traditional care on the impact of physician recommendation on patients' behavioral engagement ($PhysicianRecommendation_i TraditionalCare_i$) and allows investigation of H1b. β_5 denotes the moderating effect of patient mobile app experience on the relationship that the app integration has on the patient's behavioral engagement ($AppIntegration_i MobileAppExperience_i$) and enables investigation of H2b. The Σ_{1i},Σ_{2i}, and Σ_{3di} are error terms with intercorrelation ρ which, in line with the well-being literature [58], accounts for the interdependency between hedonic and eudaemonic well-being (see Figure 1). $Controls_{ci}$ is a vector of control variables, including patient age, gender, and duration of condition.

Because of structural equation modeling, the paths as specified in equations 1 to 2 are modeled in combination with the measurement model. The measurement model provided evidence of construct validity and discriminant validity, and additional tests revealed our data to be free from the common method and collinearity biases (See Multimedia Appendix 2 and Multimedia Appendix 3 [55,56,59-67] for more details). In addition, the model convergence was inspected and revealed evidence of a well-fitting model (see Model Convergence Assessment in Multimedia Appendix 3 for more details). Finally, the structural equation models, in line with the technology adoption literature [49], are linked between the mobile app experience and attachment to traditional care, as shown in Figure 1.

Figure 1. Proposed conceptual model. H: hypothesis; mHealth: mobile health.



Results

The chronic conditions that were the most prevalent among the respondents were orthopedic and rheumatic diseases (153/521, 29.4%), neurological diseases (133/521, 25.5%), and lung diseases (109/521, 20.9%); 130 respondents (25%) indicated comorbidity. All respondents had a smartphone, but mobile app experience varied; 240 respondents (46.1%) had no mobile app experience, 192 respondents (36.9%) had low-to-moderate mobile app experience, and 89 respondents (17.1%) had high mobile app experience. Table 1 gives an overview of the other demographics of the research participants.

Table 2 presents the model's findings. The findings of Model 1 reveal that the physician's recommendation has a positive significant influence on the patient's behavioral engagement with mHealth apps (β_1=.325, *P*=.001). H1a is thus supported. The findings also reveal that app integration has a positive effect on behavioral engagement (β_2=.225, *P*=.02), confirming H2a. As anticipated, behavioral engagement has a positive effect on both hedonic well-being (β_{10_1}=.641, *P*=.001) and eudaemonic well-being (β_{10_2}=.724, *P*=.001), thereby supporting H3a and H3b. To test the moderating effect of patients' attachment to traditional care and their mobile app experience, Model 2 in Table 2 reports the parameter estimates of the moderating effects. Model 2 reveals that a patient's attachment to traditional care only has a direct effect on behavioral engagement (β_4=–.420, *P*=.001) since the interaction terms were found to be insignificant, albeit negative as anticipated (β_3=–.157, *P*=.14). H1b is thus rejected. With regard to the moderating effect of mobile app experience, our findings show a positive significant moderating effect on the relationship between app integration and behavioral engagement (β_5=.232, *P*=.02), thereby confirming H2b. Figure 2 depicts this significant relationship, showing that there is a high positive impact of mobile app experience on behavioral engagement with mobile app integration. In addition, our model findings show that gender and condition duration have no significant effects. Interestingly, age is found to only have a negative effect on hedonic well-being (β_{11_1}=–.008, *P*=.002), whereas no such effect was observed for eudaemonic well-being. In addition, people with more mobile app experience were found to have less attachment to traditional care (β_8=–.094, *P*=.01). Finally, the proportion of variance explained (R^2) equals 21% for behavioral engagement and 52.8% and 62.2% for hedonic and eudaemonic well-being, respectively, thereby providing support for the strong effect of app integration and physician recommendation via mediation of the patients' behavioral engagement on patients' hedonic and eudaemonic well-being.

Table 1. Participant demographics.

Demographics	All respondents (N=521)	Scenario 1[a] (n=128)	Scenario 2[b] (n=123)	Scenario 3[c] (n=141)	Scenario 4[d] (n=129)	*P* value
Age (years), mean (min-max, SD)	44.24 (18-65, 13.11)	45.41 (21-65, 12.47)	44.58 (19-65, 13.46)	43.26 (18-65, 13.21)	43.84 (18-65, 13.32)	.57
Gender, n (%)						.44
Male	125 (24)	27 (21.1)	34 (27.6)	30 (21.3)	34 (26.4)	
Female	396 (76)	101 (78.9)	89 (72.4)	111 (78.7)	95 (73.6)	
Condition duration (years), mean (min-max, SD)	12.20 (0-64, 11.09)	11.95 (0-64, 11.95)	14.3 (0-64, 11.65)	11.26 (0-47, 10.83)	11.47 (1-58, 9.90)	.11

[a]Scenario 1: strong physician recommendation + integrated app.

[b]Scenario 2: weak physician recommendation + nonintegrated app.

[c]Scenario 3: strong physician recommendation + nonintegrated app.

[d]Scenario 4: weak physician recommendation + integrated app.

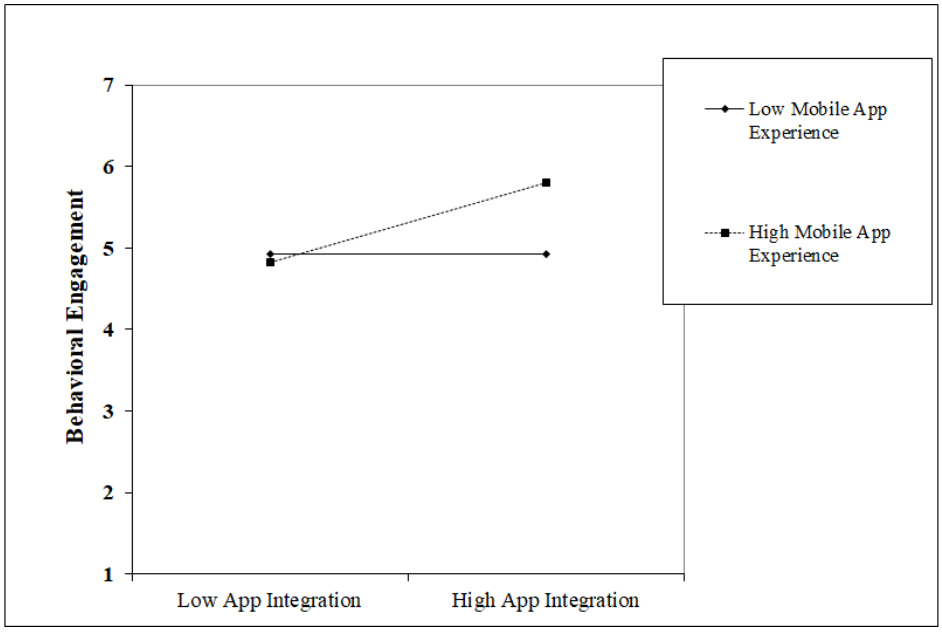
Table 2. Model findings.

	Model 1			Model 2			
	Behavioral engagement	Hedonic well-being	Eudaemonic well-being	Attachment to traditional care	Behavioral engagement	Hedonic well-being	Eudaemonic well-being
Independent variables, β (*P* value)							
Physician recommendation	.325 (.001)[a]	N/A[b]	N/A	N/A	.304 (.002)[a]	N/A	N/A
App integration	.225 (.02)[a]	N/A	N/A	N/A	.238 (.01)[a]	N/A	N/A
Behavioral engagement	N/A	.641 (.001)[a]	.724 (.001)[a]	N/A	N/A	.642 (.001)[a]	.723 (.001)[a]
Control variables, β (*P* value)							
Age	N/A	–.008 (.008)[a]	–.004 (.06)	N/A	N/A	–.008 (.002)[a]	–.004 (.06)
Gender (1=female, 0=male)	N/A	.090 (.16)	–.059 (.24)	N/A	N/A	.089 (.15)	–.070 (.19)
Condition duration	N/A	.003 (.19)	.004 (.13)	N/A	N/A	.003 (.21)	.003 (.14)
Testing moderating effects, β (*P* value)							
Physician recommendation×Attachment to traditional care	N/A	N/A	N/A	N/A	–.157 (.14)	N/A	N/A
Attachment to traditional care	N/A	N/A	N/A	N/A	–.420 (.001)[a]	N/A	N/A
App integration×Mobile app experience	N/A	N/A	N/A	N/A	.232 (.02)[a]	N/A	N/A
Mobile app experience	N/A	N/A	N/A	–.094 (.01)[a]	.196 (.01)[a]	N/A	N/A
Correlation error term (*P* value)	N/A	0.107 (.001)[a]	0.107 (.001)[a]	N/A	N/A	0.103 (.001)[a]	0.103 (.001)[a]
R^2 (proportion of variance explained; %)	2.8	52.3	61.8	1.4	20.8	52.8	62.2

[a]Effect size (β) is significant.

[b]N/A: not applicable.

Figure 2. The moderating influence of mobile app experience.



Discussion

Principal Findings

mHealth has several benefits for patients, especially for those who experience chronic conditions [17]. mHealth can enable patients to manage their condition [68], which is crucial for their well-being [69]. However, beginning to use mHealth apps can be challenging, as many patients have neither experience with nor confidence in using them [17,70]. Although health care technology adoption research has typically focused on the need for user friendliness, usefulness, and performance expectancy [16], we complement this line of research by exploring the importance of social factors (here, physician recommendation) and app-related factors (here, app integration) when behaviorally engaging patients with mHealth and when considering how this engagement impacts their well-being.

With regard to the social factors, we found that physician recommendation positively influences behavioral engagement and consequently, patient well-being. The important role laid out for the physician when behaviorally engaging patients with mHealth apps resonates with evidence about their importance in stimulating other types of patient behavior [71] and enhancing patient well-being [72]. Indeed, the physician is in a unique position to motivate their patients to use mHealth apps, especially when they have longstanding relationships with them [73].

Although we have observed that patients who are more inclined to traditional care are less inclined to use mHealth, it is remarkable to note that these patients can also be motivated to use mHealth apps to the same degree as patients who are more open to receiving modern care (with the use of mobile apps). As such, physicians should be careful not to assume that patients who are more accustomed to traditional care will be harder to motivate to use mHealth. This potential bias should not lead to neglect of this part of the patient population, particularly in the (post-)COVID world where virtual care has become a more integral part of traditional care [74]. Furthermore, although we have focused in this research on the motivating role of the physician, other parties (health care organizations, government, etc) can also be of importance in recommending patients to use mHealth apps. Future research could thus focus on marketing strategies that these parties could deploy to encourage mHealth use. Anderson et al [75] have recently called for an increased use of marketing techniques in health care. Indeed, marketing communication can be instrumental in motivating patients to use new technologies and engage in self-care by creating a positive attitude toward health care technology [76,77].

Besides social factors, our results also demonstrate the importance of app integration. Specifically, integrated apps can have an important positive impact on a patients' behavioral engagement. The major app stores already offer a staggering 350,000 mHealth apps, most of which remain unsuccessful because of limited behavioral engagement [78]. Designing yet another nonintegrated mHealth app can only add to this pile of underutilized apps. By designing mHealth apps, which have meaningful interaction with existing systems used by the patient (eg, appointment scheduler, patient file, health data), behavioral engagement can be enhanced and the continued usage of an mHealth app can be improved.

In line with previous research [22], this study also shows a moderating role for app experience. Patients with more experience in using apps will place greater value on mHealth apps that are integrated with other health platforms. Given that app experience is rising among all parts of the population [79], the importance of offering integrated apps will only increase in the future. Companies who develop mHealth apps and health care organizations who implement them should focus not only on apps' appearances and capabilities but also show great care in ensuring that apps are integrated into existing health platforms.

Finally, our data were collected in 2019, shortly before the COVID-19 pandemic. The advantages of mHealth apps during

a pandemic have been well-documented [74]. Further, numerous papers on the roles of eHealth, telehealth, and telemedicine in delivering health care services to chronically ill patients during the COVID-19 pandemic have been published [80,81]. The COVID-19 pandemic has shown the importance of mHealth as a means of interacting with patients and providing care. It would be interesting to know how the pandemic has changed the way in which patients use and feel about mHealth apps.

Limitations and Future Research

Although this study gives clear indications of the importance of physician recommendation and app integration for behavioral engagement and well-being among chronically ill patients, it is not without limitations, and the results should be interpreted accordingly. First, the respondents self-selected to participate in this study. As a result, generalization of the results to a broader population should be done with care. However, because we opted for self-selection, a larger number of participants was recruited. Second, this study uses self-reported data and not actual behavior. This makes the study more vulnerable to self-report biases such as socially desirable answers and information bias. Future research could go beyond scenario-based research and implement the proposed interventions in a randomized controlled trial. Third, this study utilized cross-sectional data, which limits the possibilities of drawing conclusions on causal relationships. Future research could benefit from a longitudinal approach by collecting data at different points in time among the same respondents [12]. It can be envisioned that the engagement of chronically ill patients, in particular, might differ in time, as the severity of their condition or their need for support fluctuates. Fourth, the strength of the relationship between the physician and patient was not included in the study design. Future studies could include measures of this relationship, since the relationship between physician and patient may act as a significant mediator of the relationship between physician recommendation, behavioral engagement with mHealth apps among chronically ill patients, and their well-being.

Conclusion

An ever increasing number of mHealth apps are being developed by both commercial enterprises and health care organizations. Although these apps can have a positive impact on patient well-being, various studies have shown that simply designing an effective app does not guarantee their adoption by users. This study focused on the importance of physician recommendation and app integration in increasing behavioral engagement and well-being among chronically ill patients. It highlights the importance of app developers considering behavioral engagement during and after the development of mHealth apps. During development, attention should be given to ensuring app integration so that communication and interaction with existing health care systems is possible. Integration is an important characteristic that can encourage patients to start using an app—especially when they are experienced app users. After development, it is important to motivate patients to adopt the mHealth app. This study has shown that physicians have an important role to play in motivating chronically ill patients to engage with mHealth apps.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Presented scenarios.
[DOCX File , 16 KB - mhealth_v10i8e33772_app1.docx]

Multimedia Appendix 2
Table on construct items (wording), factor scores, and composite reliability.
[DOCX File , 22 KB - mhealth_v10i8e33772_app2.docx]

Multimedia Appendix 3
Measurement model, common method bias, collinearity, and robustness tests.
[DOCX File , 23 KB - mhealth_v10i8e33772_app3.docx]

References

1. Kim J, Ryu B, Cho S, Heo E, Kim Y, Lee J, et al. Impact of Personal Health Records and Wearables on Health Outcomes and Patient Response: Three-Arm Randomized Controlled Trial. JMIR Mhealth Uhealth 2019 Jan 04;7(1):e12070 [FREE Full text] [doi: 10.2196/12070] [Medline: 30609978]
2. Steinhubl SR, Muse ED, Topol EJ. The emerging field of mobile health. Sci Transl Med 2015 Apr 15;7(283):283rv3 [FREE Full text] [doi: 10.1126/scitranslmed.aaa3487] [Medline: 25877894]
3. Zhou L, DeAlmeida D, Parmanto B. Applying a User-Centered Approach to Building a Mobile Personal Health Record App: Development and Usability Study. JMIR Mhealth Uhealth 2019 Jul 05;7(7):e13194 [FREE Full text] [doi: 10.2196/13194] [Medline: 31278732]

4. Ramirez V, Johnson E, Gonzalez C, Ramirez V, Rubino B, Rossetti G. Assessing the Use of Mobile Health Technology by Patients: An Observational Study in Primary Care Clinics. JMIR Mhealth Uhealth 2016 Apr 19;4(2):e41 [FREE Full text] [doi: 10.2196/mhealth.4928] [Medline: 27095507]
5. Hossain I, Ang YN, Chng HT, Wong PS. Patients' attitudes towards mobile health in Singapore: a cross-sectional study. Mhealth 2019;5:34 [FREE Full text] [doi: 10.21037/mhealth.2019.08.07] [Medline: 31620461]
6. Marcolino MS, Oliveira JAQ, D'Agostino M, Ribeiro AL, Alkmim MBM, Novillo-Ortiz D. The Impact of mHealth Interventions: Systematic Review of Systematic Reviews. JMIR Mhealth Uhealth 2018 Jan 17;6(1):e23 [FREE Full text] [doi: 10.2196/mhealth.8873] [Medline: 29343463]
7. Lee J, Choi M, Lee SA, Jiang N. Effective behavioral intervention strategies using mobile health applications for chronic disease management: a systematic review. BMC Med Inform Decis Mak 2018 Feb 20;18(1):12 [FREE Full text] [doi: 10.1186/s12911-018-0591-0] [Medline: 29458358]
8. Li J, Zhang C, Li X, Zhang C. Patients' emotional bonding with MHealth apps: An attachment perspective on patients' use of MHealth applications. International Journal of Information Management 2020 Apr;51:102054 [FREE Full text] [doi: 10.1016/j.ijinfomgt.2019.102054]
9. del Río-Lanza A, Suárez-Vázquez A, Suárez-Álvarez L, Iglesias-Argüelles V. Mobile health (mhealth): facilitators and barriers of the intention of use in patients with chronic illnesses. Journal of Communication in Healthcare 2020 Jun 12;13(2):138-146. [doi: 10.1080/17538068.2020.1777513]
10. Szinay D, Jones A, Chadborn T, Brown J, Naughton F. Influences on the Uptake of and Engagement With Health and Well-Being Smartphone Apps: Systematic Review. J Med Internet Res 2020 May 29;22(5):e17572 [FREE Full text] [doi: 10.2196/17572] [Medline: 32348255]
11. Zhang Y, Liu C, Luo S, Xie Y, Liu F, Li X, et al. Factors Influencing Patients' Intentions to Use Diabetes Management Apps Based on an Extended Unified Theory of Acceptance and Use of Technology Model: Web-Based Survey. J Med Internet Res 2019 Aug 13;21(8):e15023 [FREE Full text] [doi: 10.2196/15023] [Medline: 31411146]
12. Vaghefi I, Tulu B. The Continued Use of Mobile Health Apps: Insights From a Longitudinal Study. JMIR Mhealth Uhealth 2019 Aug 29;7(8):e12983 [FREE Full text] [doi: 10.2196/12983] [Medline: 31469081]
13. Gavin K. Health apps could help older adults, but most don't use them. mHealth Lab. URL: https://labblog.uofmhealth.org/industry-dx/health-apps-could-help-older-adults-but-most-dont-use-them [accessed 2021-09-21]
14. Dou K, Yu P, Deng N, Liu F, Guan Y, Li Z, et al. Patients' Acceptance of Smartphone Health Technology for Chronic Disease Management: A Theoretical Model and Empirical Test. JMIR Mhealth Uhealth 2017 Dec 06;5(12):e177 [FREE Full text] [doi: 10.2196/mhealth.7886] [Medline: 29212629]
15. Rahimi B, Nadri H, Lotfnezhad Afshar H, Timpka T. A Systematic Review of the Technology Acceptance Model in Health Informatics. Appl Clin Inform 2018 Jul;9(3):604-634 [FREE Full text] [doi: 10.1055/s-0038-1668091] [Medline: 30112741]
16. Ammenwerth E. Technology Acceptance Models in Health Informatics: TAM and UTAUT. Stud Health Technol Inform. Jul 2019 Jan 01;263:30-71. [doi: 10.3233/shti190111]
17. Pham Q, Graham G, Carrion C, Morita PP, Seto E, Stinson JN, et al. A Library of Analytic Indicators to Evaluate Effective Engagement with Consumer mHealth Apps for Chronic Conditions: Scoping Review. JMIR Mhealth Uhealth 2019 Jan 18;7(1):e11941 [FREE Full text] [doi: 10.2196/11941] [Medline: 30664463]
18. Ryan RM, Deci EL. On happiness and human potentials: a review of research on hedonic and eudaimonic well-being. Annu Rev Psychol 2001;52:141-166. [doi: 10.1146/annurev.psych.52.1.141] [Medline: 11148302]
19. Maruping LM, Bala H, Venkatesh V, Brown SA. Going beyond intention: Integrating behavioral expectation into the unified theory of acceptance and use of technology. Journal of the Association for Information Science and Technology 2016 Jun 03;68(3):623-637. [doi: 10.1002/asi.23699]
20. Kamal S, Shafiq M, Kakria P. Investigating acceptance of telemedicine services through an extended technology acceptance model (TAM). Technology in Society 2020 Feb;60:101212 [FREE Full text] [doi: 10.1016/j.techsoc.2019.101212]
21. Rashotte L. Social influence. The Blackwell encyclopedia of sociology 2007 Jan 01:4426-4429. [doi: 10.1002/9781405165518.wbeoss154]
22. Cajita MI, Hodgson NA, Lam KW, Yoo S, Han H. Facilitators of and Barriers to mHealth Adoption in Older Adults With Heart Failure. Comput Inform Nurs 2018 Aug;36(8):376-382 [FREE Full text] [doi: 10.1097/CIN.0000000000000442] [Medline: 29742549]
23. Apolinário-Hagen J, Menzel M, Hennemann S, Salewski C. Acceptance of Mobile Health Apps for Disease Management Among People With Multiple Sclerosis: Web-Based Survey Study. JMIR Form Res 2018 Dec 12;2(2):e11977 [FREE Full text] [doi: 10.2196/11977] [Medline: 30684408]
24. Dai B, Larnyo E, Tetteh EA, Aboagye AK, Musah AI. Factors Affecting Caregivers' Acceptance of the Use of Wearable Devices by Patients With Dementia: An Extension of the Unified Theory of Acceptance and Use of Technology Model. Am J Alzheimers Dis Other Demen 2020;35:1533317519883493 [FREE Full text] [doi: 10.1177/1533317519883493] [Medline: 31679390]
25. Zhang Y, Koch S. Mobile health apps in Sweden: what do physicians recommend? In: Studies in Health Technology and Informatics. Amesterdam, Netherlands: IOS Press; Jan 01, 2015:793-797.

26. Rogers E, Singhal A, Quinlan M. Diffusion of innovations. An integrated approach to communication theory and research. Routledge 2014 Jan 01:432-448. [doi: 10.4324/9780203710753-35]
27. Zolkepli IA, Kamarulzaman Y. Social media adoption: The role of media needs and innovation characteristics. Computers in Human Behavior 2015 Feb;43:189-209. [doi: 10.1016/j.chb.2014.10.050]
28. Arevian AC, O'Hora J, Rosser J, Mango JD, Miklowitz DJ, Wells KB. Patient and Provider Cocreation of Mobile Texting Apps to Support Behavioral Health: Usability Study. JMIR Mhealth Uhealth 2020 Jul 29;8(7):e12655 [FREE Full text] [doi: 10.2196/12655] [Medline: 32723714]
29. Kildea J, Battista J, Cabral B, Hendren L, Herrera D, Hijal T, et al. Design and Development of a Person-Centered Patient Portal Using Participatory Stakeholder Co-Design. J Med Internet Res 2019 Feb 11;21(2):e11371 [FREE Full text] [doi: 10.2196/11371] [Medline: 30741643]
30. Holmgren AJ, Patel V, Adler-Milstein J. Progress In Interoperability: Measuring US Hospitals' Engagement In Sharing Patient Data. Health Aff (Millwood) 2017 Oct 01;36(10):1820-1827. [doi: 10.1377/hlthaff.2017.0546] [Medline: 28971929]
31. Marceglia S, Fontelo P, Rossi E, Ackerman MJ. A Standards-Based Architecture Proposal for Integrating Patient mHealth Apps to Electronic Health Record Systems. Appl Clin Inform 2015;6(3):488-505 [FREE Full text] [doi: 10.4338/ACI-2014-12-RA-0115] [Medline: 26448794]
32. Tomlinson M, Rotheram-Borus MJ, Swartz L, Tsai AC. Scaling up mHealth: where is the evidence? PLoS Med 2013;10(2):e1001382 [FREE Full text] [doi: 10.1371/journal.pmed.1001382] [Medline: 23424286]
33. Lehne M, Sass J, Essenwanger A, Schepers J, Thun S. Why digital medicine depends on interoperability. NPJ Digit Med 2019;2:79 [FREE Full text] [doi: 10.1038/s41746-019-0158-1] [Medline: 31453374]
34. van Heerden A, Tomlinson M, Swartz L. Point of care in your pocket: a research agenda for the field of m-health. Bull World Health Org 2012 May 01;90(5):393-394. [doi: 10.2471/blt.11.099788]
35. Estrin D, Sim I. Health care delivery. Open mHealth architecture: an engine for health care innovation. Science 2010 Nov 05;330(6005):759-760. [doi: 10.1126/science.1196187] [Medline: 21051617]
36. Gordon W, Catalini C. Blockchain Technology for Healthcare: Facilitating the Transition to Patient-Driven Interoperability. Comput Struct Biotechnol J 2018;16:224-230 [FREE Full text] [doi: 10.1016/j.csbj.2018.06.003] [Medline: 30069284]
37. Hilliard ME, Hahn A, Ridge AK, Eakin MN, Riekert KA. User Preferences and Design Recommendations for an mHealth App to Promote Cystic Fibrosis Self-Management. JMIR Mhealth Uhealth 2014 Oct 24;2(4):e44 [FREE Full text] [doi: 10.2196/mhealth.3599] [Medline: 25344616]
38. Vichta R, Gwinner K, Collyer B. What would we use and how would we use it? Can digital technology be used to both enhance and evaluate well-being outcomes with highly vulnerable and disadvantaged young people? Evaluation Journal of Australasia 2018 Nov 28;18(4):222-233. [doi: 10.1177/1035719x18804638]
39. Yaden DB, Eichstaedt JC, Medaglia JD. The Future of Technology in Positive Psychology: Methodological Advances in the Science of Well-Being. Front Psychol 2018;9:962 [FREE Full text] [doi: 10.3389/fpsyg.2018.00962] [Medline: 29967586]
40. Botella C, Banos R, Guillen V. Positive technologies for improving health and well-being. Positive psychology interventions in practice. Springer International Publishing AG 2017 Jan 01:219-234. [doi: 10.1007/978-3-319-51787-2_13]
41. Aboelmaged M, Hashem G, Mouakket S. Predicting subjective well-being among mHealth users: a readiness – value model. International Journal of Information Management 2021 Feb;56:102247. [doi: 10.1016/j.ijinfomgt.2020.102247]
42. Van Vaerenbergh Y, De Keyser A, Larivière B. Customer intentions to invoke service guarantees: do excellence in service recovery, type of guarantee and cultural orientation matter? Managing Service Quality: An International Journal. Managing Service Quality: An International Journal 2014 Jan 07;24(1):45-62. [doi: 10.1108/msq-06-2013-0115]
43. Biswas M, Romeo M, Cangelosi A, Jones RB. Are older people any different from younger people in the way they want to interact with robots? Scenario based survey. J Multimodal User Interfaces 2019 Jul 24;14(1):61-72. [doi: 10.1007/s12193-019-00306-x]
44. Gaziel Yablowitz M, Dölle S, Schwartz DG, Worm M. Proximity-Based Emergency Response Communities for Patients With Allergies Who Are at Risk of Anaphylaxis: Clustering Analysis and Scenario-Based Survey Study. JMIR Mhealth Uhealth 2019 Aug 22;7(8):e13414 [FREE Full text] [doi: 10.2196/13414] [Medline: 31441432]
45. Smith EE, Jonides J. Storage and executive processes in the frontal lobes. Science 1999 Mar 12;283(5408):1657-1661. [doi: 10.1126/science.283.5408.1657] [Medline: 10073923]
46. O'brien MA, Rogers WA, Fisk AD. Understanding age and technology experience differences in use of prior knowledge for everyday technology interactions. ACM Trans. Access. Comput 2012 Mar;4(2):1-27. [doi: 10.1145/2141943.2141947]
47. Harris P, Taylor R, Minor B, Elliott V, Fernandez M, O'Neal L, REDCap Consortium. The REDCap consortium: Building an international community of software platform partners. J Biomed Inform 2019 Jul;95:103208 [FREE Full text] [doi: 10.1016/j.jbi.2019.103208] [Medline: 31078660]
48. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
49. Venkatesh, Thong, Xu. Consumer Acceptance and Use of Information Technology: Extending the Unified Theory of Acceptance and Use of Technology. MIS Quarterly 2012;36(1):157. [doi: 10.2307/41410412]

50. Cimperman M, Makovec Brenčič M, Trkman P. Analyzing older users' home telehealth services acceptance behavior-applying an Extended UTAUT model. Int J Med Inform 2016 Jun;90:22-31. [doi: 10.1016/j.ijmedinf.2016.03.002] [Medline: 27103194]
51. Sun P, Kong F. Affective Mediators of the Influence of Gratitude on Life Satisfaction in Late Adolescence. Soc Indic Res 2013 May 18;114(3):1361-1369. [doi: 10.1007/s11205-013-0333-8]
52. Dabholkar P, Bagozzi R. An attitudinal model of technology-based self-service: moderating effects of consumer traits and situational factors. Journal of the academy of marketing science 2002 Jan 01;30(3):184-201. [doi: 10.1177/00970302030003001]
53. Curran J, Meuter M. Self-service technology adoption: comparing three technologies. Journal of services marketing 2005 Mar 01:103-113. [doi: 10.1108/08876040510591411]
54. Raghupathi W, Raghupathi V. An Empirical Study of Chronic Diseases in the United States: A Visual Analytics Approach. Int J Environ Res Public Health 2018 Mar 01;15(3):431 [FREE Full text] [doi: 10.3390/ijerph15030431] [Medline: 29494555]
55. Zhao X, Lynch JG, Chen Q. Reconsidering Baron and Kenny: Myths and Truths about Mediation Analysis. J Consum Res 2010 Aug 01;37(2):197-206. [doi: 10.1086/651257]
56. Yuan Y, MacKinnon DP. Bayesian mediation analysis. Psychol Methods 2009 Dec;14(4):301-322 [FREE Full text] [doi: 10.1037/a0016972] [Medline: 19968395]
57. Iacobucci D. Mediation Analysis. Quantitative Applications in the Social Sciences 2008 Jan 01:1-89. [doi: 10.4135/9781412984966]
58. Henkens B, Verleye K, Larivière B. The smarter, the better?! Customer well-being, engagement, and perceptions in smart service systems. International Journal of Research in Marketing 2021 Jun;38(2):425-447. [doi: 10.1016/j.ijresmar.2020.09.006]
59. R: A language and environment for statistical computing. R Foundation for Statistical Computing. 2014 Jan 01. URL: https://www.r-project.org/foundation/ [accessed 2022-07-18]
60. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural Equation Modeling: A Multidisciplinary Journal 1999 Jan;6(1):1-55. [doi: 10.1080/10705519909540118]
61. Netemeyer R, Bearden W, Sharma S. Scaling Procedures. Issues and Applications. Personnel Psychology 2003 Dec 01;56(4):1088-1090. [doi: 10.4135/9781412985772]
62. Bagozzi RP, Yi Y. On the evaluation of structural equation models. JAMS 1988 Mar;16(1):74-94. [doi: 10.1007/bf02723327]
63. Fornell C, Larcker DF. Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. Journal of Marketing Research 1981 Feb;18(1):39. [doi: 10.2307/3151312]
64. Podsakoff PM, MacKenzie SB, Lee J, Podsakoff NP. Common method biases in behavioral research: a critical review of the literature and recommended remedies. J Appl Psychol 2003 Oct;88(5):879-903. [doi: 10.1037/0021-9010.88.5.879] [Medline: 14516251]
65. Hair J, Black W, Babin B, Anderson R. A global perspective. In: Multivariate Data Analysis. India: Cengage India Private Limited; Jan 01, 2010:640-800.
66. van Heerde HJ, Dinner IM, Neslin SA. Engaging the unengaged customer: The value of a retailer mobile app. International Journal of Research in Marketing 2019 Sep;36(3):420-438. [doi: 10.1016/j.ijresmar.2019.03.003]
67. Gelman A, Rubin DB. Inference from Iterative Simulation Using Multiple Sequences. Statist. Sci 1992 Nov 1;7(4):457-472. [doi: 10.1214/ss/1177011136]
68. Lin Y, Tudor-Sfetea C, Siddiqui S, Sherwani Y, Ahmed M, Eisingerich AB. Effective Behavioral Changes through a Digital mHealth App: Exploring the Impact of Hedonic Well-Being, Psychological Empowerment and Inspiration. JMIR Mhealth Uhealth 2018 Jun 15;6(6):e10024 [FREE Full text] [doi: 10.2196/10024] [Medline: 29907557]
69. Kidd T, Carey N, Mold F, Westwood S, Miklaucich M, Konstantara E, et al. A systematic review of the effectiveness of self-management interventions in people with multiple sclerosis at improving depression, anxiety and quality of life. PLoS One 2017;12(10):e0185931 [FREE Full text] [doi: 10.1371/journal.pone.0185931] [Medline: 29020113]
70. Holdener M, Gut A, Angerer A. Applicability of the User Engagement Scale to Mobile Health: A Survey-Based Quantitative Study. JMIR Mhealth Uhealth 2020 Jan 03;8(1):e13244 [FREE Full text] [doi: 10.2196/13244] [Medline: 31899454]
71. Riedl D, Schüßler G. The Influence of Doctor-Patient Communication on Health Outcomes: A Systematic Review. Z Psychosom Med Psychother 2017 Jun;63(2):131-150. [doi: 10.13109/zptm.2017.63.2.131] [Medline: 28585507]
72. Samuel CA, Mbah O, Schaal J, Eng E, Black KZ, Baker S, et al. The role of patient-physician relationship on health-related quality of life and pain in cancer patients. Support Care Cancer 2020 Jun;28(6):2615-2626. [doi: 10.1007/s00520-019-05070-y] [Medline: 31620925]
73. Robinson CA. Trust, Health Care Relationships, and Chronic Illness: A Theoretical Coalescence. Glob Qual Nurs Res 2016;3:2333393616664823 [FREE Full text] [doi: 10.1177/2333393616664823] [Medline: 28508016]
74. Hollander JE, Carr BG. Virtually Perfect? Telemedicine for Covid-19. N Engl J Med 2020 Apr 30;382(18):1679-1681. [doi: 10.1056/nejmp2003539]
75. Anderson S, Rayburn SW, Sierra JJ. Future thinking: the role of marketing in healthcare. EJM 2019 Aug 12;53(8):1521-1545. [doi: 10.1108/ejm-10-2017-0779]

76. Elrod JK, Fortenberry JL. Formulating productive marketing communications strategy: a major health system's experience. BMC Health Serv Res 2018 Dec 14;18(Suppl 3):926 [FREE Full text] [doi: 10.1186/s12913-018-3676-7] [Medline: 30545343]
77. Singh RP, Banerjee N. Exploring the Influence of Celebrity Credibility on Brand Attitude, Advertisement Attitude and Purchase Intention. Global Business Review 2018 Sep 23;19(6):1622-1639. [doi: 10.1177/0972150918794974]
78. Franklin R. 11 surprising mobile health statistics. Mobius. 2021 Oct 25. URL: https://mobius.md/2021/10/25/11-mobile-health-statistics/ [accessed 2021-10-25]
79. Smith A. Record shares of Americans now own smartphones, have home broadband. Pew Research Center. 2017 Jan 12. URL: https://www.pewresearch.org/fact-tank/2017/01/12/evolution-of-technology/ [accessed 2021-10-25]
80. Bitar H, Alismail S. The role of eHealth, telehealth, and telemedicine for chronic disease patients during COVID-19 pandemic: A rapid systematic review. Digit Health 2021;7:20552076211009396 [FREE Full text] [doi: 10.1177/20552076211009396] [Medline: 33959378]
81. Tarricone R, Petracca F, Ciani O, Cucciniello M. Distinguishing features in the assessment of mHealth apps. Expert Rev Pharmacoecon Outcomes Res 2021 Aug;21(4):521-526. [doi: 10.1080/14737167.2021.1891883] [Medline: 33663324]

Abbreviations

mHealth: mobile health

Edited by L Buis; submitted 22.09.21; peer-reviewed by R Marshall, C Jonassaint, O El-Shahawy, I Madujibeya; comments to author 02.02.22; revised version received 30.05.22; accepted 17.06.22; published 26.08.22.

Please cite as:
Van Baelen F, De Regge M, Larivière B, Verleye K, Schelfout S, Eeckloo K
Role of Social and App-Related Factors in Behavioral Engagement With mHealth for Improved Well-being Among Chronically Ill Patients: Scenario-Based Survey Study
JMIR Mhealth Uhealth 2022;10(8):e33772
URL: https://mhealth.jmir.org/2022/8/e33772
doi:10.2196/33772
PMID:36018618

Original Paper

The Use of Digital Health Tools for Health Promotion Among Women With and Without Chronic Diseases: Insights From the 2017-2020 Health Information National Trends Survey

Kobi V Ajayi[1], MPH, MBA; Elizabeth Wachira[2], MPH, PhD; Henry K Onyeaka[3,4], MD; Tyra Montour[1], MPH; Samson Olowolaju[5], MPH; Whitney Garney[1], MPH, PhD

[1]Department of Health & Kinesiology, Texas A&M University, College Station, TX, United States

[2]Department of Health and Human Performance, Texas A&M University, Commerce, TX, United States

[3]Department of Psychiatry, Massachusetts General Hospital, Boston, MA, United States

[4]Department of Psychiatry, McLean Hospital, Belmont, MA, United States

[5]Department of Demography, College for Health, Community and Policy, University of Texas, San Antonio, TX, United States

Corresponding Author:
Kobi V Ajayi, MPH, MBA
Department of Health & Kinesiology
Texas A&M University
2929 Research Parkway
College Station, TX, 77843
United States
Phone: 1 9797396250
Email: omo_debare@tamu.edu

Abstract

Background: In the United States, almost 90% of women are at risk of at least one chronic condition. However, the awareness, management, and monitoring of these conditions are low and present a substantial public health problem. Digital health tools can be leveraged to reduce the alarmingly high rates of chronic condition–related mortality and morbidity in women.

Objective: This study aimed to investigate the 4-year trend of digital health use for health promotion among women with chronic conditions in the United States.

Methods: Data for this study were obtained from the 2017 to 2020 iterations of the Health Information Trends Survey 5. Separate weighted logistic regression models were conducted to test the unadjusted and adjusted association of the study variables and each digital health use. The 95% CI, adjusted odds ratio (aOR), and *P* value (.05) were reported. Analysis was conducted using Stata 17 software.

Results: In total, 8573 women were included in this study. The weighted prevalence of the use of a smartphone or tablet for various activities were as follows: track health goals, 50.3% (95% CI 48.4%-52.2%; 3279/7122); make a health decision, 43.6% (95% CI 41.9%-45.3%; 2998/7101); and discuss with a provider, 40% (95% CI 38.2%-41.8%; 2834/7099). In the preceding 12 months, 33% (95% CI 30.9%-35.2%; 1395/4826) of women used an electronic wearable device, 18.7% (95% CI 17.3%-20.2%; 1532/7653) shared health information, and 35.2% (95% CI 33.2%-37.3%; 2262/6349) sent or received an SMS text message with a health professional. Between 2017 and 2020, the weighted prevalence of having 0, 1, and multiple chronic conditions were 37.4% (2718/8564), 33.4% (2776/8564), and 29.3% (3070/8564), respectively. However, slightly above half (52.2%, 95% CI 0.50%-0.53%; 4756/8564) of US women reported having at least one chronic disease. Women with multiple chronic conditions had higher odds of using their tablet or smartphone to achieve a health-related goal (aOR 1.43, 95% CI 1.16-1.77; *P*=.001) and discuss with their provider (aOR 1.55 95% CI 1.20-2.00; *P*=.001) than those without any chronic conditions. Correspondingly, in the past 12 months, the odds of using an electronic wearable device (aOR 1.40, 95% CI 1.00-1.96; *P*=.04), sharing health information (aOR 1.91, 95% CI 1.46-2.51; *P*<.001), and communicating via SMS text messaging with a provider (aOR 1.31, 95% CI 1.02-1.68; *P*=.03) were significantly higher among women with chronic conditions than those without a chronic condition.

Conclusions: This study suggests that women with chronic conditions accept and integrate digital health tools to manage their care. However, certain subpopulations experience a digital disconnect that may exacerbate existing health inequities. Implications

for research and opportunities to leverage and integrate digital health tools to prevent, monitor, manage, and treat chronic conditions in women are discussed.

(JMIR Mhealth Uhealth 2022;10(8):e39520) doi:10.2196/39520

KEYWORDS

mHealth; health promotion; chronic disease; women; digital health; USA; United States; patient engagement

Introduction

Background

Although almost 90% of women in the United States are at risk of at least one chronic condition, the awareness, management, and monitoring of these conditions are low and present a substantial public health problem [1-4]. Women bear a disproportionate burden of chronic diseases, and in 2018 alone, the prevalence of multiple chronic conditions was 28.4% versus 25.9% in men [5,6]. Moreover, some chronic conditions, such as heart disease, are the leading causes of death in women, killing 1 in every 5 women [7]. Digital health tools have the potential to advance and monitor chronic conditions, and they represent a substantial opportunity to prevent and reduce the alarmingly high rates of chronic condition–related mortality and morbidity in women [8-10].

In their review, Adedinsewo and colleagues [11] succinctly outlined how artificial intelligence and digital health tools can be leveraged to improve the screening, monitoring, educating, and managing of chronic conditions among women during the life course. The authors note that digital tools can facilitate smoking cessation and cardiometabolic health in the preconception stage, particularly for those at risk of or with preexisting chronic conditions. Similarly, during pregnancy and among postpartum and menopausal women, digital health tools can be advantageous for the remote digital monitoring of blood pressure, telehealth consultation, or other smartphone-based educational interventions [11]. This finding suggests that integrating and using technology to engage and provide care for women living with chronic conditions is key to achieving the Healthy People 2030 objective: to increase the proportion of adults using health information technology, thereby promoting the health and well-being of women [12].

Despite the prevalence of chronic conditions across the life span, less is known about the adoption of digital health use in women. Previous studies have focused on digital health use during the perinatal period or on pregnancy-related chronic conditions [13-17]. Furthermore, other studies merely compare digital health use between men and women without investigating the distinct patterns of digital health use among women [18-20]. Since it is well established that there are gender and sex differences in the development and outcomes of chronic conditions and digital health use, it is imperative to understand how the adult women population with these conditions use digital health tools to manage their health [2,11]. Understanding women's digital health use for chronic condition management may encourage the responsiveness of public health programs and interventions to women's health needs to prevent and control these conditions among this population.

Objective

Therefore, the objective of this study was to investigate the 4-year trend of digital health use for health promotion among women with noncommunicable chronic conditions in the United States, drawing from nationally representative data. This study also aimed to examine the sociodemographic and health-related factors that influence digital health use for health promotion across subpopulations of women in the United States.

Methods

Data Source

Data for this study were obtained from the Health Information Trends Survey (HINTS) [21]. HINTS is a nationally representative sample of noninstitutionalized US adults aged ≥18 years and fielded by the National Cancer Institute. HINTS collects data about respondents' health communication, digital health use, and sociodemographic characteristics. This study used data from HINTS 5, Cycles 1 (2017), 2 (2018), 3 (2019), and 4 (2020) surveys. Cycles 1 and 2 surveys were stratified by postal address to sample residential addresses randomly, whereas Cycles 3 and 4, in addition to mailed surveys, introduced a web-based option, wherein respondents were randomly grouped into (1) a web-based and paper survey option without an additional bonus and (2) a web-only option with an additional bonus. Per HINTS, there was no statistically significant difference between the 2 groups; rather, providing the bonus and using the web-based option increased the sample representativeness by allowing groups, particularly young adults, who would have been otherwise underrepresented in the mail-only option. Detailed information about the HINTS survey and methodology is reported elsewhere [21]. The total number of surveyed respondents and response rates for each year were as follows: 3285 and 32.3% (Cycle 1); 3504 and 32.9% (Cycle 2); 5438 and 30.3% (Cycle 3); and 3865 and 36.7% (Cycle 4). Unless otherwise stated, all variables analyzed in this study were surveyed across the years.

Measures

Digital Health

Following the strategy of Shan and colleagues [18], we analyzed 6 binary (no/yes) outcome variables to measure digital health use (Multimedia Appendix 1). Although Shan and colleagues [18] measured 9 digital health use variables, we deviated by only measuring digital health use among respondents who responded in the affirmative that they owned either a tablet or smartphone, because our focus is on use versus ownership. As a result, our study included respondents who have used their tablet or smartphone to (1) track health-related goals, (2) make health decisions, and (3) discuss with their provider.

Respondents were also asked if, in the past 12 months, they had (4) used an electronic wearable device to track their health-related goals, (5) shared their health information with their health provider with an electronic monitoring device or smartphone, and (6) communicated with their provider via SMS text messaging. Of the 6 digital health measures, the question regarding electronic wearable device use was asked in 2019 and 2020, whereas that regarding the use of SMS text messaging for communicating with the provider was only asked from 2017 to 2019. See Multimedia Appendix 1 for detailed information on the variables.

Key Independent Variables

Chronic conditions were measured using 2 questions and were modeled following the approach of Greenberg and colleagues [19]. Respondents answered yes-or-no questions on whether their provider told them they had diabetes or high blood sugar, high blood pressure or hypertension, heart-related conditions, lung disease–related conditions, depression or anxiety, and arthritis or rheumatism. Participants were also asked if they had ever had cancer (no/yes; Multimedia Appendix 1). The number of chronic conditions was totaled and then categorized as 0, 1, and 2 or more chronic conditions. However, in contrast to Greenberg and colleagues [19], we analyzed arthritis or rheumatism separately for all models and did not include them in the totaled chronic conditions, because these were dropped in the 2019 and 2020 HINTS iterations.

Covariates

Control variables were included based on theoretical and empirical relevance [19,22]. Sociodemographic variables included in the analyses were age (18-34, 35-49, 50-64, or ≥65 years), marital status (not married vs married), income status (<US $20,000, US $20,000-34,999, US $35,000-49,999, US $50,000-74,999, or ≥US $75,000), race/ethnicity (non-Hispanic White, non-Hispanic Black, Hispanic, non-Hispanic Asian/others, or missing), and education (high school degree or below, some college degree, and college degree or above). Health-related variables included health insurance (no/yes), self-reported health status (fair/poor, good, or excellent), regular provider (no/yes), physical activity (<150 vs >150 minutes per week), and smoking status (never, former, or current).

Statistical Analysis

All analyses used the recommended analytical strategy by the HINTS analyst and applied 200 replicate jackknife survey weights (50 jackknife survey weights for each year) to account for variance estimation and generalizability.

Initial weighted descriptive statistics were analyzed for all respondents. Chi-square tests were used to compare the characteristics of the study population to chronic conditions. Second, we summarized the weighted temporal prevalence of chronic diseases and digital health use in graphs. Separate weighted logistic regression models were then conducted to test the unadjusted association of the study variables and each digital health use variable. Third, multivariate logistic regression models were created to explore the adjusted association between the digital health use variables and all covariates.

Additionally, we examined the interactions between age and the chronic condition categories, race/ethnicity and income, and race/ethnicity and education and presented the adjusted predicted probabilities in plots. Lastly, we conducted a similar unadjusted and adjusted multivariate logistic regression analysis to examine the relationship between digital health use and individual chronic conditions.

Multicollinearity was also examined using the variance inflation factor, and there was no collinearity among the independent variables. Hosmer-Lemeshow goodness-of-fit test was conducted; models with insignificant chi-square test output suggest a good fit. We provided the unadjusted odds ratio (OR), adjusted odds ratio (aOR), and corresponding 95% CI. A 2-sided significance level of $\alpha<.05$ for statistical significance was applied. Missing or unknown observations were dropped for all analyses except for race/ethnicity. Analyses were performed with Stata statistical software (version 17 SE; StataCorp).

Results

Population Characteristics

Of the 8573 women who participated in the pooled survey, the weighted prevalence and samples of those who answered "yes" to using a tablet or smartphone for various activities were as follows: achieve or track health-related goals, 50.3% (95% CI 48.4%-52.2%; 3279/7122); make a health decision, 43.6% (95% CI 41.9%-45.3%; 2998/7101); and discuss with their provider, 40% (95% CI 38.2%-41.8%; 2834/7099). In the preceding 12 months, 33% (95% CI 30.9%-35.2%; 1395/4826) of women reported using an electronic wearable device, 18.7% (95% CI 17.3%-20.2%; 1532/7653) shared health information using an electronic monitoring device, and 35.2% (95% CI 33.2%-37.3%; 2262/6349) sent or received an SMS text message with a health professional.

As seen in Table 1, higher shares of the women in this study had health insurance (8035/8460; 92.7%), were non-Hispanic White (4871/8573; 60.6%), were aged 50-64 years (2679/8449; 29.1%), earned above US $75,000 (2937/8528; 37.5%), had never smoked (5626/8499; 67.6%), and performed <150 minutes of moderate physical activity per week (6775/8419; 69.3%). Of the 8564 women who responded to the chronic condition variable, the sample and weighted prevalence of having 0, 1, and multiple chronic conditions were 2718 (37.4%), 2776 (33.4%), and 3070 (29.3%), respectively. However, the overall weighted pooled prevalence of having any chronic diseases between 2017 and 2020 was 52.2% (95% CI 0.50%-0.53%; 4756/8564; data not shown). Furthermore, those who reported having multiple chronic conditions were predominately in 2 age groups: 50-64 years (1033/3070; 10%) and >65 years (1470/3070; 10.3%). They were also predominately non-Hispanic White (1716/3070; 18.2%), had a regular provider (2463/3070; 23.7%), and performed <150 minutes of moderate physical activity (2286/3070; 21.8%). In addition, they mostly earned <US $20,000 (902/3070; 7.9%) and had a high school degree or below (1054/3070; 11.5%).

Table 1. Characteristics of the total population and by chronic conditions in the pooled sample.

Variable	Total, n (weighted %)	Number of chronic conditions[a] (n=8564), n (weighted %)			*P* value
		0	1	≥2	
Women	8573 (100)	2718 (37.4)	2776 (33.4)	3070 (29.3)	
Age (years; n=8449)					<.001
18-34	1171 (23.7)	649 (12.5)	385 (8.4)	137 (2.8)	
35-49	1724 (26.3)	801 (12.2)	538 (7.9)	385 (6.2)	
50-64	2679 (29.1)	761 (9.1)	885 (10.1)	1033 (10)	
>65	2875 (20.9)	475 (3.7)	926 (6.9)	1470 (10.3)	
Marital status (n=8488)					.001
Not married	4409 (46.4)	1165 (16.2)	1404 (15.4)	1835 (14.9)	
Married	4079 (53.6)	1522 (21.2)	1353 (18)	1204 (14.3)	
Income (US $; n=8528)					<.001
<20,000	1749 (18.6)	346 (4.9)	497 (5.6)	902 (7.9)	
20,000-34,999	1230 (12.8)	322 (4.4)	391 (3.82)	515 (4.6)	
35,000-49,999	1145 (14)	315 (3.9)	377 (5.5)	451 (4.5)	
50,000-74,999	1467 (17.2)	471 (6.2)	509 (6.1)	487 (4.9)	
>75,000	2937 (37.5)	1246 (17.8)	990 (12.4)	700 (7.3)	
Race/ethnicity (n=8573)					<.001
Asian/others	613 (7.0)	245 (3.5)	157 (2.3)	181 (1.3)	
Hispanic	1201 (15)	510 (7.5)	336 (4.2)	355 (3.2)	
Non-Hispanic African American or Black	1275 (11.8)	334 (3.7)	420 (3.9)	520 (4.2)	
Non-Hispanic White	4871 (60.6)	1510 (21.3)	1641 (21.1)	1716 (18.2)	
Missing	613 (5.6)	119 (1.4)	192 (1.9)	298 (2.3)	
Education (n=8498)					<.001
High school or below	2230 (29.8)	523 (8.9)	651 (9.4)	1054 (11.5)	
Some college degree	2471 (37.9)	701 (14)	783 (12.6)	987 (11.3)	
College degree or above	3797 (32.3)	1475 (14.5)	1318 (11.4)	1002 (6.5)	
Insurance (n=8460)					.08
No	425 (7.25)	185 (3.2)	134 (2.3)	106 (1.7)	
Yes	8035 (92.7)	2499 (34.2)	2609 (31)	3026 (29.2)	
Self-reported health status (n=8484)					<.001
Fair/poor	1428 (15.8)	141 (2.1)	331 (4.3)	956 (9.4)	
Good	3044 (35.3)	728 (10.5)	1068 (12.8)	1248 (12.1)	
Excellent	4012 (48.9)	1810 (24.7)	1354 (16.4)	846 (7.8)	
Regular provider (n=8437)					<.001
No	2399 (32)	1087 (16.3)	766 (10.2)	543 (5.5)	
Yes	6038 (68)	1592 (21)	1978 (23.2)	2463 (23.7)	
Physical activity (n=8375)					<.001
<150 minutes per week	5961 (69.3)	1754 (24.1)	1915 (23.4)	2286 (21.8)	
>150 minutes per week	2414 (30.7)	917 (13.5)	798 (9.9)	698 (7.3)	
Smoking status (n=8499)					<.001
Never	5626 (67.6)	2020 (28.6)	1865 (22.4)	1736 (16.6)	

Variable	Total, n (weighted %)	Number of chronic conditions[a] (n=8564), n (weighted %)			*P* value
		0	1	≥2	
Former	1906 (20.3)	463 (5.6)	583 (6.7)	860 (7.9)	
Current	967 (12.1)	214 (3.2)	309 (4.2)	444 (4.7)	
Diabetes (n=8419)					<.001
No	6775 (83.7)	2654 (37.2)	2518 (30.8)	1603 (15.7)	
Yes	1644 (16.3)	0 (0)	213 (2.63)	1431 (13.6)	
High blood pressure (n=8423)					<.001
No	4847 (66.3)	2634 (37)	1647 (23)	566 (6.3)	
Yes	3576 (33.7)	0 (0)	1090 (10.5)	2486 (23.2)	
Heart condition (n=8449)					<.001
No	7774 (93.2)	2648 (37)	2693 (32.8)	2433 (23.3)	
Yes	675 (6.79)	0 (0)	53 (67.5)	622 (6.1)	
Lung disease (n=8450)					<.001
No	7200 (86.1)	2647 (37)	2485 (29.9)	2068 (19.1)	
Yes	1250 (13.9)	0(0)	264 (3.6)	986 (10.3)	
Depression or anxiety (n=8436)					<.001
No	6129 (70.7)	2639 (36.9)	1924 (20.3)	1566 (13.4)	
Yes	2307 (29.3)	0 (0)	822 (13.3)	1485 (16.1)	
Cancer (n=8531)					<.001
No	7182 (89.8)	2709 (37.4)	2432 (30.3)	2042 (22.1)	
Yes	1349 (10.2)	0 (0)	334 (3.1)	1015 (7.1)	
Arthritis[b] (n=3644)					—[c]
No	2384 (73.4)	—	—	—	—
Yes	1260 (26.6)	—	—	—	—

[a]Total diabetes, high blood pressure, heart condition, lung disease, depression/anxiety, and cancer.

[b]2017-2018 and not totaled in the chronic condition category.

[c]Not available.

Prevalence of Chronic Conditions and Digital Health Use Across Years

Figure 1 shows the weighted prevalence of chronic conditions. Between 2017 and 2020, women without any chronic diseases declined from 37.8% (569/1784) to 34.4% (601/2047). In the same period, there was an increase in the proportion of women reporting 1 and multiple chronic conditions, from 33.2% (576/1784) to 35.3% (671/2047) and from 28.9% (639/1784) to 30.4% (775/2047), respectively. Table 2 reports the weighted prevalence of digital health use among the participants who answered "yes." Between 2017 and 2020, there was an increase in the proportion of women who used a tablet or smartphone to achieve a health-related goal (from 596/1479; 44.4% to 866/1760; 53.8%) and discuss with their provider (from 492/1469; 33.9% to 779/1758; 44.2%). In the past 12 months, the use of electronic wearable devices increased from 30% (739/2794) to 35.9% (656/2032) between 2019 and 2020. However, there was a sharp decline from 17.5% (302/1578) in 2017 to 15.4% (309/1854) in 2020 among those who shared their health information with a health professional.

Figure 1. Weighted prevalence (%) of chronic disease among adult US women from 2017-2020. Chronic disease includes total diabetes, high blood pressure, heart condition, lung disease, depression or anxiety, and cancer.

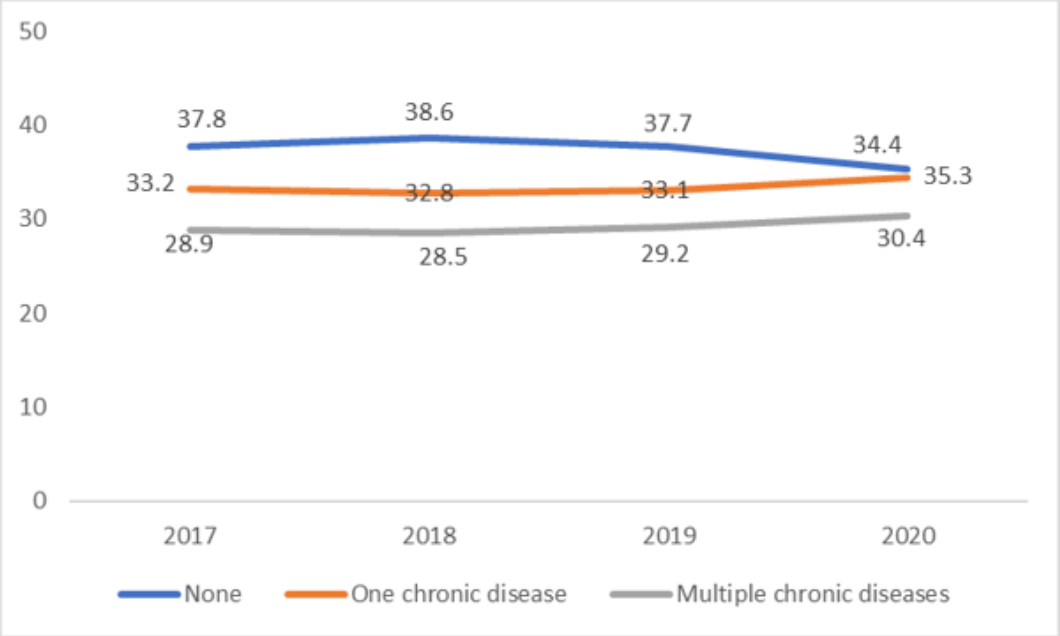


Table 2. Weighted prevalence of digital health use for health promotion among adult US women: 2017-2020.

Digital health use	Weighted prevalence (%; 95% CI)			
	2017	2018	2019	2020
Used tablet to achieve goals	44.4 (40.8-48.1)	50.6 (46.5-54.7)	52.3 (51.4-55.9)	53.8 (49.7-57.9)
Used tablet to make decision	37.7 (34.3-41.3)	45.6 (41.9-49.4)	46.4 (43.3-49.6)	44.5 (41.2-47.8)
Used wearable device[a,b]	—[c]	—	30 (27.7-32.4)	35.9 (32.4-39.6)
Used tablet to discuss with provider	33.9 (30.3-37.6)	39.4 (35.3-43.6)	42.2 (39.5-45.1)	44.2 (40.2-48.3)
Shared information with provider[b]	17.5 (14.6-20.8)	20.5 (17.8-23.4)	21.8 (18.9-25)	15.4 (12.6-18.7)
Communicated via texting with provider[b,d]	33.2 (30-36.5)	33 (29.2-37.1)	39.5 (36.1-43.1)	—

[a]Only surveyed from 2019-2020.

[b]In the past 12 months.

[c]Not available.

[d]Only surveyed from 2017-2019.

Unadjusted Odds of Digital Health Use Among Women With Chronic Conditions

In the unadjusted model (Multimedia Appendix 2), women with multiple chronic conditions were significantly less likely to use a tablet or smartphone to achieve a health goal (OR 0.74, 95% CI 0.63-0.88; *P*=.001) and use an electronic wearable device (OR 0.62, 95% CI 0.49-0.80; *P*<.001) than those with none. However, they had a higher likelihood of using their tablet or smartphone to discuss with their provider (OR 1.28, 95% CI 1.06-1.56; *P*=.01) and share health information (OR 2.02, 95% CI 1.63-2.51; *P*<.001) than women without any chronic conditions. There was a strong association between age and digital health use, and the odds were higher for women aged 18-34, 35-49, and 50-64 years (all with *P*<.001) than those aged ≥65 years. This relationship was not found when sharing their health information with a provider.

Adjusted Odds of Digital Health Use Among Women With Chronic Conditions

After adjusting for the covariates (Multimedia Appendix 3), women with multiple chronic conditions had higher odds of using their tablet or smartphone to achieve a health-related goal (aOR 1.43, 95% CI 1.16-1.77; *P*=.001) and discuss with their provider (aOR 1.55 95% CI 1.20-2.00; *P*=.001) than those without any chronic conditions. Correspondingly, in the past 12 months, the odds of using an electronic wearable device (aOR 1.40, 95% CI 1.00-1.96; *P*=.04), sharing health information (aOR 1.91, 95% CI 1.46-2.51; *P*<.001), and communicating via SMS text messaging with a provider (aOR 1.31, 95% CI 1.02-1.68; *P*=.03) were significantly higher than

women without a chronic condition. Similar to the unadjusted model, age remained a significant predictor of digital health use (P<.001). Our analysis also revealed that non-Hispanic Black women had significantly higher odds of digital health use than their non-Hispanic White counterparts (achieve a health related goal: P=.004; make a decision: P<.001; use a wearable device: P=.89; discuss with a provider: P<.001; share health information: P=.007; and communicate via SMS text messaging with a provider: P=.91).

Results from the subanalysis of the individual chronic condition and digital health use controlling for all covariates and the adjusted interaction models between age and chronic conditions, income and race/ethnicity, and education and race/ethnicity are shown in Multimedia Appendices 4-8.

Discussion

Principal Findings

This study used the latest data from the HINTS to examine the association between chronic conditions and the use of digital health tools for health promotion activities among adult US women. Our study revealed several key findings. First, we found an increasing trend of chronic condition prevalence between 2017 and 2020 from 33.2% to 35.3% and from 28.9% to 30.4% for 1 and multiple chronic conditions, respectively. We also found that slightly more than half (52.2%) of women in the United States live with a chronic disease. Following that same pattern, the overall digital health use among women has increased over time, yet interestingly, sharing information with providers has decreased. Women with 1 or multiple chronic diseases in this study had up to a 2-fold increase in using all of the digital health measures analyzed in this study compared to those without any chronic conditions. This result demonstrates that digital health technologies can provide a unique opportunity to combat chronic condition–related morbidity and mortality among women.

Overall, our results suggest that women with chronic conditions were more likely to report using digital health tools than those without these conditions. These findings are similar to previous studies showing increased digital health use, especially among those with chronic diseases [19,23-25]. A difference in digital health use was also noted based on the type of condition reported. Across the individual kinds of chronic diseases, all but cancer were strongly associated with various digital health activities. Although similar to previous studies, our results differ from 1 study showing a positive association between cancer diagnosis and digital health tools [23]. Although we may not fully understand this deviation, it is plausible that including all adult women in our model versus older women alone accounted for the difference [23]. Findings from this study suggest that although women generally accept digital health tools for numerous health activities, such as health information seeking, monitoring health conditions, patient-provider communication, or treatment, these tools are not adequately harnessed to address the growing trend of chronic diseases. In addition, since data from digital health tools can be linked to medical health records or other patient portals, our results underscore the importance of promoting its use for the continuity of medical care for chronic conditions.

Our study also explored how sociodemographic and health-related factors influence digital health use across subpopulations. Overall, younger women with higher education and income were more likely to report digital health use than older women (aged >65 years). This relationship persisted regardless of the presence of chronic conditions. Among racial/ethnic groups, non-Hispanic Black women were more likely to embrace digital health than their White counterparts. This finding should be interpreted with caution because the adjusted probability of being non-Hispanic Black, earning a higher income, and having a higher education was higher than other groups in our sample. We also found disparities in digital health use among those with more education (some college or a college degree) and incomes (>US $75,000) compared to women with lower education (high school or below) and income, respectively. These findings are similar to previous research that reports that women more likely to use digital health tools, mobile health apps, the internet, or electronic patient records are those who are younger and non-Hispanic Black and have higher education and income levels [18,19,26-28]. Age and educational differences reflect the ease, skills, and confidence in using complex digital health tools, leading to a more health-conscious and literate subpopulation that can manage their health [25,27]. This finding can guide interventions to increase the use of digital health technology to focus on digital health literacy.

Other health-related factors that increase the likelihood of digital health use were being insured, having a regular provider, being more physically active (>150 minutes/week), and being a former smoker. These findings are consistent with research on digital health use based on certain health-related factors [18,29,30]. For example, Shan and colleagues [18] found that having a regular provider was associated with at least one type of user activity. Access to a regular source of care most likely increases the likelihood of exposure to health resources and information to improve health literacy. Research supports that mobile health users are more likely to report intentions to improve diet, exercise, and lose weight, further adding to the need to increase women's access and use of digital health tools [18,27].

Overall, our study shows that despite the increase in digital health technology, certain subpopulations experience a digital disconnect leading to health inequities. For instance, older adults are an at-risk cohort with higher disease rates and more health care needs, as well as the fastest growing group, yet they are largely disconnected from the digital world [19,23]. Low-income minority groups experience barriers due to the limited availability and affordability of mobile services and internet limitations [24]. In this current age of technology, digital inclusion and literacy have been deemed "super social determinants of health" as they address all other health determinants [11,25]. For example, access to employment, housing, or medical services apps are sometimes exclusively web-based; therefore, the inability to access these apps due to literacy level or access to a smartphone or internet shapes behaviors and health outcomes [25,31]. Thus, our results are a

step in the right direction in advancing women's health in line with the Healthy People 2030 objective [12].

Limitations

We have several limitations in our study. HINTS is a cross-sectional survey of a nationally representative cohort of individuals; therefore, we cannot infer causality, and the directions of the associations in the study cannot be indicated. A second limitation is that there might be additional confounders that potentially influence the results of our analysis, such as geographic variations, individual motivation, digital literacy, privacy and security concerns, and health consciousness, that affect the use of digital health tools. Another limitation is that only noncommunicable chronic conditions were analyzed in this study. Ideally, we would have preferred to investigate whether digital health use varies by communicable (eg, COVID-19) and noncommunicable chronic disease status. Unfortunately, the HINTS data set did not directly ask questions related to the COVID-19 pandemic's impact on digital health use. Although we acknowledge that the impact of the pandemic on digital health uptake may have influenced our results, we agree that the effect would be minimal considering that we analyzed multiyear data from 2017 to 2020.

Moreover, COVID-19 only minimally affected the HINTS 5, Cycle 4 (2020) data collection (see the HINTS methodological report for more details) [32]. Nonetheless, this area should be of interest for future research considering that the pandemic exerted medical, economic, and social pressures on women. Lastly, recall bias and misrepresentation by respondents are likely, considering that the survey is self-administered. Despite these limitations, our study significantly adds to the literature and, to the best of our knowledge, is the first to comprehensively assess digital health use among women with chronic conditions.

Conclusions

As the prevalence of chronic conditions increases, especially multiple comorbidities, interventions that facilitate health promotion resulting in timely and better self-management are warranted. Despite these benefits, our study shows how women from certain subpopulations—older, low income, and uninsured—are more likely not to use digital health promotion activities. To mitigate these use disparities, Adedinsewo and colleagues [11] lay a clear blueprint on how artificial intelligence and digital tools can be harnessed across the life span to improve women's health. These findings have useful public health implications given that chronic conditions, including cardiovascular diseases, are the leading causes of death in the United States. Therefore, these findings highlight the opportunity for researchers, policy makers, and health systems, including insurers, to prescribe successfully validated digital health apps to increase digital health use and empower women to become actively engaged in their care.

Authors' Contributions

KVA contributed to conceptualization; wrote the original draft preparation; reviewed and edited the manuscript; contributed to data management, analysis, and interpretation; and approved of the submitted version. EW wrote the original draft preparation, reviewed and edited the manuscript, and approved of the submitted version. HKO wrote the original draft preparation, reviewed and edited the manuscript, and approved of the submitted version. TM reviewed and edited the manuscript and approved of the submitted version. SO contributed to data management, reviewed and edited the manuscript, and approved of the submitted version. WG reviewed and edited the manuscript, supervised the study, and approved of the submitted version.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Outcome and key independent measures.
[DOCX File , 32 KB - mhealth_v10i8e39520_app1.docx]

Multimedia Appendix 2
Unadjusted odds of digital health use for health promotion among US women: 2017-2020.
[DOCX File , 55 KB - mhealth_v10i8e39520_app2.docx]

Multimedia Appendix 3
Multivariate logistic regression models of digital health use for health promotion among US women: 2017-2020.
[DOCX File , 56 KB - mhealth_v10i8e39520_app3.docx]

Multimedia Appendix 4
Unadjusted logistic regression models of digital health use and the individual chronic conditions: 2017-2020.
[DOCX File , 55 KB - mhealth_v10i8e39520_app4.docx]

Multimedia Appendix 5
Multivariate logistic regression models of digital health use and the individual chronic conditions: 2017-2020.

[DOCX File , 54 KB - mhealth_v10i8e39520_app5.docx]

Multimedia Appendix 6
Adjusted interactions between age and chronic conditions.
[DOCX File , 253 KB - mhealth_v10i8e39520_app6.docx]

Multimedia Appendix 7
Adjusted interactions between income and race/ethnicity.
[DOCX File , 327 KB - mhealth_v10i8e39520_app7.docx]

Multimedia Appendix 8
Adjusted interactions between education and race/ethnicity.
[DOCX File , 268 KB - mhealth_v10i8e39520_app8.docx]

References

1. Brown HL, Warner JJ, Gianos E, Gulati M, Hill AJ, Hollier LM, American Heart Associationthe American College of ObstetriciansGynecologists. Promoting risk identification and reduction of cardiovascular disease in women through collaboration with obstetricians and gynecologists: a presidential advisory from the American Heart Association and the American College of Obstetricians and Gynecologists. Circulation 2018 Jun 12;137(24):e843-e852 [FREE Full text] [doi: 10.1161/CIR.0000000000000582] [Medline: 29748185]
2. O'Kelly AC, Honigberg MC. Sex differences in cardiovascular disease and unique pregnancy-associated risk factors in women. Curr Treat Options Cardio Med 2020 Oct 28;22(12):58. [doi: 10.1007/s11936-020-00860-8]
3. Cushman M, Shay CM, Howard VJ, Jiménez MC, Lewey J, McSweeney JC, American Heart Association. Ten-year differences in women's awareness related to coronary heart disease: results of the 2019 American Heart Association National Survey: a special report from the American Heart Association. Circulation 2021 Feb 16;143(7):e239-e248 [FREE Full text] [doi: 10.1161/CIR.0000000000000907] [Medline: 32954796]
4. Mosca L, Hammond G, Mochari-Greenberger H, Towfighi A, Albert MA, American Heart Association Cardiovascular Disease and Stroke in Women and Special Populations Committee of the Council on Clinical Cardiology, Council on Epidemiology and Prevention, Council on Cardiovascular Nursing, Council on High Blood Pressure Research, Council on Nutrition, Physical Activity and Metabolism. Fifteen-year trends in awareness of heart disease in women: results of a 2012 American Heart Association national survey. Circulation 2013 Mar 19;127(11):1254-63, e1 [FREE Full text] [doi: 10.1161/CIR.0b013e318287cf2f] [Medline: 23429926]
5. Boersma P, Black LI, Ward BW. Prevalence of multiple chronic conditions among US adults, 2018. Prev Chronic Dis 2020 Sep 17;17:E106 [FREE Full text] [doi: 10.5888/pcd17.200130] [Medline: 32945769]
6. Percent of U.S. adults 55 and over with chronic conditions. Centers for Disease Control and Prevention. 2015 Nov 06. URL: https://www.cdc.gov/nchs/health_policy/adult_chronic_conditions.htm [accessed 2022-05-05]
7. Women and heart disease. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/heartdisease/women.htm [accessed 2022-05-05]
8. Hamine S, Gerth-Guyette E, Faulx D, Green BB, Ginsburg AS. Impact of mHealth chronic disease management on treatment adherence and patient outcomes: a systematic review. J Med Internet Res 2015 Feb 24;17(2):e52 [FREE Full text] [doi: 10.2196/jmir.3951] [Medline: 25803266]
9. Chow PI, Showalter SL, Gerber M, Kennedy EM, Brenin D, Mohr DC, et al. Use of mental health apps by patients with breast cancer in the United States: pilot pre-post study. JMIR Cancer 2020 Apr 15;6(1):e16476 [FREE Full text] [doi: 10.2196/16476] [Medline: 32293570]
10. Bostrøm K, Børøsund E, Varsi C, Eide H, Flakk Nordang E, Schreurs KM, et al. Digital self-management in support of patients living with chronic pain: feasibility pilot study. JMIR Form Res 2020 Oct 23;4(10):e23893 [FREE Full text] [doi: 10.2196/23893] [Medline: 33094734]
11. Adedinsewo DA, Pollak AW, Phillips SD, Smith TL, Svatikova A, Hayes SN, et al. Cardiovascular disease screening in women: leveraging artificial intelligence and digital tools. Circ Res 2022 Feb 18;130(4):673-690. [doi: 10.1161/CIRCRESAHA.121.319876] [Medline: 35175849]
12. Health care access and quality. U.S. Department of Health and Human Services. URL: https://health.gov/healthypeople/objectives-and-data/browse-objectives/health-care-access-and-quality [accessed 2022-05-05]
13. Sharma AJ, Ford ND, Bulkley JE, Jenkins LM, Vesco KK, Williams AM. Use of the electronic health record to assess prevalence of anemia and iron deficiency in pregnancy. J Nutr 2021 Nov 02;151(11):3588-3595. [doi: 10.1093/jn/nxab254] [Medline: 34386820]
14. Carral F, Ayala MDC, Fernández JJ, González C, Piñero A, García G, et al. Web-based telemedicine system is useful for monitoring glucose control in pregnant women with diabetes. Diabetes Technol Ther 2015 May;17(5):349-354. [doi: 10.1089/dia.2014.0223] [Medline: 25629547]

15. O'Kelly AC, Michos ED, Shufelt CL, Vermunt JV, Minissian MB, Quesada O, et al. Pregnancy and reproductive risk factors for cardiovascular disease in women. Circ Res 2022 Feb 18;130(4):652-672. [doi: 10.1161/CIRCRESAHA.121.319895] [Medline: 35175837]
16. Ming WK, Mackillop LH, Farmer AJ, Loerup L, Bartlett K, Levy JC, et al. Telemedicine technologies for diabetes in pregnancy: a systematic review and meta-analysis. J Med Internet Res 2016 Nov 09;18(11):e290 [FREE Full text] [doi: 10.2196/jmir.6556] [Medline: 27829574]
17. Rivera-Romero O, Olmo A, Muñoz R, Stiefel P, Miranda ML, Beltrán LM. Mobile health solutions for hypertensive disorders in pregnancy: scoping literature review. JMIR mHealth uHealth 2018 May 30;6(5):e130 [FREE Full text] [doi: 10.2196/mhealth.9671] [Medline: 29848473]
18. Shan R, Ding J, Plante TB, Martin SS. Mobile health access and use among individuals with or at risk for cardiovascular disease: 2018 Health Information National Trends Survey (HINTS). J Am Heart Assoc 2019 Dec 17;8(24):e014390 [FREE Full text] [doi: 10.1161/JAHA.119.014390] [Medline: 31818220]
19. Greenberg AJ, Falisi AL, Finney Rutten LJ, Chou WS, Patel V, Moser RP, et al. Access to electronic personal health records among patients with multiple chronic conditions: a secondary data analysis. J Med Internet Res 2017 Jun 02;19(6):e188 [FREE Full text] [doi: 10.2196/jmir.7417] [Medline: 28576755]
20. Reiners F, Sturm J, Bouw LJW, Wouters EJM. Sociodemographic factors influencing the use of eHealth in people with chronic diseases. Int J Environ Res Public Health 2019 Feb 21;16(4):645 [FREE Full text] [doi: 10.3390/ijerph16040645] [Medline: 30795623]
21. HINTS data. National Cancer Institute Health Information National Trends Survey. URL: https://hints.cancer.gov/data/default.aspx [accessed 2022-05-05]
22. Sherman LD, Patterson MS, Tomar A, Wigfall LT. Use of digital health information for health information seeking among men living with chronic disease: data from the Health Information National Trends Survey. Am J Mens Health 2020;14(1):1557988320901377 [FREE Full text] [doi: 10.1177/1557988320901377] [Medline: 31973642]
23. Sedrak MS, Soto-Perez-De-Celis E, Nelson RA, Liu J, Waring ME, Lane DS, et al. Online health information-seeking among older women with chronic illness: analysis of the Women's Health Initiative. J Med Internet Res 2020 Apr 09;22(4):e15906 [FREE Full text] [doi: 10.2196/15906] [Medline: 32271152]
24. Senteio C, Murdock PJ. The efficacy of health information technology in supporting health equity for Black and Hispanic patients with chronic diseases: systematic review. J Med Internet Res 2022 Apr 04;24(4):e22124 [FREE Full text] [doi: 10.2196/22124] [Medline: 35377331]
25. Sieck CJ, Sheon A, Ancker JS, Castek J, Callahan B, Siefer A. Digital inclusion as a social determinant of health. NPJ Digit Med 2021 Mar 17;4(1):52 [FREE Full text] [doi: 10.1038/s41746-021-00413-8] [Medline: 33731887]
26. Bhuyan SS, Lu N, Chandak A, Kim H, Wyant D, Bhatt J, et al. Use of mobile health applications for health-seeking behavior among US adults. J Med Syst 2016 Jun;40(6):153. [doi: 10.1007/s10916-016-0492-7] [Medline: 27147516]
27. Mahmood A, Kedia S, Wyant DK, Ahn S, Bhuyan SS. Use of mobile health applications for health-promoting behavior among individuals with chronic medical conditions. Digit Health 2019;5:2055207619882181 [FREE Full text] [doi: 10.1177/2055207619882181] [Medline: 31656632]
28. Wang P, Abdin E, Shafie S, Chong SA, Vaingankar JA, Subramaniam M. Estimation of prevalence of osteoporosis using OSTA and its correlation with sociodemographic factors, disability and comorbidities. Int J Environ Res Public Health 2019 Jul 02;16(13):2338 [FREE Full text] [doi: 10.3390/ijerph16132338] [Medline: 31269706]
29. Adler G, Kaufman G, Simon-Tuval T. Healthcare utilization of breast cancer patients following telephone-based consultations of oncology nurse navigator via telemedical care. PLoS One 2019;14(5):e0216365 [FREE Full text] [doi: 10.1371/journal.pone.0216365] [Medline: 31048852]
30. Onyeaka H, Firth J, Kessler RC, Lovell K, Torous J. Use of smartphones, mobile apps and wearables for health promotion by people with anxiety or depression: an analysis of a nationally representative survey data. Psychiatry Res 2021 Oct;304:114120. [doi: 10.1016/j.psychres.2021.114120] [Medline: 34303946]
31. van Kessel R, Wong BLH, Clemens T, Brand H. Digital health literacy as a super determinant of health: more than simply the sum of its parts. Internet Interv 2022 Mar;27:100500 [FREE Full text] [doi: 10.1016/j.invent.2022.100500] [Medline: 35242586]
32. Westat. Health Information National Trends Survey 5 (HINTS 5): Cycle 4 Methodology Report. National Cancer Institute Health Information National Trends Survey. 2020 Dec. URL: https://hints.cancer.gov/docs/methodologyreports/HINTS5_Cycle4_MethodologyReport.pdf [accessed 2022-06-02]

Abbreviations

aOR: adjusted odds ratio
HINTS: Health Information Trends Survey
OR: odds ratio

Edited by L Buis; submitted 12.05.22; peer-reviewed by Y Xie, C Baxter; comments to author 28.06.22; revised version received 16.07.22; accepted 25.07.22; published 19.08.22.

Please cite as:
Ajayi KV, Wachira E, Onyeaka HK, Montour T, Olowolaju S, Garney W
The Use of Digital Health Tools for Health Promotion Among Women With and Without Chronic Diseases: Insights From the 2017-2020 Health Information National Trends Survey
JMIR Mhealth Uhealth 2022;10(8):e39520
URL: https://mhealth.jmir.org/2022/8/e39520
doi:10.2196/39520
PMID:35984680

Original Paper

Health Care Workers' Need for Headspace: Findings From a Multisite Definitive Randomized Controlled Trial of an Unguided Digital Mindfulness-Based Self-help App to Reduce Healthcare Worker Stress

Heather Taylor[1], BA, PhD; Kate Cavanagh[1], BA, DPhil, DClinPsychol; Andy P Field[1], BSc, DPhil; Clara Strauss[1,2], BA, DPhil, DClinPsy

[1]School of Psychology, University of Sussex, Brighton, United Kingdom

[2]Sussex Partnership NHS Foundation Trust, Hove, United Kingdom

Corresponding Author:
Clara Strauss, BA, DPhil, DClinPsy
School of Psychology
University of Sussex
Falmer
Brighton, BN1 9QH
United Kingdom
Phone: 44 01273 876638
Email: c.y.strauss@sussex.ac.uk

Abstract

Background: Health care workers experience high stress. Accessible, affordable, and effective approaches to reducing stress are lacking. In-person mindfulness-based interventions can reduce health care worker stress but are not widely available or accessible to busy health care workers. Unguided, digital, mindfulness-based self-help (MBSH) interventions show promise and can be flexibly engaged with. However, their effectiveness in reducing health care worker stress has not yet been explored in a definitive trial.

Objective: This study aimed to investigate the effectiveness of an unguided digital MBSH app (Headspace) in reducing health care worker stress.

Methods: This was a definitive superiority randomized controlled trial with 2182 National Health Service staff in England recruited on the web and allocated in a 1:1 ratio to fully automated Headspace (n=1095, 50.18%) or active control (Moodzone; n=1087, 49.82%) for 4.5 months. Outcomes were subscales of the Depression, Anxiety, and Stress (primary outcome) Scale short form; Short Warwick Edinburgh Mental Well-being Scale; Maslach Burnout Inventory; 15-item Five-Facet Mindfulness Questionnaire minus Observe items; Self-Compassion Scale–Short Form; Compassionate Love Scale; Penn State Worry Questionnaire; Brooding subscale of the Ruminative Response Scale; and sickness absence.

Results: Intention-to-treat analyses found that Headspace led to greater reductions in stress over time than Moodzone (b=–0.31, 95% CI –0.47 to –0.14; P<.001), with small effects. Small effects of Headspace versus Moodzone were found for depression (b=–0.24, 95% CI –0.40 to –0.08; P=.003), anxiety (b=–0.19, 95% CI –0.32 to –0.06; P=.004), well-being (b=0.14, 95% CI 0.05-0.23; P=.002), mindfulness (b=0.22, 95% CI 0.09-0.34; P=.001), self-compassion (b=0.48, 95% CI 0.33-0.64; P<.001), compassion for others (b=0.02, 95% CI 0.00-0.04; P=.04), and worry (b=–0.30, 95% CI –0.51 to –0.09; P=.005) but not for burnout (b=–0.19, –0.04, and 0.13, all 95% CIs >0; P=.65, .67, and .35), ruminative brooding (b=–0.06, 95% CI –0.12 to 0.00; P=.06), or sickness absence (γ=0.09, 95% CI –0.18 to 0.34). Per-protocol effects of Headspace (454/1095, 41.46%) versus Moodzone (283/1087, 26.03%) over time were found for stress, self-compassion, and compassion for others but not for the other outcomes. Engagement (practice days per week) and improvements in self-compassion during the initial 1.5-month intervention period mediated pre- to postintervention improvements in stress. Improvements in mindfulness, rumination, and worry did not mediate pre- to postintervention improvements in stress. No serious adverse events were reported.

Conclusions: An unguided digital MBSH intervention (Headspace) can reduce health care workers' stress. Effect sizes were small but could have population-level benefits. Unguided digital MBSH interventions can be part of the solution to reducing

health care worker stress alongside potentially costlier but potentially more effective in-person mindfulness-based interventions, nonmindfulness courses, and organizational-level interventions.

Trial Registration: International Standard Randomised Controlled Trial Number ISRCTN15424185; https://tinyurl.com/rv9en5kc

(JMIR Mhealth Uhealth 2022;10(8):e31744) doi:10.2196/31744

KEYWORDS

self-help; mindfulness; randomized control trial; health care worker; National Health Service; NHS; doctors; nurses; stress; mental health; burnout; mobile phone

Introduction

Background

Even before the COVID-19 pandemic, findings from meta-analyses demonstrated a high prevalence of stress in health care workers worldwide [1-3]. Stress is a vulnerability factor for work-related burnout [4], anxiety, and depression [5], all of which are disproportionately prevalent among health care workers [6-8], and stress also increases the risk of several long-term physical health conditions [9-11]. In the National Health Service (NHS) in England, which employs >1.3 million health care staff [12], 46.8% of staff reported feeling unwell because of work-related stress [12], a figure that has steadily risen since 2016. Almost one-quarter of the days lost to staff sickness in the NHS are because of stress, anxiety, depression, or other mental health problems [13], and similar concerns have been noted in health care systems worldwide [14]. Moreover, stress among health care workers can compromise patient outcomes and safety [15]. The COVID-19 pandemic has further exacerbated stress and distress for health care workers [16,17]; therefore, there is an urgent need to find effective, accessible, and affordable ways of reducing health care workers' stress.

Mindfulness involves intentionally bringing curiosity and nonjudgmental awareness to present-moment experiences such as thoughts, feelings, and physical sensations as they arise [18,19]. Mindfulness-based interventions (MBIs) typically involve teaching mindfulness in in-person group settings through 8-week courses such as mindfulness-based cognitive therapy (MBCT) [20] and mindfulness-based stress reduction (MBSR) [21], with mindfulness practice and teacher-led discussion of practice being core intervention ingredients. There is substantial evidence from meta-analyses of randomized controlled trials (RCTs) that MBCT reduces the risk of relapse in people with a history of recurrent depression [22] and that MBIs improve symptoms of a range of mental health problems [23]. The degree of engagement in mindfulness practice during MBIs is associated with treatment outcomes [24], and MBI mechanisms of action include mindfulness, rumination, worry, and self-compassion [25].

The benefits of MBIs extend beyond clinical populations, with RCTs demonstrating beneficial effects on stress in nonclinical populations [26], including working adults [27] and, specifically, health care workers [28-30]. However, there are several barriers to health care workers attending in-person MBIs, including the lack of availability [31]; high workplace demands [32,33] that make it difficult for health care workers to find the time to attend; and stigma-related concerns regarding negative social judgments and disclosure and confidentiality, which are more common among health care workers than among those working in other settings [34].

Fortunately, mindfulness-based self-help (MBSH) has the potential to increase opportunities for engagement with MBIs through a plethora of MBSH books, web-based courses, and available smartphone apps. In addition, meta-analyses of RCTs of MBSH have indicated promising effects on stress and mental health outcomes across a range of populations [35,36]. Digital MBSH using smartphone apps has the potential to be particularly accessible as it does not rely on the user having a computer or book on hand to engage with the intervention when needed. Headspace [37] is a smartphone app with >70 million users to date worldwide [38]. There is emerging empirical literature exploring the effectiveness of MBSH apps, including Headspace [39]. Preliminary findings show potential benefits in nonclinical samples, including health care workers; however, the study sample sizes were too small to draw definitive conclusions regarding this working population. Given the early stage of research in this area and studies with small sample sizes, the potential of unguided digital MBSH as a health care–wide solution to reduce health care worker stress is yet to be explored in an adequately powered trial. Although MBSH can effectively reduce stress in a range of nonclinical populations, it is possible that the particularly high demands of working in health care [32,33] will mean that when offered at scale, health care staff may struggle to engage with the intervention, leading to disappointing outcomes. The learnings available from a definitive trial of unguided digital MBSH are particularly important in the current context of rising health care worker stress during the COVID-19 pandemic.

Objectives

This study sought to overcome some of the methodological limitations of previous related studies and extend our understanding of the potential effects of unguided MBSH among health care workers. The aim of this large multisite RCT was to explore the effectiveness of unguided digital MBSH in comparison with an active control condition (it should be noted that comparisons with active controls are lacking in RCTs of MBIs [29]) for health care workers in targeting stress (primary outcome), mental health outcomes (depression, anxiety, and well-being), work-related outcomes (work-related burnout, sickness absence, and compassion for others), and proposed mechanisms of action (intervention engagement, rumination, worry, mindfulness, and self-compassion). To explore its potential as a health care–wide intervention to reduce health care worker stress, the trial recruited across the full range of NHS organization types (general practitioner or primary care,

hospital trusts, community trusts, mental health and/or learning disability trusts, and ambulance trusts), across geographically and sociodemographically diverse regions of England, and across a range of NHS job roles (medical, nursing, allied health professions, and psychological and wider health care support roles). The primary hypothesis was that participants allocated to unguided digital MBSH will show greater reductions in stress from the baseline to postintervention time points (4.5 months following randomization) in comparison with participants in the active control trial arm. The secondary hypotheses were that unguided digital MBSH will be more effective than active control in improving mental health outcomes, work-related outcomes, and potential mechanisms of action from baseline to after the initial intervention period (1.5 months after randomization) and from the baseline to postintervention time points. Analyses examining whether intervention engagement and improvements in mindfulness, self-compassion, worry, and rumination mediated the effects of the intervention on improvements in stress were planned to ascertain intervention-specific mechanisms of action.

Methods

Trial Design and Ethics Approval

This study was a 2-arm superiority definitive RCT, with a 1:1 allocation and no stratification, comparing unguided digital MBSH (Headspace [37]) with an active control group (the NHS digital platform for work-related stress, Moodzone [40]). Assessments were performed at 3 time points: baseline (time point 1 [T1]), after the initial intervention period (time point 2 [T2]; 1.5 months after randomization), and at the postintervention time point (time point 3 [T3]; 4.5 months after randomization).

Ethics approval (reference ER/HT207/8) was provided by the University of Sussex, and study approval was granted by the Health Research Authority (reference 16/HRA/5525). The study was prospectively registered on the International Standard Randomised Controlled Trial Number register (reference number: 15424185) [41].

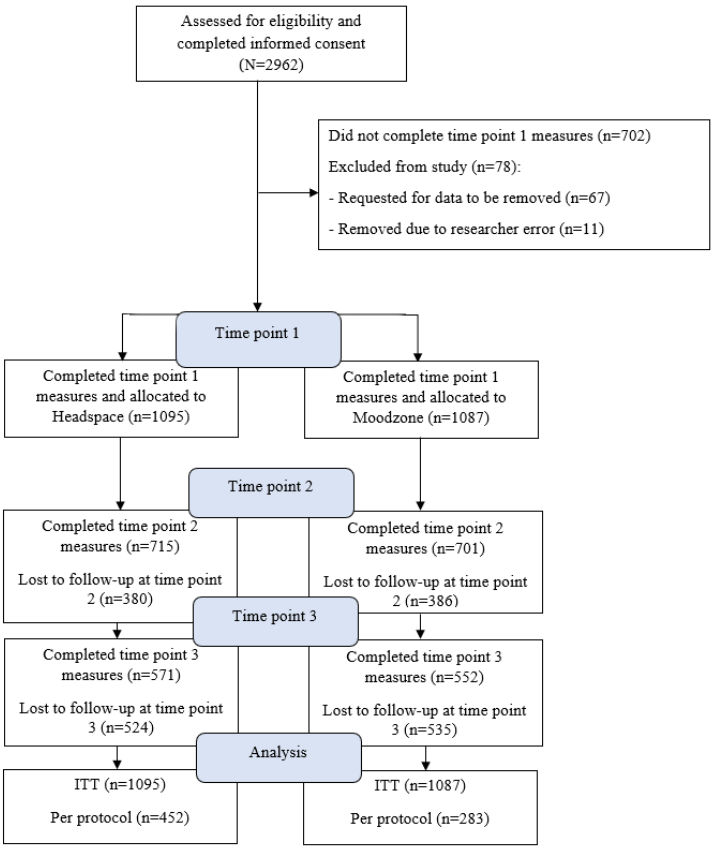
Participants and Recruitment

Participants had to (1) be employed within an NHS Trust or general practitioner practice in England, (2) be working in roles that involved direct contact with patients for a minimum of 1 day per week, (3) be currently in work (ie, not on long-term sickness absence), (4) be willing to refrain from engaging in other psychological interventions during the course of the study, (5) have regular personal access to an Apple, Android smartphone, or tablet or a computer with internet access, (6) be aged ≥18 years, and (7) have sufficient English language skills to read and understand the intervention materials. There were no additional exclusion criteria. Recruitment took place between February 21, 2017, and September 18, 2018.

Sample size calculations were conducted using G*Power [42], which indicated that 527 complete cases per study arm (1054 total) would be needed to detect a small between-group difference of Cohen d=0.20 (P=.05; 90% power; 2-tailed) on the primary outcome (stress at T3), with this estimate based on a meta-analysis of MBSH on stress outcomes [36]. A conservative estimate of a 50% study dropout rate was assumed [35], giving a total required sample size of 2108 (n=1054 per arm).

A total of 2182 participants were enrolled in the study (completed baseline measures and were randomized); 1095 (50.18%) were randomized into the Headspace arm, and 1087 (49.82%) were randomized into the Moodzone arm. The participant flow is shown in the CONSORT (Consolidated Standards of Reporting Trials) diagram (Figure 1), and further participant details are reported in the *Results* section.

Figure 1. CONSORT (Consolidated Standards of Reporting Trials) diagram showing participant flow. ITT: intention-to-treat.



Interventions

Headspace

The Headspace MBSH digital program [37] offers a range of brief mindfulness-based practices alongside psychoeducational materials. The Headspace MBSH digital program can be accessed via a website [43] or an app available on the Apple app store or Android Play store. Headspace offers a range of mindfulness-based practices and psychoeducational animations, including an introductory series that comprises daily sessions designed to teach foundational mindfulness principles and practices, as well as packs designed for more specific emotional difficulties (eg, stress and anxiety) and brief *SOS* mindfulness practices designed to be used in times of acute stress. Headspace also offers guidance on informal mindfulness practices that can be undertaken while performing everyday activities, such as running and cycling, and there is written information, including research evidence, related to mindfulness and a frequently asked questions section. At the time of the study, mindfulness practices were verbally guided by Andy Puddicombe, a founder of Headspace with many years of experience in mindfulness practice. For the introductory sessions, users were verbally guided to bring nonjudgmental awareness to the body, breath, thoughts, and feelings, with later sessions also inviting users to bring awareness to difficulties arising during practice (eg, boredom and restlessness) and behavioral choices. At the time of recruitment, users were invited to start the Headspace program by completing the *Take Ten* introductory pack, which involved undertaking guided 10-minute mindfulness practices daily for 10 consecutive days. Upon completion of the *Take Ten*, participants were provided with unlimited access to the full range of Headspace content. Although participants were free to choose the content they engaged with, they were invited to perform at least one 10-minute mindfulness practice daily for the duration of the study. Although practices range in length from 3 to 20 minutes, users can select the duration of most sessions. As Headspace is a *live* product, the program structure was nonstatic, and participants were able to access newly released and changing content as it became available.

Moodzone

The NHS Moodzone psychoeducational digital platform [40] was used as an active control. At the time of recruitment, the website offered a range of evidence-based psychosocial recommendations, advice, and guidance on how to manage work-related stress and mental health difficulties effectively. The initial web page was divided into the following sections: “What causes work stress?” “How to manage work stress,” “Learn to speak out,” “Spot the signs of work stress,” and “Who else can help with work stress?”; each provided information and recommendations or guidance relevant to the respective questions. Moodzone also included information, videos, audio tracks, podcasts, and links to other related resources. Participants were invited to engage with the Moodzone website for 10 minutes per day for the duration of the study. It should be noted that although very similar content is still available [44], the Moodzone website used in this study is no longer active. As with Headspace, a *live* nonstatic version of Moodzone was used in the study, meaning that participants could access new and changing content as it became available. Before this study,

adequately powered Moodzone trials were not undertaken. However, related evidence from a meta-analysis of RCTs identified a significantly small effect (Cohen *d*=0.20; *P*=.04) of passive psychoeducational interventions compared with control conditions in reducing depression and psychological distress at the postintervention time point [45].

Measures

Full details of the measures are shown in Multimedia Appendix 1 [4,46-55].

Participants completed the measures described in Textbox 1 at T1, T2, and T3 unless stated otherwise.

Textbox 1. Participant measures.

Participant measures

- Short version of the 21-item Depression, Anxiety, and Stress Scale [46]; the Stress subscale was the primary outcome, with time point 3 (T3) being the primary end point
- Short Warwick Edinburgh Mental Well-being Scale [47]
- Maslach Burnout Inventory [4]
- 15-item version (minus "observe") of the Five Facets of Mindfulness Questionnaire [48]
- Self-Compassion Scale–Short Form [49]
- Compassionate Love Scale [50]
- Penn State Worry Questionnaire [51]
- Brooding subscale of the Ruminative Response Scale [52]
- Sickness absence measured at time point 1 [T1] and T3 was assessed using 1 item that asked participants to report how many days they had been absent from work because of sickness during the past 3 months
- Demographic information assessed at T1 included participants' age; gender; marital status; number of children aged <18 years; number of children aged ≥18 years; National Health Service job role; trust and team; number of hours worked per week in the National Health Service job role; highest level of education; individual and household annual incomes; ethnicity; and perceived relative socioeconomic status, with response options from 1 (lowest) to 10 (highest) perceived socioeconomic status [53]
- Intervention expectancy at T1 (Credibility and Expectancy Questionnaire [54])
- Self-reported intervention engagement at time point 2 [T2] and T3:
- Formal engagement: self-reported average number of days per week spent following guided mindfulness meditation on Headspace or following a recommended stress management or well-being strategy on the Moodzone web page
- Informal engagement: self-reported average number of days per week participants brought mindfulness to daily activities or recommended stress management and well-being strategies from Moodzone into their daily lives; at T2, these questions were asked in relation to the past month, and at T3, they were asked in relation to the past 3 months
- Intervention evaluations at T2 and T3: participants asked how likely they were to recommend the intervention to friends and family, how much they really felt that their allocated intervention had helped their well-being, and how likely they were to continue practicing mindfulness (Headspace participants) or stress management and well-being strategies (Moodzone participants) over the following 6 months
- Hypothesis guess at T3: participants asked to state what they thought the purpose of the study was
- Intervention deviations at T3: participants asked to indicate whether they had engaged in the alternative study intervention during the course of the study
- Prior mindfulness experience at T3: participants asked to indicate their experiences of mindfulness before the study, including mindfulness-based cognitive therapy, stress reduction, mindfulness-based self-help, and Headspace, and how often they practiced mindfulness
- Serious adverse events were recorded in accordance with the National Institute for Health Research Good Clinical Practice guidelines [55]
- Participants were also asked to indicate the extent to which they agreed or disagreed that they had experienced "lasting bad effects" from using their allocated intervention (based on Crawford et al [56]); if the participants agreed or strongly agreed, they were asked to provide further details

Procedure

NHS staff were recruited via posters and leaflets in NHS settings, invitation emails sent through NHS organizations, and study advertisements on staff web pages or newsletters. Potential participants were directed to the study website hosted by Qualtrics XM [57], where they could read the participant information and confirm their eligibility and informed consent (Multimedia Appendix 2). After consenting, the participants were emailed a weblink along with a unique ID code and asked to self-complete the T1 measures on Qualtrics. Participants completed T1 measures, which were allocated automatically to Headspace or Moodzone using a 1:1 block randomization with a block size of 4 by Qualtrics. To ensure allocation concealment, the members of the research team responsible for collecting data and communicating with participants were blinded to the block size. Participants were informed of their random allocation and subsequently asked to indicate their views on the credibility and expectations of their assigned intervention.

Following the completion of the T1 assessment, participants were emailed information on how to access their allocated intervention. Intervention participants were given 12 months of free access to Headspace, and Moodzone was available free of charge. Allowing 5 days for participants to receive this information or download their intervention, participants were invited to engage with their allocated intervention for 10 minutes per day, every day during the initial 30-day study period. At 35 days after randomization, participants were emailed a link to complete the T2 assessments on Qualtrics and invited to continue engaging with their allocated intervention for 10 minutes per day during the remaining 90-day study period. On average, T2 was completed at 1.5 months (SD 0.57) after randomization. At 125 days after randomization, participants were emailed a link to complete the T3 assessment on Qualtrics, with T3 completed at an average of 4.5 months (SD 0.53) from randomization. At this point, the participants who completed the study were given access to the alternative intervention.

Participants who did not complete assessments within 1 week of them being sent were reminded to do so via email. One reminder email was sent for completion of the T1 assessments, and a maximum of 4 reminder emails at weekly intervals were sent for T2 and T3 assessments. The research team was available to answer technical questions or queries via email. No further support was provided.

To improve trial quality and blind participants to the study condition and direction of study hypotheses, advertisements about the study simply referred to both conditions as "online interventions to reduce NHS staff stress," and details of the alternative or nonallocated intervention were not communicated to participants until T3 assessments (after outcome and engagement measures had been taken). As all assessments were completed on the web without researchers present, the potential for researcher bias to influence assessment outcomes was minimized. All but the mediation analysis was conducted blind to the study arm.

Participants were given the option to enter a prize draw to win 1 of 5 gift vouchers for £50 (US $60).

Data Analysis Plan

Descriptive statistics are reported by trial arm and time as means and SDs (for continuous data), medians and IQRs (for ordinal data), and counts and percentages (for categorical data). Data analysis was conducted using SPSS (version 25; IBM Corp) [58] and R (version 4.0.2; R Foundation for Statistical Computing) [59] and the following packages: *emmeans* [60], *lme4* [61], *mice* [62], *papaja* [63], and *tidyverse* [64].

Handling Missing Data

A minimal number of items were missing at the item level, and missing values for missing items were imputed (using a single imputation) using predictive mean matching in *mice* [65]. At the scale level, multiple imputation was used to handle missing values. Further details are provided in Multimedia Appendix 3 [60-62].

Model Selection

As participants were nested within job roles (level 3), there are good reasons for model variations in intervention effects between job roles [66]. There is participant-level randomization to intervention arms in such a model, and job roles act as a crossed effect. We can think of time (*i*) as being nested within participants (*j*), which is nested within job roles (*k*); however, the effect of the treatment arm occurs at level 2 (the participant level), not level 3 (the job role level), of the hierarchy. This situation is described by the model given in Textbox 2.

This saturated model includes random effects for time, trial arm, and their interaction at level 3. However, this model resulted in convergence problems that yielded erratic estimates of random effects involving the trial arm in the raw sample and nearly all imputed samples. On the basis of this preanalysis, a simpler model seemed more appropriate, in which only time was treated as a random effect and only at level 2. However, to model level 3 variability in outcomes, a random intercept (at level 3) was included. This simpler model converged in all the imputed samples. The resulting model is described in Textbox 3 (notice that at level 3, a total of 2 random effects have been knocked out).

To sum up, the hypotheses were tested using a growth model fit as a general linear mixed model, with observations (level 1) nested within participants (level 2) nested within job roles (level 3). Time (time from baseline at which responses were recorded) and trial arm were predictors. The effect of the intervention was quantified and tested with the interaction between time and trial arm, which shows the degree to which the change in the outcome over time is different between the 2 trial arms. Between-group effects were reported separately at T2 and T3 in the event of significant (P<.05) trial arm × time interactions. The primary analysis was conducted on the intention-to-treat (ITT) sample with multiple imputed data sets. Secondary analysis was conducted on the per-protocol sample (formal engagement T1-T2 on at least 3 days per week [67]) with the multiple imputed data sets.

Details of the plan for reliable change analysis, mediation analysis, and randomization check can be found in Multimedia Appendix 4 [13,18,25,29,30,36,56,67-72].

Textbox 2. The saturated model showing the data structure.

Level 1

- Depression, Anxiety, and Stress Scale–Stress$_{ijk}$= π_{0jk}+ π_{1jk}Time$_{ijk}$+ ε_{ijk}

Level 2

- π_{0jk}= γ_{00k}+ γ_{01k}Trial arm$_{jk}$+ ζ_{0jk}
- π_{1jk}= γ_{10k}+γ_{11k}Trial arm$_{jk}$+ ζ_{1jk}

Level 3

- γ_{00k}= δ_{000}+υ_{0k}
- γ_{10k}= δ_{100}+υ_{1k}
- γ_{01k}= δ_{010}+υ_{2k}
- γ_{11k}= δ_{110}+υ_{3k}

Textbox 3. The fitted model.

Level 1

- Depression, Anxiety, and Stress Scale–Stress$_{ijk}$= π_{0jk}+ π_{1j}Time$_{ijk}$+ ε_{ijk}

Level 2

- π_{0jk}= γ_{00k}+ γ_{01}Trial arm$_{jk}$+ ζ_{0jk}
- π_{1j}=γ_{11}Trial arm$_{jk}$+ ζ_{1jk}

Level 3

- γ_{00k}= δ_{000} + υ_{0k}

Results

Overview

Table 1 presents the demographic characteristics of the participants by study arm, and Table 2 presents descriptive statistics on all outcome measures at all time points by study arm. Table 1 shows that participants represented a broad range of NHS Trust types and health care professions. As would be expected of a health care workforce, most participants were educated to at least an undergraduate degree level and were earning, on average, the median UK annual salary; most participants were working full-time. Participants covered the full working age spectrum, although they were disproportionately White and female. For the randomization check, all Bayes factors were very close to 0, suggesting very strong evidence for the null hypothesis: randomization was successful in balancing demographic and baseline measurements across the 2 trial arms (Multimedia Appendix 5). There were also no differences in dropout rates between the trial arms. A formal analysis using a multilevel generalized linear model, with a random intercept, predicting dropout (1=in the study and 0=dropped out) from the trial arm, study wave (as a categorical variable), and their interaction showed no significant effects.

Table 1. Demographic characteristics of participants (N=2182).

Characteristics	Moodzone	Headspace
Highest educational achieved, n (%)		
GCSE[a] or NVQ 2[b] or below (equivalent to not completing high school)	62 (2.84)	69 (3.16)
A-level or equivalent (equivalent to completing high school)	132 (6.05)	124 (5.68)
Undergraduate degree	430 (19.71)	474 (21.72)
Postgraduate degree	462 (21.17)	429 (19.66)
Other	2 (0.09)	2 (0.09)
Ethnicity, n (%)		
Black	13 (0.6)	12 (0.55)
White	998 (45.74)	1021 (46.79)
Asian	50 (2.29)	37 (1.7)
Mixed or multiple	21 (0.96)	19 (0.87)
Other	2 (0.09)	4 (0.18)
Gender, n (%)		
Female	906 (41.52)	909 (41.66)
Male	175 (8.02)	181 (8.3)
Transgender female	0 (0)	0 (0)
Transgender male	0 (0)	1 (0.05)
Nonbinary	0 (0)	0 (0)
Other	1 (0.05)	1 (0.05)
Prefer not to say	3 (0.14)	4 (0.18)
Age (years), mean (SD; range)	40.42 (10.92; 19-67)	40.64 (11.02; 18-80)
Perceived socioeconomic status (1-10), mean (SD; range)	5.66 (1.50; 1-10)	5.66 (1.49; 1-10)
Hours worked per week, n (%)		
≤30 hours	261 (11.96)	277 (12.69)
>30 hours per week	825 (37.81)	819 (37.53)
Individual income (£), median (IQR; average exchange rate at the time of the study was £1=US $1.33)	25,000-30,000 (20,000-25,000 to 35,000-40,000)	25,000-30,000 (20,000-25,000 to 35,000-40,000)
Marital status, n (%)		
Living with partner, married, or civil partnership	800 (36.66)	788 (36.11)
Single	286 (13.11)	307 (14.07)
Role, n (%)		
Allied Health Professional (eg, speech therapist and occupational therapist)	180 (8.25)	208 (9.53)
Physician	89 (4.08)	78 (3.57)
Manager	51 (2.34)	51 (2.34)
Nurse	284 (13.02)	301 (13.79)
Psychologist, psychological therapist, or practitioner	93 (4.26)	112 (5.13)
Wider health care team	216 (9.9)	193 (8.85)
Other	187 (8.57)	175 (8.02)
NHS[c] Trust type, n (%)		
Acute (hospital)	334 (15.31)	319 (14.62)
Ambulance	81 (3.71)	71 (3.25)
Combined (multiple Trust types within one Trust)	293 (13.43)	288 (13.2)

Characteristics	Moodzone	Headspace
Community	66 (3.02)	65 (2.98)
GP[d]	54 (2.47)	77 (3.53)
Mental health	245 (11.23)	264 (12.1)

[a]GCSE: General Certificate of Secondary Education.

[b]NVQ 2: National Vocational Qualification level 2.

[c]NHS: National Health Service.

[d]GP: general practitioner.

Table 2. Descriptive statistics on all outcome measures at all time points (raw complete case data; N=2182).

Measure and arm	Time point 1 (baseline)			Time point 2 (1.5 months)			Time point 3 (4.5 months)		
	Values, n (%)	Values, mean (SD)	95% CI	Values, n (%)	Values, mean (SD)	95% CI	Values, n (%)	Values, mean (SD)	95% CI
DASS-21[a] Stress (primary outcome)									
Moodzone	1087 (49.82)	16.24 (7.80)	15.78 to 16.71	701 (32.13)	13.92 (7.65)	13.36 to 14.49	552 (25.29)	14.47 (8.11)	13.79 to 15.15
Headspace	1095 (50.18)	15.67 (7.40)	15.23 to 16.11	715 (32.77)	12.86 (7.06)	12.34 to 13.38	571 (26.17)	12.39 (7.85)	11.74 to 13.03
DASS-21 Depression									
Moodzone	1087 (49.82)	10.72 (8.26)	10.23 to 11.21	701 (32.13)	9.61 (8.37)	8.99 to 10.23	552 (25.29)	9.58 (8.66)	8.86 to 10.31
Headspace	1092 (50.05)	10.29 (7.76)	9.83 to 10.75	715 (32.77)	8.34 (7.41)	7.79 to 8.88	571 (26.17)	7.87 (8.03)	7.21 to 8.53
DASS-21 Anxiety									
Moodzone	1087 (49.82)	9.06 (7.43)	8.62 to 9.51	701 (32.13)	7.42 (7.1)	6.90 to 7.95	552 (25.29)	7.45 (7.19)	6.85 to 8.05
Headspace	1095 (50.18)	8.58 (6.99)	8.16 to 8.99	716 (32.81)	6.47 (6.26)	6.02 to 6.93	571 (26.17)	5.97 (6.49)	5.43 to 6.50
SWEMWBS[b] Well-being									
Moodzone	1087 (49.82)	21.43 (3.61)	21.22 to 21.65	678 (31.07)	22.43 (4.16)	22.12 to 22.75	525 (24.06)	22.27 (4.44)	21.89 to 22.65
Headspace	1095 (50.18)	21.57 (3.68)	21.35 to 21.79	704 (32.26)	22.7 (3.99)	22.41 to 23.00	550 (25.21)	23.12 (4.41)	22.76 to 23.49
Maslach[c] Emotional Exhaustion									
Moodzone	1068 (48.95)	26.2 (11.81)	25.49 to 26.91	678 (31.07)	24.31 (12.06)	23.40 to 25.22	531 (24.34)	24.33 (12.47)	23.26 to 25.39
Headspace	1080 (49.5)	25.65 (12.08)	24.93 to 26.37	703 (32.22)	23.71 (12.15)	22.81 to 24.61	552 (25.29)	23.27 (12.69)	22.21 to 24.33
Maslach Depersonalization									
Moodzone	1067 (48.9)	5.82 (5.72)	5.47 to 6.16	677 (31.03)	5.64 (5.63)	5.21 to 6.06	530 (24.29)	5.68 (5.84)	5.18 to 6.18
Headspace	1077 (49.36)	5.75 (5.75)	5.40 to 6.09	701 (32.13)	5.38 (5.48)	4.97 to 5.79	552 (25.29)	5.51 (5.67)	5.03 to 5.98
Maslach Personal Accomplishment									
Moodzone	1065 (48.81)	36.5 (7.02)	36.08 to 36.92	677 (31.03)	37.17 (6.98)	36.64 to 37.70	529 (24.24)	36.4 (7.98)	35.72 to 37.09
Headspace	1074 (49.22)	36.42 (6.74)	36.01 to 36.82	702 (32.17)	37.2 (7.19)	36.67 to 47.73	551 (25.25)	37.39 (7.4)	36.77 to 38.01
FFMQ-15[d] (minus Observe subscale)									
Moodzone	1085 (49.73)	38.33 (7.04)	37.91 to 38.74	709 (32.49)	39.8 (7.24)	39.27 to 40.33	551 (25.25)	39.89 (7.48)	39. 27 to 40.52
Headspace	1092 (50.05)	38.22 (6.7)	37.82 to 38.62	717 (32.86)	40.17 (6.59)	39.69 to40.65	574	40.93 (6.68)	40.38 to 41.47
SCS-SF[e] Self-Compassion									
Moodzone	1085 (49.73)	34.11 (9.03)	33.58 to 34.65	688 (31.53)	36.28 (9.43)	35.57 to 36.99	544 (26.31)	36.29 (9.29)	35.51 to 37.07
Headspace	1093 (50.09)	33.86 (8.88)	33.33 to 34.38	710 (32.54)	37.3 (9.3)	36.62 to 37.99	560 (25.66)	38.22 (9.34)	37.44 to 38.99
PSWQ[f] Worry									

Measure and arm	Time point 1 (baseline)			Time point 2 (1.5 months)			Time point 3 (4.5 months)		
	Values, n (%)	Values, mean (SD)	95% CI	Values, n (%)	Values, mean (SD)	95% CI	Values, n (%)	Values, mean (SD)	95% CI
Moodzone	1086 (49.77)	54.2 (14.43)	53.34 to 55.06	677 (31.03)	51.33 (14.65)	50.22 to 52.44	526 (24.11)	51.65 (15.18)	50.35 to 52.95
Headspace	1095 (50.18)	53.53 (14.44)	52.67 to 54.38	704 (32.26)	50.28 (14.33)	49.22 to 51.34	549 (25.16)	49.37 (14.45)	48.15 to 50.58
RRS[g] Rumination (Brooding)									
Moodzone	1087 (49.82)	10.69 (3.43)	10.49 to 10.89	677 (31.03)	9.97 (3.51)	9.71 to 10.24	519 (23.79)	9.91 (3.45)	9.61 to 10.20
Headspace	1096 (50.23)	10.39 (3.35)	10.19 to 10.58	703 (32.22)	9.74 (3.19)	9.50 to 9.98	548 (25.11)	9.45 (3.35)	9.17 to 9.73
CLS[h] Compassion for Others									
Moodzone	1085 (49.73)	4.77 (1.1)	4.71 to 4.84	675 (30.93)	4.64 (1.15)	4.55 to 4.73	518 (23.74)	4.5 (1.24)	4.29 to 4.61
Headspace	1094 (50.14)	4.78 (1.09)	4.71 to 4.84	702 (32.17)	4.75 (1.12)	4.67 to 4.84	540 (24.75)	4.69 1.17 ()	4.59 to 4.79
Sickness absence (days in past month)									
Moodzone	1086 (49.77)	2.44 (7.45)	1.99 to 2.88	__[i]	—	—	573 (26.26)	2.04 (6.86)	1.48 to 2.60
Headspace	1095 (50.18)	2.35 (7.08)	1.93 to 2.77	—	—	—	593 (27.18)	2.23 (7.99)	1.58 to 2.87
Formal engagement (days/week)									
Moodzone	N/A[j]	N/A	N/A	653 (29.93)	2.33 (2.01)	2.17 to 2.48	522 (23.92)	1.35 (1.65)	1.21 to 1.49
Headspace	N/A	N/A	N/A	679 (31.12)	3.56 (2.26)	3.39 to 3.73	544 (26.31)	2.16 (1.91)	2.00 to 2.32
Informal engagement (days/week)									
Moodzone	N/A	N/A	N/A	654 (29.97)	2.2 (2.08)	2.04 to 2.36	520 (23.83)	1.4 (1.77)	1.25 to 1.55
Headspace	N/A	N/A	N/A	679 (31.12)	2.92 (2.22)	2.75 to 3.09	544 (26.31)	3 (2.18)	2.81 to 3.18
CEQ[k] credibility									
Moodzone	1080 (49.5)	−0.58[l] (2.41)	−0.72 to −0.44	—	—	—	—	—	—
Headspace	1082 (49.59)	0.58[l] (2.55)	0.43 to 0.73	—	—	—	—	—	—
Expectancy									
Moodzone	1081 (49.54)	−0.40[l] (2.70)	−0.56 to −0.24	—	—	—	—	—	—

Measure and arm	Time point 1 (baseline)			Time point 2 (1.5 months)			Time point 3 (4.5 months)		
	Values, n (%)	Values, mean (SD)	95% CI	Values, n (%)	Values, mean (SD)	95% CI	Values, n (%)	Values, mean (SD)	95% CI
Headspace	1091 (50)	0.39[l] (2.80)	0.23 to 0.56	—	—	—	—	—	—

[a]DASS-21: 21-item Depression, Anxiety, and Stress Scale.

[b]SWEMWBS: Short Warwick Edinburgh Mental Well-being Scale.

[c]Maslach Burnout Inventory.

[d]FFMQ15: 15-item Five Facets of Mindfulness Questionnaire.

[e]SCS-SF: Self-Compassion Scale–Short Form.

[f]PSWQ: Penn State Worry Questionnaire.

[g]RRS: Ruminative Response Scale.

[h]CLS: Compassionate Love Scale.

[i]Not available.

[j]N/A: not applicable.

[k]CEQ: Credibility and Expectancy Questionnaire.

[l]Means created from subscale totals of *z* scores [54].

Primary Outcome (Stress)

ITT Analysis

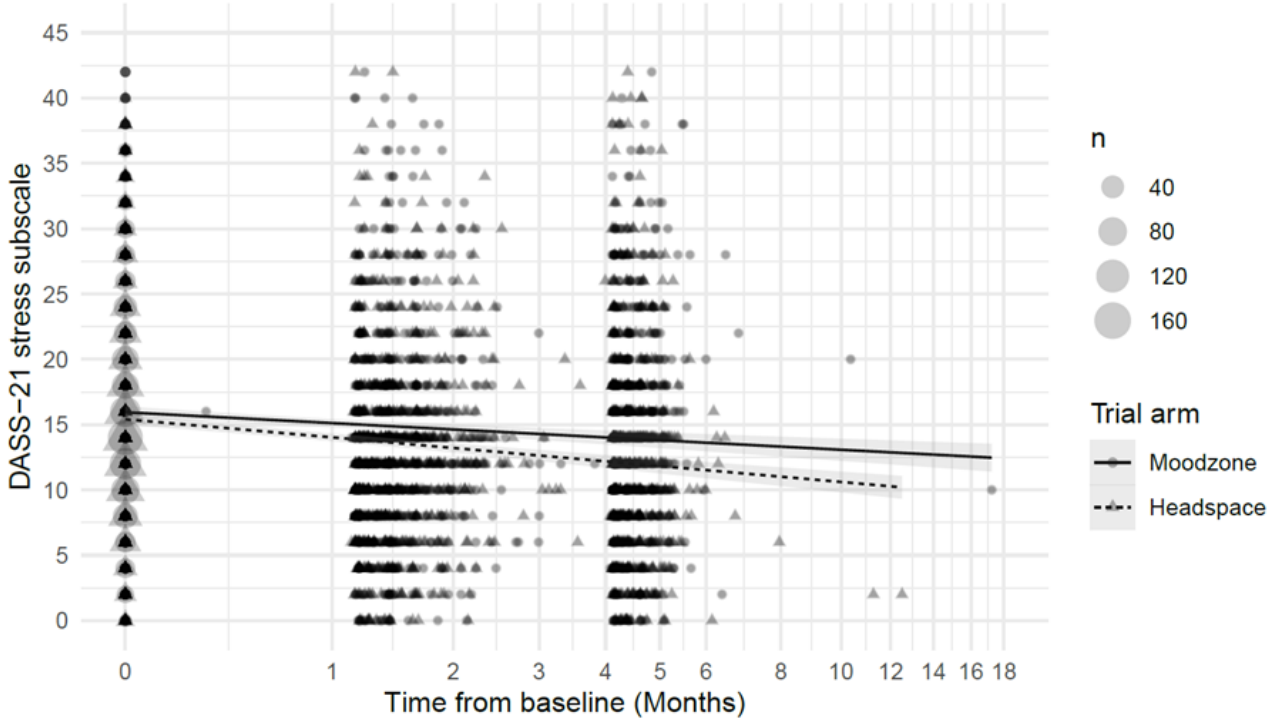
Table 3 shows that the main effects of trial arm (Headspace or Moodzone) and time (months) were significant, as was the crucial trial arm × month interaction, which indicates that the trajectories of the 21-item Depression, Anxiety, and Stress Scale (DASS-21) Stress scores over time differed significantly between the 2 trial arms for the ITT sample (Figure 2). The parameter value (b=−0.31) tells us that the rate of change (gradient) over time was −0.31 points greater on the DASS-21 Stress subscale per month in the Headspace arm than in the Moodzone arm. Specifically, for every month that passed, DASS-21 stress scores changed by −0.23 units on the scale in the Moodzone group compared with a corresponding change of −0.54 units in the Headspace group (ie, a difference between arms of −0.31 units per month).

To break down this effect, comparisons were made between the estimated marginal means of the outcome from the model at 1.5 (T2) and 4.5 (T3) months in the 2 arms. In the Moodzone arm, stress was significantly higher at baseline than at both 1.5 months (b=0.34, SE 0.09; P<.001) and 4.5 months (b=1.03, SE 0.26; P<.001). Stress was also significantly higher at 1.5 months than at 4.5 months (b=0.69, SE 0.18; P<.001). Similarly, in the Headspace arm, stress was significantly higher at baseline than at both 1.5 months (b=0.81; SE 0.08; P<.001) and 4.5 months (b=2.42, SE 0.25; P<.001), and significantly higher at 1.5 months than at 4.5 months (b=1.61, SE 0.17; P<.001). The b values represent the difference in the estimated marginal means; they show that, for example, at 4.5 months, the decrease in DASS-21 Stress compared with baseline was 1.03 points in the Moodzone arm and 2.42 points in the Headspace arm. In other words, at 4.5 months after randomization, Moodzone reduced DASS-21 Stress scores by approximately 1 point along the 42-point scale, and the equivalent change for Headspace was a reduction of approximately 2.5 points along the scale. In addition, the difference in estimated marginal means between the 2 arms was b=0.62 (SE 0.31; P=.045) at baseline, b=1.08 (SE 0.30; P<.001) at 1.5 months, and b=2.00 (SE 0.42; P<.001) at 4.5 months (the preregistered primary end point).

Table 3. Model for the 21-item Depression, Anxiety, and Stress Scale Stress (intention-to-treat sample with multiple imputation).

Effect	Unstandardized *b* (SE; 95% CI)	*t* test (*df*)	*P* value
Intercept	15.33 (0.40; 14.55 to 16.11)	38.41 (6054.49)	<.001
Trial arm	−0.62 (0.31; –1.23 to –0.01)	−2.01 (5129.90)	.045
Months	−0.23 (0.06; –0.35 to –0.11)	−3.92 (165.07)	<.001
Trial arm × month	−0.31 (0.08; –0.47 to –0.14)	−3.64 (151.13)	<.001

Figure 2. DASS-21 Stress scores over time (intention-to-treat complete case sample). Each triangle represents a Headspace participant, and each circle represents a Moodzone participant. DASS-21: 21-item Depression, Anxiety, and Stress Scale.



Per-Protocol Analysis

The per-protocol sample included only participants who formally engaged with their allocated intervention at least 3 days per week during the initial intervention period (T1-T2). Multimedia Appendix 6 shows a significant trial arm × month interaction, indicating that the trajectories of stress over time differed significantly between the 2 trial arms. The rate of change over time was −0.28 DASS-21 Stress units greater per month in the Headspace arm than in the Moodzone arm. Specifically, in the Moodzone arm, the rate of change over time was −0.42, which means that for every month that passed, DASS-21 Stress scores decreased by 0.42 points; however, in the Headspace, arm the rate of change over time was −0.70 (a difference of −0.28 between arms), which means that for every month that passed, DASS-21 Stress decreased by 0.70 points.

In the per-protocol sample in the Moodzone arm, stress was significantly higher at baseline than at both 1.5 months (T2; *b*=0.63, SE 0.16; *P*<.001) and 4.5 months (T3; *b*=1.88, SE 0.47; *P*<.001). Stress was also significantly higher at 1.5 months than at 4.5 months (*b*=1.26, SE 0.31; *P*<.001). Similarly, in the Headspace arm, stress was significantly higher at baseline than at both 1.5 months (*b*=1.05, SE 0.12; *P*<.001) and 4.5 months (*b*=3.14, SE 0.35; *P*<.001) and significantly higher at 1.5 months than at 4.5 months (*b*=2.09, SE 0.24; *P*<.001). The *b* values represent the difference in the estimated marginal means; they show that, for example, at 4.5 months the decrease in stress compared with baseline was 1.88 points in the Moodzone arm and 3.14 points in the Headspace arm. In addition, the difference in estimated marginal means between the 2 arms was not significant at baseline (*b*=0.24, SE 0.52; *P*=.65) or at 1.5 months (*b*=0.66, SE 0.48; *P*=.17) but was significant at 4.5 months (*b*=1.50, SE 0.62; *P*=.02).

Reliable Change

Multimedia Appendix 7 provides tables for the reliable change analysis. Overall, 20.5% (71/347) of Moodzone and 29.7% (102/343) of Headspace participants who scored at least in the mild stress range at T1 showed reliable improvement from T1 to T2 in stress, with 2.9% (10/347) and 2% (7/343) showing reliable deterioration, respectively. From T1 toT3, approximately 24.1% (66/247) of Moodzone and 36.8% (100/272) of Headspace participants scored at least in the mild stress range at T1 and showed reliable improvement in stress, with 2.9% (8/274) and 4% (11/272) showing reliable deterioration. The trial arm significantly predicted reliable improvement (compared with no change) at both T2 and T3. At T2, the odds of being classified as having reliable improvement were 1.45 higher in the Headspace than in the Moodzone arm, and 95% CIs did not cross 1 (95% CI 1.05-2.01). At T3, the odds of being classified as having reliable improvement were 1.48 higher in the Headspace than in the Moodzone arm, with 95% CIs not crossing 1 (95% CI1.09-2.02). The odds of being classified as showing reliable deterioration in stress were not different between arms at either T2 (odds ratio 0.71, 95% CI 0.29-1.73) or T3 (odds ratio 1.26, 95% CI 0.55- 2.92) as 95% CIs crossed 1.

Secondary Outcomes and Additional Analyses

The findings from the ITT analysis of secondary outcomes are shown in Table 4. Further details on the secondary outcomes and additional analyses are provided in Multimedia Appendix 4 [13,18,25,29,30,36,56,67-72] and details of the analysis of lasting negative effects are provided in Multimedia Appendix 8.

Table 4. Overall trial arm × time effects of the intervention on all outcomes for the intention-to-treat sample with multiple imputation (Moodzone N=1087 and Headspace N=1095 on the primary outcome)[a].

Measure	Difference between arms per month, unstandardized *b* (SE; 95% CI)	*t* test (*df*)	*P* value	Differences between arms at 1.5 months			Differences between arms at 4.5 months		
				Unstandardized[b] *b* (SE)	*P* value	Hedges *g*[c]	Unstandardized[b] *b* (SE)	*P* value	Hedges *g*[c]
DASS-21[d] Stress	−0.31 (0.08; −0.47 to −0.14)	−3.64 (151.13)	<.001	1.08 (0.30)	<.001	0.14	2.00 (0.42)	<.001	0.26
DASS-21 Anxiety	−0.19 (0.07; −0.32 to −0.06)	−2.94 (218.51)	.004	0.78 (0.27)	.04	0.14	1.36 (0.34)	<.001	0.22
DASS-21 Depression	−0.24 (0.08; −0.40 to −0.08)	−3.02 (211.23)	.003	0.92 (0.32)	.005	0.16	1.65 (0.43)	.001	0.20
SWEMWBS[e] Well-being	0.14 (0.04; 0.05 to 0.23)	3.16 (289.19)	.002	−0.35 (0.15)	.02	0.07	−0.77 (0.21)	<.001	0.19
Maslach[f] Emotional Exhaustion	−0.19 (0.10; −0.39 to 0.01)	−1.85 (372.00)	.07	N/A[g]	N/A	0.05	N/A	N/A	0.08
Maslach Depersonalization	−0.04 (0.05; −0.14 to 0.05)	−0.94 (321.54)	.35	N/A	N/A	0.05	N/A	N/A	0.03
Maslach Personal Accomplishment	0.13 (0.07; −0.01 to 0.27)	1.84 (251.87)	.07	N/A	N/A	0.00	N/A	N/A	0.13
FFMQ-15[h] (minus Observe)	0.22 (0.06; 0.09 to 0.34)	3.38 (298.64)	.001	−0.31 (0.28)	.26	0.05	−0.96 (0.35)	.006	0.15
SCS-SF[i] Self-Compassion	0.48 (0.08; 0.33 to 0.64)	6.05 (201.36)	<.001	−0.76 (0.37)	.04	0.11	−2.21 (0.46)	<.001	0.21
CLS[j] Compassion for Others	0.02 (0.01; 0.00 to 0.04)	2.07 (144.19)	.04	−0.03 (0.05)	.48	0.10	−0.09 (0.06)	.12	0.16
PSWQ[k] Worry	−0.30 (0.11; −0.51 to −0.09)	−2.83 (278.67)	.005	1.15 (0.59)	.05	0.07	2.06 (0.69)	.003	0.15
RRS[l] Rumination (Brooding)	−0.06 (0.03; −0.12 to 0.00)	−1.91 (349.8)	.06	N/A	N/A	0.07	N/A	N/A	0.14

[a]A negative value for *b* is in favor of Headspace for the DASS-21 subscales, RRS Brooding, and PSWQ Worry; a positive value for *b* is in favor of Headspace for the SWEMWBS, FFMQ-15 (minus Observe), SCS-SF Self-Compassion, and CLS Compassion for Others.

[b]Unstandardized effects at 1.5 and 4.5 months were only reported in the event of a significant trial arm × time interaction.

[c]Hedges g is the difference between trial arms at time point 2 and time point 3 based on raw data.

[d]DASS-21: 21-item Depression, Anxiety, and Stress Scale.

[e]SWEMWBS: Short Warwick Edinburgh Mental Well-being Scale.

[f]Maslach Burnout Inventory.

[g]N/A: not applicable.

[h]FFMQ15: 15-item Five Facets of Mindfulness Questionnaire.

[i]SCS-SF: Self-Compassion Scale–Short Form.

[j]CLS: Compassionate Love Scale.

[k]PSWQ: Penn State Worry Questionnaire.

[l]RRS: Ruminative Response Scale.

Intervention Engagement

Multimedia Appendix 9 shows the self-reported engagement with each intervention. Time was treated categorically (1.5 vs 4.5 months). The model was fitted is as follows:

Level 1: $DASS\text{-}21_{ij}=\pi_{0j} + \pi1 \times Time_{ij} + $  $_{ij}$

Level 2: $\pi_{0j}=\gamma00 + \gamma01 \times Trial\ arm_j + \zeta_{0j}$

In the ITT sample, Headspace participants engaged with their allocated intervention formally and informally on more days per week than Moodzone participants, both between T1 and T2 (*b*=−1.32, SE 0.11; *P*<.001 and *b*=−0.79, SE 0.11; *P*<.001,

respectively) and between T2 and T3 (b=−0.70, SE 0.10; P<.001 and b=−1.55, SE 0.12; P<.001).

Mediation Analysis

Formal engagement (practice days per week) from T1 to T2 mediated the effect of trial arm on T1 to T3 improvements in stress using complete case data within the per-protocol sample (582/2182, 26.67%) as 95% CIs did not cross 0 (95% CI −0.097 to −0.006). Similarly, improvement in self-compassion at T1 and T2 significantly mediated T1 to T3 improvement in stress for per-protocol participants (95% CI −0.144 to −0.022). However, improvements in mindfulness, worry, and rumination (brooding) at T1 and T2 did not significantly mediate improvement in stress from T1 to T3 for per-protocol participants, as all 95% CIs crossed 0 (mindfulness: 95% CI −0.107 to 0.029; worry: 95% CI −0.069 to 0.025; brooding: 95% CI −0.046 to 0.037). Overall, the mediation analysis findings suggest that the greater improvement in stress in the Headspace arm in comparison with the Moodzone arm was driven, at least in part, by engagement on more days per week in formal practices and exercises and greater improvement in self-compassion (but not in mindfulness, worry, or rumination) in the Headspace arm during the initial intervention period.

Intervention Credibility and Expectancy

At T1, between-group differences in intervention credibility and expectancy were assessed via standardized totals of the first and last 3 items of the Credibility and Expectancy Questionnaire, respectively. Headspace was rated as significantly more credible than Moodzone ($t_{2164.81}$=−10.88; P<.001; Cohen d=0.47). Significantly more positive expectancy ratings were also observed for Headspace compared with Moodzone (t_{2170}=−6.70; P<.001; Cohen d=0.29).

Awareness of Study Purpose

At T3, only 0.68% (8/1171) of the participants indicated a clear awareness of the study hypothesis. Most of these participants (7/1171, 0.59%) were allocated to Moodzone. The analysis was not conducted between the arms, given the small numbers involved.

Discussion

Principal Findings

In this study, we examined whether an unguided digital MBSH intervention (Headspace) was effective in reducing health care worker stress when compared with an active control condition (Moodzone) that was matched for duration and medium (ie, digitally delivered). In contrast to previous studies, this was a fully powered, multisite definitive RCT with patient-facing NHS staff working in a broad range of health care roles and across a broad range of health care organization types, allowing definitive conclusions to be drawn and findings to be generalized.

Primary Outcome

The stress in both arms improved over time. In comparison with Moodzone, Headspace participants showed a significantly greater reduction in stress (the preregistered primary outcome) over the 4.5-month course of the study, with significant but small differences between trial arms at 1.5 and 4.5 months (the primary endpoint). Headspace participants showed an average reduction in stress over the study period of almost 2.5 points on the 42-point scale, which was over twice the improvement in stress experienced by Moodzone participants. Compared with Moodzone participants, Headspace participants were significantly more likely to experience reliable improvements in stress, both from T1 to T2 and T1 to T3.

The between-group effect on stress at the preregistered primary end point was small (Hedges g=0.26), consistent with relevant evidence from 2 recent meta-analyses. For example, Spijkerman et al [36] identified significantly lower levels of stress for unsupported web-based mindfulness and acceptance-based self-help interventions than for control conditions at the postintervention time point among nonclinical samples, with a small effect (Hedges g=0.19), whereas a more recent systematic review and meta-analysis conducted by the study team [73] observed a similarly small and statistically significant between-group postintervention effect on stress when unguided MBSH was compared with active control conditions among nonclinical samples (mirroring the design of this study; Hedges g=0.20). As such, the modest reductions in stress observed in this study appear to be in keeping with the effects observed for unguided MBSH in the broader literature, and taken together, these observed effects suggest that a small and specific benefit may be associated with such interventions.

Medium to large between-group effects on stress have been reported for the well-established MBSR course in comparison with active and inactive control conditions (Hedges g=0.77) [29] and for a newly developed version of MBCT for the workplace, MBCT for Life (MBCT-L), in comparison with wait-list (Cohen d=0.72) [30]. Although it is not possible to directly compare with this study because of differences in control conditions, it is likely that these in-person, guided, and more intensive courses are more effective than unguided MBSH. However, there are several barriers to extending the reach of these courses. First, there are not enough mindfulness teachers working in the NHS to offer MBIs to patients in line with the National Institute of Health and Clinical Excellence guidelines [74], let alone to offer MBSR or MBCT-L courses to NHS staff. Second, stigma-related concerns among health care workers about accessing mental health support [34] may hinder uptake, even if in-person MBIs are available. Third, many health care workers struggle to commit to the highly structured and time-intensive nature of traditional MBIs [32,33].

Our study also extends the findings of meta-analyses of RCTs exploring the effects of digital interventions for stress management in the workplace more broadly. When considering smartphone apps specifically, a recent RCT of an unguided non-MBI workplace stress management app based on the Job Demands-Resources Model [75] in comparison with a wait-list found a similarly small effect on stress 6 weeks after randomization (Cohen d=0.14) [76]. When considering digital resources more broadly, Heber et al [77] examined the effects of web- and computer-based interventions based on cognitive behavioral therapy (CBT), third-wave CBT (eg, mindfulness and acceptance and commitment therapy), and non–CBT-based

interventions (eg, present control interventions and career identity training for stress management) compared with control conditions among nonclinical populations experiencing stress and found a significant between-group postintervention reduction in stress when looking at unguided interventions, with a small effect (Cohen *d*=0.33). In addition, Carolan et al [78] identified significant between-group postintervention improvements in psychological well-being (which included measures of stress), with a small effect (Hedges *g*=0.37), when comparing mainly CBT-based web-delivered interventions with control conditions in the workplace. However, many of the studies considered in these reviews used wait-list control conditions and included guided interventions, which is likely to have contributed to the magnitude of the observed effects.

Unguided digital MBSH interventions, such as Headspace, offer the potential to provide mindfulness training to NHS workers at a scale without the need for a trained mindfulness teacher on site, thus enabling workers to engage with an MBI at a time, place, and pace that suits them. However, to optimize the benefits available from such interventions, it is important that they are offered in a supportive workplace context; are aligned with organizational values, goals, and practices; and protected time and space are available for such self-care [79].

We do not contend that MBSH could or should replace in-person MBIs for NHS workers, given the likely larger effect of in-person courses; however, unguided MBSH interventions could be part of a solution to widening access to mindfulness training while simultaneously endeavoring to find ways of increasing the availability of in-person MBIs. Additional costs associated with providing trained practitioners also put unguided MBSH at an advantage over guided MBSH interventions, as they have the potential to be made more widely available. However, a disadvantage is that effectiveness similarly appears to be reduced, with Spijkerman et al [36] finding significantly smaller between-group effects for mindfulness- and acceptance-based self-help interventions that were unguided (Hedges *g*=0.19) compared with guided interventions (Hedges *g*=0.89). Therefore, what is gained in the widening reach may be lost in reducing the benefits. However, there is emerging evidence that book-based unguided MBSH may produce larger effect sizes than digital MBSH and a direct head-to-head comparison of MBSH formats (especially book vs digital) is warranted [73].

Intervention Engagement

In comparison with Moodzone, Headspace participants reported a significantly greater number of days spent formally engaging with mindfulness practice. Self-reported practice engagement in the Headspace arm averaged 3.5 days per week during the initial intervention period and 2 days per week during the follow-up period. As such, our findings suggest that sustained commitment to even brief mindfulness practice is challenging for many health care workers; therefore, the reduced practice times afforded by MBSH may provide a more viable alternative to mindfulness training. Interestingly, although daily practice at home is encouraged in MBCT or MBSR, it appears that greater benefits for mental health are seen when people practice at least 3 days a week during the initial intervention period, as compared with people who practice <3 days a week [67]. In this study, 66.6% (452/679) and 37.9% (206/544) of Headspace participants said that they practiced at least 3 days a week at T2 and T3, respectively.

Per-protocol analyses were also conducted to examine the effects of Headspace compared with Moodzone for only those participants who reported formally engaging with their allocated intervention ≥3 days per week during the initial intervention period (based on Crane et al [67]). This shows the overall beneficial effects of Headspace over time in comparison with Moodzone. However, although there were significant between-group effects at T3 in favor of Headspace, between-group effects at T2 were no longer significant. Moreover, most effects of secondary outcomes over time were nonsignificant in the per-protocol analysis. If Headspace engagement is the active ingredient of change, per-protocol effects might be expected to be larger than ITT effects and remain statistically significant, despite the relatively smaller sample contributing to the per-protocol analysis. Therefore, further research is needed to explore the relationship between engagement with Headspace and the magnitude of outcomes.

Given that formal engagement with Headspace (days per week) was greater than that with Moodzone, it could be that once the formal engagement is accounted for in the per-protocol sample (ie, all included participants formally engaged for at least 3 days per week during the initial intervention period), the relative benefits of Headspace over Moodzone are somewhat diminished. However, finding ways of encouraging engagement in unguided digital well-being interventions is a well-recognized challenge [80], and greater engagement with Headspace in comparison with an NHS-developed digital well-being offer is important in itself, as, in the real world, it is the ITT benefits that are realized rather than the per-protocol effects.

Multimedia Appendix 4 [13,18,25,29,30,36,56,67-72] provides a discussion of the findings on secondary outcomes and additional analyses.

Strengths and Limitations

Although the adequately powered sample size and rigorous study design represent the key strengths of our study, the findings should be considered within the context of several limitations. In this trial, the NHS's digital workplace stress resource, Moodzone, was selected as the active control condition, inviting study participants to engage with a range of evidence-based recommendations for a minimum of 10 minutes each day as a time match to the Headspace intervention.

However, as previously discussed, intervention engagement was significantly greater for Headspace than for Moodzone; therefore, it is plausible that the active ingredient was intervention engagement rather than intervention content. However, even if Headspace is more effective than Moodzone simply because it is more engaging, this will have implications for real-world effectiveness. To determine the effectiveness of intervention content specifically, future research should compare Headspace with an equally engaging active control. In addition, after providing participants with postrandomization information about their allocated intervention, Headspace received

significantly higher credibility and expectancy ratings than Moodzone. Expectancy effects can affect psychotherapeutic outcomes [81], and it is plausible that the greater credibility and expectancy of Headspace than that of Moodzone could explain the study findings. However, the beneficial effects of Headspace on stress outcomes in comparison with Moodzone were retained in a post hoc analysis where credibility and expectancy ratings were entered as covariates, suggesting that the intervention effects cannot be purely explained by the greater credibility and expectancy of Headspace. Future studies should consider the role of credibility and expectancy in more depth and compare Headspace with an intervention matched for credibility and expectancy.

Models were fitted for 11 secondary outcomes, each with 3 predictors (trial arm, time, and their interaction), yielding 33 *P* values. To control for the type I error rate across these models, the reported *P* values for the interaction effects for secondary outcomes in Table 4 were evaluated against a critical *P* value of .002 (ie, .05/33). When evaluating against this stricter criterion, all the interaction effects for secondary outcomes were nonsignificant, except for self-compassion. However, the main goal of *P* value correction is to mitigate fishing expeditions, and all models were preplanned; in addition, the trade-off in controlling type I errors is losing control of type II errors, and there is no inherent reason why controlling type I errors is more desirable. In addition, where the interaction is significant, we tried to carefully evaluate the raw effect size, which adds to the important context of the real-world importance of the effect irrespective of the *P* value.

Recent attention has been paid to the concept of a "digital placebo effect," whereby nontherapeutic elements of digital interventions are thought to engender either real or imagined improvements in mental health outcomes [82]. As such, it is perhaps also of note that although Headspace was delivered via a sophisticated smartphone app that offered structured daily guidance, Moodzone was delivered via a series of web pages that participants were expected to navigate independently. Therefore, it is possible that the observed effects are, at least in part, because of participants' more favorable expectations of Headspace relative to Moodzone because of differences in content delivery. Future research should compare Headspace with an active control matched for delivery format and style.

For reasons beyond our control, Headspace was temporarily advertised on the Moodzone web page (notwithstanding the widespread advertising of Headspace on social media and other platforms), which may explain why, despite apparently successful blinding of the study hypotheses, a proportion of Moodzone participants completing measures at T3 reported using Headspace during the study period. However, this is only likely to have diluted between-group differences, and, at worst, our findings can be considered to reflect a conservative estimate of the difference between groups. Moreover, although minor design, platform, and content changes are unlikely to have affected our results [83], it is also worth noting that both Headspace and Moodzone were examined as *live* resources, and as such, both were subject to changes during the study period.

Our study suggests the benefits of an invitation for brief mindfulness-based practices using unguided digital MBSH; however, a *class effect* (ie, the translation of these benefits to any unguided digital MBSH resource) cannot be assumed. Further research is required to identify and optimize the active ingredients of unguided MBSH.

Further limitations of this study are that all outcomes and measures of engagement were self-reported and that dropout at T3 was relatively high, although not atypical for RCTs of digital interventions. Finally, although we recruited a large sample of health care staff working in a variety of job roles and across a variety of NHS organization types across England, our sample was not entirely representative of the NHS workforce. For example, 83.22% (1815/2181) of participants identified as female compared with 77% of NHS staff more broadly [84], and our sample underrepresented Black, Asian, and minority ethnic staff, with 92.74% (2019/2177) White participants in comparison with 77.9% in the NHS workforce [85]. Future studies could monitor demographic characteristics as recruitment progresses and adjust recruitment strategies accordingly to target underrepresented groups.

Future Research

Future research should match unguided digital MBSH to equally credible active control conditions with equal expectations of benefits. Doing so would help enable greater confidence in conclusions about the relative benefits of mindfulness-based content. Moreover, dismantling trials would also be beneficial to unpick the active ingredients of digital resources such as Headspace.

Another important avenue for future research involves identifying the moderators of engagement. Identifying moderators of engagement with unguided digital MBSH interventions may facilitate the targeted intervention of barriers to and facilitators of regular mindfulness practice to promote engagement and, in turn, potentially boost the effects.

Guided mindfulness- and acceptance-based self-help has larger effects on stress outcomes than unguided approaches [36]. There is a balance to be struck between providing MBSH at scale to more health care workers (without guidance and its associated costs) and providing maximally effective MBSH to potentially fewer health care workers (with guidance). Few head-to-head trials exist, and a well-designed study comparing the clinical effectiveness and cost-effectiveness of guided digital MBSH with unguided digital MBSH for health care workers is warranted to explore the relative advantages and disadvantages of each approach. Future research could also explore the clinical effectiveness and cost-effectiveness of different methods of providing MBSH support and guidance at different levels of intensity (eg, automated but personalized, regular email or text guidance; an MBSH support helpline; asynchronous email support from a trained practitioner; and weekly support sessions with a mindfulness teacher). For interventions that incur a cost to the individual or organization, it is particularly important to have a good understanding of the balance between economic costs (eg, funding a subscription for health care staff in an organization) and economic benefits (eg, sickness absence). Future research should include a full health economic evaluation

to examine not only the clinical effectiveness of different MBSH interventions but also their cost-effectiveness. In addition, future research should also examine naturalistic, real-world outcomes of Headspace in specific populations to complement RCT findings.

Implementation

Overall, the findings suggest that an unguided digital MBSH program appears to be a safe intervention for health care workers, which can yield small but significant improvements in stress and other mental health outcomes with minimal time investment from users. However, it is important to consider that a wide range of non-MBI digital interventions is effective in improving stress and mental health both within [78] and outside the workplace [77] and may be preferred by some health care workers. Furthermore, our findings should be considered within the context of significantly larger effects on stress (in various populations) in guided versus unguided mindfulness- and acceptance-based self-help interventions [36] and larger effects on health care worker stress with MBSR [29] and MBCT-L [30], although this does not directly compare like for like. Although unguided digital MBSH interventions can offer a potential solution to some of the barriers associated with accessing guided MBSH and MBSR or MBCT-L, the smaller effects indicate that a careful balance needs to be struck between effectiveness and accessibility.

It is also worth considering that Headspace was not beneficial for the workplace outcomes of burnout and sickness absence, and as such, alternative strategies will be needed to identify appropriate solutions to these problems. Given the greater effects of MBSR and MBCT-L on health care workers, unguided digital MBSH could also be considered as the first MBI step, with some users moving on to more intensive, as well as more effective, in-person courses. However, this does not dismiss the potential of unguided MBSH, given its scalability. We found that 36.8% (100/272) of Headspace participants showed a reliable improvement in stress over the course of the study compared with 24.1% (66/274) in the Moodzone arm (the NHS digital well-being offer at the time of recruitment). If this difference in reliable improvement were replicated across, for example, 10% of the 1.2 million NHS workforce, this would translate into >15,000 NHS workers showing a reliable improvement in stress if offered Headspace rather than Moodzone.

Conclusions

Unguided use of a digital MBSH intervention appears safe and is effective in reducing stress in health care workers compared with an active control condition, with improvements in self-compassion and formal intervention engagement explaining, at least in part, its beneficial effects. Effect sizes were small in comparison with in-person MBIs; however, unguided digital MBSH has the potential to be offered as part of a package of approaches to support health care workers' stress, mental health, and well-being. The findings support offering unguided MBSH as an addition to the ecosystem of evidence-based approaches to support health care workers' well-being, which offers choices and solutions at different levels of intensity and with different levels of guidance. Unguided MBSH must be contextualized within a supportive environment that promotes self-care at work [79]. Prioritizing the well-being and mental health of health care workers is critical, now more than ever, as we seek to find ways of supporting health care workers to live with the projected aftereffects of the COVID-19 pandemic.

Acknowledgments

The authors would like to extend their thanks to all the National Health Service staff who participated in this study. In addition, the authors are grateful to Jenny Gu and James Moon who supported recruitment to the study, and they would also like to thank the Clinical Research Networks and National Health Service Trusts who supported the study.

This study was supported by a doctoral studentship awarded to the first author jointly funded by Headspace and the Economic and Social Research Council (ES/J500173/1). Data analysis and writing of the manuscript were conducted independently from Headspace, and Headspace did not have any influence on the decision to publish the findings.

Authors' Contributions

HT contributed to the conceptualization, methodology, formal analysis, investigation, writing (original draft; review and editing) of the manuscript, and project administration. KC contributed to the conceptualization, methodology, writing (review and editing) of the manuscript, supervision, and funding acquisition. APF conducted the formal analysis and contributed to the writing (review and editing) of the manuscript. CS contributed to the conceptualization, methodology, formal analysis, writing (original draft; review and editing) of the manuscript, supervision, project administration, and funding acquisition.

Conflicts of Interest

HT was awarded a doctoral studentship for this work by Headspace. KC received research and consultancy funding from digital health care companies, including Headspace. CS is the Research Lead for Sussex Mindfulness Centre and has received research funding from the National Institute for Health Research and Headspace to evaluate mindfulness-based interventions.

Multimedia Appendix 1
Measures.
[DOCX File , 56 KB - mhealth_v10i8e31744_app1.docx]

Multimedia Appendix 2
Participant information, consent, and debriefing.
[PDF File (Adobe PDF File), 547 KB - mhealth_v10i8e31744_app2.pdf]

Multimedia Appendix 3
Handling missing data.
[DOCX File , 26 KB - mhealth_v10i8e31744_app3.docx]

Multimedia Appendix 4
Additional data analysis plan, results, and discussion.
[DOCX File , 88 KB - mhealth_v10i8e31744_app4.docx]

Multimedia Appendix 5
Bayes factors for assessing randomization success.
[DOCX File , 15 KB - mhealth_v10i8e31744_app5.docx]

Multimedia Appendix 6
Per-protocol findings.
[DOCX File , 25 KB - mhealth_v10i8e31744_app6.docx]

Multimedia Appendix 7
Reliable change tables.
[DOCX File , 17 KB - mhealth_v10i8e31744_app7.docx]

Multimedia Appendix 8
Lasting negative effects.
[DOCX File , 22 KB - mhealth_v10i8e31744_app8.docx]

Multimedia Appendix 9
Self-reported formal and informal intervention engagement.
[DOCX File , 21 KB - mhealth_v10i8e31744_app9.docx]

Multimedia Appendix 10
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 3820 KB - mhealth_v10i8e31744_app10.pdf]

References

1. Medisauskaite A, Kamau C. Prevalence of oncologists in distress: systematic review and meta-analysis. Psychooncology 2017 Nov 17;26(11):1732-1740. [doi: 10.1002/pon.4382] [Medline: 28116833]
2. Onigbogi C, Banerjee S. Prevalence of psychosocial stress and its risk factors among health-care workers in Nigeria: a systematic review and meta-analysis. Niger Med J 2019;60(5):238. [doi: 10.4103/nmj.nmj_67_19]
3. Gheshlagh R, Parizad N, Dalvand S, Zarei M, Farajzadeh M, Karami M, et al. The prevalence of job stress among nurses in Iran: a meta-analysis study. Nurs Midwifery Stud 2017;6(4):143. [doi: 10.4103/nms.nms_33_17]
4. The Maslach Burnout Inventory manual. In: Evaluating Stress: A Book of Resources. Minneapolis, MN: University of Minnesota; 1997.
5. MELCHIOR M, CASPI A, MILNE BJ, DANESE A, POULTON R, MOFFITT TE. Work stress precipitates depression and anxiety in young, working women and men. Psychol Med 2007 Apr 04;37(8):1119-1129. [doi: 10.1017/s0033291707000414]
6. O'Connor K, Muller Neff D, Pitman S. Burnout in mental health professionals: a systematic review and meta-analysis of prevalence and determinants. Eur Psychiatry 2018 Sep;53:74-99. [doi: 10.1016/j.eurpsy.2018.06.003] [Medline: 29957371]
7. Petrie K, Milligan-Saville J, Gayed A, Deady M, Phelps A, Dell L, et al. Prevalence of PTSD and common mental disorders amongst ambulance personnel: a systematic review and meta-analysis. Soc Psychiatry Psychiatr Epidemiol 2018 Sep;53(9):897-909. [doi: 10.1007/s00127-018-1539-5] [Medline: 29869691]
8. Rotenstein LS, Torre M, Ramos MA, Rosales RC, Guille C, Sen S, et al. Prevalence of burnout among physicians: a systematic review. JAMA 2018 Sep 18;320(11):1131-1150 [FREE Full text] [doi: 10.1001/jama.2018.12777] [Medline: 30326495]

9. Backé EM, Seidler A, Latza U, Rossnagel K, Schumann B. The role of psychosocial stress at work for the development of cardiovascular diseases: a systematic review. Int Arch Occup Environ Health 2012 Jan;85(1):67-79 [FREE Full text] [doi: 10.1007/s00420-011-0643-6] [Medline: 21584721]
10. Kelly S, Ismail M. Stress and type 2 diabetes: a review of how stress contributes to the development of type 2 diabetes. Annu Rev Public Health 2015 Mar 18;36:441-462. [doi: 10.1146/annurev-publhealth-031914-122921] [Medline: 25581145]
11. Thoits P. Stress and health: major findings and policy implications. J Health Soc Behav 2010;51 Suppl:S41-S53. [doi: 10.1177/0022146510383499] [Medline: 20943582]
12. NHS staff survey. NHS Digital. URL: https://www.nhsstaffsurveys.com/static/b3377ce95070ce69e84460fe210a55f0/ST21_National-briefing.pdf [accessed 2022-08-19]
13. Copeland A. NHS sickness absence: let's talk about mental health. The King's Fund. 2019 Oct 1. URL: https://www.kingsfund.org.uk/blog/2019/10/nhs-sickness-absence [accessed 2021-04-01]
14. Health care workers may be the nation's most stressed employees. Advisory Board. 2014 Feb 13. URL: https://www.advisory.com/daily-briefing/2014/02/13/health-care-workers-may-be-the-nations-most-stressed-employees [accessed 2021-06-15]
15. Hall L, Johnson J, Watt I, Tsipa A, O'Connor DB. Healthcare staff wellbeing, burnout, and patient safety: a systematic review. PLoS One 2016;11(7):e0159015 [FREE Full text] [doi: 10.1371/journal.pone.0159015] [Medline: 27391946]
16. Bohlken J, Schömig F, Lemke M, Pumberger M, Riedel-Heller S. [COVID-19 Pandemic: stress experience of healthcare workers - a short current review]. Psychiatr Prax 2020 May;47(4):190-197 [FREE Full text] [doi: 10.1055/a-1159-5551] [Medline: 32340048]
17. Lai J, Ma S, Wang Y, Cai Z, Hu J, Wei N, et al. Factors associated with mental health outcomes among health care workers exposed to coronavirus disease 2019. JAMA Netw Open 2020 Mar 02;3(3):e203976 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.3976] [Medline: 32202646]
18. Baer RA. Mindfulness training as a clinical intervention: a conceptual and empirical review. Clin Psychol Sci Pract 2003;10(2):125-143. [doi: 10.1093/clipsy.bpg015]
19. Kabat-Zinn J. Mindfulness-based interventions in context: past, present, and future. Clin Psychol Sci Pract 2003 11;10(2):144-156. [doi: 10.1093/clipsy.bpg016]
20. Mindfulness-Based Cognitive Therapy for Depression: A New Approach to Preventing Relapse. New York, United States: Guilford Press; 2012.
21. Kabat-Zinn J. Full Catastrophe Living; Using the Wisdom of Your Body and Mind to Face Stress, Pain and Illness. London: Piatkus; 1996.
22. Kuyken W, Warren FC, Taylor RS, Whalley B, Crane C, Bondolfi G, et al. Efficacy of mindfulness-based cognitive therapy in prevention of depressive relapse: an individual patient data meta-analysis from randomized trials. JAMA Psychiatry 2016 Jun 01;73(6):565-574 [FREE Full text] [doi: 10.1001/jamapsychiatry.2016.0076] [Medline: 27119968]
23. Goldberg S, Tucker R, Greene P, Davidson R, Wampold B, Kearney D, et al. Mindfulness-based interventions for psychiatric disorders: a systematic review and meta-analysis. Clin Psychol Rev 2018 Feb;59:52-60 [FREE Full text] [doi: 10.1016/j.cpr.2017.10.011] [Medline: 29126747]
24. Parsons C, Crane C, Parsons L, Fjorback L, Kuyken W. Home practice in mindfulness-based cognitive therapy and mindfulness-based stress reduction: a systematic review and meta-analysis of participants' mindfulness practice and its association with outcomes. Behav Res Ther 2017 Aug;95:29-41 [FREE Full text] [doi: 10.1016/j.brat.2017.05.004] [Medline: 28527330]
25. Gu J, Strauss C, Bond R, Cavanagh K. How do mindfulness-based cognitive therapy and mindfulness-based stress reduction improve mental health and wellbeing? A systematic review and meta-analysis of mediation studies. Clin Psychol Rev 2015 Apr;37:1-12. [doi: 10.1016/j.cpr.2015.01.006] [Medline: 25689576]
26. Chiesa A, Serretti A. Mindfulness-based stress reduction for stress management in healthy people: a review and meta-analysis. J Altern Complement Med 2009 May;15(5):593-600. [doi: 10.1089/acm.2008.0495] [Medline: 19432513]
27. Virgili M. Mindfulness-based interventions reduce psychological distress in working adults: a meta-analysis of intervention studies. Mindfulness 2013 Dec 13;6(2):326-337. [doi: 10.1007/s12671-013-0264-0]
28. Lomas T, Medina JC, Ivtzan I, Rupprecht S, Eiroa-Orosa FJ. A systematic review and meta-analysis of the impact of mindfulness-based interventions on the well-being of healthcare professionals. Mindfulness 2018 Nov 29;10(7):1193-1216. [doi: 10.1007/s12671-018-1062-5]
29. Spinelli C, Wisener M, Khoury B. Mindfulness training for healthcare professionals and trainees: a meta-analysis of randomized controlled trials. J Psychosom Res 2019 May;120:29-38. [doi: 10.1016/j.jpsychores.2019.03.003] [Medline: 30929705]
30. Strauss C, Gu J, Montero-Marin J, Whittington A, Chapman C, Kuyken W. Reducing stress and promoting well-being in healthcare workers using mindfulness-based cognitive therapy for life. Int J Clin Health Psychol 2021 May;21(2):100227 [FREE Full text] [doi: 10.1016/j.ijchp.2021.100227] [Medline: 33680001]
31. Rycroft-Malone J, Gradinger F, Owen Griffiths H, Anderson R, Crane R, Gibson A, et al. 'Mind the gaps': the accessibility and implementation of an effective depression relapse prevention programme in UK NHS services: learning from

mindfulness-based cognitive therapy through a mixed-methods study. BMJ Open 2019 Sep 08;9(9):e026244 [FREE Full text] [doi: 10.1136/bmjopen-2018-026244] [Medline: 31501097]
32. Mackenzie CS, Poulin PA, Seidman-Carlson R. A brief mindfulness-based stress reduction intervention for nurses and nurse aides. Appl Nurs Res 2006 May;19(2):105-109. [doi: 10.1016/j.apnr.2005.08.002] [Medline: 16728295]
33. Shapiro SL, Astin JA, Bishop SR, Cordova M. Mindfulness-based stress reduction for health care professionals: results from a randomized trial. Int J Stress Manag 2005;12(2):164-176. [doi: 10.1037/1072-5245.12.2.164]
34. Clement S, Schauman O, Graham T, Maggioni F, Evans-Lacko S, Bezborodovs N, et al. What is the impact of mental health-related stigma on help-seeking? A systematic review of quantitative and qualitative studies. Psychol Med 2014 Feb 21;45(1):11-27. [doi: 10.1017/s0033291714000129]
35. Cavanagh K, Strauss C, Forder L, Jones F. Can mindfulness and acceptance be learnt by self-help?: a systematic review and meta-analysis of mindfulness and acceptance-based self-help interventions. Clin Psychol Rev 2014 Mar;34(2):118-129. [doi: 10.1016/j.cpr.2014.01.001] [Medline: 24487343]
36. Spijkerman M, Pots W, Bohlmeijer E. Effectiveness of online mindfulness-based interventions in improving mental health: a review and meta-analysis of randomised controlled trials. Clin Psychol Rev 2016 Apr;45:102-114 [FREE Full text] [doi: 10.1016/j.cpr.2016.03.009] [Medline: 27111302]
37. Headspace homepage. Headspace. URL: https://headpsace.com/ [accessed 2017-02-13]
38. About Headspace. Headspace. URL: https://www.headspace.com/about-us [accessed 2022-08-18]
39. Lau N, O'Daffer A, Colt S, Yi-Frazier J, Palermo T, McCauley E, et al. Android and iPhone mobile apps for psychosocial wellness and stress management: systematic search in app stores and literature review. JMIR Mhealth Uhealth 2020 May 22;8(5):e17798 [FREE Full text] [doi: 10.2196/17798] [Medline: 32357125]
40. NHS: Moodzone. NHS. URL: https://www.mentallyhealthyschools.org.uk/resources/nhs-moodzone/ [accessed 2017-02-13]
41. Investigation of wellbeing interventions in NHS staff. ISRCTN. 2019 Mar 22. URL: https://www.isrctn.com/ISRCTN15424185 [accessed 2021-06-28]
42. Faul F, Erdfelder E, Lang A, Buchner A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods 2007 May;39(2):175-191 [FREE Full text] [doi: 10.3758/bf03193146] [Medline: 17695343]
43. Meditation and sleep made simple ? Headspace. URL: https://www.headspace.com/ [accessed 2022-05-09]
44. 10 stress busters. NHS. URL: https://www.nhs.uk/conditions/stress-anxiety-depression/reduce-stress [accessed 2022-05-09]
45. Donker T, Griffiths KM, Cuijpers P, Christensen H. Psychoeducation for depression, anxiety and psychological distress: a meta-analysis. BMC Med 2009 Dec 16;7:79 [FREE Full text] [doi: 10.1186/1741-7015-7-79] [Medline: 20015347]
46. Lovibond P, Lovibond S. The structure of negative emotional states: comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. Behav Res Ther 1995 Mar;33(3):335-343. [doi: 10.1016/0005-7967(94)00075-U]
47. Stewart-Brown S, Tennant A, Tennant R, Platt S, Parkinson J, Weich S. Internal construct validity of the Warwick-Edinburgh Mental Well-being Scale (WEMWBS): a Rasch analysis using data from the Scottish Health Education Population Survey. Health Qual Life Outcomes 2009 Feb 19;7(1):15 [FREE Full text] [doi: 10.1186/1477-7525-7-15] [Medline: 19228398]
48. Gu J, Strauss C, Crane C, Barnhofer T, Karl A, Cavanagh K, et al. Examining the factor structure of the 39-item and 15-item versions of the Five Facet Mindfulness Questionnaire before and after mindfulness-based cognitive therapy for people with recurrent depression. Psychol Assess 2016 Jul;28(7):791-802 [FREE Full text] [doi: 10.1037/pas0000263] [Medline: 27078186]
49. Raes F, Pommier E, Neff KD, Van Gucht D. Construction and factorial validation of a short form of the Self-Compassion Scale. Clin Psychol Psychother 2011 Jun 08;18(3):250-255. [doi: 10.1002/cpp.702] [Medline: 21584907]
50. Sprecher S, Fehr B. Compassionate love for close others and humanity. J Social Personal Relationship 2016 Jun 30;22(5):629-651. [doi: 10.1177/0265407505056439]
51. Meyer T, Miller ML, Metzger RL, Borkovec TD. Development and validation of the Penn State Worry Questionnaire. Behav Res Ther 1990;28(6):487-495. [doi: 10.1016/0005-7967(90)90135-6] [Medline: 2076086]
52. Nolen-Hoeksema S, Morrow J. A prospective study of depression and posttraumatic stress symptoms after a natural disaster: The 1989 Loma Prieta earthquake. J Personality Social Psychol 1991;61(1):115-121. [doi: 10.1037/0022-3514.61.1.115]
53. Adler N, Stewart J. The MacArthur scale of subjective social status. Psychosocial Research Notebook. 2007. URL: http://www.macses.ucsf.edu/research/psychosocial/subjective.php [accessed 2021-06-16]
54. Devilly GJ, Borkovec TD. Psychometric properties of the credibility/expectancy questionnaire. J Behav Ther Experimental Psychiatry 2000 Jun;31(2):73-86. [doi: 10.1016/s0005-7916(00)00012-4]
55. Clinical Trials Guide. National Institute for Health and Care Excellence. URL: https://www.nihr.ac.uk/documents/clinical-trials-guide/20595 [accessed 2021-06-16]
56. Crawford MJ, Thana L, Farquharson L, Palmer L, Hancock E, Bassett P, et al. Patient experience of negative effects of psychological treatment: results of a national survey†. Br J Psychiatry 2016 Mar 02;208(3):260-265. [doi: 10.1192/bjp.bp.114.162628] [Medline: 26932486]
57. Qualtrics XM homepage. Qualtrics XM. URL: https://www.qualtrics.com/uk/ [accessed 2017-02-13]
58. IBM SPSS Statistics for Windows, Version 25.0. New York: IBM Corp; 2017.

59. R: A Language and Environment for Statistical Computing. Vienna: R Foundation for Statistical Computing; 2020.
60. Lenth RV, Buerkner P, Herve M, Jung M, Love J, Miguez F, et al. Estimated Marginal Means, aka Least-Squares Means. CRAN. URL: https://cran.r-project.org/web/packages/emmeans/emmeans.pdf [accessed 2021-06-16]
61. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. J Stat Softw 2015;67(1):1-48. [doi: 10.18637/jss.v067.i01]
62. Buuren SV, Groothuis-Oudshoorn K. mice: multivariate imputation by chained equations in R. J Stat Softw 2011;45(3):1-67. [doi: 10.18637/jss.v045.i03]
63. Aust F, Barth M. Prepare APA Journal Articles with R Markdown. PAPAJA. URL: https://cran.r-project.org/web/packages/papaja/vignettes/papaja.pdf [accessed 2021-06-16]
64. Tidyverse easily install and load. Tidyverse. URL: https://tidyverse.tidyverse.org/ [accessed 2021-06-16]
65. van Buuren S, Groothuis-Oudshoorn K, Vink G, Schouten R, Robitzsch A, Rockenschaub P, et al. Package 'mice' Multivariate Imputation by Chained Equations. CRAN. URL: https://cran.r-project.org/web/packages/mice/mice.pdf [accessed 2021-06-15]
66. Magnusson K, Nilsson A, Carlbring P. Modeling longitudinal gambling data: challenges and opportunities. PsyArXiv 2019 [FREE Full text] [doi: 10.31234/osf.io/uvxk2]
67. Crane C, Crane RS, Eames C, Fennell MJ, Silverton S, Williams JM, et al. The effects of amount of home meditation practice in Mindfulness Based Cognitive Therapy on hazard of relapse to depression in the Staying Well after Depression Trial. Behav Res Ther 2014 Dec;63:17-24 [FREE Full text] [doi: 10.1016/j.brat.2014.08.015] [Medline: 25261599]
68. Jacobson NS, Truax P. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. J Consult Clin Psychol 1991;59(1):12-19. [doi: 10.1037/0022-006x.59.1.12]
69. Morey RD, Rouder JN, Jamil T, Urbanek S, Forner K, Ly A. BayesFactor: computation of bayes factors for common designs. rdrr. URL: https://rdrr.io/cran/BayesFactor/ [accessed 2022-02-27]
70. Rouder JN, Speckman PL, Sun D, Morey RD, Iverson G. Bayesian t tests for accepting and rejecting the null hypothesis. Psychon Bull Rev 2009 Apr;16(2):225-237. [doi: 10.3758/PBR.16.2.225] [Medline: 19293088]
71. Gunel E, Dickey J. Bayes factors for independence in contingency tables. Biometrika 1974 Dec;61(3):545. [doi: 10.2307/2334738]
72. Jamil T, Ly A, Morey RD, Love J, Marsman M, Wagenmakers E. Default "Gunel and Dickey" Bayes factors for contingency tables. Behav Res Methods 2017 Apr 20;49(2):638-652 [FREE Full text] [doi: 10.3758/s13428-016-0739-8] [Medline: 27325166]
73. Taylor H, Strauss C, Cavanagh K. Can a little bit of mindfulness do you good? A systematic review and meta-analyses of unguided mindfulness-based self-help interventions. Clin Psychol Rev 2021 Nov;89:102078. [doi: 10.1016/j.cpr.2021.102078] [Medline: 34537665]
74. Mindful Nation UK. Mindfulness All-Party Parliamentary Group (MAPPG). 2015 Oct. URL: https://mindfulnessinschools.org/wp-content/uploads/2017/09/Mindfulness-APPG-Report_Mindful-Nation-UK_Oct2015-1.pdf [accessed 2021-06-15]
75. Bakker AB, Demerouti E. The Job Demands ‐ Resources model: state of the art. J Managerial Psych 2007 Apr 03;22(3):309-328. [doi: 10.1108/02683940710733115]
76. Weber S, Lorenz C, Hemmings N. Improving stress and positive mental health at work via an app-based intervention: a large-scale multi-center randomized control trial. Front Psychol 2019 Dec 6;10:2745 [FREE Full text] [doi: 10.3389/fpsyg.2019.02745] [Medline: 31866915]
77. Heber E, Ebert D, Lehr D, Cuijpers P, Berking M, Nobis S, et al. The benefit of web- and computer-based interventions for stress: a systematic review and meta-analysis. J Med Internet Res 2017 Feb 17;19(2):e32 [FREE Full text] [doi: 10.2196/jmir.5774] [Medline: 28213341]
78. Carolan S, Harris P, Cavanagh K. Improving employee well-being and effectiveness: systematic review and meta-analysis of web-based psychological interventions delivered in the workplace. J Med Internet Res 2017 Jul 26;19(7):e271 [FREE Full text] [doi: 10.2196/jmir.7583] [Medline: 28747293]
79. Micklitz K, Wong G, Howick J. Mindfulness-based programmes to reduce stress and enhance well-being at work: a realist review. BMJ Open 2021 Mar 19;11(3):e043525 [FREE Full text] [doi: 10.1136/bmjopen-2020-043525] [Medline: 33741667]
80. Borghouts J, Eikey E, Mark G, De Leon C, Schueller SM, Schneider M, et al. Barriers to and facilitators of user engagement with digital mental health interventions: systematic review. J Med Internet Res 2021 Mar 24;23(3):e24387 [FREE Full text] [doi: 10.2196/24387] [Medline: 33759801]
81. Tambling RB. A literature review of therapeutic expectancy effects. Contemp Fam Ther 2012 Jun 20;34(3):402-415. [doi: 10.1007/s10591-012-9201-y]
82. Torous J, Firth J. The digital placebo effect: mobile mental health meets clinical psychiatry. Lancet Psychiatry 2016 Feb;3(2):100-102. [doi: 10.1016/S2215-0366(15)00565-9]
83. Torous J, Andersson G, Bertagnoli A, Christensen H, Cuijpers P, Firth J, et al. Towards a consensus around standards for smartphone apps and digital mental health. World Psychiatry 2019 Feb;18(1):97-98 [FREE Full text] [doi: 10.1002/wps.20592] [Medline: 30600619]
84. Women in the NHS homepage. Women in the NHS. URL: https://www.womeninthenhs.co.uk/ [accessed 2021-06-16]
85. Ethnicity facts and figures. GOV.UK. URL: https://www.ethnicity-facts-figures.service.gov.uk/ [accessed 2021-06-16]

Abbreviations

CBT: cognitive behavioral therapy
CONSORT: Consolidated Standards of Reporting Trials
DASS-21: 21-item Depression, Anxiety, and Stress Scale
ITT: intention-to-treat
MBCT: mindfulness-based cognitive therapy
MBCT-L: mindfulness-based cognitive therapy for Life
MBI: mindfulness-based intervention
MBSH: mindfulness-based self-help
MBSR: mindfulness-based stress reduction
NHS: National Health Service
RCT: randomized controlled trial
T1: time point 1
T2: time point 2
T3: time point 3

Edited by L Buis; submitted 02.07.21; peer-reviewed by S Goldberg, H Hsin; comments to author 28.09.21; revised version received 27.02.22; accepted 08.06.22; published 25.08.22.

Please cite as:
Taylor H, Cavanagh K, Field AP, Strauss C
Health Care Workers' Need for Headspace: Findings From a Multisite Definitive Randomized Controlled Trial of an Unguided Digital Mindfulness-Based Self-help App to Reduce Healthcare Worker Stress
JMIR Mhealth Uhealth 2022;10(8):e31744
URL: https://mhealth.jmir.org/2022/8/e31744
doi:10.2196/31744
PMID:36006668

Original Paper

System Usability Scale Benchmarking for Digital Health Apps: Meta-analysis

Maciej Hyzy[1,2], BSc; Raymond Bond[1], BSc, PhD; Maurice Mulvenna[1], BSc, MPhil, PhD; Lu Bai[1], BEng, PhD; Alan Dix[3], BA, DPhil; Simon Leigh[2,4], BSc, MSc, PhD; Sophie Hunt[2], BSc

[1]School of Computing, Ulster University, Newtownabbey, United Kingdom

[2]Organisation for the Review of Care and Health Applications, Daresbury, United Kingdom

[3]Computational Foundry, Swansea University, Swansea, United Kingdom

[4]Institute of Digital Healthcare, University of Warwick, Coventry, United Kingdom

Corresponding Author:
Maciej Hyzy, BSc
School of Computing
Ulster University
Ulster University School Office
16G24 Shore Road
Newtownabbey, BT37 0QB
United Kingdom
Phone: 44 7526852505
Email: maciejmarekzych@gmail.com

Abstract

Background: The System Usability Scale (SUS) is a widely used scale that has been used to quantify the usability of many software and hardware products. However, the SUS was not specifically designed to evaluate mobile apps, or in particular digital health apps (DHAs).

Objective: The aim of this study was to examine whether the widely used SUS distribution for benchmarking (mean 68, SD 12.5) can be used to reliably assess the usability of DHAs.

Methods: A search of the literature was performed using the ACM Digital Library, IEEE Xplore, CORE, PubMed, and Google Scholar databases to identify SUS scores related to the usability of DHAs for meta-analysis. This study included papers that published the SUS scores of the evaluated DHAs from 2011 to 2021 to get a 10-year representation. In total, 117 SUS scores for 114 DHAs were identified. R Studio and the R programming language were used to model the DHA SUS distribution, with a 1-sample, 2-tailed *t* test used to compare this distribution with the standard SUS distribution.

Results: The mean SUS score when all the collected apps were included was 76.64 (SD 15.12); however, this distribution exhibited asymmetrical skewness (–0.52) and was not normally distributed according to Shapiro-Wilk test (P=.002). The mean SUS score for "physical activity" apps was 83.28 (SD 12.39) and drove the skewness. Hence, the mean SUS score for all collected apps excluding "physical activity" apps was 68.05 (SD 14.05). A 1-sample, 2-tailed *t* test indicated that this health app SUS distribution was not statistically significantly different from the standard SUS distribution (P=.98).

Conclusions: This study concludes that the SUS and the widely accepted benchmark of a mean SUS score of 68 (SD 12.5) are suitable for evaluating the usability of DHAs. We speculate as to why physical activity apps received higher SUS scores than expected. A template for reporting mean SUS scores to facilitate meta-analysis is proposed, together with future work that could be done to further examine the SUS benchmark scores for DHAs.

(JMIR Mhealth Uhealth 2022;10(8):e37290) doi:10.2196/37290

KEYWORDS

mHealth SUS scores meta-analysis; SUS for digital health; digital health apps usability; mHealth usability; SUS meta-analysis; mHealth; mobile app; mobile health; digital health; System Usability Scale

Introduction

According to Nielsen [1], "usability is a quality attribute that assesses how easy user interfaces are to use. The word 'usability' also refers to methods for improving ease-of-use during the design process." In Nielsen's [1] model, usability consists of a number of components, including the system's learnability, efficiency, memorability, errors, and satisfaction.

According to the International Organization for Standardization, "usability is the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use" [2].

The public is increasingly searching for digital health apps (DHAs) in app stores to help self-manage their health and well-being [3]. With the uptake of DHAs, national health care organizations such as the National Health Service in the United Kingdom are offering curated access to health care apps as part of social prescription and related services [4].

The usability of DHAs is important as inferior usability could negatively impact the adoption of such technologies, and potentially, their users' health [5]. For example, a study conducted in 2019 found that self-management DHAs with higher rated usability (rated based on heuristic usability testing) lead to increased exercise engagement and quality of life in patients with breast cancer [6]. Reliably measuring the usability of DHAs can be used to distinguish between usable and less usable DHAs and help identify DHAs that may require improved usability.

The System Usability Scale (SUS), commonly described as a "quick and dirty" way of measuring usability, is a short 10-item questionnaire (each question with a Likert scale ranging from strongly agree to strongly disagree) designed to measure the usability of a system [7]. The SUS is a well-designed, balanced survey consisting of 5 questions with positive statements and 5 questions with negative statements, with scores ranging from 0 to 100. The current literature suggests that a score of 68 is a useful benchmark (mean SUS score), where 50% of apps fall below and above it [8]. Sauro and Lewis [8] discuss using data from 446 studies and 5000 individual SUS responses that indicate a mean SUS score of 68 (SD 12.5) [8]. Hence, the standard normal SUS distribution is said to be 68 (SD 12.5).

The SUS has become a common method for measuring the usability for different digital products or systems (including DHAs) since its development in 1986 [9]. According to a scoping review from 2019 [10], SUS was the most frequently used questionnaire for evaluating the usability of DHAs. However, the normal SUS distribution evaluated by Sauro and Lewis [8] (68 SD 12.5) was not likely representative of SUS scores achieved by mobile apps or DHAs.

The mHealth App Usability Questionnaire (MAUQ) is a validated alternative to SUS for measuring usability that is tailored to mobile health (mHealth) apps [10]. Although MAUQ may be more suitable for measuring the usability of DHAs, it is a relatively new scale developed in 2019. SUS has been used to evaluate DHAs since their inception; however, it remains to be seen whether the mean 68 (SD 12.5) benchmarking distribution represents the SUS scores achieved by DHAs.

The aim of this study was to determine if the widely accepted benchmark and SUS distribution of mean 68 (SD 12.5) is reliable for evaluating the usability of DHAs. This work is important given that the SUS benchmarking distribution that is being used is assumed to represent the usability of DHAs even though this standard SUS distribution was developed based on the usability of systems more generally (well beyond the genre of DHAs). Given that SUS is a frequently used tool for measuring the usability of DHAs, this study is needed to reassure researchers if the mean 68 (SD 12.5) distribution benchmark is reliable when evaluating DHAs using SUS and discover if a different SUS benchmark should be used for different genres of DHAs. To determine these findings, a comparison of published SUS scores from evaluated DHAs with the standard SUS distribution was conducted.

Methods

SUS Score

A SUS score is computed using the 10 Likert ratings that is typically completed by a user after having been exposed to the system for a period of time. The process for computing a SUS score is as follows:

1. Subtract 1 from the user's Likert ratings for odd-numbered items or questions.
2. Subtract the user's Likert ratings from 5 for even-numbered items.
3. Each item score will range from 0 to 4.
4. Sum the numbers and multiply the total by 2.5.
5. This calculation will provide a range of possible SUS scores from 0 to 100 [7].

Data Collection

Table 1 provides the criteria and search strategy for selecting the research papers that were used to conduct the meta-analysis on SUS scores. In this study, we aimed to collect papers that published the SUS scores of the evaluated DHAs after 2011. This criterion allowed us to curate a relatively "modern" set of SUS scores from DHA evaluations with a 10-year representation. A total of 114 DHAs producing 117 SUS scores were collected to conduct this meta-analysis.

Table 2 provides the number of papers and SUS scores that were used in this study to populate a DHA SUS data set.

Table 1. Population, Intervention, Comparator, Outcome, and Study Design framework for the data collection of digital health app (DHA) System Usability Scale (SUS) scores.

Frame	Inclusion criterion	Exclusion criterion
Population	Members of the general population—globally	Developers or designers of DHA that conducted SUS on their own product
Intervention	DHA	Not a DHA and research papers published before 2011
Comparator	N/A[a]	N/A
Outcome	SUS score or mean SUS score for DHA	SUS score not conducted by end users
Study design	The data set of SUS scores for measuring the usability of DHAs was collected using 5 search engines: ACM Digital Library, IEEE Xplore, CORE, PubMed, and Google Scholar. The keywords and queries used in the search included: “health app SUS,” “mhealth SUS,” “digital health apps SUS,” “mobile health SUS,” “mhealth apps usability,” and “mental health apps SUS.”	N/A

[a]N/A: not applicable.

Table 2. Number of papers and System Usability Scale (SUS) scores per year.

Year	Paper (N=19), n (%)	SUS score (N=117), n (%)
2014	2 (11)	14 (12)
2015	2 (11)	2 (1.7)
2016	2 (11)	3 (2.6)
2017	1 (5)	2 (1.7)
2018	3 (16)	71 (60.1)
2019	3 (16)	9 (7.7)
2020	3 (16)	12 (10.2)
2021	3 (16)	4 (3.4)

Study Screening

The research papers included in this study were screened by title and abstract. If the research paper included a SUS score for a DHA and the SUS evaluation was conducted by end users, it was included in this study.

Risk of Bias

SUS is a simple method of measuring the usability of hardware and software that should be conducted by end users. When conducting this study, the exclusion criterion was set to not include SUS evaluation scores that were provided by the developers or designers of the DHA, due to potential bias. However, none of the SUS scores collected met that exclusion criterion.

There may also be a bias if there are more SUS scores published for DHAs of a particular genre, or there could be a publication bias, as researchers are more likely to publish studies that achieved “good” (above the 68 benchmark) SUS scores. This is related to the file drawer effect [11], where researchers withhold studies that show nonsignificant or negative results (P>.05). Literature indicates that about 95% of studies in the file drawer contain nonsignificant results, whereas journals contain a disproportionate number of studies with type 1 errors. When there are more SUS scores published for DHAs of a particular genre, they could be overrepresented in a general health app SUS distribution and perhaps skew the distribution. This bias could be avoided by conducting this study on a data set where the different genres of DHAs are balanced. Publication bias could be countered by collecting new data sets where end users complete SUSs when viewing a large random sample of DHAs.

SUS has been developed in English to be used by English-speaking users. Using SUS with non-English speakers requires a new version of SUS that needs to be adapted and validated. Otherwise, there could be language and cultural bias in the assessment. Cross-cultural adaptation guidelines [12] could be used to adapt SUS; previously, these guidelines have been used to develop the Indonesian version of SUS [13]. Moreover, a study conducted in 2020 examined the Arabic, Chinese, French, German, and Spanish versions of the SUS [14]. The study found that these SUS versions were adequately adapted; however, cultural differences had to be highlighted [14]. Furthermore, the different devices and genres of DHAs may need their own, more specific SUS benchmarks.

Data Extraction

The study-specific data that were extracted from the research papers included first author’s name, DHA’s focused health area, DHA’s name, device that the DHA was used on, platform the

DHA is available on, sample size used to calculate the mean DHA SUS score, year the research paper was published in, and DHA SUS score.

Data Analysis

The data were separated into 3 subsets: (1) a SUS distribution including all DHAs, (2) a SUS distribution with only SUS scores from physical activity apps, and (3) a SUS distribution including all apps except physical activity apps. This separation was done due to the large frequency of physical activity apps that are present in the data set and the high mean of these apps (83.28, SD 12.39), which dominated the shape of the probability distribution.

R statistical software (version 4.0.3; R Foundation for Statistical Computing) was used to conduct the meta-analysis, compute statistics, and produce graphs. Shapiro-Wilk normality tests were used to test whether the SUS distributions were normally distributed (where $P<.05$ denotes that the distribution is not normal). Skewness and kurtosis were computed to determine how symmetrical (or unsymmetrical) and heavy- or light-tailed the data distributions are. The data were also visually explored using density plots, histograms, and boxplots to interrogate the distribution of SUS scores.

Wilcoxon signed rank tests and 1-sample, 2 tailed *t* tests were used to compare the mean SUS scores of DHAs with the widely accepted SUS distribution (mean 68, SD 12.5) that is typically used for benchmarking usability. *p* values <.05 were considered statistically significant in this study.

Results

Table 3 provides the mean, SD, and frequency of DHAs for each category. The "physical activity" category mainly included fitness apps. The "health care" category included DHAs that help with self-managing health and well-being, including living with and the treatment of obesity, allergies, suicide prevention, depression, and smoking cessation. The category "first aid, CPR, and choking" mainly included DHAs that assist with first aid and cardiopulmonary resuscitation. The category "diet, food, and nutrition" included diet apps and food and nutrition apps. The category "health information" included DHAs that provide health-related information and educational content. See Multimedia Appendix 1 [5,15-32] for more information.

Table 4 provides a summary of the characteristics of the 3 SUS distributions: (1) a SUS distribution from all categories of DHAs, (2) a SUS distribution from physical activity apps only, and (3) a SUS distribution from all categories excluding the physical activity apps. It is clear that the SUS distributions from all DHAs and the SUS distribution from physical activity apps only are not normally distributed. However, the distribution of SUS scores from all DHAs excluding physical activity apps is more akin to a normal distribution. The participant sample sizes used to collect the SUS scores have distribution of 6 (SD 6.16; range 2-31). See Multimedia Appendix 1 for the sample size of each SUS score collected.

Table 5 provides a summary of the 1-sample, 2-tailed *t* tests. The table indicates that the SUS distribution from all DHAs and the SUS distribution from physical activity apps only are statistically different distributions compared to the accepted mean 68 (SD 12.5) SUS distribution ($P=.002$). However, when excluding physical activity apps, the 1-sample, 2-tailed *t* test suggests that the distribution is comparable to the standard SUS distribution of mean 68 (SD 12.5).

Table 3. Category and frequency of apps included in this study.

Category	App (N=117), n (%)	SUS[a] score, mean (SD)
Physical activity	66 (56.4)	83.28 (12.39)
Health care	25 (21.4)	71.30 (12.72)
First aid, CPR[b], and choking	16 (13.7)	61.29 (15.08)
Diet, food, and nutrition	8 (6.8)	71.06 (14.55)
Health information	2 (1.7)	69.45 (5.30)

[a]SUS: System Usability Scale.

[b]CPR: cardiopulmonary resuscitation.

Table 4. Characteristics of System Usability Scale (SUS) probability distributions for the 3 categories.

Characteristic	SUS scores from all categories	SUS scores from all categories excluding physical activity apps	SUS scores from physical activity apps only
P value (Shapiro-Wilk)	.002	.24	.001
Mean (SD)	76.64 (15.12)	68.05 (14.05)	83.28 (12.39)
Median	78.75	68.30	86.00
Skewness	–0.52	–0.39	–0.69
Kurtosis	2.67	2.74	2.55
Standard error	1.4	1.97	1.53

Table 5. Results from hypothesis test.

Hypothesis, test	*P* value	95% CI
All categories versus standard SUS[a] distribution		
1-sample, 2-tailed *t* test	<.001	73.87-79.41
Wilcoxon signed rank test with continuity correction	<.001	74.50-80.00
All categories excluding physical activity apps versus standard SUS distribution		
1-sample, 2-tailed *t* test	.98	64.10-72.00
Wilcoxon signed rank test with continuity correction	.86	64.30-72.60
Physical activity apps only versus standard SUS distribution		
1-sample, 2-tailed *t* test	<.001	80.23-86.33
Wilcoxon signed rank test with continuity correction	<.001	80.50-87.50

[a]SUS: System Usability Scale.

The graphs in Figure 1 show that there is an unexpected peak in SUS scores for the range of 80-90, and the frequency in this range is greater than that for the range of 60-70. Table 3 shows the frequency of SUS scores for each category and indicates that the physical activity category has the highest frequency, which could be responsible for the peak in the 80-90 SUS score range/bin.

Figure 1 visually demonstrates that the SUS distribution for all DHAs is asymmetrical. For example, when all categories are included, the cumulative distribution function indicates that there is a 28.39% probability that the SUS score will be 68 or less, whereas the accepted standard probability is 50% that the SUS score will be 68 or less [8]. Figure 2 indicates that physical activity apps are responsible for the second "peak" in Figure 2A and B. The mean of 83.28 is much greater than the expected mean of 68. The SUS scores for physical activity apps could be inflated or that these apps typically have a greater degree of usability, which would need to be determined by conducting further studies. Figure 2 shows that there is a probability of 10.88% that the SUS score in the category of physical activity will be 68 or less, indicating that this distribution is very different compared to the expected SUS distribution of mean 68 (SD 12.5). Figure 3 shows that the mean and median are both very close to 68 after removing SUS scores from physical activity apps. This finding helps confirm that the SUS score distribution of DHAs is similar to that of the accepted standard SUS distribution. When using this distribution, Figure 3D shows that there is a probability of 49.85% that the SUS score will be 68 or less, making it very similar to the standard.

Figure 1. Analysis of SUS distribution for all categories of digital health apps: A) histogram of SUS scores, B) density plot of SUS scores, C) boxplot of SUS scores, and D) normal curve probabilities of SUS scores for all categories (mean 76.64, SD 15.12; shaded area: 0.2839). Blue line=68 (average SUS score for apps), red line=78.75 (median), orange line=76.64 (mean). SUS: System Usability Score.

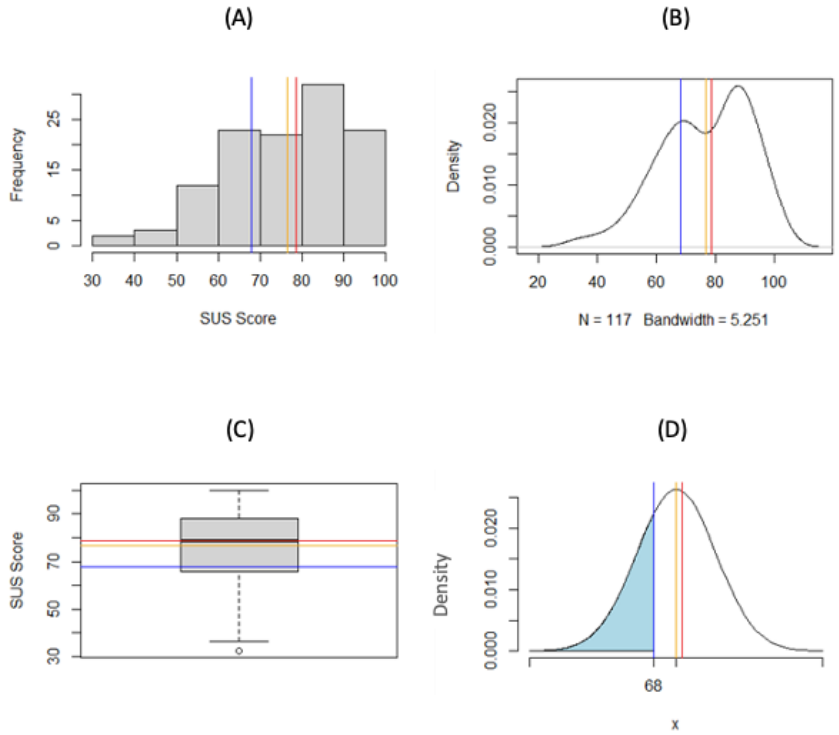


Figure 2. Analysis of SUS distribution for physical activity apps only: A) histogram of SUS scores, B) density plot of SUS scores, C) boxplot of SUS scores, and D) normal curve probabilities of SUS scores for all categories (mean 83.28, SD 12.39; shaded area: 0.1088). Blue line=68 (average SUS score for apps), red line=86 (median), orange line=83.28 (mean). SUS: System Usability Score.

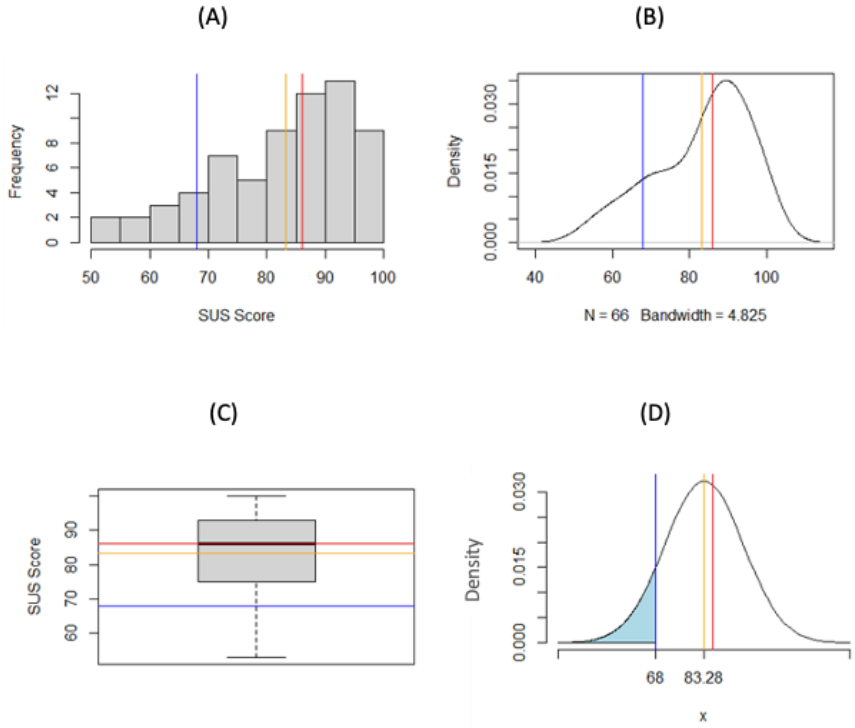
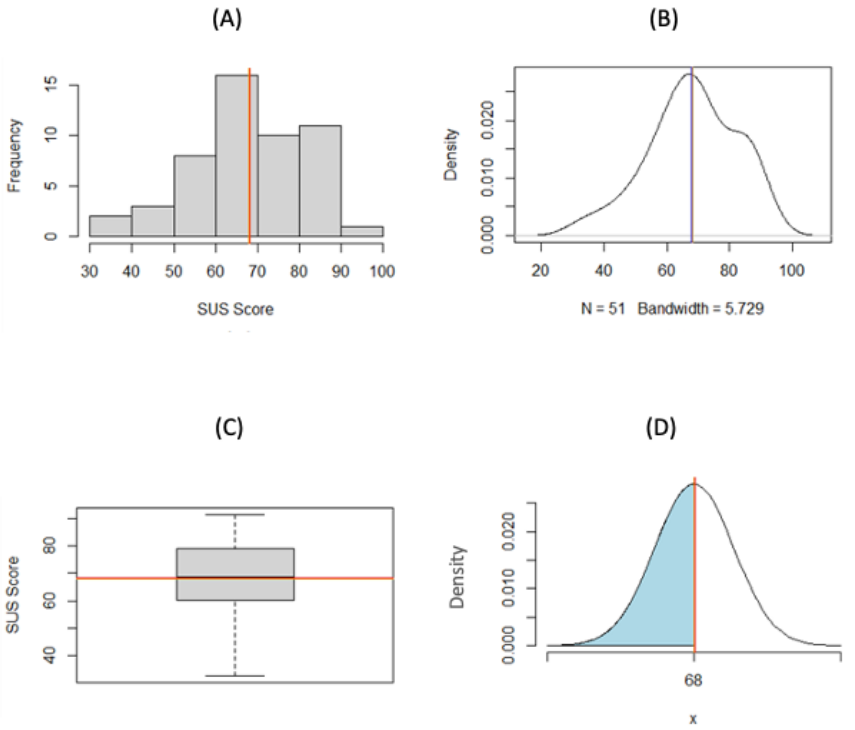


Figure 3. Analysis of SUS distribution for all categories excluding physical activity apps: A) histogram of SUS scores, B) density plot of SUS scores, C) boxplot of SUS scores, and D) normal curve probabilities of SUS scores for all categories (mean 68.05, SD 14.05; shaded area: 0.4985). Blue line=68 (average SUS score for apps), red line=68.30 (median), orange line=68.05 (mean). SUS: System Usability Score.



Discussion

Principal Findings

The data set used for this study contained 117 SUS scores collected from 114 DHAs (some apps were assessed by different end users, such as clinicians, researchers, or participants, that gave them different SUS scores, which were included in this study). The SUS mean when all of the apps are included is 76.64; however, this mean score lies between 2 peaks, as seen in Figure 1B. Thus, this mean may not be suitable for benchmarking DHAs. In Figure 1B, the blue line indicates the mean SUS score of 68 when all SUS scores are included in the distribution, which is exactly in line with the first peak in the distribution. This finding indicates that many of the DHAs follow a similar SUS distribution to that in the expected standard.

When investigating the results in Figure 1, we explored the cause of the second peak in Figure 1B. Hence, due to frequency of physical activity apps (66 DHAs) in the data set and the mean of 83.28 (SD 12.39; Table 3), a distribution of only physical activity apps was examined (Figure 2). We discovered that the second peak in Figure 1B was driven by the SUS scores of physical activity apps.

When the SUS scores of physical activity apps are excluded from the data set, the SUS score distribution for DHAs become normally distributed (mean 68.05, SD 14.05) and is similar to the widely used SUS distribution (mean 68, SD 12.5). Although the SUS distribution of DHAs have a slightly greater SD (14.05 vs 12.5), this finding could be due to the small sample size in this study. The results indicate that the standard SUS score benchmark of 68 can be used when evaluating DHAs. This assumption was important to test given that the accepted distribution of mean 68 (SD 12.5) was not primarily based on SUS scores from mobile apps, or in particular DHAs. The usability of systems may generally improve over time, which could change the average SUS score that would be achieved by digital systems. Moreover, given that DHAs can be critically important apps to users (nonrecreational or nonhedonic), their usability could be greater, hence achieving higher SUS scores.

The paper that published the SUS scores of these 65 physical activity apps focused on the most popular apps available to conduct their SUS evaluation, which could indicate that more popular apps are perhaps more usable. Further research is needed to determine if there is a link between app popularity and the usability of DHAs. Other possibilities are inflated SUS scores, popularity in the market [33] leading to better usability, and greater budgets to invest into usability. More familiar design has been shown to influence usability, as stated by Jakob's law: "users spend most of their time on other sites. This means that users prefer your site to work the same way as all the other sites they already know" [34].

Developers of physical activity apps appear to be investing a lot into usability. For example, to encourage physical activity for those with low socioeconomic status and youths, the prototyping for a smartphone user-centric framework for developing game-based physical activity apps has been created [35]. A study from 2017, where the top 50 health and fitness apps were downloaded from the Apple app store, found that physical activity and weight loss apps most frequently (97%) used gamification [36]. Gamification has been shown to improve the use of physical activity apps [37], which could explain the higher-than-expected usability of physical activity apps and indicates that a different benchmark may need to be used when dealing with physical activity apps.

Set of Guidelines for Presenting SUS Analysis to Facilitate High-Quality Meta-analyses

When conducting the meta-analysis for this paper, we encountered a couple of problems when gathering the SUS scores from research papers. Some papers used the word "expert" when stating the sample size of reviewers who used SUS to assess a DHA. It was unclear as to whether the word "expert" referred to an expert usability reviewer or expert in the health area for which a DHA has been developed. Clearly stating who the reviewer is would be useful when conducting a rigorous meta-analysis for SUS.

Textbox 1 recommends a standard template for reporting SUS analysis and scores that could be helpful when presenting an SUS analysis to facilitate high-quality meta-analyses.

Textbox 1. Recommended template for reporting mean System Usability Scale (SUS) scores to facilitate meta-analyses.

Participants

- Novice users (those with no experience in using the system being assessed)
- Expert users (those who already have experience in using the system)
- Expert user-experience evaluators
- Representative users (those who are likely to use the app; eg, recruiting doctors when testing a medical system) and nonrepresentative users (anyone outside the domain of interest; eg, recruiting any person to test the usability of a fitness app)

Context

- include information such as a usability testing session with prescribed tasks, a usability testing session without prescribed tasks, SUS scores collected after a trial (lasting n days, weeks, or months), or other details (eg, remote usability test and lab-based or in-situ [eg, workplace or "in the wild"])

Sample size (n)

Mean (SD) score (rounded to 2 decimal places)

Median score (min/max; rounded to 2 decimal places)

Standard error of the mean (rounded to 2 decimal places)

95% CI (lower to upper)

Test (eg, 1-sample, 2-tailed *t* test)

SUS grade (A-F)

Related and Future Works

Although this study assessed SUS for evaluating DHAs, there are other scales that could be used, which includes the previously mentioned MAUQ. Currently, there are 4 versions of the MAUQ, 2 for stand-alone apps (provider and patient versions) and 2 for interactive mHealth apps (provider and patient versions). The SUS and MAUQ are correlated, but the correlation is not strong (r=0.6425) [38].

A systematic literature review [39] evaluated the methodologies of usability analyses, domains of usability being assessed, and results of usability analyses. The paper concluded that out of the 3 usability domains in MAUQ, only satisfaction is regularly assessed. A similar meta-analysis to the one conducted in this study could be done with the MAUQ.

The usability of DHAs can be improved; in the study by Liew et al [40], researchers provided insight and suggestions for improving the usability of health and wellness mobile apps. The paper concluded that better connectivity between mHealth suppliers and users will have a positive outcome for the mHealth app ecosystem and increase the uptake of mHealth apps.

Improving usability is important as the lack of it can slow down the adoption of DHAs. Islam et al [5] investigated the usability of mHealth apps in Bangladesh using a heuristic evaluation and the SUS. The paper concluded that the usability of DHAs in Bangladesh is not satisfactory and could be a barrier for the wider adoption of DHAs.

As the SUS scores for physical activity apps were higher than other apps in this study, future work is needed to explore how these scores could be inflated or whether these apps have a greater degree of usability.

The study conducted in this paper could be expanded in the following ways. Future studies could be done by comparing the SUS scores evaluated by experts and nonexperts. The meta-analysis conducted here could be repeated on a bigger data set. A SUS meta-analysis could be conducted for a wide range of health app categories to validate if all follow the standard SUS distribution (mean 68, SD 12.5). A study with randomly selected apps could be conducted with several recruited end users completing the SUS questionnaire that would allow for a more unbiased distribution of SUS scores.

The paper with 65 physical activity apps [15] focused specifically on the most popular apps. Research could be done to determine if there is a link between popularity and the usability of DHAs when using the SUS or MAUQ.

Limitations

This study has a few limitations. This meta-analysis collected SUS results from 19 papers—some of which used a mean SUS score resulting from as few as 2 or 3 reviewers. Some of the reviewers could have been "generous" when filling the SUS questionnaire, resulting in inflated SUS scores. The data set used for this study is small (SUS scores: n=117). Moreover, 65 of the physical activity apps used in this study came from the same paper [15]. This paper used 2 reviewers when evaluating each of the apps. A speculation can be made that since 65 physical activity apps were being evaluated, it is possible that the reviewers had limited time to spend on each of the app evaluations, although no information is provided to support this.

This study was conducted in 2021, and some of the apps may have been updated. Various changes to the design could have been made since their SUS score was evaluated, and thus, the SUS score may no longer be applicable to the app.

Conclusion

The aim of this study was to conduct a meta-analysis to determine if the standard SUS distribution (mean 68, SD 12.5) for benchmarking is applicable to evaluating DHAs. This study compared the standard SUS score distribution to the distribution for different categories of DHAs. The data for this study were collected from different research papers that were found using different search engines or research repositories. This study indicates that the SUS distribution of DHAs (when excluding physical activity apps) is similar to the widely used SUS distribution. This work implies that the SUS and existing benchmarking approaches could be used to evaluate DHAs and that the SUS could be used by health care departments and organizations such as the National Health Service or Organisation for the Review of Care and Health Applications to validate and assure the quality of DHAs in terms of their usability. Readers of this work may also choose to use our SUS distribution (mean 68.05, SD 14.05) for benchmarking the SUS scores of DHAs.

Acknowledgments

This study was conducted as part of a doctoral Co-operative Awards in Science and Technology award, with funding from the Department for the Economy in Northern Ireland and the Organisation for the Review of Care and Health Applications in the United Kingdom.

Conflicts of Interest

None declared

Multimedia Appendix 1
Collected System Usability Scale scores.
[DOCX File , 38 KB - mhealth_v10i8e37290_app1.docx]

References

1. Nielsen J. Usability 101: introduction to usability. Nielson Norman Group. 2012 Jan 03. URL: https://www.nngroup.com/articles/usability-101-introduction-to-usability/ [accessed 2021-11-21]
2. ISO/TS 20282-2:2013(en) usability of consumer products and products for public use — part 2: summative test method. International Organization for Standardization. 2013. URL: https://www.iso.org/obp/ui/#iso:std:iso:ts:20282:-2:ed-2:v1:en [accessed 2021-12-17]
3. Leigh S, Daly R, Stevens S, Lapajne L, Clayton C, Andrews T, et al. Web-based internet searches for digital health products in the United Kingdom before and during the COVID-19 pandemic: a time-series analysis using app libraries from the Organisation for the Review of Care and Health Applications (ORCHA). BMJ Open 2021 Oct 11;11(10):e053891 [FREE Full text] [doi: 10.1136/bmjopen-2021-053891] [Medline: 34635531]
4. Apps and tools for patient care. National Health Service. URL: https://www.nhsx.nhs.uk/key-tools-and-info/apps-and-tools-patient-care/ [accessed 2022-06-13]
5. Islam MN, Karim MM, Inan TT, Islam AKMN. Investigating usability of mobile health applications in Bangladesh. BMC Med Inform Decis Mak 2020 Feb 03;20(1):19 [FREE Full text] [doi: 10.1186/s12911-020-1033-3] [Medline: 32013965]
6. Ebnali M, Shah M, Mazloumi A. How mHealth apps with higher usability effects on patients with breast cancer? 2019 Sep 15 Presented at: 2019 International Symposium on Human Factors and Ergonomics in Health Care; March 24-27, 2019; Chicago, IL p. 81-84 URL: https://doi.org/10.1177/2327857919081018, [doi: 10.1177/2327857919081018]
7. Brooke J. SUS: a quick and dirty usability scale. ResearchGate. 1995 Nov. URL: https://www.researchgate.net/publication/228593520_SUS_A_quick_and_dirty_usability_scale [accessed 2021-11-21]
8. Sauro J, Lewis J. Quantifying the User Experience: Practical Statistics for User Research. 2nd ed. Burlington, MA: Elsevier/Morgan Kaufmann; Jul 12, 2016.
9. Brooke J. SUS: a retrospective. J Usability Stud 2013 Feb;8(2):29-40 [FREE Full text]
10. Maramba I, Chatterjee A, Newman C. Methods of usability testing in the development of eHealth applications: a scoping review. Int J Med Inform 2019 Jun;126:95-104. [doi: 10.1016/j.ijmedinf.2019.03.018] [Medline: 31029270]
11. Rosenthal R. The file drawer problem and tolerance for null results. Psychol Bull 1979;86(3):638-641. [doi: 10.1037/0033-2909.86.3.638]
12. Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. Spine (Phila Pa 1976) 2000 Dec 15;25(24):3186-3191. [doi: 10.1097/00007632-200012150-00014] [Medline: 11124735]
13. Sharfina Z, Santoso HB. An Indonesian adaptation of the System Usability Scale (SUS). 2017 Mar 09 Presented at: 2016 International Conference on Advanced Computer Science and Information Systems (ICACSIS); October 15-16, 2016; Malang, Indonesia p. 145-148. [doi: 10.1109/icacsis.2016.7872776]

14. Gao M, Kortum P, Oswald FL. Multi-language toolkit for the System Usability Scale. Int J Hum Comput Interact 2020 Aug 19;36(20):1883-1901. [doi: 10.1080/10447318.2020.1801173]
15. Bondaronek P, Alkhaldi G, Slee A, Hamilton FL, Murray E. Quality of publicly available physical activity apps: review and content analysis. JMIR mHealth uHealth 2018 Mar 21;6(3):e53 [FREE Full text] [doi: 10.2196/mhealth.9069] [Medline: 29563080]
16. Ferrara G, Kim J, Lin S, Hua J, Seto E. A focused review of smartphone diet-tracking apps: usability, functionality, coherence with behavior change theory, and comparative validity of nutrient intake and energy estimates. JMIR mHealth uHealth 2019 May 17;7(5):e9232 [FREE Full text] [doi: 10.2196/mhealth.9232] [Medline: 31102369]
17. Isaković M, Sedlar U, Volk M, Bešter J. Usability pitfalls of diabetes mHealth apps for the elderly. J Diabetes Res 2016;2016:1604609 [FREE Full text] [doi: 10.1155/2016/1604609] [Medline: 27034957]
18. Metelmann B, Metelmann C, Schuffert L, Hahnenkamp K, Brinkrolf P. Medical correctness and user friendliness of available apps for cardiopulmonary resuscitation: systematic search combined with guideline adherence and usability evaluation. JMIR mHealth uHealth 2018 Nov 06;6(11):e190 [FREE Full text] [doi: 10.2196/mhealth.9651] [Medline: 30401673]
19. Gowarty MA, Longacre MR, Vilardaga R, Kung NJ, Gaughan-Maher AE, Brunette MF. Usability and acceptability of two smartphone apps for smoking cessation among young adults with serious mental illness: mixed methods study. JMIR Ment Health 2021 Jul 07;8(7):e26873 [FREE Full text] [doi: 10.2196/26873] [Medline: 34255699]
20. Morey SA, Barg-Walkow LH, Rogers WA. Managing heart failure on the go: usability issues with mHealth apps for older adults. Proc Hum Factors Ergon Soc Annu Meet 2017 Sep 28;61(1):1-5. [doi: 10.1177/1541931213601496]
21. Gibosn A, McCauley C, Mulvenna M, Ryan A, Laird L, Curran K, et al. Assessing usability testing for people living with dementia. 2016 Oct 13 Presented at: REHAB '16: Proceedings of the 4th Workshop on ICTs for improving Patients Rehabilitation Research Techniques; October 13-14, 2016; Lisbon, Portugal p. 25-31. [doi: 10.1145/3051488.3051492]
22. O'Grady C, Melia R, Bogue J, O'Sullivan M, Young K, Duggan J. A mobile health approach for improving outcomes in suicide prevention (SafePlan). J Med Internet Res 2020 Jul 30;22(7):e17481 [FREE Full text] [doi: 10.2196/17481] [Medline: 32729845]
23. Browne S, Kechadi M, O'Donnell S, Dow M, Tully L, Doyle G, et al. Mobile health apps in pediatric obesity treatment: process outcomes from a feasibility study of a multicomponent intervention. JMIR mHealth uHealth 2020 Jul 08;8(7):e16925 [FREE Full text] [doi: 10.2196/16925] [Medline: 32673267]
24. Teixeira LC, Beça P, Freitas J, Pinto I, Oliveira C, Lousada M. Usability and acceptability of an online tool to promote health of the teacher's voice: pilot study. 2019 Jul 15 Presented at: 2019 14th Iberian Conference on Information Systems and Technologies (CISTI); June 19-22, 2019; Coimbra, Portugal p. 1-6. [doi: 10.23919/cisti.2019.8760678]
25. Kalz M, Lenssen N, Felzen M, Rossaint R, Tabuenca B, Specht M, et al. Smartphone apps for cardiopulmonary resuscitation training and real incident support: a mixed-methods evaluation study. J Med Internet Res 2014 Mar 19;16(3):e89 [FREE Full text] [doi: 10.2196/jmir.2951] [Medline: 24647361]
26. Banos O, Moral-Munoz J, Diaz-Reyes I, Arroyo-Morales M, Damas M, Herrera-Viedma E, et al. mDurance: a novel mobile health system to support trunk endurance assessment. Sensors (Basel) 2015 Jun 05;15(6):13159-13183 [FREE Full text] [doi: 10.3390/s150613159] [Medline: 26057034]
27. Goldsmith JV, Wittenberg E, Ferrell B. An app to support difficult interactions among providers, patients, and families. J Adv Pract Oncol 2015 Sep 14;6(5):481-485 [FREE Full text] [doi: 10.6004/jadpro.2015.6.5.8] [Medline: 27069740]
28. Kizakevich PN, Eckhoff R, Weger S, Weeks A, Brown J, Bryant S, et al. A personal health information toolkit for health intervention research. Stud Health Technol Inform 2014;199:35-39. [Medline: 24875686]
29. Hoevenaars D, Holla JFM, Te Loo L, Koedijker JM, Dankers S, Houdijk H, WHEELS Study Group. Mobile app (WHEELS) to promote a healthy lifestyle in wheelchair users with spinal cord injury or lower limb amputation: usability and feasibility study. JMIR Form Res 2021 Aug 09;5(8):e24909 [FREE Full text] [doi: 10.2196/24909] [Medline: 34379056]
30. Fuller-Tyszkiewicz M, Richardson B, Klein B, Skouteris H, Christensen H, Austin D, et al. A mobile app-based intervention for depression: end-user and expert usability testing study. JMIR Ment Health 2018 Aug 23;5(3):e54 [FREE Full text] [doi: 10.2196/mental.9445] [Medline: 30139722]
31. Salamah Y, Asyifa RD, Asfarian A. Improving the usability of personal health record in mobile health application for people with autoimmune disease. 2021 Sep 07 Presented at: CHI '21: CHI Conference on Human Factors in Computing Systems; May 8-13, 2021; Yokohama, Japan p. 180-188. [doi: 10.1145/3429360.3468207]
32. Cameron G, Cameron D, Megaw G, Bond R, Mulvenna M, O'Neill S, et al. Assessing the usability of a chatbot for mental health care. In: Lecture Notes in Computer Science, vol 11551. 2019 Apr 17 Presented at: INSCI 2018 International Workshops; October 24-26, 2018; St. Petersburg, Russia p. 121-132. [doi: 10.1007/978-3-030-17705-8_11]
33. Leading health and fitness apps in the Google Play Store worldwide in January 2021, by number of downloads. Statista. 2021 Feb 24. URL: https://www.statista.com/statistics/690887/leading-google-play-health-worldwide-downloads/ [accessed 2022-01-28]
34. Jon Y. Jakob's Law. Sebastopol, CA: O'Reilly Media; Apr 2020.
35. Blackman KC, Zoellner J, McCrickard DS, Harlow J, Winchester III WW, Hill JL, et al. Developing mobile apps for physical activity in low socioeconomic status youth. J Mob Technol Med 2016 Mar 26;5(1):33-44. [doi: 10.7309/jmtm.5.1.6]

36. Cotton V, Patel MS. Gamification use and design in popular health and fitness mobile applications. Am J Health Promot 2019 Mar;33(3):448-451 [FREE Full text] [doi: 10.1177/0890117118790394] [Medline: 30049225]
37. Mazeas A, Duclos M, Pereira B, Chalabaev A. Evaluating the effectiveness of gamification on physical activity: systematic review and meta-analysis of randomized controlled trials. J Med Internet Res 2022 Jan 04;24(1):e26779 [FREE Full text] [doi: 10.2196/26779] [Medline: 34982715]
38. Zhou L, Bao J, Setiawan IMA, Saptono A, Parmanto B. The mHealth App Usability Questionnaire (MAUQ): development and validation study. JMIR mHealth uHealth 2019 Apr 11;7(4):e11500 [FREE Full text] [doi: 10.2196/11500] [Medline: 30973342]
39. Patel B, Thind A. Usability of mobile health apps for postoperative care: systematic review. JMIR Perioper Med 2020 Jul 20;3(2):e19099 [FREE Full text] [doi: 10.2196/19099] [Medline: 33393925]
40. Liew MS, Zhang J, See J, Ong YL. Usability challenges for health and wellness mobile apps: mixed-methods study among mHealth experts and consumers. JMIR mHealth uHealth 2019 Jan 30;7(1):e12160 [FREE Full text] [doi: 10.2196/12160] [Medline: 30698528]

Abbreviations

DHA: digital health app
MAUQ: mHealth App Usability Questionnaire
mHealth: mobile health
SUS: System Usability Scale

Edited by L Buis; submitted 15.02.22; peer-reviewed by EM Schomakers, G Dosovitsky; comments to author 08.04.22; revised version received 06.05.22; accepted 25.07.22; published 18.08.22.

Please cite as:
Hyzy M, Bond R, Mulvenna M, Bai L, Dix A, Leigh S, Hunt S
System Usability Scale Benchmarking for Digital Health Apps: Meta-analysis
JMIR Mhealth Uhealth 2022;10(8):e37290
URL: https://mhealth.jmir.org/2022/8/e37290
doi:10.2196/37290
PMID:35980732

Original Paper

Factors Associated With Using the COVID-19 Mobile Contact-Tracing App Among Individuals Diagnosed With SARS-CoV-2 in Amsterdam, the Netherlands: Observational Study

Feiko Ritsema[1], MSc; Jizzo R Bosdriesz[1,2], PhD; Tjalling Leenstra[1], MD, PhD; Mariska W F Petrignani[1], MD; Liza Coyer[1,2], MSc; Anja J M Schreijer[1], MD, PhD; Yvonne T H P van Duijnhoven[1], PhD; Janneke H H M van de Wijgert[3], MD, PhD; Maarten F Schim van der Loeff[1,2], MD, PhD; Amy Matser[1,2], PhD

[1]Department of Infectious Diseases, GGD Amsterdam, Amsterdam, Netherlands

[2]Department of Internal Medicine, Amsterdam Institute for Infection and Immunity, Amsterdam UMC, Academic Medical Center, Amsterdam, Netherlands

[3]Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, Utrecht, Netherlands

Corresponding Author:
Jizzo R Bosdriesz, PhD
Department of Infectious Diseases
GGD Amsterdam
Nieuwe Achtergracht 100
Amsterdam, 1018WT
Netherlands
Phone: 31 649616442
Email: jbosdriesz@ggd.amsterdam.nl

Abstract

Background: Worldwide, efforts are being made to stop the COVID-19 pandemic caused by SARS-CoV-2. Contact tracing and quarantining are key in limiting SARS-CoV-2 transmission. Mathematical models have shown that the time between infection, isolation of cases, and quarantining of contacts are the most important components that determine whether the pandemic can be controlled. Mobile contact-tracing apps could accelerate the tracing and quarantining of contacts, including anonymous contacts. However, real-world observational data on the uptake and determinants of contact-tracing apps are limited.

Objective: The aim of this paper is to assess the use of a national Dutch contact-tracing app among notified cases diagnosed with SARS-CoV-2 infection and investigate which characteristics are associated with the use of the app.

Methods: Due to privacy regulations, data from the app could not be used. Instead, we used anonymized SARS-CoV-2 routine contact-tracing data collected between October 28, 2020, and February 26, 2021, in the region of Amsterdam, the Netherlands. Complete case logistic regression analysis was performed to identify which factors (age, gender, country of birth, municipality, number of close contacts, and employment in either health care or education) were associated with using the app. Age and number of close contacts were modelled as B-splines due to their nonlinear relationship.

Results: Of 29,766 SARS-CoV-2 positive cases, 4824 (16.2%) reported app use. Median age of cases was 41 (IQR 29-55) years, and 46.7% (n=13,898) were male. In multivariable analysis, males (adjusted odds ratio [AOR] 1.11, 95% CI 1.04-1.18) and residents of municipalities surrounding Amsterdam were more likely to use the app (Aalsmeer AOR 1.34, 95% CI 1.13-1.58; Ouder-Amstel AOR 1.96, 95% CI 1.54-2.50), while people born outside the Netherlands, particularly those born in non-Western countries (AOR 0.33, 95% CI 0.30-0.36), were less likely to use the app. Odds of app use increased with age until the age of 58 years and decreased sharply thereafter (P<.001). Odds of app use increased with number of contacts, peaked at 8 contacts, and then decreased (P<.001). Individuals working in day care, home care, and elderly nursing homes were less likely to use the app.

Conclusions: Contact-tracing app use among people with confirmed SARS-CoV-2 infection was low in the region of Amsterdam. This diminishes the potential impact of the app by hampering the ability to warn contacts. Use was particularly low among older people, people born outside the Netherlands, and people with many contacts. Use of the app was also relatively low compared to those from some other European countries, some of which had additional features beyond contact tracing, making them potentially more appealing. For the Dutch contact-tracing app to have an impact, uptake needs to be higher; therefore, investing more into promotional efforts and additional features could be considered.

(JMIR Mhealth Uhealth 2022;10(8):e31099) doi:10.2196/31099

KEYWORDS

COVID-19; contact tracing; mobile contact tracing app; pandemic; mHealth; digital health; contact tracing app; mobile applications; health applications; public health; surveillance

Introduction

The COVID-19 pandemic, caused by SARS-CoV-2, has had a major impact. Two years into the pandemic, as of December 2021, over 260 million people have been infected worldwide, of whom more than 5 million have died [1]. Large-scale control measures are necessary to limit transmission of an emerging infectious disease such as COVID-19, for which a vaccine or treatment is (initially) unavailable [2]. Nonpharmaceutical interventions have been implemented by many countries, including face masking, physical distancing, travel restrictions, large-scale testing, and contact tracing [3]. To prevent the onward transmission of SARS-CoV-2, it is key to identify, test, and isolate infectious cases.

Contact tracing is a targeted approach to identify individuals who have been in close contact with confirmed cases [2]. The contacts of cases should be quarantined as soon as possible because the incubation period is short, and individuals can become infectious even before the onset of symptoms [4]. Contact tracing is a labor-intensive and time-consuming process. Its effect largely depends on the speed of contact tracing and the proportion of contacts that index cases are willing and able to identify from the start of probable infectiousness [5]. This is complicated by the fact that many of these contacts might be anonymous. Mathematical models have shown that the time between infection and isolation of cases, on the one hand, and quarantining of contacts, on the other, are the most important components that determine whether the pandemic can be controlled [6-8]. They also show that reducing delays in testing and contact tracing could reduce the spread of the virus, especially when there is no delay between case notification and quarantining of contacts. The models suggest that tracing apps for mobile phones have the potential to speed up the contact-tracing process and help identify unknown contacts, thereby significantly curbing SARS-CoV-2 spread [7-9]. However, these mathematical models rely on several assumptions, some of which might be violated by real world data, making it necessary to complement these studies with observational research.

Many countries have implemented tracing apps to identify and notify contacts of SARS-CoV-2 cases with various levels of success [10-17]. This fits in with a more generalized trend of increasing use of mobile apps for tracking and managing many aspects of health and behavior, providing users with more (sense of) control [18]. In the Netherlands, a tracing app developed by the Dutch government (CoronaMelder) was launched on October 10, 2020. The Dutch app uses Bluetooth to register other mobile phones on which the app is installed, their Bluetooth is active, and are within a 1.5-meter radius for at least 15 minutes. Data are stored locally on mobile phones for 14 days. When someone tests positive for SARS-CoV-2, the Public Health Service (PHS) will initiate contact tracing. As part of that process, the index case is asked whether they are using the app and are willing to notify the contacts that were registered by the app via the app. The registered contacts will subsequently receive a notification that they have been close to someone with a SARS-CoV-2 infection and the date on which this happened. In this notification, the app users are advised to quarantine themselves with immediate effect and to get tested. From October 10, 2020, to December 1, 2020, app users who received a notification were only allowed to be tested free of charge at a PHS facility if they were symptomatic. However, from December 1, 2020, onwards, asymptomatic users were also allowed free testing from the 5th day after the most recent exposure listed in the app notification.

Introduction of the app required an amendment to Dutch law [19] and generated much political and societal discussion about safeguarding the privacy of users. Controlling the spread of SARS-CoV-2 and protecting personal health are mentioned as main determinants of the willingness to use contact-tracing apps [20,21]. Conversely, safety and privacy concerns were associated with lower willingness to use the app. In general, 41% to 66% of participants were willing to use the app [20,21], which could be sufficient to reduce SARS-CoV-2 spread [8]. These figures are comparable to other Western European countries, where over 40% of participants said they would definitely install such an app, and an additional 35% of participants would probably install it [22]. However, the willingness to use an app might not lead to actual use. Nevertheless, reported app uptake numbers are encouraging (around 60% in Australia, Denmark, France, and the UK; 75% in the United States; and 90% in Japan [23]). More data are needed on actual app use in practice to complement theoretical models of app impact and willingness to use. Moreover, as research on other mobile health app has shown, there might be significant differences in uptake by age, income, education, health literacy, self-reported health, and intention to engage in healthy behavior [18]. To what extent those findings apply to an app such as CoronaMelder remains to be seen, since after installing it, no further active use of the app is required.

In this study, we therefore aimed to study the self-reported use of the Dutch CoronaMelder app and determinants of use in a real-life setting. As data from the app itself are not available due to privacy policies, we used data registered in the source and contact-tracing system after notification of a positive SARS-CoV-2 case instead. We evaluated which proportion of individuals who tested positive for SARS-CoV-2 (between October 28, 2020, and February 26, 2021, in the Amsterdam region) had used the mobile Dutch national contact-tracing app. Furthermore, we examined whether there were any significant differences in app uptake by several sociodemographic factors.

Methods

Population

In the Netherlands, SARS-CoV-2 tests are performed at publicly funded testing facilities of the PHS and hospitals, free of charge, or by commercial providers for a fee. SARS-CoV-2 is a notifiable infection, which means that all confirmed SARS-CoV-2 cases must be reported to the PHS regardless of where the testing took place. In this analysis, we included all adults (≥18 years old) who live in the Amsterdam region and were approached by the PHS of Amsterdam between October 28, 2020, and February 26, 2021, for contact tracing after a SARS-CoV-2 diagnosis.

Using data directly from the CoronaMelder app itself was not possible due to the anonymous nature of those data and privacy regulations. Therefore, we used data collected by PHS staff during routine contact tracing by phone and stored in HPZone (inFact UK Ltd). Routine procedure stipulates that PHS staff call persons diagnosed with SARS-CoV-2 (ie, cases) in the Netherlands to inform them about the diagnosis and isolation measures, and to initiate contact tracing. The case and a PHS staff member together systematically make an inventory of all identifiable persons that the case had been in contact with, 2 days prior to the date of symptom onset (if symptomatic) or positive test result (if asymptomatic). Moreover, PHS staff members are instructed to ask if the case used the CoronaMelder app, to note the answer in a standard format in a text field template in HPZone, and to activate the contact notification function of the app.

Variables

Data for this study were extracted from HPZone and anonymized before analysis. We extracted age in years at symptom onset (for symptomatic cases) or at the time of initiating contact tracing (for asymptomatic individuals) as a continuous variable. Other variables of interest were gender, categorized as male and female (other or nonbinary was not available in the system, was regarded as missing, and was therefore excluded from the analyses), and the municipality of residence (Aalsmeer, Amstelveen, Ouder-Amstel, Diemen, Uithoorn, or Amsterdam). During contact tracing, contacts were categorized into household contacts, close contacts, or other contacts. For this study, we extracted the number of close contacts, defined as contacts with whom a case had been within 1.5 meters for more than 15 minutes, excluding household contacts. Self-reported country of birth was recorded and later categorized as the Netherlands, other Western country, or non-Western country, in accordance with the definition used by Statistics Netherlands [24]. Employment in health care was categorized as “not,” “hospital,” “nursing home for elderly,” “other 24-hour care home,” “in-home care,” and “other health care.” Employment in education was categorized as “not,” “day care,” “elementary school,” and “secondary or higher education.” Data on CoronaMelder app use was extracted using a regular expression (“Gebruik coronamelder:”) from the free text notes. For those who used the app, we also extracted data on the reason of requesting a SARS-CoV-2 test.

Ethical Considerations

The medical ethics committee of the Amsterdam University Medical Centers deemed it not necessary to fully review the study, because the study does not fall under the scope of the Medical Research Involving Human Subjects Act (W20_432#20.479). No data from the app are used in this paper; therefore, the privacy regulations of the app were not reviewed for the purposes of this study, though they can be found on the web [25].

Analysis

Differences in characteristics between individuals who reported to use the mobile app and individuals who did not were assessed with chi-squared tests for categorical variables and Kruskal-Wallis tests for continuous variables. Trends over time in data availability on app use, as well as app use itself, were tested with the Pettitt test. Logistic regression analyses were performed to identify determinants of mobile app use. First, in univariable models, we tested for each independent variable (age, gender, country of birth, municipality, number of close contacts, employment in health care, and employment in education) whether they were associated with the dependent variable—self-reported use of the CoronaMelder app. Second, we combined all aforementioned independent variables and the dependent variable in 1 multivariable model. Age in years and the number of close contacts were added as continuous variables. As these variables were found to have a nonlinear relation to the outcome variable in exploratory analysis and regressions, B-splines were used with respectively 4 and 2 knots and a degree of 2. Gender and self-reported use of the CoronaMelder app were added as dichotomous variables, and all other variables were added as categorical variables. A complete cases analysis was performed; cases with missing data were excluded from the analysis. Outliers in the continuous variables age in years (above 100 years old) and number of close contacts (more than 12 close contacts, 99th percentile) were removed. In sensitivity analysis, multiple imputation using Multivariate Imputation by Chained Equations was carried out to impute missing outcomes and independent variables [26,27]. Analysis was performed using the statsmodels library in Python3 (Python Software Foundation) [28].

Results

From October 28, 2020, until February 26, 2021, the PHS of Amsterdam contacted 34,591 cases who were ≥18 years old and lived in the region of Amsterdam for contact tracing. We excluded 3354 (9.7%) cases because data on app use were not available, 1310 (3.8%) cases because they had missing values in one of the explanatory variables (such as gender), and 161 (0.47%) cases because they were outliers (>100 years old or >12 close contacts). Missing data on app use were caused by either invalid entries (anything except “yes/no” and variants of this) or missing entries, and they were higher in the first weeks after the introduction of the app (Figure 1). Cases with missing data on app use were older and more often born in a non-Western country.

The median age of the 29,766 included cases was 41 years (IQR 29-55); 13,898 (46.7%) were male, and 18,798 (63.2%) were

born in the Netherlands (Table 1). At the time of diagnosis, 4824 (16.2%) cases reported using the app. The number of cases reporting app use decreased significantly over time, especially after the first week of 2021—until January 4, 2021, a total of 5120 (17.2%) cases used the app, while this was 12,799 (14.3%) after that date (*P*=.001).

In total, 2494 (51.7%) out of 4824 app users and 15,442 (61.9%) out of 24,942 nonusers did not report any close contacts during the probable infectious period. The median number of reported close contacts among app users with at least one contact was 2 (IQR 1-4), and 2 (IQR 1-3) among nonusers. Among app users, 314 (6.5%) cases reported to have received a notification by the app that they had been in contact with a person diagnosed with SARS-CoV-2. The number of reported close contacts did not differ significantly between app users who received a notification and app users who did not receive a notification (*P*=.07, median 0; IQR 0-1; 90th percentile=3 for both groups). In total, 506/3227 (15.7%) individuals working in health care and 187/1154 (16.2%) individuals working in education used the app.

In multivariable logistic regression, the odds of reporting app use increased with increasing age (Figure 2a), until about the age of 58 years, after which the odds decreased sharply (*P*<.001). Men were slightly more likely to report app use than women (adjusted odds ratio [AOR] 1.11; 95% CI 1.04-1.18; Table 2). Cases who were born in other Western countries (AOR 0.74; 95% CI 0.65-0.84), and cases born in non-Western countries (AOR 0.33; 95% CI 0.30-0.36) were less likely to report app use compared with cases born in the Netherlands. Compared to cases living in the municipality of Amsterdam, cases living in most of the surrounding municipalities were more likely to report app use (eg, AOR 1.96; 95% CI 1.54-2.50 for cases living in Ouder-Amstel). Furthermore, there was a positive association between reporting more close contacts and reporting app use (Figure 2b), up to 8 reported close contacts, above which app use was less likely. Compared to cases not working in health care, cases working in elderly nursing homes (AOR 0.48; 95% CI 0.36-0.63) and home care (AOR 0.61; 95% CI 0.42-0.90) were less likely to report app use. The AOR for cases working in day care was 0.39 (95% CI 0.26-0.59) compared to cases not working in education.

The results after multiple imputation were similar to the results of complete case analyses (data not shown).

Figure 1. The percentage of cases with available data on the use of the contact tracing app (a) and the percentage of cases who used the mobile contact tracing app (b) by week (w) among SARS-CoV-2 positive cases in the region of Amsterdam (October 28, 2020, to February 26, 2021).

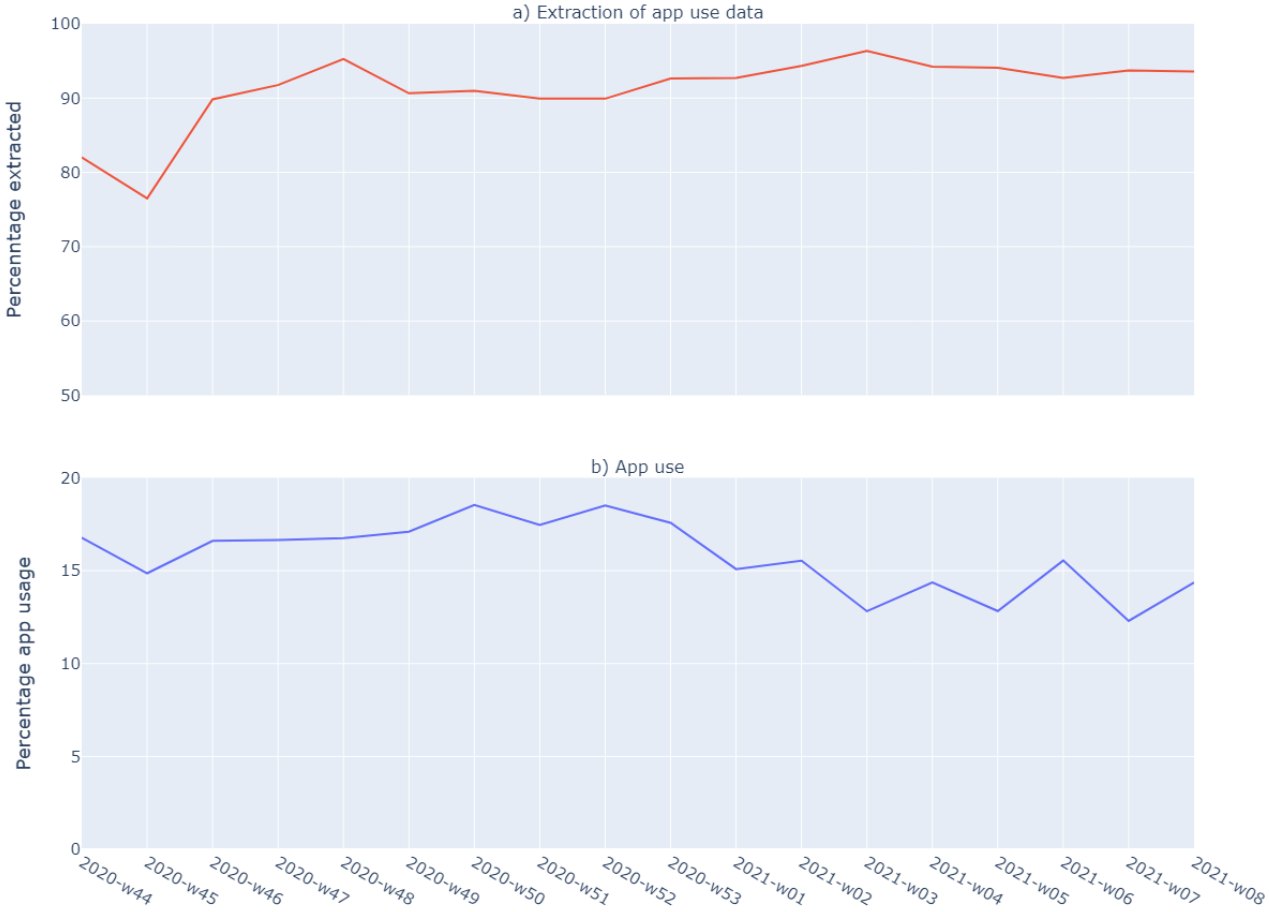


Table 1. Characteristics of individuals (≥18 years old) diagnosed with SARS-CoV-2 in the region of Amsterdam by reported mobile app use (October 28, 2020, to February 26, 2021).

Characteristics	Total[a] (N=29,766)	App users (n=4824)	Nonusers (n=24,942)	*P* value[b]
Age (years), mean (IQR)	41 (29-55)	42 (29-54)	41 (29-55)	.89
Gender, n (%)				<.001
Female	15,868 (53.3)	2437 (15.4)	13,431 (84.6)	
Male	13,898 (46.7)	2387 (17.2)	11,511 (82.8)	
Country of birth[c], n (%)				<.001
Netherlands	18,798 (63.2)	3803 (20.2)	14,995 (79.8)	
Non-Western	9116 (30.6)	730 (8.0)	8386 (92.0)	
Other Western	1852 (6.2)	291 (15.7)	1561 (84.3)	
Municipality, n (%)				<.001
Amsterdam	24,970 (83.9)	3832 (15.4)	21,138 (84.7)	
Aalsmeer	852 (2.9)	197 (23.1)	655 (76.9)	
Amstelveen	1921 (6.5)	408 (21.2)	1513 (78.8)	
Diemen	895 (3.0)	145 (16.2)	750 (83.8)	
Ouder-Amstel	340 (1.1)	99 (29.1)	241 (70.9)	
Uithoorn	788 (2.7)	143 (18.2)	645 (81.9)	
Median close contacts, mean (IQR)	0 (0-1)	0 (0-2)	0 (0-1)	<.001
Close contacts, n (%)				<.001
0	17,936 (60.3)	2494 (13.9)	15,442 (86.1)	
1-3	9133 (30.7)	1736 (19.0)	7397 (81.0)	
4-6	2024 (6.8)	449 (22.2)	1575 (77.8)	
>6	673 (2.3)	145 (21.6)	528 (78.5)	
Employment in health care, n (%)				<.001
No	26,539 (89.2)	4318 (16.3)	22,221 (83.7)	
Hospital	845 (2.8)	162 (19.2)	683 (80.8)	
Nursing home for elderly	664 (2.2)	57 (8.6)	607 (91.4)	
Other 24-hour care home	331 (1.1)	49 (14.8)	282 (85.2)	
Home care	261 (0.9)	31 (11.9)	230 (88.1)	
Other health care	1126 (3.8)	207 (18.4)	919 (81.6)	
Employment in education, n (%)				<.001
No	28,612 (96.1)	4637 (16.2)	23,975 (83.8)	
Yes, day care	340 (1.1)	26 (7.7)	314 (92.4)	
Yes, elementary school	637 (2.1)	130 (20.4)	507 (79.6)	
Yes, secondary or higher education	177 (0.6)	31 (17.5)	146 (82.5)	

[a]From the total sample, the following have been excluded: 3354 cases because of missing data on app use, 1310 cases because of missing values on an independent variable, and 161 cases because of outliers on continuous variables.

[b]*P* values for differences between app users and nonusers were assessed with Kruskal-Wallis tests for age and number of close contacts, and with chi-squared tests for all other variables.

[c]For the categorization of country of birth into non-Western or other Western, the definition from Statistics Netherlands was used [24].

Figure 2. Predicted probability of reporting CoronaMelder app use by (a) age in years and (b) the reported number of close contacts, resulting from multivariable logistic regression analysis using B-splines among 29,766 SARS-CoV-2 positive cases in the region of Amsterdam (October 28, 2020, to February 26, 2021).

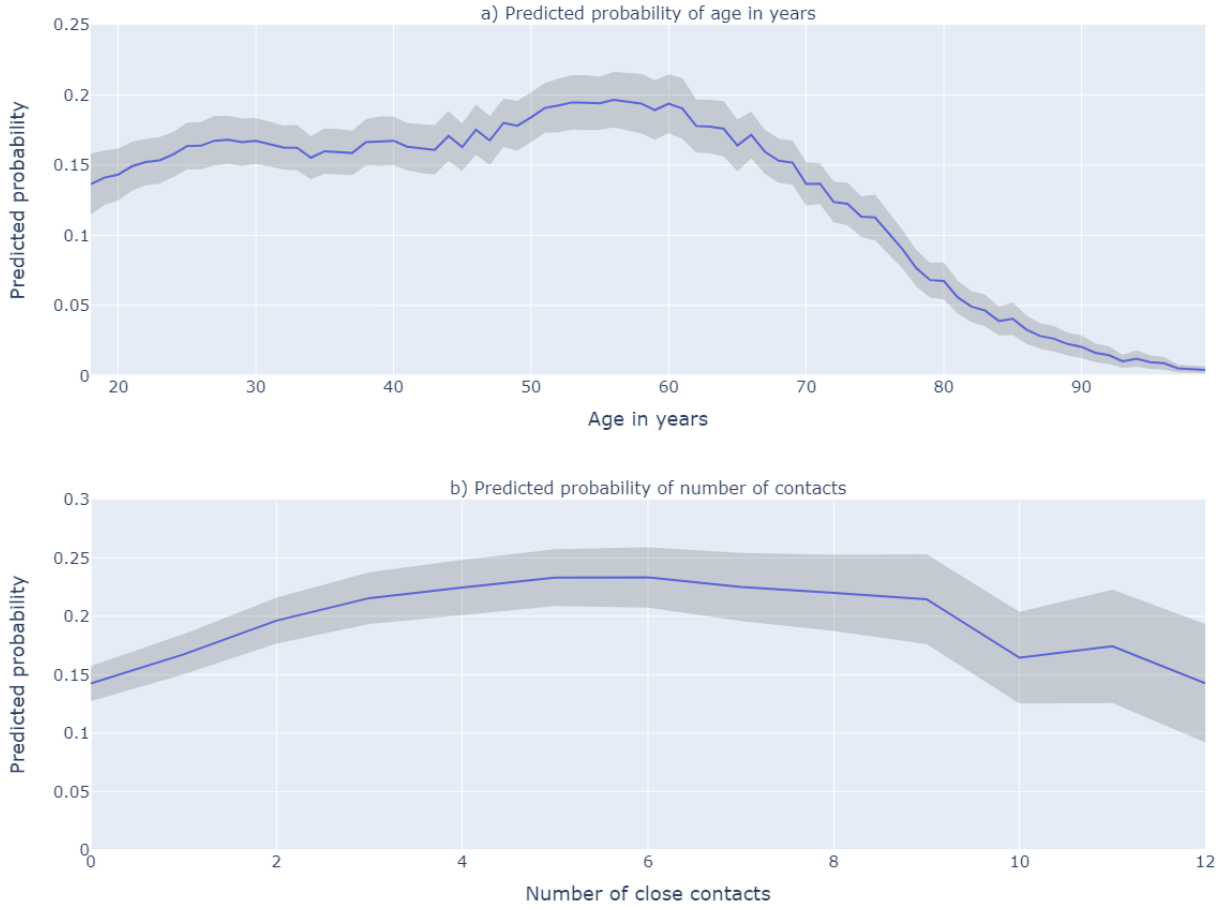


Table 2. Factors associated with mobile app use among 29,283 individuals (≥18 years old) diagnosed with SARS-CoV-2 in the region of Amsterdam (October 28, 2020, to February 26, 2021).

Characteristics	OR[a] (95% CI)[b]	*P* value	AOR[c] (95% CI)[b]	*P* value
Age[d]	—[e]	<.001	—	<.001
Gender		<.001		.002
Female	1		—	
Male	*1.14 (1.07-1.22)*		*1.11 (1.04-1.18)*	
Country of birth[d]		<.001		<.001
Netherlands	1		—	
Other Western	*0.74 (0.65-0.84)*		*0.74 (0.65-0.84)*	
Non-Western	*0.34 (0.32-0.37)*		*0.33 (0.30-0.36)*	
Municipality		<.001		<.001
Amsterdam	1		1	
Aalsmeer	*1.66 (1.41-1.95)*		*1.34 (1.13-1.58)*	
Amstelveen	*1.49 (1.33-1.67)*		*1.43 (1.27-1.61)*	
Diemen	1.07 (0.89-1.28)		1.02 (0.85-1.23)	
Ouder-Amstel	*2.27 (1.79-2.87)*		*1.96 (1.54-2.50)*	
Uithoorn	*1.22 (1.02-1.47)*		1.03 (0.85-1.25)	
Number of close contacts[f]	—	<.001	—	<.001
Employment in health care		<.001		<.001
No	1		1	
Hospital	*1.22 (1.03-1.45)*		1.02 (0.85-1.22)	
Nursing home for elderly	*0.48 (0.37-0.64)*		*0.48 (0.36-0.63)*	
Other 24-hour care home	0.89 (0.66-1.21)		0.78 (0.57-1.06)	
Home care	0.69 (0.48-1.01)		*0.61 (0.42-0.90)*	
Other health care	1.16 (0.99-1.35)		0.95 (0.81-1.12)	
Employment in education		<.001		<.001
No	1		1	
Yes, day care	*0.43 (0.29-0.64)*		*0.39 (0.26-0.59)*	
Yes, elementary school	*1.33 (1.09-1.61)*		1.07 (0.88-1.31)	
Yes, secondary or higher education	1.1 (0.74-1.62)		0.91 (0.61-1.35)	

[a]OR: odds ratio.

[b]Significant associations are italicized.

[c]AOR: adjusted odds ratio.

[d]For the categorization of country of birth into non-Western or Other Western, the definition from Statistics Netherlands was used [24].

[e]Not applicable.

[f]Variables modelled as B-splines (Figure 2).

Discussion

Principal Findings

In this study, we found that fewer than 1 in 6 individuals diagnosed with SARS-CoV-2 in the region of Amsterdam reported using the CoronaMelder contact-tracing app. As 24,942 (84%) out of 29,766 cases were not using the app, their close contacts could never receive a notification through the app, even though they might have installed it themselves. Only 6.5% (1935/29,766) of the positive cases with the app had received an app notification themselves. Reporting app use was associated with being middle-aged, having a few (ie, 3-8) close contacts during the infectious period, living in municipalities surrounding Amsterdam (rather than the city itself), and being born in the Netherlands. App use was less often reported by individuals with more than 8 close contacts and individuals who are born outside the Netherlands.

Limitations

Caution is warranted when interpreting these results and the potential explanations and implications. Moreover, these results cannot be directly extrapolated to the general population, including those who did not test positive for SARS-CoV-2. According to national data, over 4.5 million people have downloaded the app during the study period [29], which is approximately 26% of the Dutch population. In our study population of cases, however, this percentage was only 16% (4824/29,766). It is possible that the cases included in our sample represent a population that is less likely to take any preventive measures. It is also possible that PHS staff did not consistently ask cases about app use, as Amsterdam has been a region with high infection numbers, leading to high work pressure for the contact-tracing team. On the other hand, the national number is a cumulative number that does not account for app removals or inactivation, multiple app downloads by one person, or underreporting of app use during contact tracing, while the number in our study represents prevalent use. A second limitation is that the routine PHS data were not collected for the purpose of scientific research. This limits the number of variables and thus the potential to explain our observations. Furthermore, ascertainment bias may have been introduced because data may not have been collected consistently and uniformly. However, sensitivity analysis showed that bias caused by missing data was very small.

Comparison With Prior Work

In Dutch acceptability studies performed prior to the introduction of the app, in April 2020, younger individuals reported to be more willing to download the app once available [20,21]. A survey performed in France, Germany, Italy, the United Kingdom, and the United States showed the same age trend [22]. Willingness to download the app was associated with positive attitudes toward technology and with fear for COVID-19 [21]. However, in our study, middle-aged individuals were more likely to use the app compared to younger individuals. Fear for COVID-19 might have played a role in these older age groups, in line with their higher risk of more serious disease once infected. Conversely, the absence of fear, privacy concerns, and a lower willingness to obey COVID-19 control measures might have been more important among younger individuals. The oldest individuals in our study were less likely to use the app, which may relate to lower smartphone and app usage among elderly people in general [18]. This is supported by another evaluation of the same CoronaMelder app, which showed that elderly people had problems with understanding why, when, and how to use the app [30].

Cases living in the municipalities of Amsterdam or Diemen, the latter being geographically strongly connected to Amsterdam, were less likely to use the app compared with cases in the surrounding municipalities. Improving app use in more densely populated urban settings might be worthwhile because the app is especially useful to identify anonymous close contacts who cannot be traced otherwise. Additionally, we found strong associations with being born outside the Netherlands and not using the app. If national app usage trends reflect those found in our sample, this would be worrisome given that previous studies in Amsterdam and internationally have shown that some ethnic minority groups are disproportionately affected by SARS-CoV-2 [31-36]. Cultural differences or distrust in the authorities may underlie this observation, but other more practical issues might be important as well. Even though the app itself and information on the CoronaMelder app website are available in 10 different languages, language barriers might still exist, and communication about the app might not reach all groups. Unfortunately, the routine data used in this study do not contain information on language skills or parental birth country, and thus we cannot investigate app use among second-generation immigrants. Further research in this group is therefore needed to reduce health inequalities between ethnic groups [35,36].

While the likelihood of using the app increased with the number of close contacts in the range of 0 to 8 close contacts, it decreased with higher number of contacts. The advantages of using an app for contact tracing include speed, the fact that anonymous close contacts can be reached, and that there is no recall bias, which is especially beneficial among individuals with many contacts. Thus, it might be worthwhile to study barriers for use and promote app use among individuals with many close contacts.

Lastly, we saw moderate differences in app use among people working in high-risk professions during which many contacts may be unavoidable. In the health care sector, precautions are taken to prevent infection (eg, use of personal protective equipment). Using the app during working hours may result in *false* notifications that are indistinguishable from notifications after real risk contacts. This might explain why people working in nursing homes and home care were using the app less often compared to individuals not working in health care. However, a pause button was introduced to the app (to be used in situations such as when the phone is left in a locker) to allow people to keep the app but reduce the chance of receiving *false* notifications [37]. Place of work as a reason for not downloading the app was mentioned in a survey in the United Kingdom [38]. Individuals working in day care centers were also less likely to have the app. For this group, the app could be of added value because they encounter parents of children without full protection.

If contact-tracing apps are used efficiently and uptake is high, they have the potential to speed up contact tracing, identify contacts that would otherwise go unnoticed, and prevent infections. For instance, the app of the National Health Service in the United Kingdom has been downloaded by 49% of the eligible population with compatible smartphones [39], which is >30% of the total population. A modelling study showed that this app averted about one case per index case willing to send the notification to their contacts [39]. The percentage of the population who downloaded tracing apps was also high in countries such as Germany, Switzerland, and Finland (ie, 32%-45%), but much lower in Spain, Italy, and France (ie, 15%-19%) [40]. This high level of adoption in some countries could be driven by the fact that the National Health Service app and the German app are among several apps that combine the tracing function with other features such as local area risk indicators and a link for booking a test, or by differences in

promotion efforts. The Dutch Ministry of Health decided not to equip the CoronaMelder app with such additional features that might appeal to users, and soon after introduction, it stopped actively promoting app use. Instead, the Dutch Ministry of Health developed a second national app to function as a COVID-19 passport, registering vaccinations, recovery from infection, and negative test results. Combining the contact-tracing app and the corona passport app might have increased use of the tracing app. This knowledge, combined with the observed low uptake of the app in our sample, suggests that, to yield its potential effect on the control of the COVID-19 epidemic, the app needs to be used by more people. Based on our findings, app promotion efforts should particularly target younger individuals, individuals with >8 close contacts, and individuals who are not born in the Netherlands.

Conclusions

This study shows that app use is low; only 4824 (16.2%) out of 29,766 individuals who tested positive for SARS-CoV-2 in the Amsterdam region. Moreover, we observed significant differences in app uptake by sociodemographic factors. Elderly persons, women, people not born in the Netherlands, and those either reporting none or many close contacts were less likely to have installed the CoronaMelder app. If confirmed in a nationally representative sample, this would mean the app is unlikely to have the impact on SARS-CoV-2 spread it could potentially have. Moreover, app uptake seems to be lower in certain subgroups of the population, indicating that more targeted efforts to improve uptake are necessary.

Conflicts of Interest

None declared.

References

1. WHO Coronavirus Disease (COVID-19) Dashboard. World Health Organization. 2021. URL: https://covid19.who.int/ [accessed 2021-12-13]
2. Müller J, Kretzschmar M. Contact tracing - Old models and new challenges. Infect Dis Model 2021;6:222-231 [FREE Full text] [doi: 10.1016/j.idm.2020.12.005] [Medline: 33506153]
3. Bo Y, Guo C, Lin C, Zeng Y, Li HB, Zhang Y, et al. Effectiveness of non-pharmaceutical interventions on COVID-19 transmission in 190 countries from 23 January to 13 April 2020. Int J Infect Dis 2021 Jan;102:247-253 [FREE Full text] [doi: 10.1016/j.ijid.2020.10.066] [Medline: 33129965]
4. Koh WC, Naing L, Chaw L, Rosledzana MA, Alikhan MF, Jamaludin SA, et al. What do we know about SARS-CoV-2 transmission? A systematic review and meta-analysis of the secondary attack rate and associated risk factors. PLoS One 2020;15(10):e0240205 [FREE Full text] [doi: 10.1371/journal.pone.0240205] [Medline: 33031427]
5. Braithwaite I, Callender T, Bullock M, Aldridge RW. Automated and partly automated contact tracing: a systematic review to inform the control of COVID-19. The Lancet Digital Health 2020 Nov;2(11):e607-e621. [doi: 10.1016/s2589-7500(20)30184-9]
6. Hellewell J, Abbott S, Gimma A, Bosse NI, Jarvis CI, Russell TW, et al. Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts. The Lancet Global Health 2020 Apr;8(4):e488-e496 [FREE Full text] [doi: 10.1016/S2214-109X(20)30074-7] [Medline: 32119825]
7. Kretzschmar ME, Rozhnova G, Bootsma MCJ, van Boven M, van de Wijgert JHHM, Bonten MJM. Impact of delays on effectiveness of contact tracing strategies for COVID-19: a modelling study. The Lancet Public Health 2020 Aug;5(8):e452-e459. [doi: 10.1016/s2468-2667(20)30157-2]
8. Ferretti L, Wymant C, Kendall M, Zhao L, Nurtay A, Abeler-Dörner L, et al. Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. Science 2020 May 08;368(6491):eabb6936 [FREE Full text] [doi: 10.1126/science.abb6936] [Medline: 32234805]
9. Hinch R, Probert W, Nurtay A, Kendall M, Wymant C, Hall M, et al. Effective configurations of a digital contact tracing app: a report to NHSX. Academic Journalism Society. 2020. URL: https://cdn.theconversation.com/static_files/files/1009/Report_-_Effective_App_Configurations.pdf [accessed 2021-12-13]
10. Cho H, Ippolito D, Yu Y. Contact tracing mobile apps for COVID-19: privacy considerations and related trade-offs. arXiv Preprint posted online March 30, 2020. [doi: 10.48550/arXiv.2003.11511]
11. Polanco Walters F, Anyane-Yeboa A, Landry AM. The not-so-silent killer missing in medical-training curricula: racism. Nat Med 2020 Aug;26(8):1160-1161. [doi: 10.1038/s41591-020-0984-3] [Medline: 32661392]
12. Majority of Americans say they won't use COVID contact tracing apps. Avira. URL: https://www.avira.com/en/covid-contact-tracing-app-report [accessed 2021-01-20]
13. Only 3.1% of French people have downloaded the StopCovid application, according to a study. World Today News. 2020 Jul 17. URL: https://www.world-today-news.com/only-3-1-of-french-people-have-downloaded-the-stopcovid-application-according-to-a-study/ [accessed 2021-01-20]
14. COVIDSafe app. Australian Government: Department of Health and Aged Care. URL: https://www.health.gov.au/resources/apps-and-tools/covidsafe-app [accessed 2021-03-17]
15. Aarogya Setu mobile app. Government of India. 2020. URL: https://www.mygov.in/aarogya-setu-app/ [accessed 2021-03-17]

16. Coronavirus. Helsenorge. 2020. URL: https://helsenorge.no/coronavirus/smittestopp [accessed 2021-03-17]
17. How TraceTogether works. Government of Singapore. 2020. URL: https://www.tracetogether.gov.sg/ [accessed 2021-03-17]
18. Grundy Q. A Review of the Quality and Impact of Mobile Health Apps. Annu Rev Public Health 2022 Apr 05;43(1):117-134. [doi: 10.1146/annurev-publhealth-052020-103738] [Medline: 34910582]
19. Staatsblad van het Koninkrijk der Nederlanden. Ministry of Health, Welfare, and Sport. URL: https://zoek.officielebekendmakingen.nl/stb-2020-374.html [accessed 2021-03-17]
20. Jonker M, de Bekker-Grob E, Veldwijk J, Goossens L, Bour S, Rutten-Van Mölken M. COVID-19 Contact Tracing Apps: Predicted Uptake in the Netherlands Based on a Discrete Choice Experiment. JMIR Mhealth Uhealth 2020 Oct 09;8(10):e20741 [FREE Full text] [doi: 10.2196/20741] [Medline: 32795998]
21. Jansen-Kosterink S, Hurmuz M, den Ouden M, van Velsen L. Predictors to Use Mobile Apps for Monitoring COVID-19 Symptoms and Contact Tracing: Survey Among Dutch Citizens. JMIR Form Res 2021 Dec 20;5(12):e28416 [FREE Full text] [doi: 10.2196/28416] [Medline: 34818210]
22. Altmann S, Milsom L, Zillessen H, Blasone R, Gerdon F, Bach R, et al. Acceptability of App-Based Contact Tracing for COVID-19: Cross-Country Survey Study. JMIR Mhealth Uhealth 2020 Aug 28;8(8):e19857 [FREE Full text] [doi: 10.2196/19857] [Medline: 32759102]
23. Grekousis G, Liu Y. Digital contact tracing, community uptake, and proximity awareness technology to fight COVID-19: a systematic review. Sustain Cities Soc 2021 Aug;71:102995 [FREE Full text] [doi: 10.1016/j.scs.2021.102995] [Medline: 34002124]
24. Migratieachtergrond. Statistics Netherlands. 2021. URL: https://www.cbs.nl/nl-nl/onze-diensten/methoden/begrippen/migratieachtergrond [accessed 2021-12-13]
25. CoronaMelder privacy statement. Ministry of Health, Welfare, and Sport. URL: https://coronamelder.nl/en/privacy [accessed 2021-12-22]
26. The R Project for Statistical Computing. The R Foundation. URL: https://www.r-project.org/ [accessed 2022-07-21]
27. Buuren SV, Groothuis-Oudshoorn K. mice: Multivariate Imputation by Chained Equations in R. J. Stat. Soft 2011;45(3):1-67. [doi: 10.18637/jss.v045.i03]
28. Seabold S, Perktold J. Statsmodels: Econometric and Statistical Modeling with Python. SciPy. 2010. URL: https://conference.scipy.org/proceedings/scipy2010/pdfs/seabold.pdf [accessed 2022-07-21]
29. Factsheet Coronamelder. Government of the Netherlands. 2021. URL: https://coronamelder.nl/media/Factsheet_Corona_latest.pdf [accessed 2021-01-28]
30. Bente BE, van 't Klooster JWJR, Schreijer MA, Berkemeier L, van Gend JE, Slijkhuis PJH, et al. The Dutch COVID-19 Contact Tracing App (the CoronaMelder): Usability Study. JMIR Form Res 2021 Mar 26;5(3):e27882 [FREE Full text] [doi: 10.2196/27882] [Medline: 33724198]
31. Indseth T, Grøsland M, Arnesen T, Skyrud K, Kløvstad H, Lamprini V, et al. COVID-19 among immigrants in Norway, notified infections, related hospitalizations and associated mortality: A register-based study. Scand J Public Health 2021 Feb;49(1):48-56 [FREE Full text] [doi: 10.1177/1403494820984026] [Medline: 33406993]
32. Sze S, Pan D, Nevill CR, Gray LJ, Martin CA, Nazareth J, et al. Ethnicity and clinical outcomes in COVID-19: A systematic review and meta-analysis. EClinicalMedicine 2020 Dec;29:100630 [FREE Full text] [doi: 10.1016/j.eclinm.2020.100630] [Medline: 33200120]
33. Webb Hooper M, Nápoles AM, Pérez-Stable EJ. COVID-19 and Racial/Ethnic Disparities. JAMA 2020 Jun 23;323(24):2466-2467. [doi: 10.1001/jama.2020.8598] [Medline: 32391864]
34. Disparities in the risk and outcomes of COVID-19. Public Health England. 2020. URL: https://tinyurl.com/395c33kh [accessed 2021-01-28]
35. Coyer L, Boyd A, Schinkel J, Agyemang C, Galenkamp H, Koopman ADM, et al. SARS-CoV-2 antibody prevalence and correlates of six ethnic groups living in Amsterdam, the Netherlands: a population-based cross-sectional study, June-October 2020. BMJ Open 2022 Jan 06;12(1):e052752 [FREE Full text] [doi: 10.1136/bmjopen-2021-052752] [Medline: 34992110]
36. Coyer L, Boyd A, Schinkel J, Agyemang C, Galenkamp H, Koopman ADM, et al. Differences in SARS-CoV-2 infections during the first and second wave of SARS-CoV-2 between six ethnic groups in Amsterdam, the Netherlands: A population-based longitudinal serological study. Lancet Reg Health Eur 2022 Feb;13:100284 [FREE Full text] [doi: 10.1016/j.lanepe.2021.100284] [Medline: 34927120]
37. Update CoronaMelder: nieuwe richtlijn verwerkt én pauzefunctie beschikbaar. Government of the Netherlands. 2021. URL: https://www.rijksoverheid.nl/actueel/nieuws/2021/02/03/update-coronamelder-nieuwe-richtlijnen-verwerkt-en-pauzefunctie-beschikbaar [accessed 2021-12-21]
38. Panchal M, Singh S, Rodriguez-Villegas E. Analysis of the factors affecting the adoption and compliance of the NHS COVID-19 mobile application: a national online questionnaire survey in England. medRxiv Preprint posted online March 08, 2021. [doi: 10.1101/2021.03.04.21252924]
39. Wymant C, Ferretti L, Tsallis D, Charalambides M, Abeler-Dörner L, Bonsall D, et al. The epidemiological impact of the NHS COVID-19 app. Nature 2021 Jun 12;594(7863):408-412. [doi: 10.1038/s41586-021-03606-z] [Medline: 33979832]

40. Kozyreva A, Lorenz-Spreen P, Lewandowsky S, Garrett PM, Herzog SM, Pachur T, et al. Psychological factors shaping public responses to COVID-19 digital contact tracing technologies in Germany. Sci Rep 2021 Sep 21;11(1):18716 [FREE Full text] [doi: 10.1038/s41598-021-98249-5] [Medline: 34548550]

Abbreviations

AOR: adjusted odds ratio
PHS: Public Health Service

Edited by L Buis, A Mavragani; submitted 09.06.21; peer-reviewed by K Lutrick, L Welch; comments to author 22.11.21; revised version received 22.12.21; accepted 04.04.22; published 24.08.22.

Please cite as:
Ritsema F, Bosdriesz JR, Leenstra T, Petrignani MWF, Coyer L, Schreijer AJM, van Duijnhoven YTHP, van de Wijgert JHHM, Schim van der Loeff MF, Matser A
Factors Associated With Using the COVID-19 Mobile Contact-Tracing App Among Individuals Diagnosed With SARS-CoV-2 in Amsterdam, the Netherlands: Observational Study
JMIR Mhealth Uhealth 2022;10(8):e31099
URL: https://mhealth.jmir.org/2022/8/e31099
doi:10.2196/31099
PMID:35867842

Original Paper

Predicting the Next-Day Perceived and Physiological Stress of Pregnant Women by Using Machine Learning and Explainability: Algorithm Development and Validation

Ada Ng[1], PhD; Boyang Wei[1], MS; Jayalakshmi Jain[1], MS; Erin A Ward[2], MA, MSW; S Darius Tandon[2], PhD; Judith T Moskowitz[2], MPH, PhD; Sheila Krogh-Jespersen[2], PhD; Lauren S Wakschlag[2], PhD; Nabil Alshurafa[2], PhD

[1]McCormick School of Engineering, Northwestern University, Evanston, IL, United States

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

Corresponding Author:
Ada Ng, PhD
McCormick School of Engineering
Northwestern University
633 Clark St
Evanston, IL, 60208
United States
Phone: 1 8474913741
Email: adang@u.northwestern.edu

Abstract

Background: Cognitive behavioral therapy–based interventions are effective in reducing prenatal stress, which can have severe adverse health effects on mothers and newborns if unaddressed. Predicting next-day physiological or perceived stress can help to inform and enable pre-emptive interventions for a likely physiologically and perceptibly stressful day. Machine learning models are useful tools that can be developed to predict next-day physiological and perceived stress by using data collected from the previous day. Such models can improve our understanding of the specific factors that predict physiological and perceived stress and allow researchers to develop systems that collect selected features for assessment in clinical trials to minimize the burden of data collection.

Objective: The aim of this study was to build and evaluate a machine-learned model that predicts next-day physiological and perceived stress by using sensor-based, ecological momentary assessment (EMA)–based, and intervention-based features and to explain the prediction results.

Methods: We enrolled pregnant women into a prospective proof-of-concept study and collected electrocardiography, EMA, and cognitive behavioral therapy intervention data over 12 weeks. We used the data to train and evaluate 6 machine learning models to predict next-day physiological and perceived stress. After selecting the best performing model, Shapley Additive Explanations were used to identify the feature importance and explainability of each feature.

Results: A total of 16 pregnant women enrolled in the study. Overall, 4157.18 hours of data were collected, and participants answered 2838 EMAs. After applying feature selection, 8 and 10 features were found to positively predict next-day physiological and perceived stress, respectively. A random forest classifier performed the best in predicting next-day physiological stress (F1 score of 0.84) and next-day perceived stress (F1 score of 0.74) by using all features. Although any subset of sensor-based, EMA-based, or intervention-based features could reliably predict next-day physiological stress, EMA-based features were necessary to predict next-day perceived stress. The analysis of explainability metrics showed that the prolonged duration of physiological stress was highly predictive of next-day physiological stress and that physiological stress and perceived stress were temporally divergent.

Conclusions: In this study, we were able to build interpretable machine learning models to predict next-day physiological and perceived stress, and we identified unique features that were highly predictive of next-day stress that can help to reduce the burden of data collection.

(JMIR Mhealth Uhealth 2022;10(8):e33850) doi:10.2196/33850

KEYWORDS

explainability; just-in-time interventions; machine learning; prenatal stress; stress prediction; wearable; mobile phone

Introduction

Background

Welcoming a new member to the family is cause for celebration but can also lead to substantial stress, particularly for mothers. A systematic review of perinatal depression (PD) predictors identified prenatal stressors as either episodic (eg, life events or daily hassles) or chronic (eg, parenting stress, perceived stress, and chronic strain) [1]. Lack of social support, stressful life events, domestic violence, low socioeconomic status, and past history of depression contribute significantly to increased prenatal stress [2,3]. Maternal stress can lead to preterm birth or low birth weight [4,5], which are leading causes of infant mortality in the United States [6,7], or structural malformations [8] and psychosocial impairment [9].

To mitigate these negative outcomes, a number of interventions have been developed and tested to reduce stress in pregnant women, often using the principles of mindfulness [10-15] and cognitive behavioral therapy (CBT) [16-19] in group or individual format. A key characteristic of many CBT-based interventions is the inclusion of personal practice or homework between intervention sessions to facilitate adoption of newly learned skills [20]. Such homework can take the form of technologically supported just-in-time (JIT) interventions [21], which in the case of maternal stress can enhance effectiveness of stress-reducing techniques. Incorporation of JIT interventions is facilitated by technology that participants can receive on mobile phones. Use of JIT interventions is associated with improvement in mental health symptoms and conditions that these interventions target [22-24]. However, the timing of interventions may affect participation [25], especially given that JIT interventions typically require individuals to perform an action in the moment to achieve desired outcomes. To appropriately target stress with JIT interventions, it is necessary to identify factors that are most predictive of stress to develop a mechanism to proactively detect and deliver a timely preventive intervention.

However, there is no singular definition of stress and the mechanisms underlying physiological and perceived stress are different, requiring different means of detection and prediction [26]. Physiological stress that persists from one day to the next, or *residual stress*, can be most damaging to neurovascular health and lead to chronic diseases [27,28]. Although it is unclear how perceived stress maps onto future disease state, perceived stress can be debilitating and linked to poor life satisfaction [29]. The ability to predict next-day stress, whether physiological or perceived, and to understand predictors of either type of stress may allow for advanced scheduling of JIT interventions that help to reduce or prevent next-day stress.

Machine learning models have been used to successfully predict both physiological and perceived stress; however, few models predict beyond the near future while also explaining the driving forces behind the predictions. Several sensing systems have been designed to forecast physiological stress in the future [30,31]. However, studies predicting physiological stress are often performed in a laboratory owing to the limited feasibility of frequent stress assessments in the wild [32]. Other studies have captured both perceived and physiological stress but only consider physiological stress when determining ground truth for the machine-learned models [33]. The few examples of next-day stress prediction using prior days' data [34,35] either focused on testing the difference between generalized and personalized models or focused primarily on prediction of perceived stress.

Being able to predict stress earlier and with minimal data collection burden while assessing the interpretability of the model will allow researchers to improve their understanding of the learned model, increasing their understanding of how the model determines stress the next day and informing the design of JIT intervention. Models can use global explanations, which attempt to describe the overall functionality of the learned model (eg, feature importance), or local explanations, which are aimed at explaining the model's reasoning for a specific instance. Some types of explanations such as Shapley Additive Explanations (SHAP) [36] enable greater interpretability of models and are considered model agnostic, providing both global and local explanations. SHAP can be used to create global explanations by aggregating SHAP values to create feature importance, summary, and dependence plots. SHAP values are feature attributions that act as driving forces either contributing to the prediction or not. Ultimately, these results can inform means of low-burden early stress detection, which in turn enables scheduling of JIT intervention content that prevents future stress and its correlates.

Objectives

In this pilot study, we aimed to predict next-day stress in pregnant women who participated in a perinatal stress reduction course. Specifically, we obtained data from sensors and participant self-report and then used several machine learning models to find the best performer. We evaluated the potential of our model to predict next-day stress and applied an explainability model to provide meaning to our predictions.

Methods

Study Design

We collaborated with a private university's obstetrics and gynecology clinic to recruit pregnant women into our study. To be eligible for enrollment, women had to be aged ≥18 years, enrolled at 10 to 18 weeks' gestation with a singleton pregnancy, and own a smartphone. Women were excluded if they had a known medical or pregnancy complication that may place their infant at risk for neurological disorders or significant mental health disorders.

Upon enrollment, participants received a 12-week person-to-person intervention called the *Mothers and Babies* (MB) course [19] from a master's-level social worker and wore a mobile electrocardiography (ECG) sensor, BioStampRC

(MC10), to capture heart rate (HR) data (Figure S1 in Multimedia Appendix 1). Throughout the study, participants received SMS text messages on their mobile phones in the form of ecological momentary assessment (EMA) surveys for self-reported stress assessment. During the first intervention session, participants were shown the BioStampRC sensor and trained to use it. Subsequent MB intervention sessions (1:1 interventions) were delivered every 1 to 2 weeks by the same social worker, either in person or through the phone. At the end of 12 intervention sessions, participants returned the sensor and were asked to provide feedback on the usability and wearability of the sensor, as well as the acceptability of the EMA surveys through a semistructured exit interview. The women received US $200 compensation for completing the study.

After the completion of the study, we performed data extraction and preprocessing and then applied machine learning models and a SHAP explainability model to identify predictors of next-day physiological and perceived stress. Overviews of the study design and MB program are shown in Figures 1 and 2, respectively.

Figure 1. Data collection and processing pipeline for predicting next-day stress. CBT: cognitive behavioral therapy; EMA: ecological momentary assessment.

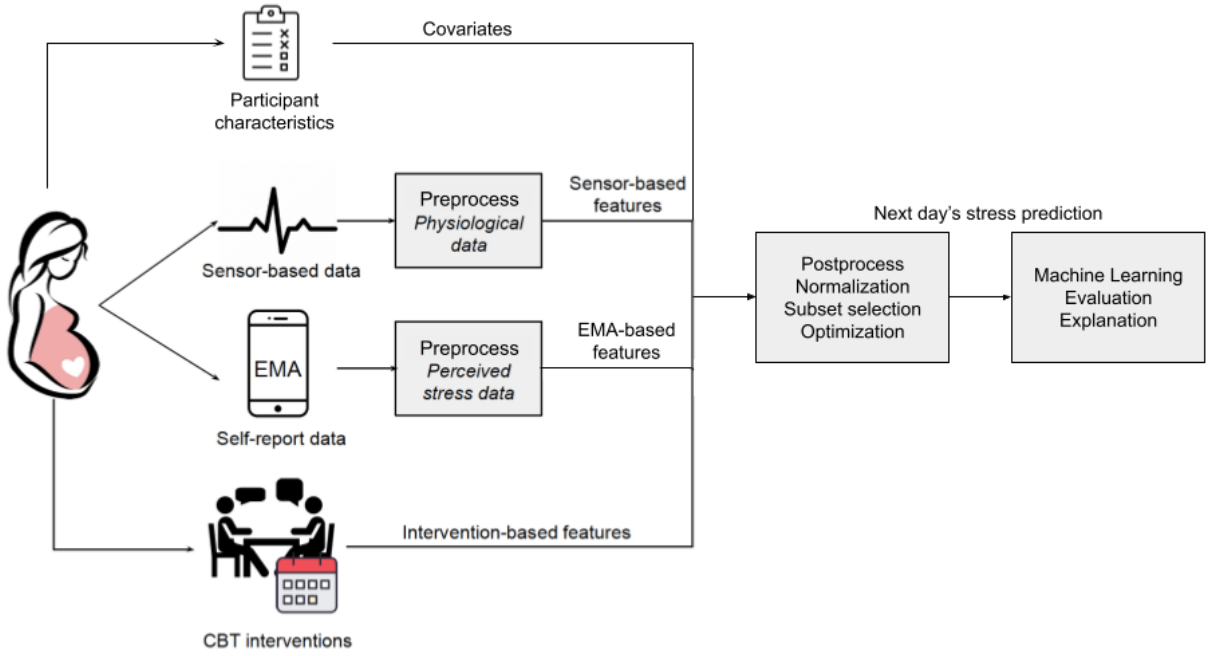


Figure 2. Mothers and Babies program intervention schedule and content. JIT: just-in-time.

Ethics Approval

The study was approved by Northwestern University's institutional review board (approval number: STU00205776), and all women provided written informed consent before enrollment.

MB Course

The MB course is an effective, evidence-based intervention originally developed for preventing postpartum depression [37]. The MB course comprises a dozen 1:1 sessions, each designed to last for 15 to 20 minutes. The intervention provides a toolkit of cognitive behavior approaches to promote increasing healthy behaviors, helpful thoughts, and social support within the

context of parenting and bonding with one's baby. Throughout the course, a variety of mindfulness practices are introduced along with mindfulness tips to support integration into daily life. The first 2 MB sessions introduce the cognitive behavior model and discuss the relationship between one's mood and stress and one's behaviors, thoughts, and social interactions. The pleasant activities module (sessions 3-5) focuses on identifying and increasing engagement in pleasant activities alone, with others, and with one's baby. The thoughts module (sessions 6-8) focuses on strategies to increase helpful thoughts and decrease unhelpful thoughts. The contact with others module (sessions 9-11) focuses on increasing positive and supportive interactions with other people.

The MB sessions were delivered weekly or every other week, occasionally with 2 sessions delivered in 1 visit to facilitate timely completion of the intervention, allowing for participant scheduling needs. In addition, participants received a JIT intervention, consisting of 4 SMS text messages, every other day at 7:45 PM on their mobile phones after each completed 1:1 intervention. The SMS text messages included brief messages and links to external content (eg, worksheets, videos, and guided mindfulness practices) and were designed to reinforce the most recent 1:1 intervention content and to encourage skill practice so that the skills become more frequently used to manage stress in one's daily life [38].

Assessments

Baseline demographics and pregnancy history were collected. Depression symptoms at baseline were measured using the Edinburgh Postnatal Depression Scale (EPDS), a validated 10-item self-report assessment that is the most frequently used PD screening tool. The EPDS assesses symptoms of anxiety and depression, both of which are frequent features in perinatal mood disorders and excludes symptoms that are commonly experienced during pregnancy and the postpartum period, such as changes in sleep and appetite. Individual responses are scored on a scale from 0 to 3, with 0 indicating no symptoms of depression and 3 indicating high frequency of depression symptoms. The total score ranges from 0 to 30, with higher scores indicating increased severity and frequency of depression symptoms.

We collected ECG data using a patch-like flexible sensor, BioStampRC, that was placed on the left side of the participant's chest. The BioStampRC is effective in using HR-based features to predict physiological and perceived stress in pregnant women [39]. Participants used a study-provided tablet to start and stop sensor recording and to upload completed recordings to a secure cloud platform at the end of the day. Participants were asked to wear the device during waking hours throughout the 12-week study and could take a few days off to prevent adverse effects of wearing a strong adhesive in the same location every day.

At the time of study enrollment, all participants were asked to identify their usual daily wake and sleep times, and daily EMA questionnaires were programmed to be sent 5 times a day at evenly distributed intervals within each participant's waking hours. Each EMA questionnaire consisted of 12 questions (Table S1 in Multimedia Appendix 1).

Model Development

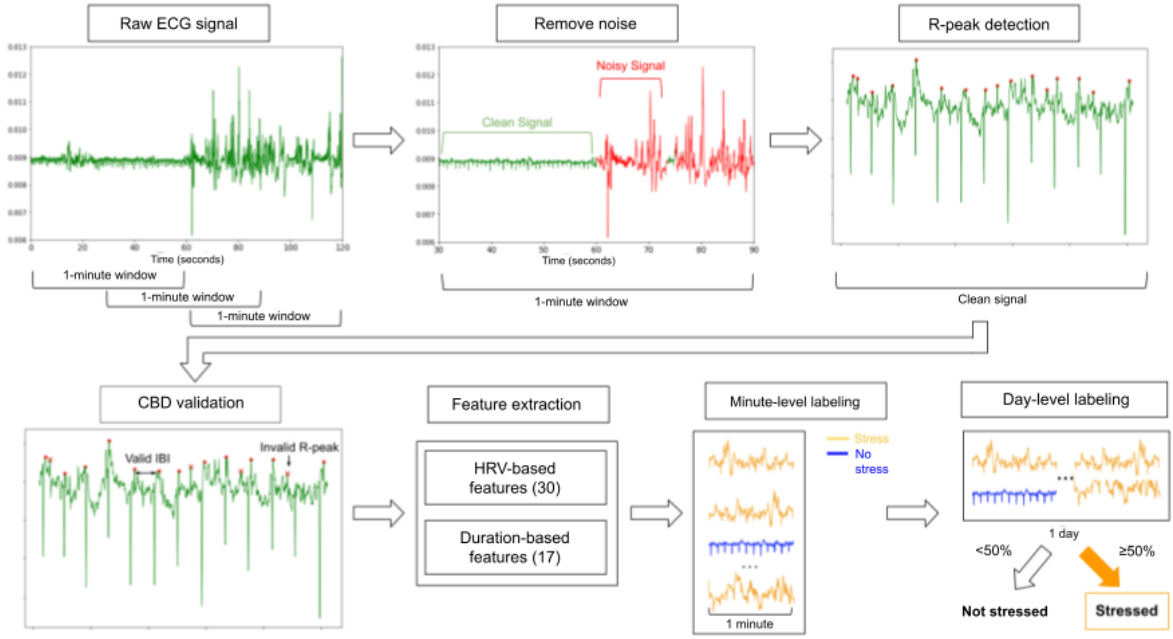
Sensor Data Processing

During the processing pipeline, we first filtered out noisy segments of the ECG signal and calculated interbeat intervals (IBIs) for each 1-minute segment. We then extracted HR variability (HRV)–based features and classified each minute as physiological stress positive or physiological stress negative (Figure 3).

To remove noisy signals caused by sensor deformation because of skin stretching, we first segmented the cleaned ECG signal using a window size of 1 minute with 30 seconds of overlap. Noise was filtered using an ensemble support vector machine (SVM) and neural network noise model described by Zhang et al [40]. The model involves further segmenting of each 1-minute ECG signal into 0.6-second intervals, extracting 3 HRV-based features from the R peaks detected, running both pretrained SVM and neural network classifiers, and classifying each interval as *clean* or *noisy* based on agreement between both models. Within each segment, we discarded segments with >20% noise. We further analyzed the cleaned 1-second segments using a reliability metric: the ratio of the number of data points collected in 1 second divided by the expected sampling rate. The segments were defined as high quality if the reliability was >80%, and low-quality segments were discarded.

Next, we repeated segmenting of clean ECG signals by 1-minute windows with 30 seconds of overlap to extract R peaks and IBIs. We then ran a Shannon energy based algorithm with modifications of nonlinear transformation and first-order Gaussian differentiator for extracting the initial set of R peaks [41,42]. Subsequently, we used the criterion beat difference [43] to filter out R peaks that were inconceivable (ie, out of normal heart rhythm range for humans). We then extracted timestamps between each pair of consecutive R peaks to calculate the IBIs.

Figure 3. Sensor data processing pipeline. The red lines denote noisy segments of the signal by noise model. Red dots are R peaks, and red crosses are invalid peaks filtered by CBD. CBD: criterion beat difference; ECG: electrocardiography; HRV: heart rate variability; IBI: interbeat interval.



Feature Extraction

A full list of features extracted for the model is provided in Table S2 in Multimedia Appendix 1.

Using the minute-level IBIs for each participant per day, we extracted 30 HRV-based features, and from these, we extracted 17 duration-based features. We calculated the average of each feature within a single day to define the day-level feature value. Given that prolonged stress may have different lasting effects compared with brief periods of physiological stress, we calculated an additional 17 duration-based features, which calculated the time spent physiologically stressed while wearing the sensor. As there is no formal duration that defines a *stressful event*, we crafted features that captured a range of minimum consecutive stress-positive window sizes (1, 2, 5, and 10 minutes). To create the features, we first adapted a pretrained SVM grid-search model from King et al [39] that classifies minute-level ECG signal as physiologically stress positive or physiologically stress negative. Specifically, the pretrained SVM model takes 30 statistical features extracted from R-R intervals (the time elapsed between 2 successive R-waves of the QRS signal on the electrocardiogram) and peaks by each minute of ECG signal and outputs the ground truth of stress positive or stress negative. Next, we used 1-, 2-, 5-, and 10-minute consecutive windows to derive total consecutive minutes and episodes from minute-level stress minutes.

We extracted 13 EMA-based features. A total of 12 questions were sent, with 7 questions inquiring about negative emotions and 5 questions inquiring about positive emotions (Table S1 in Multimedia Appendix 1).

These questions included the 4-item Perceived Stress Scale (PSS; ie, the PSS-4) [44], a widely used perceived stress evaluation questionnaire. The response options for each question of the PSS-4 range from 0 to 4, with the final range between 0 and 16, with 0 indicating no stress and 16 indicating very high stress. We calculated the mean score of the responses for each question per day to obtain the 12 daily scores as EMA-based features. By averaging responses to all the PSS-4 questions in a day, we derived the 13th EMA-based feature.

The following 4 intervention-based features were extracted from the 1:1 interventions and JIT interventions: intervention day (ie, whether the prior day was 1:1 intervention day), count intervention (cumulative number of 1:1 interventions received up to the prior day), JIT intervention day (ie, whether a JIT intervention was sent the prior day), and count JIT intervention (cumulative number of JIT interventions received up to the prior day). As the timing of intervention distribution varied, the *count JIT intervention* and *count intervention* variables enabled us to factor in the cumulative number of interventions. If either cumulative variable negatively predicted next-day stress, we were able to suggest that length of participation was negatively correlated with stress levels. If JIT intervention day and intervention day negatively predicted next-day stress, we were able to suggest short-term effectiveness of the interventions because they only indicated prior-day information.

We factored in the following 5 participant characteristics as covariates: age, gestational age at enrollment, number of prior pregnancies, number of prior children, and EPDS score (Table S3 in Multimedia Appendix 1 shows all participant characteristics).

Physiological Stress: Ground Truth

To establish ground truth for next-day physiological stress, we first fed the 30 HRV-based features to the model described by King et al [39] to determine the minute-level stress classification. From the model output, we calculated total

consecutive stress minutes from all 1-minute–level classification results. A day was labeled as physiologically stress positive if the total number of consecutive stress minutes was >50%, based on previously published literature [45].

Perceived Stress: Ground Truth

Ground truth for perceived stress was labeled by calculating the average value of PSS-4 scores throughout a given day. PSS-4 was calculated using the Cohen Perceived Stress Scale by combining the 4 PSS items (PSS-Control, PSS-Overcome, PSS-Confident, and PSS–Your Way). A day was labeled as perceived stress positive if the PSS-4 score was >4.7 [46].

Results

Participants

A total of 16 pregnant women enrolled in the study. Demographics and data captured are shown in Table 1. In total, the participants collected 4157.18 hours of data over a total of 344 days, of which 256 (74.4%) were consecutive, with 114 (44.5%) being nonstressed days and 142 (55.5%) being stressed days. After filtering out noise, 89.2% (3708/4157.18) of the data remained clean for prediction. Participants wore the sensor for a mean of 21.5 (SD 5.21) days, with a mean of 16 (SD 3.28) days being consecutive days. Of the 16 participants, 14 (88%) had consecutive events of perceived stress recorded. Across all participants, a total of 956 days with EMA records were collected, with 881 (92.2%) being consecutive days. Among these 881 consecutive days, there were 412 days (46.7%) being nonstressed and 469 days (53.3%) being stressed. Participants answered a mean of 2.9 (SD 0.89) EMA questions per day. These results are supported by Wakschlag et al [45], who found that pregnant women experience stress an average of 49.9% of the day. Of the 16 participants, 3 (19%) did not wear the sensor for any consecutive days (necessary to predict next-day stress), and therefore data from these participants were discarded. Tables S4 and S5 in Multimedia Appendix 1 shows the quantity of sensor data and EMA data, respectively, collected by each participant.

Table 1. Participant characteristics and the amount of sensor and ecological momentary assessment (EMA) data captured (N=16).

Data captured	Values
Age (years), median (range)	35 (30-39)
Gestational age[a] (weeks), median (range)	11.5 (10-17)
Number of prior children, median (range)	1 (0-2)
Number of prior pregnancies, median (range)	2 (1-5)
EPDS[b] score, mean (SD)[a]	7.2 (3.4)
Sensor data captured[c]	
Days worn, median (range)	23 (5-68)
Consecutive days worn, median (range)	16 (4-59)
Total wear time (hours), median (range)	245.6 (65.6-797.9)
Clean data (%), median (range)	87.9 (64.9-98.9)
Hours worn per day, mean (SD; range)	13.0 (1.55; 9.2-14.5)
EMA data captured[c]	
Days answered, median (range)	71 (30-94)
Consecutive days answered, median (range)	65.5 (24-93)
Total EMAs answered, median (range)	190 (64-346)
EMAs answered per day, mean (SD; range)	2.9 (0.88; 1.2-4.6)

[a]At enrollment.

[b]EPDS: Edinburgh Postnatal Depression Scale.

[c]Data from 3 participants were outliers in wear time and thus excluded from analysis.

Model Validation

Baseline Model Evaluation

We tested 6 widely used machine learning models using the scikit-learn Python package to evaluate the importance of input variables on next-day prediction of perceived and physiological stress [47-49]. In baseline models, we included the following models with default hyperparameters: gradient boost machine (min_samples_split: 5; min_samples_leaf: 2; max_depth: 3), SVM (kernel: rbf; C: 1.0; gamma: “scale”), adaptive boosting (n_estimators: 50), naïve Bayes, decision tree (min_samples_split: 5; min_samples_leaf: 2; max_depth: 3), and random forest (n_estimators: 10; min_samples_split: 5; min_samples_leaf: 2; max_depth: 3). In all baseline models, we used all 69 features as input to predict next day’s physiological and perceived stress and applied 5-fold cross-validation on each model. Each fold consisted of 80% training data and 20% testing data randomly selected from all

participants combined. We adopted commonly used evaluation metrics (precision, recall, and F1 score) for binary classification [50]. The random forest classifier performed best across both types of stress, with an average F1 score of 81.9% when predicting physiological stress and 72.5% when predicting perceived stress.

Correlation-Based Feature Subset Selection

We used correlation-based feature selection (CFS) [51] on our set of 69 features (4 intervention related, 5 covariates, 30 HRV based, 17 duration based, and 13 EMA based). CFS helps evaluate the intrinsic correlations within features to avoid redundancy and high feature-class correlation to maintain or increase predictive power. CFS helped to reduce the number of features, which allowed us to understand which data we may not need to collect in the future or to explain which features are not contributing to the resulting prediction.

Bayesian Optimization

In addition to CFS for selecting the optimal subset of features, we adapted Bayesian optimization based on the work of Snoek et al [52] and the Python implementation package built by Nogueira [53]. In Bayesian optimization, the general performance of the selected machine learning algorithms was modeled as a sample from a Gaussian process, and the nature of the Gaussian process helped to optimize and tune the hyperparameters to further improve the model performance.

Combination of Feature Types

To reduce the burden of data collection by removing features without sacrificing significant predictive power (as measured by the F1 score), we ran random forest with 5-fold cross-validation with various combinations of sensor-, EMA-, and intervention-based data. For both types of stress, we used 6 combinations of data: sensor only, EMA only, intervention only, sensor with EMA, intervention with EMA, and intervention with sensor. We then compared the results with those of a model that used all types of data. For physiological stress predictions, with any combination of data types, the average F1 score remained >73% (Figure 4). For perceived stress predictions, only combinations with EMA data continued to perform well (Figure 5).

Figure 4. 5-fold cross-validation of next-day physiological stress by subset of feature types. EMA: ecological momentary assessment.

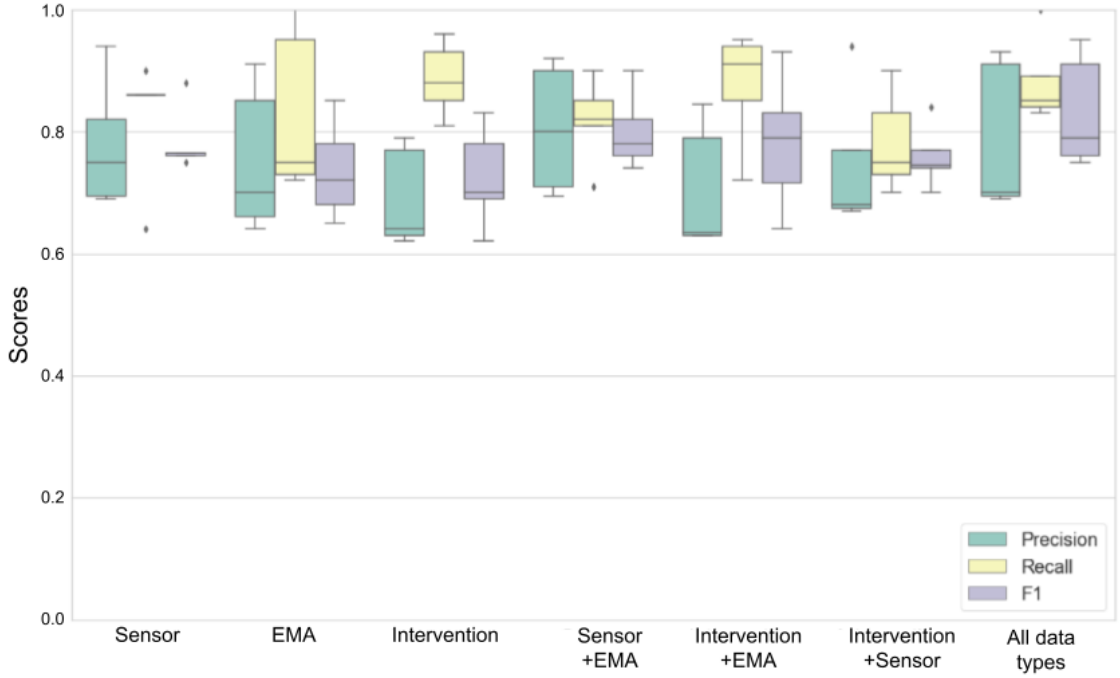
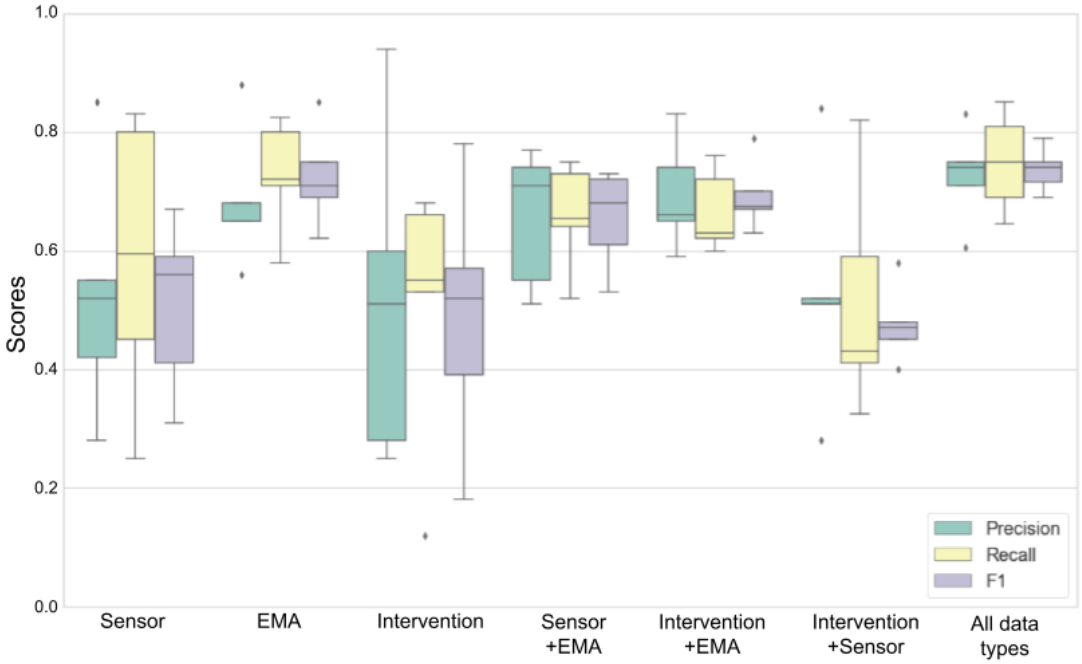


Figure 5. 5-fold cross-validation of next-day perceived stress by subset of feature types. EMA: ecological momentary assessment.



Model Performance

First, CFS was applied to select the subset of features used to build physiological and perceived stress models. Next, we applied Bayesian optimization to all the baseline models. Random forest outperformed the rest of the models after the hyperparameters were optimized (n_estimators, criterion, max_depth, min_samples_split, and max_features) using a range of 200 values for each continuous hyperparameter (eg, n_estimators) and the maximum number of options for each categorical hyperparameter (eg, criterion). The detailed hyperparameters after optimization are presented in Table S6 in Multimedia Appendix 1. The resulting F1 score increased to 83.6% when predicting physiological stress and 74.4% when predicting perceived stress.

Figures 6 and 7 show the results of using the 6 different classifiers to predict next-day physiological stress and perceived stress, respectively, and show F1 scores in descending order using the associated subset of features identified by CFS. These data in table format are shown in Tables S7 and S8 in Multimedia Appendix 1.

Figure 6. Predicting next-day physiological stress by using 6 different machine learning models using 5-fold cross-validation. Boxes indicate the IQR, whiskers indicate the minimum and maximum, and solid lines indicate the median. AdaBoost: adaptive boosting; GBM: gradient boosting machine; SVM: support vector machine.

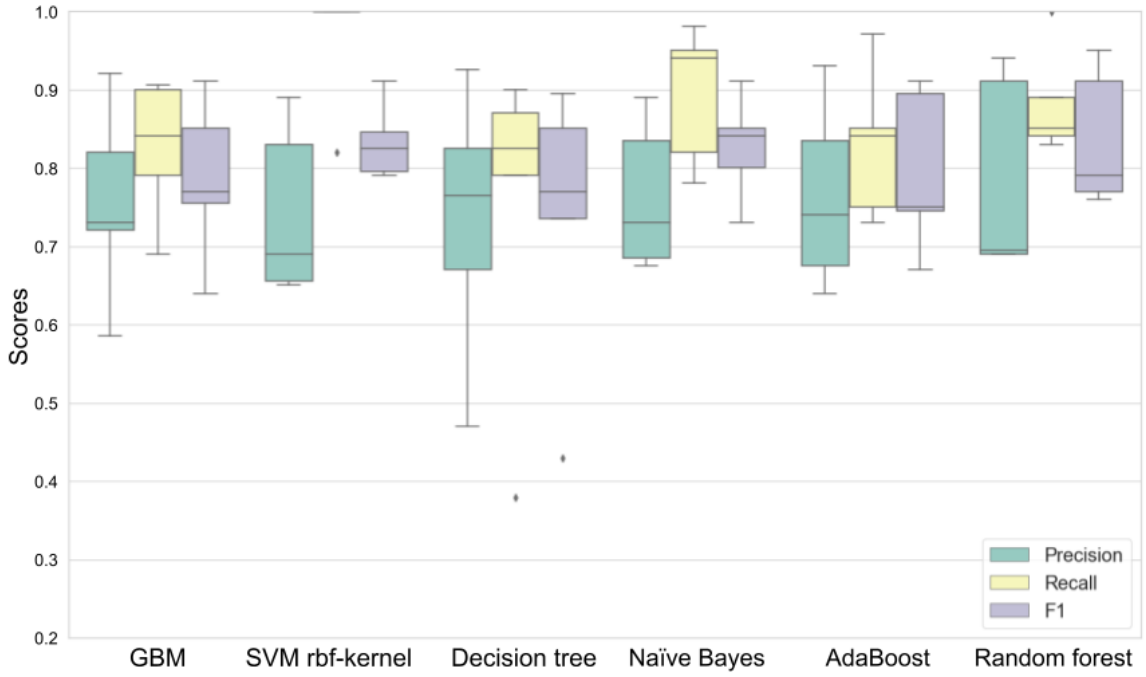
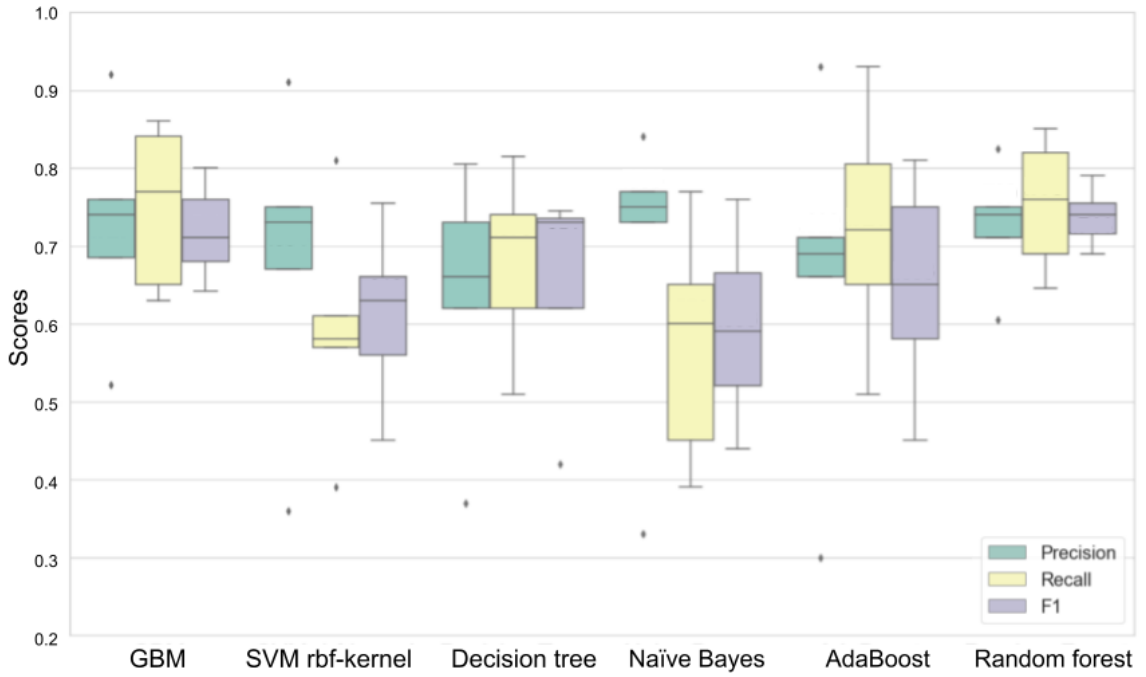


Figure 7. Predicting next-day physiological stress by using 6 different machine learning models using 5-fold cross-validation. Boxes indicate the IQR, whiskers indicate the minimum and maximum, and solid lines indicate the median. AdaBoost: adaptive boosting; GBM: gradient boosting machine; SVM: support vector machine.



Feature Importance

We applied SHAP on the 8 features selected by the random forest model (because random forest performed best with the highest F1 score) to predict physiological stress. The top 5 features ranked by mean absolute SHAP values were as follows: number of consecutive stress minutes by 10-minute minimum threshold, count of interventions, number of consecutive stress minutes percentage by 10-minute minimum threshold, number of children, and PSS-Overcome (Figure 8). The following were also predictive but exhibited mean SHAP values of <0.05, suggesting lower predictive value: JIT intervention day, intervention day, and binary stress.

Similarly, SHAP was applied on 10 features selected by the random forest model to predict perceived stress. The top 6 features ranked by mean absolute SHAP values were as follows: PSS-4, PSS-Control, PSS-Overcome, number of children, happy stress, and content stress (Figure 9). The following features were also predictive but exhibited mean SHAP values of <0.05: worried stress, binary stress, JIT intervention day, and intervention day.

The SHAP analysis of physiological stress (Figure 10) showed that the greater the number of consecutive stress episodes (minimum of 10 minutes per stress event), the more likely the following day would also be a physiologically stressful day. Conversely, the greater the number of count interventions (cumulative number of 1:1 interventions received up to the prior day), the lower the physiological stress the next day.

The SHAP analysis of perceived stress (Figure 11) shows that low values of PSS-4 (<4.0) are characterized by negative SHAP values; this suggests that lower PSS-4 scores are related to a lower likelihood of next-day perceived stress prediction. High values of PSS-Control (>2.0; ie, not feeling as though one can control important things) were generally associated with positive SHAP values: predictions of higher perceived stress the next day.

Figure 8. SHAP summary plot for feature importance in predicting physiological stress. JIT: just-in-time; PSS: Perceived Stress Scale; SHAP: Shapley Additive Explanations.

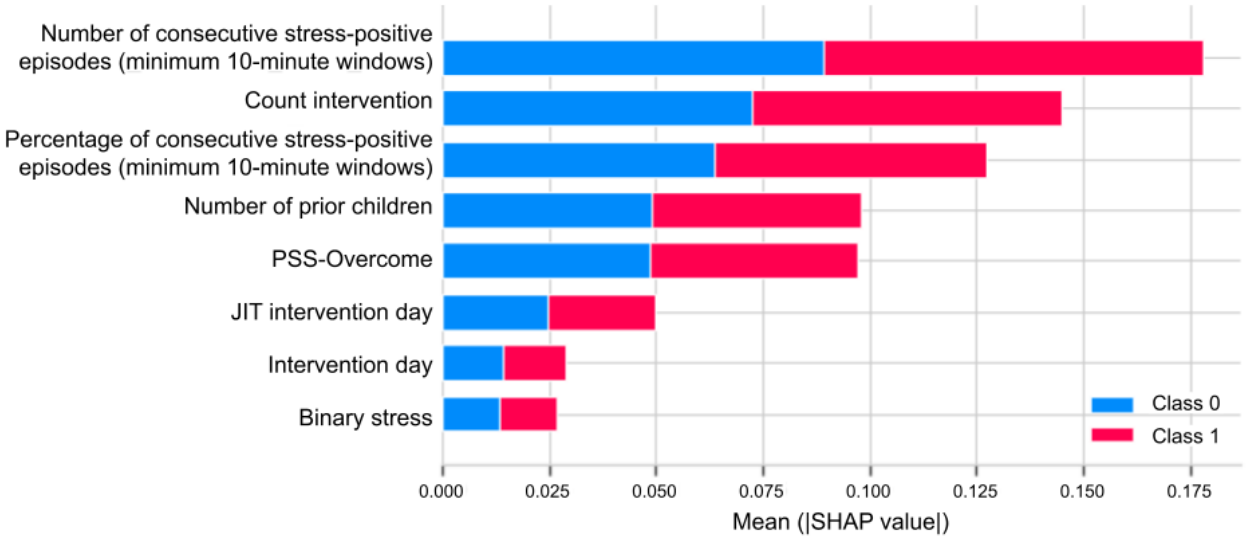


Figure 9. SHAP summary plot for feature importance in predicting perceived stress. JIT: just-in-time; PSS: Perceived Stress Scale; SHAP: Shapley Additive Explanations.

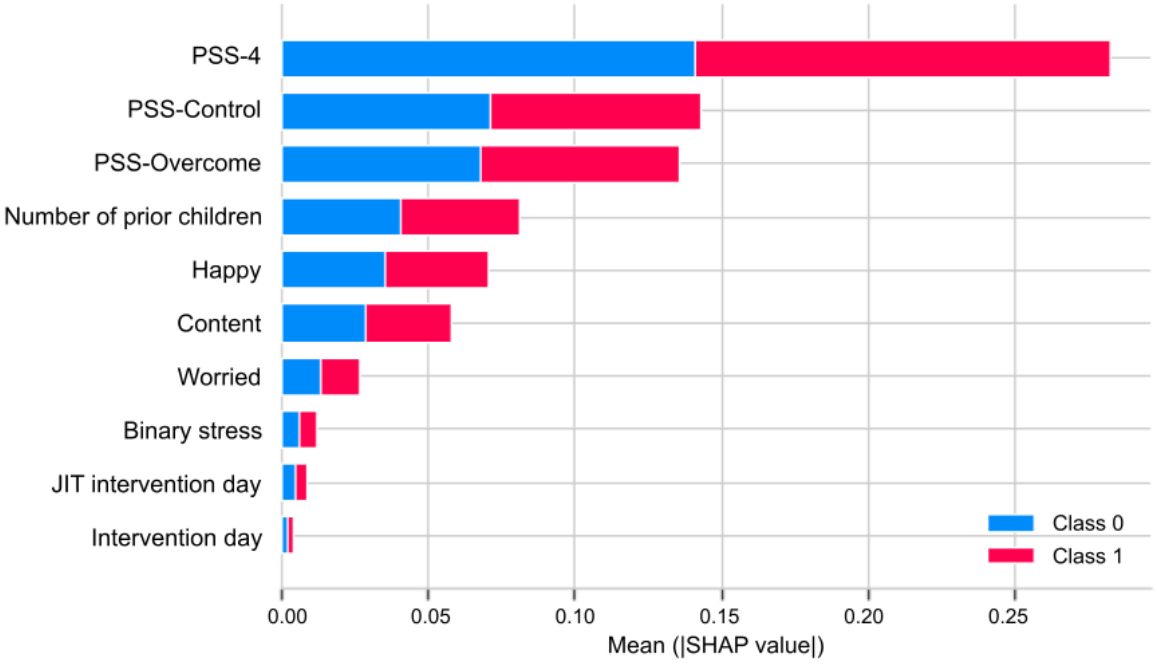


Figure 10. SHAP dependence plot for physiological stress features with values >0.05. SHAP: Shapley Additive Explanations.

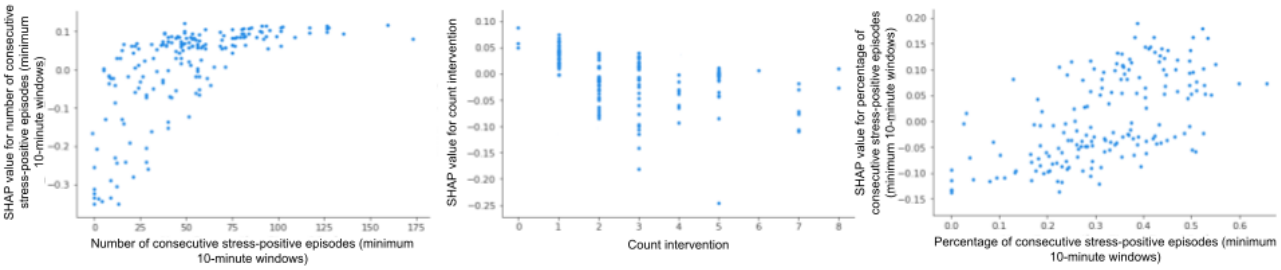
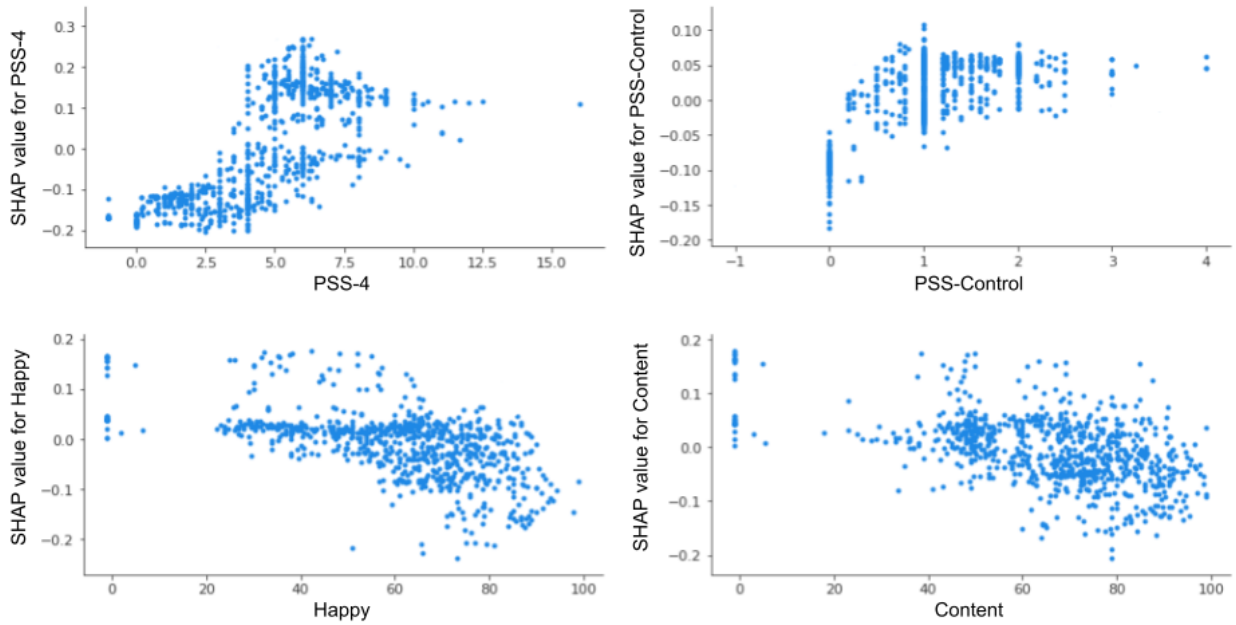


Figure 11. SHAP dependence plot for perceived stress features with values >0.05. PSS: Perceived Stress Scale; SHAP: Shapley Additive Explanations.



Furthermore, we observed a distribution of negative SHAP values for happy stress scores >20 and content stress scores >40, suggesting that higher scores in these areas tend to drive predictions of less stress the following day. However, some observations with positive SHAP values were distributed across a wide range of scores; we suspect these were due to interactions with other features.

The EMA-based feature PSS-Overcome ("Did you feel difficulties piling up so you cannot overcome them?") generally seemed to predict lower levels of next-day physiological stress but higher levels of next-day perceived stress (Figure 12). Although the feature *number of children* scored highly important when predicting next-day physiological stress and perceived stress (Figures 8 and 9, respectively), according to the summary plot (Figure 13), we see variability in how the number of children a mother has influenced physiological and perceived stress. Physiologically, having no child (and therefore being pregnant with a first child) is positively associated with an increase in physiological stress and having 2 children (and therefore being pregnant with a third child) is associated with a reduced probability of physiological stress for the next day. Perceptually, it seems that the first time a mother is pregnant while caring for her first child, she experiences greater stress than when she has no existing children or has already gone through the experience.

Figure 12. SHAP dependence plot for the shared physiological stress (left plot) and perceived stress (right plot) feature PSS-Overcome. PSS: Perceived Stress Scale; SHAP: Shapley Additive Explanations.

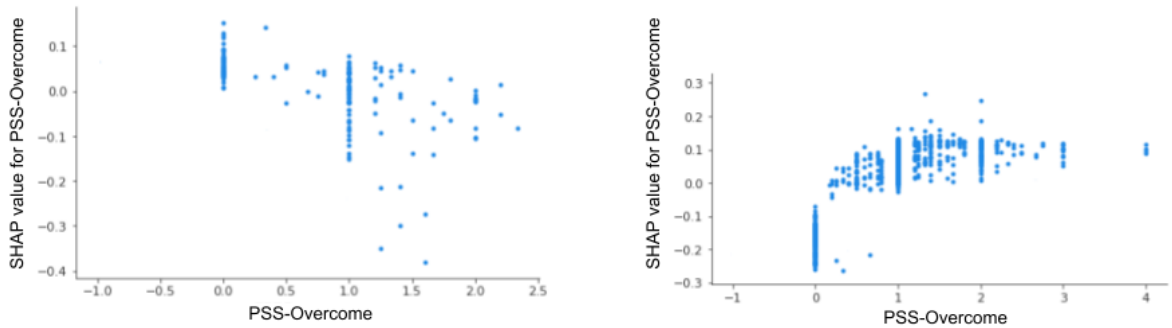
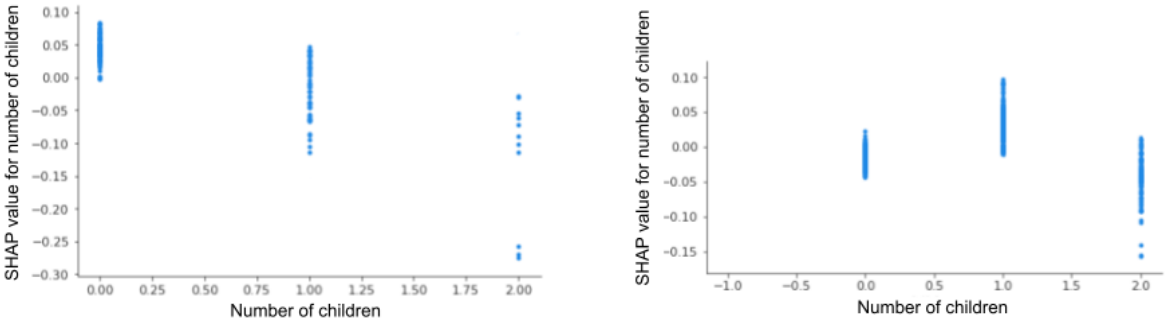


Figure 13. SHAP dependence plot for the shared physiological stress (left plot) and perceived stress feature (right plot) number of children. SHAP: Shapley Additive Explanations.



User Feedback

Of the 14 participants, 10 (71%) completed the full feedback survey on wearability and usability of the sensor (Figure 14). Most (7/10, 70%) of the respondents reported that the device was not painful; however, 20% (2/10) reported extreme pain because of the accompanying strong adhesive of the device and repeated application to the same location every day. When considering physical discomfort when the device was worn, 50% (5/10) of the respondents reported discomfort to be *a little bit* or *not at all*, whereas 50% (5/10) reported discomfort to be ranging from *somewhat* to *extreme*. Most (8/10, 80%) of the respondents found the device easy to use.

To measure the burden of self-report, we hypothesized that frequent EMA surveys were burdensome to participants in longitudinal studies [32] and because of burnout response rates would diminish as participants continued in the study. In this study, participant responsiveness peaked at the second week and continuously decreased throughout the following weeks, with the lowest response rates in the last week of the study (Figure 15). In addition, many participants reported concerns about the frequency of the EMAs. Of the 10 participants who completed the survey, 4 (40%) stated that EMAs were sent too often and 2 (20%) stated that the timing was not always convenient, which may have been because of sending multiple surveys that were ≥2 hours apart in a day.

Figure 14. Participants' feedback regarding BioStampRC usability and wearability.

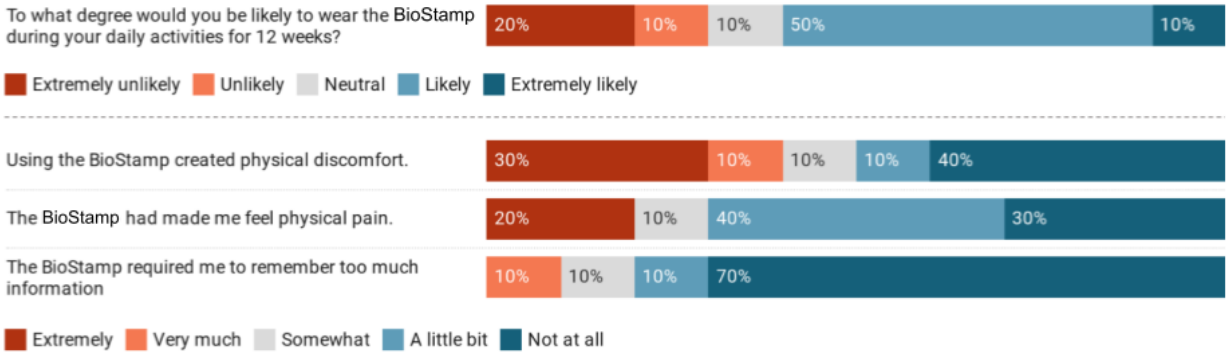
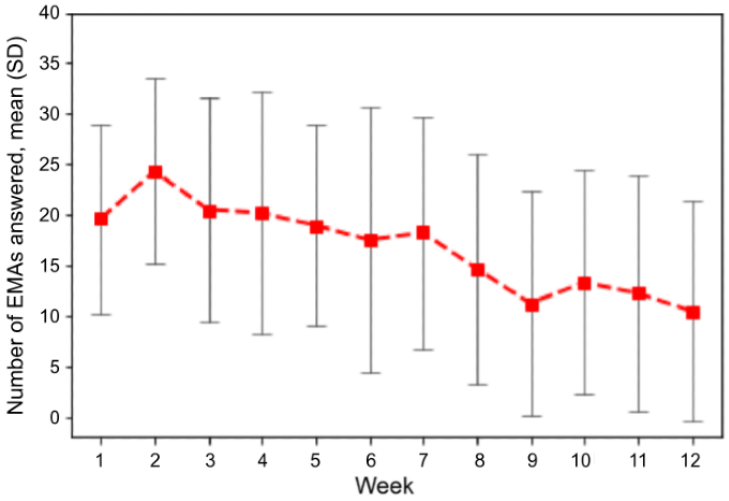


Figure 15. Average response rate to daily EMA surveys per study week. EMA: ecological momentary assessment.



Discussion

Principal Findings

A total of 16 pregnant women enrolled in our pilot study for predicting next-day stress in pregnant women participating in a CBT-based course aimed at reducing stress. Overall, 4157.18 hours of data were collected and participants answered 2838 EMAs. Approximately half (142/256, 55.5%) of the days were determined to be stressed days, which is in line with the results from Wakschlag et al [45], who found that pregnant women experience stress an average of 49.9% of the day. In our study, we identified sensor-based features that best predicted next-day physiological stress and EMA-based features that best predicted next-day perceived stress. Notably, 2 features emerged for predicting both physiological stress and perceived stress: the EMA question PSS-Overcome and the number of children the participant already had.

Our results inform opportunities and challenges with using various measures to predict perceived and physiological stress 1 day in advance and highlight the temporal and relational differences between perceived stress and physiological stress. The significant input variables noted provide opportunity for predictive systems, including machine learning models, to be tailored using these variables in scheduling future interventions. For instance, perhaps sensor data should not be used to predict next-day perceived stress, which can be better predicted using EMA data, the low-burden intervention-based data, or covariates.

Explaining Important Features in Predicting Next-Day Physiological Stress

After feature reduction using only sensor-based data with covariates (average F1 score of 78.3%), all the important features were consistently related to sustained attributes: number of consecutive stress episodes of a minimum duration of 10 minutes, number of children, percentage of wear time episodes spent stress positive (10-minute minimum threshold), number of prior pregnancies, gestation week, age, and depression score. Most features identified to be predictive are not prone to quick change, suggesting that the carryover of physiological stress can be a reflection of chronic stress.

The features with a SHAP value >0.05 that predicted next-day physiological stress were as follows: number of consecutive stress episodes of a minimum duration of 10 minutes, intervention count, number of consecutive perceived stress episodes of a minimum duration of 10 minutes, number of children, and the EMA question PSS-Overcome. Overall, the greater the percentage of episodes classified as stress positive, the more likely the following day was classified as stress positive. This suggests that when there is prolonged physiological stress during a day, it is more likely that the next day will continue to be physiologically stressful. This could also indicate that a minimum duration of 5 minutes is required to reduce the influence of false positives and capture more substantial episodes of stress.

The high feature importance of count intervention, or the cumulative number of interventions the participant received, further suggests the lasting effects of physiological stress and interventions. The more interventions the participant received, the more likely they were to have lower stress the following day up until 4 interventions were received, at which point the effect flattened. A randomized controlled trial would be useful to distinguish whether the interventions were actually effective in lowering next-day stress and whether the mindfulness-based skills were effective in lowering physiological stress. Moreover, in-person visits are costly and affect the scalability of the intervention. Our findings suggest that there may be an optimal number of in-person visits needed to reduce stress. Further research may aim to compare the effects of adjusting the number of in-person visits and its effects on cost and physiological stress reduction.

Explaining Important Features in Predicting Next-Day Perceived Stress

Similar to our analysis of physiological stress, we analyzed the dependencies of features that predicted next-day perceived stress based on SHAP mean values >0.05. The highly predictive features were as follows: PSS-4, PSS-Control, PSS-Overcome, number of children, happy stress, and content stress. Our results show that a participant's perceived ability to control important things were predictive of perceived stress the following day, whereas happy stress and content stress were predictive of lower perceived stress the following day. These results suggest that there is a lasting effect of perceived stress into the following day. The findings that perceived stress can carry over because it may linger in the mind has been identified in previous studies [54], and our findings confirm that this applies to pregnant women. In addition, when studying the different combinations of variable types for their ability to predict next-day perceived stress, we found that models that excluded EMA features performed poorly, suggesting that, unlike prediction of next-day physiological stress, alternative data collection methods do not replace EMA features. However, our user feedback survey showed that participants considered the EMAs in our study to be burdensome, confirming the need for lower-burden self-report models for detection and prediction of stress; thus, future work must confirm the most predictive features of next-day stress before incorporating them into stress-reducing interventions to ensure that the intervention does not lead to excess user burden and thus increased stress.

In terms of next-day perceived stress, increased perceptions of stress can carry over to the following day, as shown in our prior findings and in previous literature [54]. However, it was surprising that, generally, higher values of PSS-Overcome were associated with lower probabilities of next-day physiological stress. A prior study showed a moderate positive correlation (r=0.48) between intended in-laboratory stressors and PSS-Overcome [39]. Our findings suggest that PSS-Overcome may have a reverse predictive value over time (ie, the next day). Future research should investigate the predictive value of these features over time.

Number of Children and Feeling Unable to Overcome Difficulties

The covariate number of children and the EMA question PSS-Overcome were ranked as highly important in predicting

next-day physiological and perceived stress but in opposing valence. This reinforces the narrative that physiological stress and perceived stress are conceptually and temporally divergent. Our results suggest that generally, for any given day, the more children a participant had, the lower their physiological stress was predicted to be the following day. This may be explained by the initial stressful transition into motherhood and gradual acclimation or accumulation of resilience with each new child. Simultaneously, when it comes to predicting next-day perceived stress, participants about to have their second child were generally more likely to feel stressed the following day. Recent research suggests that having a second child does worsen parents' mental health, in particular for women who often bear the brunt of child-rearing tasks [55]. Our results suggest that parents often do not expect that the work in raising a second child would be exponentially greater after already having a child. However, a drop in next-day perceived stress during a third pregnancy may signify that coping strategies for handling pregnancy and an additional child are learned. These findings offer a unique perspective into physiological and perceived experiences of stress through a cross-section of stages throughout the journey of parenthood. Future studies are needed to verify our findings through a longitudinal study.

Future of Mobile Health Systems in Mental Health Prediction and Intervention

Future work in developing mobile health systems that detect physiological and perceived incidences of problematic mental health episodes should investigate and compare the predictive value of sensor-captured data, self-reported measures, and other incidentally captured data through means such as intervention schedules and covariates. In this study, we investigated 2 types of chronic stress: perceived stress and chronic strain, which most consistently predict PD, making it an appropriate target for PD prevention both psychologically and physiologically [1]. Using a multimodal detection system allowed us to not only identify features that most strongly predicted physiological and perceived stress but also allowed us to discover how to minimize the features that are necessary to make next-day predictions. Other developers of mobile health systems for mental health detection may also consider identifying how specific episodes manifest physiologically and perceptually. Ultimately, to create an effective predictive mobile health system with JIT interventions, each component—sensor, EMA, and intervention—must be sustainable and usable.

The accompanying strong adhesive used in the sensors for physiological stress detection was reported by 20% (2/10) of the participants to cause extreme physical pain because of repeated application to the same location every day, which is a barrier to creating such a sustainable system. Prolonged wear and comfort are critical because our findings highlight the importance of predicting 10-minute bouts of physiological stress through sensor data as potential indicators of chronic strain that is likely to persist the following day. In addition, short battery lifetime is a barrier to extended wear. To collect more robust data, increasing battery lifetime and sensor size may stand in opposition to the comfort of participants. Future work should investigate finding a balance between the sensor's robustness and user willingness to wear it.

In the absence of costly and possibly uncomfortable sensors, EMAs were a pathway to predicting next-day perceived or physiological stress. However, responsiveness to EMAs decreased after the first 2 weeks. Studies have shown that user-friendly interfaces and directly useful features such as allowing the data to be viewable to participants and increasing their self-awareness and tracking their progress can increase engagement with EMAs [56,57]. Creating a sustainable system that incorporates the collection of perceived mental health status will require directly providing more value to users. Furthermore, as perceived stress is malleable to psychological intervention whereas chronic stress is not, perceived stress is a viable intervention target [45,58].

Incidentally captured data in the form of interventions, covariates, or other nonsensor passively captured data offer additional opportunity to predict next-day mental health concerns. This category of features is low burden for individuals to collect and may still be strong predictors. These data may also offer more data in the form of contextualization to create a robust system of sensing and intervention at the most opportune times [59,60]. For example, knowing the timing and effectiveness of a recently completed intervention may allow a system to recommend related interventions in the future.

Limitations

To learn about the feasibility of wearing an ECG sensor longitudinally, we conducted a study in a natural setting, but this presents unavoidable natural variations in wear time. For instance, over the 12-week period when participants wore the device, the average wear time was 11.5 hours per day. Although most participants found the device easy to use, albeit somewhat painful (because of the repeated application of the strong adhesive in the same location), the disagreement between perceived stress and physiological stress could be a consequence of participants not wearing the device specifically during stressful moments of the day. Although minutes-ahead predictors may hold for short-term studies, challenges with wearability [61] and sensor quality [62] over time will still likely affect prediction accuracy. Our analysis is limited by the distribution of the participants' ages, which were in the range of 30 to 39 years. Although this was not intentional, it allowed us to collect data from a distribution of first-time pregnancies and non–first-time pregnancies. However, these data do not reflect mothers who may have had children earlier or later in life.

Conclusions

In this work, we used machine learning and SHAP to predict and explain the relationships between potential predictors of next-day physiological stress and perceived stress. We built interpretable models to predict next-day physiological stress with an F1 score of 83.6% and next-day perceived stress with an F1 score of 74.4%. We further identified unique features such as number and percentage of consecutive stress episodes of a minimum duration of 10 minutes to be predictive of next-day physiological stress. Using this technique, we evaluated the feature space of intervention-, sensor-, and EMA-based data to find features that can predict next-day physiological stress and perceived stress. Our results show that it is possible to predict next-day physiological and perceived stress while

reducing the burden of data collection. Our study is the first of its kind in terms of assessing pregnant women over a period of 12 weeks (vs a single day in the study by King et al [39]); however, future studies should validate our models with a larger participant sample. Although tomorrow's stress is imminent, future stress research should consider predicting further future stress at other time points such as the following week to understand the sustained predictive value of these features.

Acknowledgments

This work was supported by Ann & Robert H Lurie Children's Hospital of Chicago and Stanley Manne Children's Research Institute Perinatal Origins of Disease Strategic Research Initiative and Northwestern University's Institute for Innovations in Developmental Sciences. The authors would also like to acknowledge support from the National Institute of Diabetes and Digestive and Kidney Diseases (grants K25DK113242 and R03DK127128) and National Institute of Biomedical Imaging and Bioengineering (grants R21EB030305). The authors also gratefully acknowledge the contributions of colleagues and collaborators Zachary King, Dr Begum Egilmez, Dr William Grobman, Dr Amelie Petitclerc, and Dr Michael Bass. The authors would also like to thank Cristina Barkowski of the intervention team, MC10, Dr Roozbeh Jafari, and Dr John Rogers for their support in using the wearable patch, as well as Dr Krystina Neuman for her constructive feedback.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary figure and tables.
[DOCX File , 648 KB - mhealth_v10i8e33850_app1.docx]

References

1. Yim IS, Tanner Stapleton LR, Guardino CM, Hahn-Holbrook J, Dunkel Schetter C. Biological and psychosocial predictors of postpartum depression: systematic review and call for integration. Annu Rev Clin Psychol 2015;11:99-137 [FREE Full text] [doi: 10.1146/annurev-clinpsy-101414-020426] [Medline: 25822344]
2. McCormick MC, Brooks-Gunn J, Shorter T, Holmes JH, Wallace CY, Heagarty MC. Factors associated with smoking in low-income pregnant women: relationship to birth weight, stressful life events, social support, health behaviors and mental distress. J Clin Epidemiol 1990;43(5):441-448. [doi: 10.1016/0895-4356(90)90132-9] [Medline: 2324784]
3. Satyanarayana VA, Lukose A, Srinivasan K. Maternal mental health in pregnancy and child behavior. Indian J Psychiatry 2011 Oct;53(4):351-361 [FREE Full text] [doi: 10.4103/0019-5545.91911] [Medline: 22303046]
4. Glynn LM, Schetter CD, Hobel CJ, Sandman CA. Pattern of perceived stress and anxiety in pregnancy predicts preterm birth. Health Psychol 2008 Jan;27(1):43-51. [doi: 10.1037/0278-6133.27.1.43] [Medline: 18230013]
5. Lobel M, Cannella DL, Graham JE, DeVincent C, Schneider J, Meyer BA. Pregnancy-specific stress, prenatal health behaviors, and birth outcomes. Health Psychol 2008 Sep;27(5):604-615. [doi: 10.1037/a0013242] [Medline: 18823187]
6. Hoyert DL, Xu J. Deaths: preliminary data for 2011. Natl Vital Stat Rep 2012 Oct 10;61(6):1-51 [FREE Full text] [Medline: 24984457]
7. Xu J, Murphy SL, Kochanek KD, Bastian B, Arias E. Deaths: final data for 2016. Natl Vital Stat Rep 2018 Jul;67(5):1-76 [FREE Full text] [Medline: 30248015]
8. Mulder EJ, Robles de Medina PG, Huizink AC, Van den Bergh BR, Buitelaar JK, Visser GH. Prenatal maternal stress: effects on pregnancy and the (unborn) child. Early Hum Dev 2002 Dec;70(1-2):3-14. [doi: 10.1016/s0378-3782(02)00075-0] [Medline: 12441200]
9. Nimby GT, Lundberg L, Sveger T, McNeil TF. Maternal distress and congenital malformations: do mothers of malformed fetuses have more problems? J Psychiatr Res 1999;33(4):291-301. [doi: 10.1016/s0022-3956(99)00004-7] [Medline: 10404467]
10. Beddoe AE, Paul Yang CP, Kennedy HP, Weiss SJ, Lee KA. The effects of mindfulness-based yoga during pregnancy on maternal psychological and physical distress. J Obstet Gynecol Neonatal Nurs 2009;38(3):310-319. [doi: 10.1111/j.1552-6909.2009.01023.x] [Medline: 19538619]
11. Carolan M, Barry M, Gamble M, Turner K, Mascareñas O. The Limerick Lullaby project: an intervention to relieve prenatal stress. Midwifery 2012 Apr;28(2):173-180. [doi: 10.1016/j.midw.2010.12.006] [Medline: 21371795]
12. Duncan LG, Bardacke N. Mindfulness-based childbirth and parenting education: promoting family mindfulness during the perinatal period. J Child Fam Stud 2010 Apr;19(2):190-202 [FREE Full text] [doi: 10.1007/s10826-009-9313-7] [Medline: 20339571]
13. Dunn C, Hanieh E, Roberts R, Powrie R. Mindful pregnancy and childbirth: effects of a mindfulness-based intervention on women's psychological distress and well-being in the perinatal period. Arch Womens Ment Health 2012 Apr;15(2):139-143. [doi: 10.1007/s00737-012-0264-4] [Medline: 22382281]

14. Guardino CM, Dunkel Schetter C, Bower JE, Lu MC, Smalley SL. Randomised controlled pilot trial of mindfulness training for stress reduction during pregnancy. Psychol Health 2014;29(3):334-349 [FREE Full text] [doi: 10.1080/08870446.2013.852670] [Medline: 24180264]
15. Muzik M, Hamilton SE, Lisa Rosenblum K, Waxler E, Hadi Z. Mindfulness yoga during pregnancy for psychiatrically at-risk women: preliminary results from a pilot feasibility study. Complement Ther Clin Pract 2012 Nov;18(4):235-240 [FREE Full text] [doi: 10.1016/j.ctcp.2012.06.006] [Medline: 23059438]
16. Dennis CL, Dowswell T. Psychosocial and psychological interventions for preventing postpartum depression. Cochrane Database Syst Rev 2013 Mar 28(2):CD001134. [doi: 10.1002/14651858.CD001134.pub3] [Medline: 23450532]
17. Hofmann SG, Asnaani A, Vonk IJ, Sawyer AT, Fang A. The efficacy of cognitive behavioral therapy: a review of meta-analyses. Cognit Ther Res 2012 Oct 01;36(5):427-440 [FREE Full text] [doi: 10.1007/s10608-012-9476-1] [Medline: 23459093]
18. O'Connor E, Senger CA, Henninger ML, Coppola E, Gaynes BN. Interventions to prevent perinatal depression: evidence report and systematic review for the US Preventive Services Task Force. JAMA 2019 Feb 12;321(6):588-601. [doi: 10.1001/jama.2018.20865] [Medline: 30747970]
19. Tandon SD, Ward EA, Hamil JL, Jimenez C, Carter M. Perinatal depression prevention through home visitation: a cluster randomized trial of mothers and babies 1-on-1. J Behav Med 2018 Oct;41(5):641-652. [doi: 10.1007/s10865-018-9934-7] [Medline: 29766470]
20. Kazantzis N, L'Abate L. Handbook of Homework Assignments in Psychotherapy: Research, Practice, and Prevention. Berlin, Germany: Springer Science & Business Media; 2007.
21. Hardeman W, Houghton J, Lane K, Jones A, Naughton F. A systematic review of just-in-time adaptive interventions (JITAIs) to promote physical activity. Int J Behav Nutr Phys Act 2019 Apr 03;16(1):31 [FREE Full text] [doi: 10.1186/s12966-019-0792-7] [Medline: 30943983]
22. Firth J, Torous J, Nicholas J, Carney R, Rosenbaum S, Sarris J. Can smartphone mental health interventions reduce symptoms of anxiety? A meta-analysis of randomized controlled trials. J Affect Disord 2017 Aug 15;218:15-22 [FREE Full text] [doi: 10.1016/j.jad.2017.04.046] [Medline: 28456072]
23. Heron KE, Smyth JM. Ecological momentary interventions: incorporating mobile technology into psychosocial and health behaviour treatments. Br J Health Psychol 2010 Mar;15(Pt 1):1-39 [FREE Full text] [doi: 10.1348/135910709X466063] [Medline: 19646331]
24. Proudfoot J, Parker G, Hadzi Pavlovic D, Manicavasagar V, Adler E, Whitton A. Community attitudes to the appropriation of mobile phones for monitoring and managing depression, anxiety, and stress. J Med Internet Res 2010 Dec 19;12(5):e64 [FREE Full text] [doi: 10.2196/jmir.1475] [Medline: 21169174]
25. Smyth JM, Heron KE. Is providing mobile interventions "just-in-time" helpful? An experimental proof of concept study of just-in-time intervention for stress management. In: Proceedings of the 2016 IEEE Wireless Health. 2016 Presented at: WH '16; October 25-27, 2016; Bethesda, MD, USA p. 1-7 URL: https://sites.psu.edu/shadelab/files/2016/11/Smyth_Heron_2016_JIT_IEEE-1vleki4.pdf [doi: 10.1109/wh.2016.7764561]
26. Goodday SM, Friend S. Unlocking stress and forecasting its consequences with digital technology. NPJ Digit Med 2019 Jul 31;2:75 [FREE Full text] [doi: 10.1038/s41746-019-0151-8] [Medline: 31372508]
27. Sapolsky RM. Why stress is bad for your brain. Science 1996 Aug 09;273(5276):749-750. [doi: 10.1126/science.273.5276.749] [Medline: 8701325]
28. von Känel R. Acute mental stress and hemostasis: when physiology becomes vascular harm. Thromb Res 2015 Feb;135 Suppl 1:S52-S55. [doi: 10.1016/S0049-3848(15)50444-1] [Medline: 25903538]
29. Hamarat E, Thompson D, Zabrucky KM, Steele D, Matheny KB, Aysan F. Perceived stress and coping resource availability as predictors of life satisfaction in young, middle-aged, and older adults. Exp Aging Res 2001;27(2):181-196. [doi: 10.1080/036107301750074051] [Medline: 11330213]
30. Jaimes LG, Llofriu M, Raij A. A stress-free life: just-in-time interventions for stress via real-time forecasting and intervention adaptation. In: Proceedings of the 9th International Conference on Body Area Networks. 2014 Presented at: BodyNets '14; September 29-October 1, 2014; London, UK p. 197-203. [doi: 10.4108/icst.bodynets.2014.258237]
31. Munoz-Organero M, Corcoba-Magana V. Predicting upcoming values of stress while driving. IEEE Trans Intell Transport Syst 2017 Jul;18(7):1802-1811. [doi: 10.1109/tits.2016.2618424]
32. Musthag M, Raij A, Ganesan D, Kumar S, Shiffman S. Exploring micro-incentive strategies for participant compensation in high-burden studies. In: Proceedings of the 13th International Conference on Ubiquitous Computing. 2011 Presented at: UbiComp '11; September 17-21, 2011; Beijing, China p. 435-444. [doi: 10.1145/2030112.2030170]
33. Sarsenbayeva Z, van Berkel N, Hettiachchi D, Jiang W, Dingler T, Velloso E, et al. Measuring the effects of stress on mobile interaction. Proc ACM Interact Mob Wearable Ubiquitous Technol 2019 Mar 29;3(1):1-18. [doi: 10.1145/3314411]
34. Rozet A, Kronish IM, Schwartz JE, Davidson KW. Using machine learning to derive just-in-time and personalized predictors of stress: observational study bridging the gap between nomothetic and ideographic approaches. J Med Internet Res 2019 Apr 26;21(4):e12910 [FREE Full text] [doi: 10.2196/12910] [Medline: 31025942]

35. Umematsu T, Sano A, Taylor S, Picard R. Improving students' daily life stress forecasting using LSTM neural networks. In: Proceedings of the 2019 IEEE International Conference on Biomedical and Health Informatics. 2019 Presented at: BHI '19; May 19-22, 2019; Chicago, IL, USA. [doi: 10.1109/bhi.2019.8834624]
36. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. In: Proceedings of the 31st International Conference on Neural Information Processing Systems. 2017 Presented at: NIPS '17; December 4-9, 2017; Long Beach, CA, USA p. 4768-4777. [doi: 10.5555/3295222.3295230]
37. Le HN, Perry DF, Mendelson T, Tandon SD, Muñoz RF. Preventing perinatal depression in high risk women: moving the mothers and babies course from clinical trials to community implementation. Matern Child Health J 2015 Oct;19(10):2102-2110. [doi: 10.1007/s10995-015-1729-7] [Medline: 25673369]
38. Barrera AZ, Hamil J, Tandon D. Integrating SMS text messages into a preventive intervention for postpartum depression delivered via in-home visitation programs: feasibility and acceptability study. JMIR Form Res 2021 Nov 18;5(11):e30995 [FREE Full text] [doi: 10.2196/30995] [Medline: 34792478]
39. King ZD, Moskowitz J, Egilmez B, Zhang S, Zhang L, Bass M, et al. Micro-stress EMA: a passive sensing framework for detecting in-the-wild stress in pregnant mothers. Proc ACM Interact Mob Wearable Ubiquitous Technol 2019 Sep 09;3(3):91 [FREE Full text] [doi: 10.1145/3351249] [Medline: 32432212]
40. Zhang L, King Z, Egilmez B, Reeder J, Ghaffari R, Rogers J, et al. Measuring fine-grained heart-rate using a flexible wearable sensor in the presence of noise. In: Proceedings of the IEEE 15th International Conference on Wearable and Implantable Body Sensor Networks. 2018 Presented at: BSN '18; March 4-7, 2018; Las Vegas, NV, USA p. 160-164. [doi: 10.1109/bsn.2018.8329683]
41. Kathirvel P, Sabarimalai Manikandan M, Prasanna SR, Soman KP. An efficient R-peak detection based on new nonlinear transformation and first-order Gaussian differentiator. Cardiovasc Eng Tech 2011 Oct 12;2(4):408-425. [doi: 10.1007/s13239-011-0065-3]
42. Manikandan MS, Soman KP. A novel method for detecting R-peaks in electrocardiogram (ECG) signal. Biomed Signal Process Control 2012 Mar;7(2):118-128. [doi: 10.1016/j.bspc.2011.03.004]
43. Berntson GG, Quigley KS, Jang JF, Boysen ST. An approach to artifact identification: application to heart period data. Psychophysiology 1990 Sep;27(5):586-598. [doi: 10.1111/j.1469-8986.1990.tb01982.x] [Medline: 2274622]
44. Cohen S. Perceived stress scale. Mind Garden. 1994. URL: https://www.northottawawellnessfoundation.org/wp-content/uploads/2018/04/PerceivedStressScale.pdf [accessed 2021-05-02]
45. Wakschlag LS, Tandon D, Krogh-Jespersen S, Petitclerc A, Nielsen A, Ghaffari R, et al. Moving the dial on prenatal stress mechanisms of neurodevelopmental vulnerability to mental health problems: a personalized prevention proof of concept. Dev Psychobiol 2021 May;63(4):622-640. [doi: 10.1002/dev.22057] [Medline: 33225463]
46. Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. J Health Soc Behav 1983 Dec;24(4):385-396. [Medline: 6668417]
47. Chung H, Ko H, Kang WS, Kim KW, Lee H, Park C, et al. Prediction and feature importance analysis for severity of COVID-19 in South Korea using artificial intelligence: model development and validation. J Med Internet Res 2021 Apr 19;23(4):e27060 [FREE Full text] [doi: 10.2196/27060] [Medline: 33764883]
48. Jaques N, Taylor S, Azaria A, Ghandeharioun A, Sano A, Picard R. Predicting students' happiness from physiology, phone, mobility, and behavioral data. Int Conf Affect Comput Intell Interact Workshops 2015 Sep;2015:222-228 [FREE Full text] [doi: 10.1109/ACII.2015.7344575] [Medline: 28515966]
49. Sang S, Sun R, Coquet J, Carmichael H, Seto T, Hernandez-Boussard T. Learning from past respiratory infections to predict COVID-19 outcomes: retrospective study. J Med Internet Res 2021 Feb 22;23(2):e23026 [FREE Full text] [doi: 10.2196/23026] [Medline: 33534724]
50. Ferrario A, Demiray B, Yordanova K, Luo M, Martin M. Social reminiscence in older adults' everyday conversations: automated detection using natural language processing and machine learning. J Med Internet Res 2020 Sep 15;22(9):e19133 [FREE Full text] [doi: 10.2196/19133] [Medline: 32866108]
51. Hall MA. Correlation-based feature selection for machine learning. University of Waikato. 1999 Apr. URL: https://www.cs.waikato.ac.nz/~mhall/thesis.pdf [accessed 2021-05-02]
52. Snoek J, Larochelle H, Adams RP. Practical Bayesian optimization of machine learning algorithms. In: Proceedings of the Advances in Neural Information Processing Systems 25. 2012 Presented at: NeurIPS '12; December 3-8, 2012; Lake Tahoe, NV, USA URL: https://proceedings.neurips.cc/paper/2012/file/05311655a15b75fab86956663e1819cd-Paper.pdf [doi: 10.1007/978-1-4842-6579-6_4]
53. Nogueira F. Bayesian Optimization: Open source constrained global optimization tool for Python. GitHub. 2014. URL: https://github.com/fmfn/BayesianOptimization [accessed 2021-05-02]
54. Hovsepian K, Al'Absi M, Ertin E, Kamarck T, Nakajima M, Kumar S. cStress: towards a gold standard for continuous stress assessment in the mobile environment. In: Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing. 2015 Presented at: UbiComp '15; September 7-11, 2015; Osaka, Japan p. 493-504. [doi: 10.1145/2750858.2807526]
55. Ruppanner L, Perales F, Baxter J. Harried and unhealthy? Parenthood, time pressure, and mental health. J Marriage Fam 2019 Apr;81(2):308-326. [doi: 10.1111/jomf.12531]

56. Hartmann R, Sander C, Lorenz N, Böttger D, Hegerl U. Utilization of patient-generated data collected through mobile devices: insights from a survey on attitudes toward mobile self-monitoring and self-management apps for depression. JMIR Ment Health 2019 Apr 03;6(4):e11671 [FREE Full text] [doi: 10.2196/11671] [Medline: 30942693]
57. Ortiz A, Grof P. Electronic monitoring of self-reported mood: the return of the subjective? Int J Bipolar Disord 2016 Dec;4(1):28 [FREE Full text] [doi: 10.1186/s40345-016-0069-x] [Medline: 27900735]
58. Brown H, Krogh-Jespersen S, Tandon D, Graham A, Mackiewicz Seghete K, Wakschlag L. Looking ahead: pre-and perinatal interventions for maternal distress to prevent neurodevelopmental vulnerability. In: Wazana A, Székely E, Oberlander TF, editors. Prenatal Stress and Child Development. Cham, Switzerland: Springer; 2021:595-622.
59. Liao P, Dempsey W, Sarker H, Hossain SM, Al'absi M, Klasnja P, et al. Just-in-time but not too much: determining treatment timing in mobile health. Proc ACM Interact Mob Wearable Ubiquitous Technol 2018 Dec;2(4):179 [FREE Full text] [doi: 10.1145/3287057] [Medline: 30801052]
60. Mishra V, Künzler F, Kramer JN, Fleisch E, Kowatsch T, Kotz D. Detecting receptivity for mHealth interventions in the natural environment. Proc ACM Interact Mob Wearable Ubiquitous Technol 2021 Jun;5(2):74 [FREE Full text] [doi: 10.1145/3463492] [Medline: 34926979]
61. Sarker H, Hovsepian K, Chatterjee S, Nahum-Shani I, Murphy S, Spring B. From markers to interventions: the case of just-in-time stress intervention. In: Rehg JM, Murphy SA, Kumar S, editors. Mobile Health: Sensors, Analytic Methods, and Applications. Cham, Switzerland: Springer; 2017:411-433.
62. Evenson KR, Goto MM, Furberg RD. Systematic review of the validity and reliability of consumer-wearable activity trackers. Int J Behav Nutr Phys Act 2015 Dec 18;12:159 [FREE Full text] [doi: 10.1186/s12966-015-0314-1] [Medline: 26684758]

Abbreviations

CBT: cognitive behavioral therapy
CFS: correlation-based feature selection
ECG: electrocardiography
EMA: ecological momentary assessment
EPDS: Edinburgh Postnatal Depression Scale
HR: heart rate
HRV: heart rate variability
IBI: interbeat interval
JIT: just-in-time
MB: Mothers and Babies
PD: perinatal depression
PSS: Perceived Stress Scale
SHAP: Shapley Additive Explanations
SVM: support vector machine

Edited by L Buis; submitted 26.09.21; peer-reviewed by M Rabbi, A Ferrario; comments to author 08.12.21; revised version received 02.02.22; accepted 13.05.22; published 02.08.22.

Please cite as:
Ng A, Wei B, Jain J, Ward EA, Tandon SD, Moskowitz JT, Krogh-Jespersen S, Wakschlag LS, Alshurafa N
Predicting the Next-Day Perceived and Physiological Stress of Pregnant Women by Using Machine Learning and Explainability: Algorithm Development and Validation
JMIR Mhealth Uhealth 2022;10(8):e33850
URL: https://mhealth.jmir.org/2022/8/e33850
doi:10.2196/33850
PMID:35917157

Original Paper

Mediators of Effects on Physical Activity and Sedentary Time in an Activity Tracker and Behavior Change Intervention for Adolescents: Secondary Analysis of a Cluster Randomized Controlled Trial

Simone Johanna Josefa Maria Verswijveren[1], PhD; Gavin Abbott[1], PhD; Samuel K Lai[1], MHealthProm; Jo Salmon[1], PhD; Anna Timperio[1], PhD; Helen Brown[1], PhD; Susie Macfarlane[2], PhD; Nicola D Ridgers[1,3], PhD

[1]Institute for Physical Activity and Nutrition, School of Exercise and Nutrition Sciences, Deakin University, Burwood, Australia

[2]Learning Futures, Deakin University, Geelong, Australia

[3]Alliance for Research in Exercise, Nutrition and Activity, University of South Australia, Adelaide, Australia

Corresponding Author:
Simone Johanna Josefa Maria Verswijveren, PhD
Institute for Physical Activity and Nutrition
School of Exercise and Nutrition Sciences
Deakin University
221 Burwood Hwy
Burwood, 3125
Australia
Phone: 61 03 9244 6100
Email: s.verswijveren@deakin.edu.au

Abstract

Background: Adolescence is a critical age where steep declines in physical activity and increases in sedentary time occur. Promoting physical activity should therefore be a priority for short- and long-term health benefits. Wearable activity trackers in combination with supportive resources have the potential to influence adolescents' physical activity levels and sedentary behavior. Examining the pathways through which such interventions work can inform which mediators to target in future studies.

Objective: The aim of this paper is to examine the impact of the Raising Awareness of Physical Activity (RAW-PA) intervention on potential mediators of behavior change after intervention, and whether these mediated the intervention effects on physical activity and sedentary time at 6-month follow-up.

Methods: RAW-PA was a 12-week intervention, grounded in social cognitive theory and behavioral choice theory, aimed at increasing physical activity among inactive adolescents through combining a wearable activity tracker with digital resources delivered via a private Facebook group (n=159 complete cases). The targeted potential mediators were identified from previous studies conducted in adolescents and included self-efficacy, peer support, family support, teacher support, self-regulation strategies, barriers, and enjoyment. Outcomes included sedentary time as well as light- and moderate-to-vigorous–intensity physical activity. A series of mixed linear models were used to estimate intervention effects on physical activity and sedentary behavior at follow-up and on potential mediators after intervention and to test whether there were indirect effects of the intervention on physical activity and sedentary behavior via mediators.

Results: Adolescents in the intervention group (n=75) engaged in higher sedentary time and lower light intensity at 6-month follow-up compared to the wait-list controls (n=84). There were no intervention effects for moderate-to-vigorous–intensity physical activity. The intervention group perceived more barriers to physical activity than the wait-list control group at 6-month follow-up (mean adjusted difference=1.77; 95% CI 0.19-3.34; *P*=.03). However, indirect effects for each outcome were not statistically significant, indicating that perceived barriers to physical activity did not mediate intervention effects for physical activity or sedentary time.

Conclusions: RAW-PA did not beneficially impact hypothesized mediators in these inactive adolescents, despite strategies being designed to target them. This suggests that the lack of overall intervention effects on physical activity and sedentary time observed in the RAW-PA study could be due to the limited impact of the intervention on the targeted mediators. Future studies should consider different strategies to target theoretically informed potential mediators and identify intervention strategies that

effectively target key mediators to improve physical activity among inactive adolescents. Finally, intervention effects according to level of wearable tracker use or level of engagement with the intervention should be explored. This may provide important insights for designing successful wearable activity tracker interventions.

Trial Registration: Australian New Zealand Clinical Trials Registry ACTRN12616000899448; https://anzctr.org.au/Trial/Registration/TrialReview.aspx?id=370716&isReview=true

International Registered Report Identifier (IRRID): RR2-10.1186/s12889-016-3945-5

(JMIR Mhealth Uhealth 2022;10(8):e35261) doi:10.2196/35261

KEYWORDS

movement behavior; youth; accelerometry; Fitbit; correlates; correlate; physical activity; exercise; randomized controlled trial; RCT; control trial; Australia; adolescent; adolescence; teenager; sedentary; cognitive theory; behavioral theory; wearable; tracker; tracking device; clinical trial

Introduction

Most adolescents worldwide do not engage in sufficient physical activity to benefit health, including cardiometabolic health, and cognitive development [1,2]. This high prevalence of inactivity is evident across low-, middle-, and high-income countries [1]. By contrast, emerging research has shown that excessive sedentary behavior, which includes activities such as TV viewing, may be detrimentally associated with adolescents' health [3]. As adolescence is a critical age where steep declines in physical activity occur [4], particularly in adolescents living in areas of socioeconomic disadvantage [5], and as youth engage more in total sedentary time when they grow older [6], promoting physical activity should be a priority for short- and long-term health benefits [7].

The past decade has seen advances in the development of commercially available wearable activity trackers. Such devices can measure physical activity, enable self-monitoring, and provide real-time feedback. Wearable activity trackers are increasingly being used in physical activity research as both a measurement and intervention tool, while some devices have the capability to provide feedback and alerts for periods of sedentary time [8,9]. Emerging evidence suggests that these trackers have the potential to be used as intervention tools as adolescents perceive them as easy to use and useful [10], and that they increase their motivation to be physically active [11]. However, surprisingly few studies have used activity trackers in physical activity interventions for adolescents, and the scarce evidence is mixed. While one review showed that these devices are feasible for increasing physical activity and decreasing sedentary behavior [12], others have found that effectiveness for physical activity, particularly long-term, has not yet been established [9,11]. More research is required to understand the effect of using activity trackers on adolescents' daily behaviors, which include physical activity of different intensities and sedentary time [13].

Previous technology-based physical activity interventions using activity trackers have shown that multicomponent interventions may be more effective than using wearables as a stand-alone strategy [9,11]. That is, combining activity trackers with additional resources may be a useful strategy to change physical activity behaviors [14]. Examples include web-based advice to increase social support [14,15], tailored advice [16], incentives [17], and counselling [18] in combination with activity trackers. Although some recent activity tracker interventions have started to incorporate additional components [9,12], no previous activity tracker studies have examined potential mediators of change in physical activity and sedentary behavior. Such evidence would provide insights into why the interventions may or may not have been effective. Examining mediators of behavior change may help to identify the pathway through which activity tracker interventions work in adolescents [19,20] and may inform the design and delivery of such interventions in the future.

The Raising Awareness of Physical Activity (RAW-PA) study was a multicomponent intervention for adolescents living in areas of socioeconomic disadvantage that aimed to integrate more physical activity into their day by combining an activity tracker with digital resources that specifically targeted evidence-based behavior-change techniques. Grounded in social cognitive theory [21] and behavioral choice theory [13], the intervention targeted multiple potential mediators of adolescent physical activity. While this intervention did not increase inactive adolescents' accelerometer-derived and self-reported physical activity levels immediately after intervention [8], it provides a unique opportunity to explore potential cognitive, behavioral, and interpersonal mediators of physical activity and sedentary behavior. This is important as identifying which factors are on the mediating pathway between the intervention and the targeted outcomes can inform the development of future physical activity interventions. Therefore, the aim of this study was to examine the impact of the Raising Awareness of Physical Activity (RAW-PA) intervention on potential mediators of behavior change after intervention, and whether these mediated the intervention effects on physical activity and sedentary time at 6-month follow-up.

Methods

Ethics Approval

Ethical approval was obtained from the Deakin University Human Research Ethics Committee (2016–179) and the Victorian Department of Education and Training.

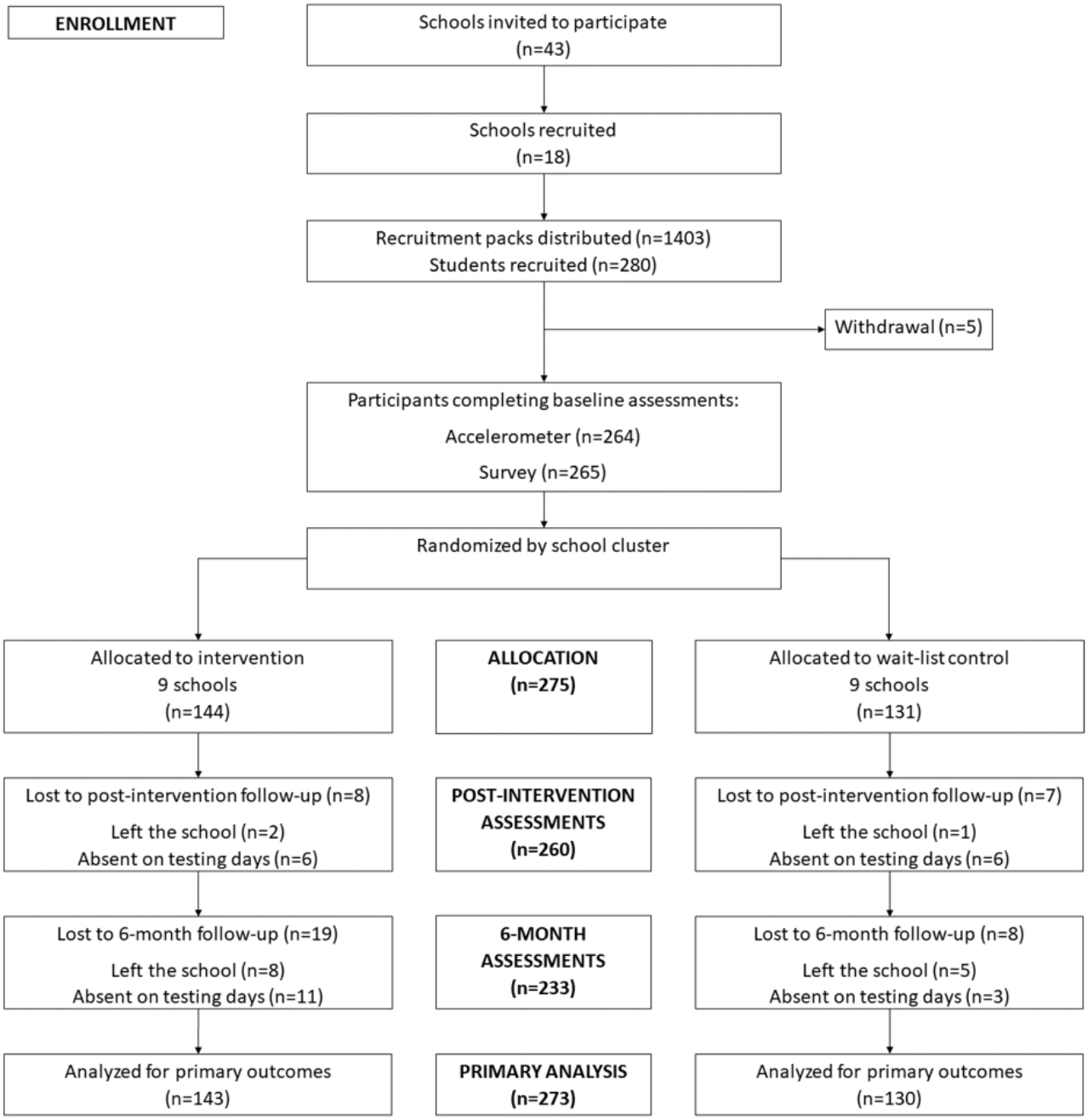
Study Design, Recruitment, and Participants

The RAW-PA study was evaluated using a cluster-randomized controlled trial design, with schools being the unit of randomization [14]. The trial is registered with the Australian

and New Zealand Clinical Trials Registry (ACTRN12616000899448). The protocol for this trial has been previously published [14]. In brief, RAW-PA combined a Fitbit Flex (wearable activity tracker; core component) and the accompanying app with digital behavior change resources that were accessible via social media. Schools were eligible to participate if they were located in socioeconomically disadvantaged suburbs based on a Socio-Economic Index for Areas (SEIFA) [22] score of ≤5. In total, 18 schools were recruited and assigned to the intervention group (n=9) or wait-list control (n=9) using a computer-based random number generator.

The eligibility criteria for participants included the following: (1) being ≥13 years old; (2) having access to the internet outside of school; (3) not engaging in regular organized physical activity or sport outside of school; (4) not meeting physical activity guidelines of ≥60 minutes of moderate to vigorous physical activity (MVPA) every day; and (5) not a current or past owner of an activity tracker. All eligible students (n=280) who returned informed written parental consent and student assent were recruited into the study [8]. As 5 students withdrew before baseline data collection, a total of 275 students participated. Of those, 2 participants did not provide any data; thus, 273 participants were considered for analyses (depending on whether they provided complete data) [8]. A flow diagram of the participants is shown in Figure 1.

Figure 1. Participant flow diagram.



Intervention

Developed using participatory research principles [14], the 12-week intervention combined Fitbit Flex and the accompanying app with interactive individual or weekly *missions*, including behavior change resources (eg, infographics, videos, and social forums) that were accessible via a private, researcher-moderated Facebook group, and new content alerts in the student's own time [14]. Each weekly mission focused on how to integrate low-cost, everyday physical activity into daily life to facilitate real-world implementation [23] and provided participants with opportunities to learn and practice behavior change techniques. The weekly *missions* centered on different intervention objectives (eg, social support for physical activity), and the digital behavior change resources targeted a range of behavior change techniques and mediators [14]. The behavior change techniques were mapped against key determinants or potential mediators identified from social cognitive theory [21] and behavioral choice theory [13]. These recognize that physical activity and sedentary behavior are influenced by factors operating at multiple levels including intrapersonal (eg, self-efficacy) and interpersonal (eg, peer support and influences) [14]. Seven potential mediators of physical activity and sedentary behavior were targeted, which were as follows: self-efficacy, peer support, family support, teacher support, self-regulation strategies, perceived barriers to physical activity, and physical activity enjoyment [14,19,20,24,25]. These were selected as they have been previously identified as potential mediators of physical activity (perceived barriers to physical activity, teacher support, and physical activity enjoyment) [14,19,20,24] or both physical activity and sedentary behavior (self-efficacy, peer support, family support, and self-regulation strategies) [25]. The resources matched the weekly missions to reinforce key messages. For example, in "Week 3," social support was targeted through encouraging participants to support their peers to increase their activity levels. The intervention also incorporated 1 week of content on breaking up sitting time with physical activity. The participants in the wait-list control group were provided access to all resources after the 6-month follow-up assessments.

Measures

All assessments were conducted at baseline (T0), immediately after intervention (T1), and at 6-month follow-up (T2) by trained research assistants using standardized protocols.

Accelerometry

Physical activity and sedentary time were measured using GT3X+ ActiGraph accelerometers (ActiGraph, Pensacola, Florida, USA). The participants were instructed to wear the accelerometer on the hip for 8 consecutive days at each time point during waking hours (except during water-based activities). Raw acceleration data were downloaded and processed into 15-second epochs using manufacturer software (ActiLife). Data were then processed in a customized Microsoft Excel macro to identify sedentary time and time in light-intensity physical activity (LPA) and MVPA. Nonwear time was defined as 60 minutes of consecutive zeroes [26]. Sedentary time was calculated as the total time spent in ≤25 counts per epoch [27]. Age-specific thresholds were used to determine time spent in MVPA [28]. The time spent in between the thresholds for sedentary time and MVPA was classified as LPA. Total sedentary time, LPA, and MVPA were averaged across valid days. A valid day was defined as ≥8 hours on weekdays and ≥7 hours on weekends [29]. Adolescents who provided valid data on any 3 or more days were included for analysis [29].

Demographic Characteristics

The participants self-reported sex, age, and date of birth at each time point via a web-based survey. Area-level socioeconomic status (SES) of the school location using the participants' postcode was obtained via SEIFA [22]. The SEIFA scores were categorized into decile data for use in the analyses.

Mediators

The 7 potential mediators (self-efficacy, peer support, family support, teacher support, self-regulation strategies, perceived barriers to physical activity, and physical activity enjoyment [14,19,20,24,25]) were assessed at all time points using items that were validated for youth [30-32]. The scales used to assess self-efficacy, peer support, family support, teacher support, and self-regulation strategies were adapted from a previously designed instrument for assessing social cognitive measures related to physical activity behaviors [30]. This instrument was developed based on constructs from social cognitive theory by Bandura [21]. The scales have acceptable reliability in adolescents (Cronbach α=.69-.79; test-retest intraclass correlation=0.86-0.91) [30]. The self-efficacy scale included five items (eg, "I do not feel comfortable using local facilities to be active" [30]), which were assessed using a 6-point Likert-type scale ranging from "strongly disagree" (1) to "strongly agree" (6). Social support from peers and family was assessed using 9 items (eg, "In the past three months, how often did your friends encourage you to be active?" [30]) on a 5-point Likert-type scale ranging from "never" (1) to "always" (5). These items were then adapted to assess social support from teachers (4 items). Self-regulation strategies were assessed using 6 items with a 5-point Likert-type scale, from "never" (1) to "always" (5). Example items included "In the past three months, how often did you keep track of how much physical activity you did?" [30].

Perceived barriers to physical activity were assessed using 9 items drawn from the Adolescent Physical Activity Perceived Barriers and Benefits Scale (eg, "I have minor aches and pains from activity") [32]. The participants responded to each item using a 4-point Likert-type scale ranging from "not at all true" (1) to "very true" (4). This scale has acceptable test-retest reliability (r=0.71) and internal consistency (Cronbach α=.79) [32] in adolescents. Enjoyment of physical activity was assessed using the 16-item Physical Activity Enjoyment Scale [31]. The participants used a 5-point Likert-type scale ranging from "Disagree a lot" (1) to "Agree a lot" (5) to answer questions such as "When I am active, I enjoy it". The Physical Activity Enjoyment Scale has been validated for use in adolescents [31].

The overall scores for each of these variables were created by summing individual item scores. These scores were then used as indicators for self-efficacy (range: 5-30), peer support (range:

5-25), family support (range: 4-20), teacher support (range: 4-20), self-regulation strategies (range: 6-30), perceived barriers to physical activity (range: 9-36), and physical activity enjoyment (range: 16-80).

Analytical Sample

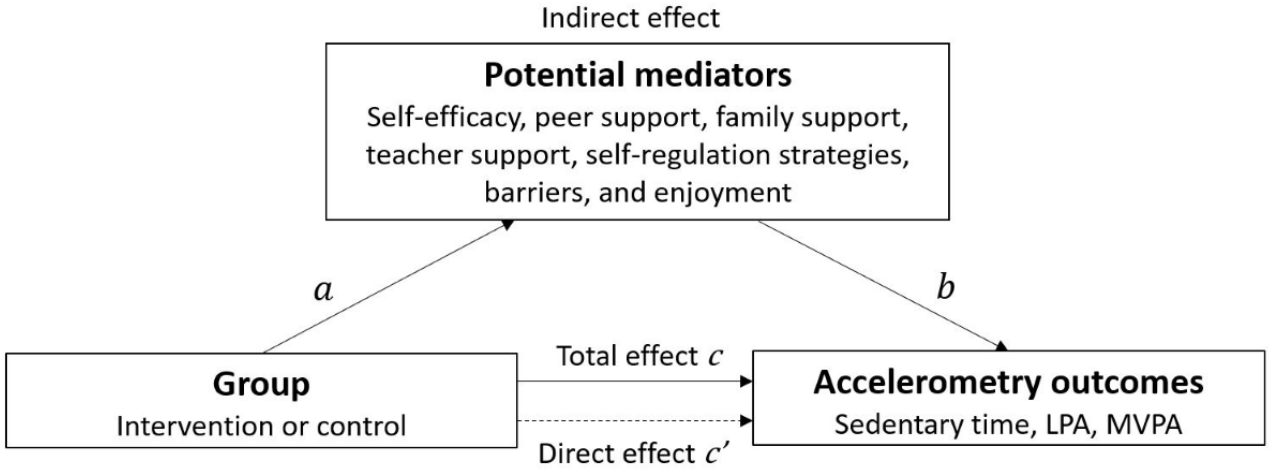
Of the 273 participants who were considered for inclusion in the analyses, baseline accelerometer data were available for 264 (96.7%) students and survey assessments for 265 (97.1%). Valid accelerometer data were available for 246 (90.1%) participants at baseline (T0), 198 (72.5%) after intervention (T1), and 193 (70.7%) at 6-month follow-up (T2). The complete case analysis sample (n=159; 58.2%) included those with full data for covariates (sex and SES), mediators at baseline (T0) and after intervention (T1), and valid accelerometry at baseline (T0) and 6-month follow-up (T2).

Statistical Analysis

Baseline (T0) demographic characteristics as well as baseline (T0) and 6-month follow-up (T2) accelerometry variables (mean [SD]) of the analytic sample were calculated and descriptively compared with excluded participants. As per the Consolidated Standards of Reporting Trials recommendations [33], baseline characteristics for the wait-list control and intervention groups were presented separately but were not compared using inferential tests. Inspection of histograms of the physical activity and sedentary time variables indicated all were normally distributed. Moreover, the median, 25th, and 75th percentile values were quite consistent with values expected for a normal distribution (ie, mean and mean [SD 0.67], respectively). For the mediation analyses, linear mixed models were used, with random intercepts for schools to account for clustering by schools. All models adjusted for sex, school-area SES, and average accelerometer wear time. The covariates included in this study were selected due to potential sex differences and socioeconomic differences in the adolescents' physical activity and sedentary behavior. Wear time was also included due to differences in accelerometer wear time by adolescents in the study. Stata SE 15 (StataCorp) was used to conduct analyses, and statistical significance was set at $P<.05$.

Mediation analysis of randomized controlled trials is important both when an intervention effect on the outcome is observed and when it is not [34]. In the latter case, which is relevant to this intervention [8], the inspection of intervention effects on potential mediators can help to elucidate why the intervention did not affect the outcome (ie, some mediators may play a *suppression* role), for example, because the intervention failed to impact important intermediate factors (*a* path) or that the intermediate factors did not impact the outcome as hypothesized (*b* path). Accordingly, modern approaches to mediation emphasize the importance of testing the *a* and *b* paths irrespective of the total intervention effect on the outcome. A visual representation of the hypothesized mediator model, including the several pathways assessed, is depicted in Figure 2. In this framework, the total effect is the overall effect of the exposure (intervention) on physical activity and sedentary behavior (outcome variables), while the direct effect is the effect of the exposure (intervention) on physical activity and sedentary behavior that operates independently of the mediator in question [35]. First, the models were fitted to estimate the intervention effect on each potential mediator after intervention (T1; *a* path), while adjusting for baseline levels of the mediator. For potential mediators on which there was a statistically significant intervention effect at T1, single-mediator analyses were conducted to determine the indirect effect of the intervention on physical activity and sedentary time at 6-month follow-up (T2) via the mediator (assessed after intervention [T1]). This involved fitting a model while simultaneously estimating the effects of the intervention (direct effect; *c'* path) and the potential mediator (*b* path) on each outcome variables at 6-month follow-up (T2). The mediated, or indirect, effect (ie, the portion of the exposure-outcome relationship that occurs via the mediator) was calculated as $a \times b$, following the "product of coefficients" method, and this quantity was bootstrapped with 1000 resamples to produce percentile-based 95% CIs [35]. The models were also fitted to estimate the intervention effect on each outcome (*c* path). All models in this stage were adjusted for baseline levels of both the mediator and total physical activity or sedentary time (depending on outcome of interest).

Figure 2. Visual representation of the hypothesized mediator model. Pathway a represents the intervention effect on potential mediators, b is the effect of the potential mediator on the behavioural outcome while adjusting for (independent of) intervention group, c is the total effect of the intervention on the outcome, c' (direct effect) is the effect of the intervention on the behavior outcome independent of the potential mediator, while the indirect effect is calculated as a×b and represents the portion of intervention effect on the outcome, which occurs via the mediator. LPA: light-intensity physical activity; MVPA: moderate-to-vigorous–intensity physical activity.



Results

Participants

Baseline (T0) participant characteristics and accelerometry variables at baseline and 6-month follow-up (T2) are shown in Table 1. The participants were approximately 14 years old at baseline, and there were slightly more females than males in both groups (wait-list control: 38% [32/84]; intervention: 45% [34/75]). On average, the participants spent approximately 70% of accelerometer wear time at baseline being sedentary, 25% in LPA and 5% in MVPA. The participants in the analytic sample (n=159) were more likely to be female (93/159, 58% vs 48/114, 42%) and come from the highest eligible school SES decile (47/159, 30% vs 12/114, 11%) compared to the participants who were excluded from the analyses. However, the participants were similar in terms of baseline mediators, physical activity, and sedentary time (data not shown).

Table 1. Baseline (T0) demographic characteristics as well as baseline (T0) and 6-month follow-up (T2) accelerometry variables (n=159).

Characteristics	Control (n=84)	Intervention (n=75)
Baseline age (years), mean (SD)	13.7 (0.4)	13.7 (0.4)
Sex, n (%)		
Male	32 (38)	34 (45)
Female	52 (62)	41 (55)
School-area SES[a] decile, n (%)		
1	21 (25)	22 (29)
2	0 (0)	25 (33)
3	10 (12)	17 (23)
4	12 (14)	5 (7)
5	41 (49)	6 (8)
Activity variables, mean (SD)		
Baseline (T0)		
Average daily wear time (min)	744.6 (99.0)	753.9 (149.2)
Average daily sedentary time (min)	512.3 (90.9)	541.3 (146.6)
Average daily LPA[b] (min)	192.3 (53.6)	178.2 (40.8)
Average daily MVPA[c] (min)	39.9 (18.8)	34.5 (21.0)
6-month follow-up (T2)		
Average daily wear time (min)	769.1 (105.9)	751.3 (155.1)
Average daily sedentary time (min)	538.7 (100.1)	558.7 (149.6)
Average daily LPA (min)	193.8 (53.4)	163.5 (36.2)
Average daily MVPA (min)	36.6 (18.6)	29.0 (14.5)

[a]SES: socioeconomic status.

[b]LPA: light-intensity physical activity.

[c]MVPA: moderate-to-vigorous–intensity physical activity.

Mediation Analysis

The mean scores for each mediator at baseline (T0) and after intervention (T1) are shown in Table 2, along with estimated intervention effects after intervention (*a* path). There was evidence of an intervention effect on the perceived barriers to physical activity score after intervention, with adolescents in the intervention group perceiving more barriers to physical activity compared to those in the wait-list control group (mean adjusted difference=1.77; 95% CI 0.19, 3.34; *P*=.03). There was little evidence of intervention effects for the remaining potential mediators.

Table 2. Estimated effects of the intervention on potential mediators immediately after intervention (a path; n=159). Models were adjusted for sex, school socioeconomic status, accelerometer wear time, and baseline levels of potential mediators.

Mediators	Baseline (T0)		After intervention (T1)		After intervention MAD[a]	
	Control, mean (SD)	Intervention, mean (SD)	Control, mean (SD)	Intervention, mean (SD)	β (95% CI)	*P* value
Self-efficacy	19.79 (4.25)	18.60 (5.01)	19.87 (4.79)	19.29 (4.71)	0.21 (–1.12, 1.53)	.76
Peer support	15.73 (4.10)	14.83 (4.01)	16.30 (4.26)	15.40 (4.54)	0.00 (–1.44, 1.43)	.99
Family support	12.90 (3.61)	11.85 (3.88)	12.82 (3.98)	11.28 (4.28)	–0.93 (–2.14, 0.29)	.14
Teacher support	11.87 (3.49)	10.71 (3.73)	12.40 (3.82)	11.33 (4.01)	–0.45 (–1.70, 0.80)	.48
Self-regulation strategies	17.02 (3.88)	16.37 (4.71)	18.49 (4.88)	18.56 (4.72)	1.00 (–0.62, 2.61)	.23
Perceived barriers to physical activity	20.39 (4.45)	20.32 (4.59)	18.79 (5.35)	20.19 (4.88)	1.77 (0.19, 3.34)	.03
Physical activity enjoyment	64.42 (10.09)	60.93 (10.43)	63.86 (10.88)	62.03 (10.54)	–0.68 (–3.79, 2.43)	.67

[a]MAD: mean adjusted difference, where values above 0 indicate higher estimated means for the intervention group.

As the perceived barriers to physical activity were the only potential mediators for which there was a statistically significant intervention effect, formal mediation analyses were only conducted for this factor. Table 3 shows the estimated total (*c* path) and direct (*c'* path; while adjusting for the perceived barriers to physical activity score) effects of the intervention on physical activity and sedentary behavior at 6-month follow-up (T2; *c'* path). The total effects indicated that adolescents in the intervention group engaged in higher sedentary time and lower LPA at 6-month follow-up (T2) compared to the wait-list controls. These differences were still observed while adjusting for the perceived barriers to physical activity score (ie, direct effect).

Table 3 also presents the estimated indirect effects, defined as the effect of the intervention on physical activity and sedentary time at 6-month follow-up (T2) that occurred via the perceived barriers to physical activity score after intervention (T1). However, the indirect effects for each outcome were not statistically significant, indicating that the perceived barriers to physical activity did not mediate the intervention effects for physical activity or sedentary time.

Table 3. Estimated total (c path), direct (c' path), and indirect (a × b) effects of the intervention on 6-month follow-up (T2) physical activity and sedentary behavior, and indirect effects of the intervention on 6-month follow-up (T2) physical activity and sedentary behavior via the perceived barriers to physical activity mediator immediately after intervention (T1; n=159).

Activity variables	Total effect[a]		Direct effect		Indirect effect	
	c path (95% CI)	*P* value[b]	*c'* path (95% CI)	*P* value	*a* × *b* (95% CI)	*P* value[b]
Sedentary time	21.23 (3.57, 38.89)	.02	22.52 (4.76, 40.28)	.01	–1.28 (–4.93, 1.16)	>.05
Light-intensity physical activity	–15.87 (–27.86, –3.89)	<.01	–16.41 (–28.57, –4.24)	<.01	0.54 (–1.45, 3.49)	>.05
Moderate-to-vigorous–intensity physical activity	–6.33 (–13.97, 1.31)	.10	–6.74 (14.28, 0.80)	.08	0.48 (–0.16, 1.83)	>.05

[a]Direct and indirect effects may not exactly add up to the total effect due to variations in school-level effects between models.

[b]Exact *P* values not readily available for asymmetric percentile-based confidence intervals.

Discussion

Principal Results

This study aimed to examine the impact of the RAW-PA intervention on potential mediators of behavior change after intervention, and whether these mediated the intervention effects on physical activity and sedentary time at 6-month follow-up. While the intervention had a significant adverse effect on the perceived barriers to physical activity after intervention (*a* path), with adolescents in the intervention group perceiving more barriers than those in the wait-list control group, this did not mediate the intervention effects on physical activity or sedentary time at 6-month follow-up. No intervention effects were observed on any of the remaining factors identified as potential mediators, despite the intervention aiming to change target variables (eg, self-efficacy) hypothesized to be causally related to youth physical activity and sedentary time [14,20].

While there has been some variability of intervention effects on the barriers assessed in previous studies conducted in youth [20,36,37], the intervention effects on the barriers to physical activity in this study are consistent with several studies that have reported an increase in perceived barriers following the

intervention [38-40]. Within RAW-PA, adolescents were asked to reflect on the potential barriers to engaging in physical activity, and a range of strategies for overcoming barriers and integrating activity into daily life were targeted via the weekly missions and the accompanying behavior change resources [14]. It is possible, however, that this may have increased the adolescents' awareness of potential barriers, or they may have encountered barriers more frequently when attempting to increase their activity levels [39]. It is also possible that the intervention dose may not have been sufficient to overcome these barriers. Alternatively, the activity tracker itself may have affected perceived barriers. A key component of this technology is the ability to self-monitor physical activity, which has been shown to be an effective behavior change technique and critical for changing behavior [41]. However, this may have increased the adolescents' awareness of low activity levels, thus reinforcing barriers to being active.

While the analyses found evidence of an intervention effect on the perceived barriers to physical activity, there was no evidence of an indirect mediating effect. That is, the perceived barriers to physical activity after intervention did not mediate physical activity and sedentary time at 6-month follow-up. This suggests that the change in perceived barriers may not be on the causal pathway [42], which is consistent with the findings of several previous studies conducted in adolescents [39,40], albeit using different intervention strategies. The total effects observed may be explained by variables that were not assessed within this intervention, such as those that may be specific to the use of the activity tracking technology (eg, goal focus) [43], which was a core component within RAW-PA. Future research projects using activity trackers should consider examining a broader range of potential mechanisms drawn from different theories (eg, the technology acceptance model [44]) that focus on how adolescents perceive and use such devices.

To date, few activity tracker interventions, regardless of population group, have examined mediators of physical activity and sedentary time. A recent review highlighted that activity tracker studies have instead focused on mediators of outcomes such as task motivation in adult populations, with intervention effects observed for mediators including self-efficacy and self-awareness [43]. Despite this, the findings from this study are consistent with those from previous physical activity interventions conducted with adolescents who have reported few or inconsistent intervention effects on assessed mediators [15,20,25]. This suggests that the combination of activity trackers and digital resources may not beneficially change the targeted mediators or that that the targeted mediators were not the most effective for changing physical activity and sedentary behavior among inactive adolescents. It may be that other mediators of physical activity (eg, emotion [41]) should be targeted. Nevertheless, this supports previous research [15] that also showed few intervention effects on the determinants of activity but did not use mediation analysis. However, it should be noted that the RAW-PA implementation evaluation found that engagement with digital resources (eg, social media posts and challenges) was low across the intervention, with only 36% of adolescents in the intervention group reporting having completed the weekly challenges [23]. This may have contributed to the lack of effects observed, as the resources related to key behavior change techniques may not have been accessed and therefore utilized by adolescents in the intervention group. This study did not collect data concerning potential mediators of activity tracker use specifically, and future studies may investigate this by exploring intervention effects according to the level of wearable tracker use or the level of engagement with the intervention. This will provide greater insights into what elements worked best and what to target in future strategies.

Strengths and Limitations

This is the first study, to the authors' knowledge, to examine potential mediators of change in physical activity and sedentary time during an activity tracker intervention conducted in adolescents. The mediators were targeted within weekly *missions* [14] and were assessed using items that have demonstrated acceptable reliability in adolescent populations [25]. However, there are several limitations that should be acknowledged. First, just over half of the adolescents participating in the study provided complete, valid data at each time point, leading to exclusion from the current data analyses. It is possible that the sample size was not sufficient to detect the changes in the assessed variables [24]. Second, due to the sample size, it was not possible to examine mediators separately for males and females, even though there may have been differential effects. Third, the intervention was conducted in low-SES neighborhoods. It must be noted that these results may not apply to the general population in inactive adolescents.

Conclusions

In conclusion, an intervention effect was only observed for barriers to physical activity, with adolescent's perceiving more barriers than those in the wait-list control immediately after intervention. However, this did not mediate changes in physical activity and sedentary time at 6-month follow-up. This suggests that the lack of overall intervention effects on physical activity and sedentary time observed in the RAW-PA study could be due to the limited impact of the intervention on the targeted mediators. Future studies should identify intervention strategies that effectively target key mediators to improve physical activity among inactive adolescents. They should also explore additional potential mediators that may explain changes in the use of activity trackers and digital resources over time. Finally, intervention effects according to the level of wearable tracker use or the level of engagement with the intervention should be explored. This information is critical for the design of future successful interventions to increase physical activity in adolescents.

Acknowledgments

The RAW-PA (Raising Awareness of Physical Activity) study was supported by a VicHealth Innovation Research Grant. SV is supported by an Alfred Deakin Postdoctoral Research Fellowship. NDR is supported by a National Heart Foundation of Australia Future Leader Fellowship (Award ID 101895). AT was supported by a National Heart Foundation of Australia Future Leader Fellowship (100046). JS is supported by a NHMRC (National Health and Medical Research Council) Leadership Level 2 Fellowship (APP1176885). We would like to thank the schools and students who participated in this study. The authors also gratefully acknowledge the contribution of Kylie Ball as well as all project staff, especially Kate Dullaghan, Felicity J Pendergast, Stephanie Renehan, Stephanie Chappel, Sophie Aarts, Kara Richards, Tom Steele, Sarah Tighe, and Holly Beswick.

Conflicts of Interest

NDR and JS declare involvement in a start-up technological company. The remaining authors SJJMV, GA, SL, AT, HB, and SM declare no conflicts of interest.

References

1. Guthold R, Stevens GA, Riley LM, Bull FC. Global trends in insufficient physical activity among adolescents: a pooled analysis of 298 population-based surveys with 1·6 million participants. The Lancet Child & Adolescent Health 2020 Jan;4(1):23-35. [doi: 10.1016/s2352-4642(19)30323-2]
2. Australian 24-hour movement guidelines for children and young people (5 to 17 years): An integration of physical activity, sedentary behaviour, and sleep. Department of Health. 2019. URL: https://bit.ly/3QiS3U7 [accessed 2022-08-02]
3. Carson V, Hunter S, Kuzik N, Gray CE, Poitras VJ, Chaput J, et al. Systematic review of sedentary behaviour and health indicators in school-aged children and youth: an update. Appl Physiol Nutr Metab 2016 Jun;41(6 Suppl 3):S240-S265 [FREE Full text] [doi: 10.1139/apnm-2015-0630] [Medline: 27306432]
4. Nader PR, Bradley RH, Houts RM, McRitchie SL, O'Brien M. Moderate-to-vigorous physical activity from ages 9 to 15 years. JAMA 2008 Jul 16;300(3):295-305. [doi: 10.1001/jama.300.3.295] [Medline: 18632544]
5. Drummond M, Drummond C, Dollman J, Abery L. Physical activity from early childhood to adolescence: a literature review of issues and interventions in disadvantaged populations. JSW 2011 Jan 31;4(2):17. [doi: 10.21913/jsw.v4i2.722]
6. Stierlin AS, De Lepeleere S, Cardon G, Dargent-Molina P, Hoffmann B, Murphy MH, DEDIPAC consortium. A systematic review of determinants of sedentary behaviour in youth: a DEDIPAC-study. Int J Behav Nutr Phys Act 2015 Oct 09;12(1):133 [FREE Full text] [doi: 10.1186/s12966-015-0291-4] [Medline: 26453175]
7. Global recommendations on physical activity for health. World Health Organization. 2010. URL: https://www.who.int/publications/i/item/9789241599979 [accessed 2022-08-02]
8. Ridgers ND, Timperio A, Ball K, Lai SK, Brown H, Macfarlane S, et al. Effect of commercial wearables and digital behaviour change resources on the physical activity of adolescents attending schools in socio-economically disadvantaged areas: the RAW-PA cluster-randomised controlled trial. Int J Behav Nutr Phys Act 2021 Apr 12;18(1):52 [FREE Full text] [doi: 10.1186/s12966-021-01110-1] [Medline: 33845853]
9. Ridgers ND, McNarry MA, Mackintosh KA. Feasibility and Effectiveness of Using Wearable Activity Trackers in Youth: A Systematic Review. JMIR Mhealth Uhealth 2016 Nov 23;4(4):e129 [FREE Full text] [doi: 10.2196/mhealth.6540] [Medline: 27881359]
10. Ridgers ND, Timperio A, Brown H, Ball K, Macfarlane S, Lai SK, et al. Wearable Activity Tracker Use Among Australian Adolescents: Usability and Acceptability Study. JMIR Mhealth Uhealth 2018 Apr 11;6(4):e86 [FREE Full text] [doi: 10.2196/mhealth.9199] [Medline: 29643054]
11. Creaser AV, Clemes SA, Costa S, Hall J, Ridgers ND, Barber SE, et al. The Acceptability, Feasibility, and Effectiveness of Wearable Activity Trackers for Increasing Physical Activity in Children and Adolescents: A Systematic Review. Int J Environ Res Public Health 2021 Jun 08;18(12):6211 [FREE Full text] [doi: 10.3390/ijerph18126211] [Medline: 34201248]
12. Cajita MI, Kline CE, Burke LE, Bigini EG, Imes CC. Feasible but Not Yet Efficacious: A Scoping Review of Wearable Activity Monitors in Interventions Targeting Physical Activity, Sedentary Behavior, and Sleep. Curr Epidemiol Rep 2020 Mar 28;7(1):25-38 [FREE Full text] [doi: 10.1007/s40471-020-00229-2] [Medline: 33365227]
13. Epstein LH. Integrating theoretical approaches to promote physical activity. American Journal of Preventive Medicine 1998 Nov;15(4):257-265. [doi: 10.1016/s0749-3797(98)00083-x]
14. Ridgers ND, Timperio A, Brown H, Ball K, Macfarlane S, Lai SK, et al. A cluster-randomised controlled trial to promote physical activity in adolescents: the Raising Awareness of Physical Activity (RAW-PA) Study. BMC Public Health 2017 Jan 04;17(1):6 [FREE Full text] [doi: 10.1186/s12889-016-3945-5] [Medline: 28052773]
15. Slootmaker SM, Chinapaw MJ, Seidell JC, van Mechelen W, Schuit AJ. Accelerometers and Internet for physical activity promotion in youth? Feasibility and effectiveness of a minimal intervention [ISRCTN93896459]. Prev Med 2010 Jul;51(1):31-36. [doi: 10.1016/j.ypmed.2010.03.015] [Medline: 20380847]
16. Hooke MC, Gilchrist L, Tanner L, Hart N, Withycombe JS. Use of a Fitness Tracker to Promote Physical Activity in Children With Acute Lymphoblastic Leukemia. Pediatr Blood Cancer 2016 Apr 12;63(4):684-689. [doi: 10.1002/pbc.25860] [Medline: 26756736]

17. Evans EW, Abrantes AM, Chen E, Jelalian E. Using Novel Technology within a School-Based Setting to Increase Physical Activity: A Pilot Study in School-Age Children from a Low-Income, Urban Community. Biomed Res Int 2017;2017:4271483-4271487 [FREE Full text] [doi: 10.1155/2017/4271483] [Medline: 29670894]
18. Remmert JE, Woodworth A, Chau L, Schumacher LM, Butryn ML, Schneider M. Pilot Trial of an Acceptance-Based Behavioral Intervention to Promote Physical Activity Among Adolescents. J Sch Nurs 2019 Dec 13;35(6):449-461. [doi: 10.1177/1059840518786782] [Medline: 30004269]
19. Salmon J, Brown H, Hume C. Effects of strategies to promote children's physical activity on potential mediators. Int J Obes (Lond) 2009 Apr 13;33 Suppl 1(S1):S66-S73. [doi: 10.1038/ijo.2009.21] [Medline: 19363512]
20. Lubans DR, Foster C, Biddle SJ. A review of mediators of behavior in interventions to promote physical activity among children and adolescents. Prev Med 2008 Nov;47(5):463-470. [doi: 10.1016/j.ypmed.2008.07.011] [Medline: 18708086]
21. Bandura A. Social Foundations of Thought and Action: A Social-Cognitive View. Englewood Cliffs, NJ, US: Prentice-Hall; Jan 1987.
22. Census of population and housing: Socio-economic indexes for areas (SEIFA), Australia, 2011. Australian Bureau of Statistics. 2011. URL: https://www.abs.gov.au/ausstats/abs@.nsf/DetailsPage/2033.0.55.0012011 [accessed 2022-08-02]
23. Koorts H, Salmon J, Timperio A, Ball K, Macfarlane S, Lai SK, et al. Translatability of a Wearable Technology Intervention to Increase Adolescent Physical Activity: Mixed Methods Implementation Evaluation. J Med Internet Res 2020 Aug 07;22(8):e13573 [FREE Full text] [doi: 10.2196/13573] [Medline: 32763872]
24. Brown H, Hume C, Pearson N, Salmon J. A systematic review of intervention effects on potential mediators of children's physical activity. BMC Public Health 2013 Feb 23;13(1):165 [FREE Full text] [doi: 10.1186/1471-2458-13-165] [Medline: 23433143]
25. Dewar DL, Morgan PJ, Plotnikoff RC, Okely AD, Batterham M, Lubans DR. Exploring changes in physical activity, sedentary behaviors and hypothesized mediators in the NEAT girls group randomized controlled trial. J Sci Med Sport 2014 Jan;17(1):39-46. [doi: 10.1016/j.jsams.2013.02.003] [Medline: 23506657]
26. Cain K, Sallis JF, Conway TL, Van Dyck D, Calhoon L. Using accelerometers in youth physical activity studies: a review of methods. J Phys Act Health 2013 Mar;10(3):437-450 [FREE Full text] [doi: 10.1123/jpah.10.3.437] [Medline: 23620392]
27. Ridgers ND, Salmon J, Ridley K, O'Connell E, Arundell L, Timperio A. Agreement between activPAL and ActiGraph for assessing children's sedentary time. Int J Behav Nutr Phys Act 2012 Feb 19;9(1):15 [FREE Full text] [doi: 10.1186/1479-5868-9-15] [Medline: 22340137]
28. Freedson P, Pober D, Janz KF. Calibration of accelerometer output for children. Med Sci Sports Exerc 2005 Nov;37(11 Suppl):S523-S530. [doi: 10.1249/01.mss.0000185658.28284.ba] [Medline: 16294115]
29. Smith M, Taylor S, Iusitini L, Stewart T, Savila F, Tautolo E, et al. Accelerometer data treatment for adolescents: Fitting a piece of the puzzle. Prev Med Rep 2017 Mar;5:228-231 [FREE Full text] [doi: 10.1016/j.pmedr.2016.12.010] [Medline: 28101444]
30. Dewar DL, Lubans DR, Morgan PJ, Plotnikoff RC. Development and evaluation of social cognitive measures related to adolescent physical activity. J Phys Act Health 2013 May;10(4):544-555. [doi: 10.1123/jpah.10.4.544] [Medline: 22975481]
31. Motl RW, Dishman RK, Saunders R, Dowda M, Felton G, Pate RR. Measuring enjoyment of physical activity in adolescent girls. American Journal of Preventive Medicine 2001 Aug;21(2):110-117. [doi: 10.1016/s0749-3797(01)00326-9]
32. Robbins LB, Wu T, Sikorskii A, Morley B. Psychometric assessment of the Adolescent Physical Activity Perceived Benefits and Barriers Scales. J Nurs Meas 2008 Sep 01;16(2):98-112. [doi: 10.1891/1061-3749.16.2.98] [Medline: 18939715]
33. Schulz KF, Altman DG, Moher D, CONSORT Group. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. Trials 2010 Mar 24;11(1):32 [FREE Full text] [doi: 10.1186/1745-6215-11-32] [Medline: 20334632]
34. Whittle R, Mansell G, Jellema P, van der Windt D. Applying causal mediation methods to clinical trial data: What can we learn about why our interventions (don't) work? Eur J Pain 2017 Apr 14;21(4):614-622 [FREE Full text] [doi: 10.1002/ejp.964] [Medline: 27739626]
35. MacKinnon DP, Lockwood CM, Hoffman JM, West SG, Sheets V. A comparison of methods to test mediation and other intervening variable effects. Psychol Methods 2002 Mar;7(1):83-104 [FREE Full text] [doi: 10.1037/1082-989x.7.1.83] [Medline: 11928892]
36. van Stralen MM, Yildirim M, te Velde SJ, Brug J, van Mechelen W, Chinapaw MJM, ENERGY-consortium. What works in school-based energy balance behaviour interventions and what does not? A systematic review of mediating mechanisms. Int J Obes (Lond) 2011 Oct 12;35(10):1251-1265 [FREE Full text] [doi: 10.1038/ijo.2011.68] [Medline: 21487398]
37. Kelly S, Stephens J, Hoying J, McGovern C, Melnyk BM, Militello L. A systematic review of mediators of physical activity, nutrition, and screen time in adolescents: Implications for future research and clinical practice. Nurs Outlook 2017 Sep;65(5):530-548 [FREE Full text] [doi: 10.1016/j.outlook.2017.07.011] [Medline: 28823500]
38. Lytle LA, Murray DM, Evenson KR, Moody J, Pratt CA, Metcalfe L, et al. Mediators affecting girls' levels of physical activity outside of school: findings from the trial of activity in adolescent girls. Ann Behav Med 2009 Oct 12;38(2):124-136 [FREE Full text] [doi: 10.1007/s12160-009-9127-2] [Medline: 20012810]
39. Dunton GF, Schneider M, Cooper DM. An investigation of psychosocial factors related to changes in physical activity and fitness among female adolescents. Psychology & Health 2007 Dec;22(8):929-944. [doi: 10.1080/14768320601124915]

40. Haerens L, Cerin E, Maes L, Cardon G, Deforche B, De Bourdeaudhuij I. Explaining the effect of a 1-year intervention promoting physical activity in middle schools: a mediation analysis. Public Health Nutr 2007 Sep 06;11(5):501-512. [doi: 10.1017/s136898000700078x]
41. Michie S, Johnston M, Francis J, Hardeman W, Eccles M. From Theory to Intervention: Mapping Theoretically Derived Behavioural Determinants to Behaviour Change Techniques. Applied Psychology 2008;57(4):660-680. [doi: 10.1111/j.1464-0597.2008.00341]
42. Bauman AE, Sallis JF, Dzewaltowski DA, Owen N. Toward a better understanding of the influences on physical activity. American Journal of Preventive Medicine 2002 Aug;23(2):5-14. [doi: 10.1016/s0749-3797(02)00469-5]
43. Jin D, Halvari H, Maehle N, Olafsen AH. Self-tracking behaviour in physical activity: a systematic review of drivers and outcomes of fitness tracking. Behaviour & Information Technology 2020 Aug 05;41(2):242-261. [doi: 10.1080/0144929x.2020.1801840]
44. Davis FD. A technology acceptance model for empirically testing new end-user information systems: Theory and results. Massachusetts Institute of Technology. 1985. URL: https://dspace.mit.edu/bitstream/handle/1721.1/15192/14927137-MIT.pdf?sequence=2&isAllowed=y [accessed 2022-08-02]

Abbreviations

LPA: light-intensity physical activity
MVPA: moderate-to-vigorous physical activity
RAW-PA: Raising Awareness of Physical Activity
SEIFA: Socio-Economic Index for Areas
SES: Socioeconomic status

Edited by L Buis; submitted 30.11.21; peer-reviewed by S Compernolle, J Bianchi-Hayes; comments to author 27.01.22; revised version received 04.04.22; accepted 28.04.22; published 16.08.22.

Please cite as:
Verswijveren SJJM, Abbott G, Lai SK, Salmon J, Timperio A, Brown H, Macfarlane S, Ridgers ND
Mediators of Effects on Physical Activity and Sedentary Time in an Activity Tracker and Behavior Change Intervention for Adolescents: Secondary Analysis of a Cluster Randomized Controlled Trial
JMIR Mhealth Uhealth 2022;10(8):e35261
URL: https://mhealth.jmir.org/2022/8/e35261
doi:10.2196/35261
PMID:35972777

Corrigenda and Addenda

Correction: Technical Guidance for Clinicians Interested in Partnering With Engineers in Mobile Health Development and Evaluation

Lochan M Shah[1], MD; William E Yang[1], MD; Ryan C Demo[2], BSc; Matthias A Lee[2], PhD; Daniel Weng[1], BA; Rongzi Shan[1,3], BSc; Shannon Wongvibulsin[1,2], PhD; Erin M Spaulding[4], RN, BSN; Francoise A Marvel[1], MD; Seth S Martin[1,2], MD, MHS

[1]Johns Hopkins University School of Medicine, Baltimore, MD, United States

[2]Johns Hopkins University Whiting School of Engineering, Baltimore, MD, United States

[3]David Geffen School of Medicine at University of California, Los Angeles, Los Angeles, CA, United States

[4]Johns Hopkins University School of Nursing, Baltimore, MD, United States

Corresponding Author:
Lochan M Shah, MD
Johns Hopkins University School of Medicine
733 N Broadway
Baltimore, MD, 21231
United States
Phone: 1 410 550 3350
Email: lochanshah2019@gmail.com

Related Article:

Correction of: https://mhealth.jmir.org/2019/5/e14124/

(JMIR Mhealth Uhealth 2022;10(8):e41813) doi:10.2196/41813

In “Technical Guidance for Clinicians Interested in Partnering With Engineers in Mobile Health Development and Evaluation” (JMIR Mhealth Uhealth 2019;7(5):e14124) the authors made two updates.

The corresponding author’s telephone number was changed to 1 410 550 3350, and their email address was changed to lochanshah2019@gmail.com.

The corrections will appear in the online version of the paper on the JMIR Publications website on August 18, 2022, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Submitted 09.08.22; this is a non–peer-reviewed article;accepted 09.08.22; published 18.08.22.

Please cite as:
Shah LM, Yang WE, Demo RC, Lee MA, Weng D, Shan R, Wongvibulsin S, Spaulding EM, Marvel FA, Martin SS
Correction: Technical Guidance for Clinicians Interested in Partnering With Engineers in Mobile Health Development and Evaluation
JMIR Mhealth Uhealth 2022;10(8):e41813
URL: https://mhealth.jmir.org/2022/8/e41813
doi:10.2196/41813
PMID:35981322

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org