# Contents

## Review

Effects of Smartphone-Based Interventions on Physical Activity in Children and Adolescents: Systematic Review and Meta-analysis (e22601)
Zihao He, Hua Wu, Fengyu Yu, Jinmei Fu, Shunli Sun, Ting Huang, Runze Wang, Delong Chen, Guanggao Zhao, Minghui Quan. 2

## Original Papers

Addressing Implementation Challenges to Digital Care Delivery for Adults With Multiple Chronic Conditions: Stakeholder Feedback in a Randomized Controlled Trial (e23498)
Kelly Williams, Sarah Markwardt, Shannon Kearney, Jordan Karp, Kevin Kraemer, Margaret Park, Paul Freund, Andrew Watson, James Schuster, Ellen Beckjord. 14

Examining an Integrative Cognitive Model of Predicting Health App Use: Longitudinal Observational Study (e24539)
Kwanho Kim, Chul-Joo Lee. 27

Digital Medicine Community Perspectives and Challenges: Survey Study (e24570)
Brinnae Bent, Ida Sim, Jessilyn Dunn. 39

WeChat as a Platform for Baduanjin Intervention in Patients With Stable Chronic Obstructive Pulmonary Disease in China: Retrospective Randomized Controlled Trial (e23548)
Junjie Bi, Wei Yang, Ping Hao, Yongmei Zhao, Dan Wei, Yipeng Sun, Yuhua Lin, Meng Sun, Xuan Chen, Xuming Luo, Shangun Li, Wei Zhang, Xiongbiao Wang. 48

Heart Rate Variability and Firstbeat Method for Detecting Sleep Stages in Healthy Young Adults: Feasibility Study (e24704)
Lisa Kuula, Anu-Katrina Pesonen. 61
Effects of Smartphone-Based Interventions on Physical Activity in Children and Adolescents: Systematic Review and Meta-analysis

Zihao He1, MA; Hua Wu2, MD; Fengyu Yu1, BA; Jinmei Fu3, MS; Shunli Sun3, MS; Ting Huang1, MA; Runze Wang1, MA; Delong Chen1, MA; Guanggao Zhao1, PhD; Minghui Quan4, PhD
1School of Physical Education, NanChang University, NanChang, China
2Rehabilitation Medicine Center, The Second Affiliated Hospital of Jiaxing University, Jiaxing, China
3JiangXi Institute of Sport Science, NanChang, China
4School of Kinesiology, Shanghai University of Sport, Shanghai, China
*these authors contributed equally

Corresponding Author:
Minghui Quan, PhD
School of Kinesiology
Shanghai University of Sport
No 399 Chang Hai Road
Shanghai, 200438
China
Phone: 86 021 65507367
Email: quanminghui@163.com

Abstract

Background: About 70% of children and adolescents worldwide do not meet the recommended level of physical activity (PA), which is closely associated with physical, psychological, and cognitive well-being. Nowadays, the use of technologies to change PA is of interest due to the need for novel, more effective intervention approaches. The previous meta-analyses have examined smartphone-based interventions and their impact on PA in adults, but evidence in children and adolescents still needs further research.

Objective: This systematic review and meta-analysis aimed to determine the effectiveness of smartphone-based interventions for improving PA in children and adolescents.

Methods: Five electronic databases (PubMed, Web of Science, OVID, Scopus, and the China National Knowledge Infrastructure) were searched up to June 29, 2020. Randomized controlled trials with a control group that examine the effect of smartphone interventions on PA among children and adolescents were included. Bias risks were assessed using the Cochrane collaboration tool. Meta-analysis was performed to assess the pooled effect on PA using a random effects model. Subgroup analyses were conducted to examine the potential modifying effects of different factors (eg, types of intervention, intervention duration, age, measurement, study quality).

Results: A total of 9 studies were included in this review, including 4 mobile app interventions, 3 SMS text messaging interventions, and 2 app + SMS text messaging interventions. In general, the risk of bias of included studies was low. Compared with the control group, the use of smartphone intervention significantly improved PA (standardized mean difference [SMD] 0.44, 95% CI 0.11-0.77, P=.009), especially for total PA (TPA; weighted mean difference [WMD] 32.35, 95% CI 10.36-54.33, P=.004) and daily steps (WMD 1185, 95% CI 303-2068, P=.008), but not for moderate-to-vigorous PA (WMD 3.91, 95% CI –1.99 to 9.81, P=.19). High statistical heterogeneity was detected (I²=73.9%, P<.001) for PA. Meta-regression showed that duration (β=–.08, 95% CI –0.15 to –0.01, n=16) was a potential factor for high heterogeneity. The results of subgroup analyses indicated that app intervention (SMD 0.76, 95% CI 0.23-1.30, P=.005), children (SMD 0.64, 95% CI 0.10-1.18, P=.02), “≤8 weeks” (SMD 0.76, 95% CI 0.23-1.30, P=.005), objective measurement (SMD 0.50, 95% CI 0.09-0.91, P=.02), and low risk of bias (SMD 0.96, 95% CI 0.38-1.54, P=.001) can significantly improve PA.

Conclusions: The evidence of meta-analysis shows that smartphone-based intervention may be a promising strategy to increase TPA and steps in children and adolescents. Currently, app intervention may be a more effective strategy among smartphone
intervention technologies. To extend the promise of smartphone intervention, the future needs to design comparative trials among different smartphone technologies.

**Trial Registration:** PROSPERO CRD42019148261; https://tinyurl.com/y5modsrd

**KEYWORDS**
adolescents; children; mHealth; physical activity; smartphone

### Introduction

Childhood and adolescence are critical periods of growth. Engaging in enough physical activity (PA) has been demonstrated to benefit children’s physical and mental health, such as reducing health risks, preventing obesity, and developing cognitive function [1,2]. To achieve health benefits through PA, the World Health Organization (WHO) recommends that children and adolescents accumulate moderate-to-vigorous intensity PA (MVPA) exceeding 60 minutes per day [3]. However, the rising prevalence of physical inactivity is a serious concern worldwide. Globally, about 70% of children and adolescents do not meet the recommendations on PA [4]. For example, a Chinese PA and fitness survey showed that two-thirds of children and adolescents did not meet the recommended PA [5]. Insufficient PA is closely related to obesity, coronary heart disease, and other health problems [6-8]. Hence, it is of paramount importance to promote and facilitate PA safely and effectively during this critical period. In response to this difficult situation, researchers have carried out a series of intervention studies on PA. However, many intervention strategies not only suffer from high cost, but are also difficult to maintain and implement on a large scale [9-11]. Therefore, how to use cost-effective and innovative intervention strategies to improve the PA level of children and adolescents effectively remains a major public health problem.

To date, the popularity of smartphones in the world is extremely high: 73.1% of children and adolescents own a smartphone in China [12], and this trend can also be seen in the United States [13] and other countries [14]. Given the global scale of noncommunicable diseases, there is a need to provide preventative interventions to reach a large population at a low cost. Therefore, many researchers have applied smartphone technologies, such as mobile apps and SMS text messaging, to health-related fields and have achieved rich research results, such as weight management, cancer nursing, and chronic obstructive pulmonary disease self-monitoring [15-17]. It is gratifying that more researchers have tried to introduce smartphone technology into the field of PA. The participants included not only adults [18-21], but also children and adolescents who urgently need attention [22-30]. This undoubtedly provides a new perspective for solving the aforementioned problems. Therefore, at the 65th Annual Meeting of American College of Sports Medicine (ACSM) and the 9th World Congress on Exercise is Medicine held in the United States in 2018, the promotion of smartphones for PA was highlighted [31].

To date, many researchers have explored the effect of smartphone interventions on improving the PA of children and adolescents through randomized controlled trials (RCT), but there are controversies about inconsistent research results. Some studies have found that smartphone interventions can significantly improve the level of PA relative to their baseline more than the control group, such as Garde et al [22] (1758 steps/day, 95% CI 133-3384; 31.3 total PA [TPA] minutes/day, 95% CI 3.9 to 58.9), Chen et al [23] (0.4 PA day per week, 95% CI 0.15-0.66), Garde et al [25] (2934 steps/day, 95% CI 1434-4434; 46 TPA minutes/day, 95% CI 20-72), but other studies have not found a significant positive effect, such as Mendoza et al [24] (MVPA, –4.5 minutes/day, 95% CI –35.9 to 27), Direito et al [26] (MVPA, –1.82 minutes/day, 95% CI –16 to 12.36), Armstrong et al [28] (MVPA, 10 minutes/day, 95% CI –2.5 to 30), Thompson et al [29] (MVPA, 1.73 minutes/day, 95% CI –5.1 to 8.5; step, 318 steps/day, 95% CI –466 to 1102), andNewton et al [30] (step, –22 steps/day, 95% CI –1407 to 1364). Although there was 1 meta-analysis of smartphone intervention on adolescents to improve PA [32] and found a significant improvement on MVPA (standardized mean difference [SMD] 0.341, 95% CI 0.02-0.66), only 5 studies were included. Also, 2 of the 5 studies were multicomponent interventions (including smartphone and other components), which made it difficult to identify the true smartphone effect. Furthermore, this review missed some studies in the database [25,27,30]. Given the fact that there have been many new RCTs in recent years [22,24,28], and the previous reviews include comprehensive intervention strategies, it is unclear whether intervention effects were truly due to the smartphone itself, or rather the other intervention components [18]. Therefore, conducting a new meta-analysis on this topic is necessary.

The objective of this review is to evaluate the effectiveness of smartphone interventions to promote PA in children and adolescents, by using systematic review and meta-analysis to combine the most comprehensive and up-to-date literature. The findings of this study are expected to provide insights and practice for the development of future smartphone interventions.

### Methods

#### Registration and Approval

This research program has been registered on the PROSPERO System Evaluation Registration Platform, registration number: CRD42019148261. This study has been reported according to the Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) guidelines [33].

#### Search Strategy

A systematic literature search was conducted to find out relevant studies in 5 electronic databases: PubMed, Web of Science, and other databases (Google Scholar, CINAHL, Embase, PsyINFO). The search terms were used as follows:智能手机 (smartphone) AND 健康行为 (healthy behavior) AND 学生 (students). The search was conducted between January 2008 and December 2018. The inclusion criteria were randomized controlled trials and population-based studies on children and adolescents. The exclusion criteria were studies on adults, systematic reviews, meta-analyses, and non-English language articles. A total of 11 studies were included, and the results are presented in Table 1.
Two authors (GZ and ZH) extracted information and data independently, including study characteristics (the first author, publication year, country, study design, contents of intervention, study duration), subject characteristics (age, sex, sample size), and outcome (measurement strategy, statistical analysis, results). Disagreement was resolved through discussion until a consensus decision was reached. In the case of missing data, this information was requested from the authors a minimum of 3 times over 4 weeks.

### Risk of Bias and Quality Assessment

The Cochrane Collaboration risk of bias tool was used to categorize the risk of bias in six domains [35]: (1) sequence generation, (2) allocation sequence concealment, (3) blinding of outcome assessment, (4) incomplete outcome data, (5) selective outcome reporting, and (6) other potential threats to validity. The item blinding of participants and personnel were excluded because it is not feasible in these types of studies [20]. In addition, the risk of bias assessment for blinding of outcome assessment was based on the method of outcome assessment (objective or subjective) [20]. Each domain was scored as low, unclear, or high risk of bias. Overall classification of low, unclear, or high risk of bias in each study was based on the combination of the domains. Figures were generated by Review Manager software (RevMan 5.3; Nordic Cochrane). Disagreement about the risk of bias assessments was resolved by consensus or consulting the third author.

### Statistical Analyses

Random-effects models were used in this study for the meta-analysis of the included studies. For studies that only presented data through graphs (eg, Boxplot), we estimated mean and SDs using the y-axis and length of the graphs [22,25,27]. For studies that reported standard errors, CI, or quartile, we converted these data to SDs [36]. We compared the changes from baseline to endpoint data between groups. The formulas for the mean and SD pre- to post-change values were as follows: $\text{Mean}_{\text{change}} = \text{Mean}_{\text{post}} - \text{Mean}_{\text{pre}}$ and $\text{SD}_{\text{change}} = \sqrt{\text{SD}_{\text{pre}}^2 + \text{SD}_{\text{post}}^2 - 2 \times \text{Corr} \times \text{SD}_{\text{pre}} \times \text{SD}_{\text{post}}}$, where the correlation coefficient was set to 0.5 based on the Cochrane Collaboration Handbook guidelines [35]. SMD and 95% CI were calculated in this study because the outcomes of the included studies are measured using different methods [37].

In the following cases, specific statistical procedures were employed: (1) When there were several publications from the same project, the study with the longest follow-up was selected; if there was no intervention during the follow-up, the result of the last intervention was selected as statistical analysis data [22]. (2) If there were multiple intervention groups in the same studies, the data were considered as independent samples for analysis. Moreover, sample size from control groups were evenly allocated to each intervention group in the meta-analysis to avoid artificial inflation of the true sample size [26,29]. Similarly, if a study measured 2 or more PA domains (ie, TPA, MVPA, or steps), the sample size of the control group was divided by the number of domains in which the study was measured [22,25,27,29,30]. (3) Studies that reported PA in other forms (eg, counts per minute, days per week) were included only in the systematic review, but not for meta-analysis, because the data cannot be converted into minutes per day [23].

**Selection Criteria of Studies**

### Inclusion Criteria

Inclusion criteria, according to PICOS (population, intervention, comparison, outcomes, and study) [34], were as follows:

1. Participants: children and adolescents aged 6-18 years, based on the PubMed MeSH definition of children (6-12 years) and adolescents (13-18 years).
2. Interventions: smartphone as the intervention tool, which used either app or SMS text messaging or both to promote PA.
3. Control groups included participants not using smartphone technology.
4. Outcomes: PA including daily steps or any intensities of PA. To be included in the meta-analysis, the outcome should be reported as steps, minutes, or hours. Studies that reported PA in other forms (eg, counts per minute, days per week) were included only in the systematic review.
5. The study design was RCTs.

### Exclusion Criteria

1. Studies where the intervention technologies were not smartphone based (computer) or incorporated other components (eg, physical education, school seminar).
2. Studies did not report data on PA level (eg, PA score, self-efficacy on PA).
3. Studies were not written in English or Chinese.

### Data Extraction

Two authors (GZ and ZH) extracted information and data independently, including study characteristics (the first author, population, intervention, study duration), subject characteristics (age, sex, sample size), and outcome (measurement strategy, statistical analysis, results). Disagreement was resolved through discussion until a consensus decision was reached. In the case of missing data, this information was requested from the authors a minimum of 3 times over 4 weeks.
Additionally, subgroup analysis was based on the characteristics of the review, that is, outcomes (MVPA, TPA versus steps), types of intervention (app, SMS text messaging versus app + SMS text messaging), age (children versus adolescents), intervention duration (“≤ 8 weeks” versus “> 8 weeks”), measurement (objective versus subjective), and risk of bias (low; unclear versus high). Given the consistency of variable units between the same outcome indicator among the continuous variables in TPA, MVPA, and steps, weighted mean difference (WMD) was calculated in this subgroup for statistical analysis.

The statistical heterogeneity was examined using $I^2$ between included studies and Cochran Q-test; it was defined as very low, low, medium, and high heterogeneity when $I^2$ values were <25%, 25% to <50%, 50% to <75%, and ≥75%, respectively [38]. Potential sources of heterogeneity were investigated using meta-regression (eg, duration, age, BMI). Egger test was adopted to detect publication bias [39]. Additionally, the “trim and fill” method was performed to estimate the impact of publication bias on the results [40]. Furthermore, to test the robustness of the results of this study, the following methods were used to conduct sensitivity analyses: 1 article was removed each time to examine whether each article had a significant influence on the overall effect ($P<.05$).

All statistical calculations were performed using the statistical software STATA 15.1 (Release 15.1 College Station, TX, USA); $P<.05$ was defined as a significant difference.

**Results**

**Overview**

There were 3263 studies produced from the electronic database search, and the titles and abstracts of 2149 of them were screened after deleting duplicates. In the screening process, a total of 2004 records were excluded, so 145 full-text studies remained to be assessed. From these, manual searches were conducted for studies that met the inclusion criteria. Finally, 9 studies were included in this review. A flow chart of the systematic literature search is displayed in Figure 1.

![Flow chart of study selection](https://mhealth.jmir.org/2021/2/e22601)

**Characteristics of the Included Studies**

All included studies were published after 2009, 8 of which were after 2015. The study areas were distributed in 3 different countries: America (n=4) [23,24,28,29], Canada (n=3) [22,25,27], and New Zealand (n=2) [26,30]. The sample size was 558, the mean age of the participants was 13.2 years, 4 studies included children [22,25,27,28], and 5 studies included adolescents [23,24,26,29,30]. The intervention content is mainly based on smartphone technologies, app, and SMS text messaging, including 4 studies based on app [22,25-27], 3 studies based on SMS text messaging [28-30], and 2 studies based on app + SMS text messaging [23,24]. The study designs were all RCTs. The duration of interventions ranged from 2 weeks to 6 months. In addition, 1 study reporting PA days per week was not included in the meta-analysis because the data cannot be converted into minutes per day [23]. For TPA and MVPA, 6 studies objectively measured PA with an...
accelerometer or Tractivity activity monitor [22,24-27,29], and 3 studies used subjective assessments (questionnaires or self-reports) [23,28,30]. For the measurement of steps, 2 studies used pedometers [29,30], and 3 studies used Tractivity activity monitor [22,25,27] (Multimedia Appendix 2).

Risk of Bias

Figures 2 and 3 show the risk of bias assessment of the 9 included studies; of these, 3 studies were classified as having a low risk of bias, 4 studies were classified as having an unclear risk of bias, and 2 had a high risk of bias rating. Three studies were subjective measurement methods, so the blinded outcome assessment was rated as high risk of bias.

Figure 2. Risk of bias graph: each risk of bias item is presented as percentages.

Figure 3. Risk of bias of included studies. Green: low risk of bias; yellow: unclear risk of bias; red: high risk of bias.

Result of Meta-analysis on PA

The Summary Effect Analysis

A random-effects meta-analysis, including 8 studies (16 effects), demonstrated that there was a significant improvement in PA in the intervention group compared to the control group (SMD 0.44, 95% CI 0.11-0.77, \( P=0.009 \)), and high statistical heterogeneity was detected (\( I^2=73.9\% \), \( P<.001 \); Figure 4). Meta-regression showed that duration (\( \beta=-0.08 \), 95% CI –0.15 to –0.01, n=16) was a potential factor for high heterogeneity. The Egger test showed that there was no significant publication bias between the studies (\( P=0.28 \)).
Figure 4. Meta-analysis of effects of intervention versus control on physical activity (PA). MVPA: moderate-to-vigorous-intensity physical activity; SMD: standardized mean difference; TPA: total physical activity.

Subgroup Analysis

The results of subgroup analysis of the effects on outcomes are shown in Table 1. Compared with the control group, subgroups of TPA (WMD 32.35, 95% CI 10.36-54.33, \(P=0.004\)), step (WMD 1185, 95% CI 303-2068, \(P=0.008\)), app intervention (SMD 0.76, 95% CI 0.23-1.30, \(P=0.005\)), children (SMD 0.64, 95% CI 0.10-1.18, \(P=0.002\)), “≤8 weeks” (SMD 0.76, 95% CI 0.23-1.30, \(P=0.005\)), objective measurement (SMD 0.50, 95% CI 0.09-0.91, \(P=0.02\)), and low risk of bias (SMD 0.96, 95% CI 0.38-1.54, \(P=0.001\)) can significantly increase PA.
Table 1. Subgroup analyses on the effect of intervention versus control on PA in children and adolescents.

<table>
<thead>
<tr>
<th>Potential modifiers</th>
<th>Studies, n</th>
<th>Effect size (95% CI)</th>
<th>$I^2$ (%)</th>
<th>P-value heterogeneity</th>
</tr>
</thead>
<tbody>
<tr>
<td>All studies</td>
<td>8</td>
<td>0.44 (0.11 to 0.77)</td>
<td>73.9</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>Outcome</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TPA</td>
<td>3</td>
<td>32.35 (10.36 to 54.33)</td>
<td>61.8</td>
<td>.07</td>
</tr>
<tr>
<td>MVPA</td>
<td>7</td>
<td>3.91 (-1.99 to 9.81)</td>
<td>0.0</td>
<td>.94</td>
</tr>
<tr>
<td>Step</td>
<td>6</td>
<td>1185 (303 to 2068)</td>
<td>43.0</td>
<td>.12</td>
</tr>
<tr>
<td><strong>Intervention</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>App</td>
<td>4</td>
<td>0.76 (0.23 to 1.30)</td>
<td>76.4</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>SMS text messaging</td>
<td>3</td>
<td>0.18 (-0.06 to 0.42)</td>
<td>0.0</td>
<td>.99</td>
</tr>
<tr>
<td>App + SMS text messaging</td>
<td>1</td>
<td>-0.03 (-0.54 to 0.49)</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td><strong>Age</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Children</td>
<td>3</td>
<td>0.64 (0.10, 1.18)</td>
<td>74.1</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Adolescents</td>
<td>5</td>
<td>0.32 (-0.12, 0.75)</td>
<td>74.8</td>
<td>.002</td>
</tr>
<tr>
<td><strong>Duration</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤8 weeks</td>
<td>4</td>
<td>0.76 (0.23, 1.30)</td>
<td>76.4</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>&gt;8 weeks</td>
<td>4</td>
<td>0.14 (-0.07, 0.36)</td>
<td>0.0</td>
<td>.99</td>
</tr>
<tr>
<td><strong>Measurement</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Objective</td>
<td>6</td>
<td>0.50 (0.09, 0.91)</td>
<td>76.6</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Subjective</td>
<td>2</td>
<td>0.22 (-0.08, 0.51)</td>
<td>0.0</td>
<td>.78</td>
</tr>
<tr>
<td><strong>Risk of bias</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>3</td>
<td>0.96 (0.38, 1.54)</td>
<td>74.9</td>
<td>.001</td>
</tr>
<tr>
<td>Unclear</td>
<td>3</td>
<td>0.09 (-0.19, 0.37)</td>
<td>0.0</td>
<td>.99</td>
</tr>
<tr>
<td>High</td>
<td>2</td>
<td>0.22 (-0.08, 0.51)</td>
<td>0.0</td>
<td>.78</td>
</tr>
</tbody>
</table>

aOutcome: There are studies reporting 2 outcomes, so the total exceeds the total number of included studies; besides, only this subgroup was calculated using weighted mean difference (WMD), whereas for others SMD is reported.

bTPA: total physical activity.

cMVPA: moderate-vigorous intensity physical activity.

Robustness of the Results

Sensitivity analyses were conducted to test the robustness of the findings. One study was removed each time to perform a meta-analysis again. The results of the effect did not change significantly, which indicates that the results of the meta-analysis in this study were reliable (Multimedia Appendix 3).

Discussion

Principal Findings

The primary objective of this study was to determine the effectiveness of the smartphone-based intervention in improving PA in children and adolescents. The results of this study indicated that smartphone-based intervention has a significant effect on PA, especially for TPA and steps, but not for MVPA.

Comparison With Previous Systematic Review and Meta-analysis

The findings of this study indicated that smartphone-based intervention has a positive effect on PA in children and adolescents, and our results are a valuable extension of recently published systematic reviews and meta-analysis. Previous similar studies mainly focused on the intervention effect of smartphone, app, and a combination of app and wearables on MVPA and step counts in adults, but the results of the studies were not consistent. Gal et al [20] (age range 19-79 years) reported that smartphone-based intervention was effective in promoting MVPA (SMD 0.43, 95% CI 0.03-0.82), whereas a nonsignificant difference on MVPA was observed in Romeo et al [18] (age range 22-63 years; mean difference [MD] –2.16, 95% CI –15.68 to 11.36; MD –3.16, 95% CI –7.85 to 0.63), Flores et al [19] (mean 39 years; SMD 0.40, 95% CI –0.07 to 0.87), and Direito et al’s [41] study (age range 8.4-71.7 years; SMD 0.37, 95% CI –0.03 to 0.77). Besides, Gal et al [20] (19-79 years) and Feter et al [21] (mean 40.7 [SD 14.4] years) reported that smartphone-based intervention has a significant positive effect on steps in adults (SMD 0.51, 95% CI 0.12-0.91; MD 735, 95% CI 28-1243, respectively). However, Romeo et al [18] (age range 22-63 years) and Direito et al [41] (age range 8.4-71.7 years) did not find these results (MD 477, 95% CI –230 to 1183 and SMD 0.14, 95% CI –0.01 to 0.29, respectively).
The possible explanation is that the intervention effects of smartphone, app, or app plus other components are the difference [21,42]. It is necessary to conduct controlled trials between different interventions. In addition, although a significant MVPA increase was not observed in most studies, we cannot ignore the potential health-promoting effects of increased other intensity PA by smartphone interventions. Recent epidemiological evidence indicated the potential benefits of increasing light-intensity PA (LPA), including association with decreased systolic blood pressure, diastolic blood pressure, markers of lipid, and glucose metabolism [43,44]. Therefore, how to improve LPA is also the focus of future research.

At present, only Shin et al’s [32] study (10-19 years) focused on children and adolescents, and a significant improvement effect was found on MVPA (SMD 0.34, 95% CI 0.02-0.66). However, Shin et al [32] only included 5 studies, and it is difficult to identify the real smartphone intervention effect because 2 of these 5 studies were multicomponent interventions (including smartphones and other components). Hence, to fill up the research gaps from the previous meta-analysis, this study included more studies published in recent years and determined the actual effect of smartphone-based intervention alone on PA in children and adolescents which may provide additional information and be a valuable contribution to this area of inquiry.

The Intervention of Two Smartphone Technologies and Their Effects

At present, the number of smartphone apps on the Chinese market monitored is 4.49 million, and youth per capita under 10, 10-14, and 15-19 years is as high as 30, 44, and 59, respectively [45]. These show that app technologies are mature enough to provide technical guarantees for the development of different interventions. Indeed, subgroup analysis found that app intervention can significantly improve PA. This finding is similar to previous meta-analyses on the adult population [14,21]. The advantage of app lies in its convenience and novelty. Through the app, you can receive feedback in real time, communicate, and self-monitor, among other possibilities. At present, an increasing number of children and adolescents are searching for health-related information and guiding their fitness via app [46,47]. Therefore, an app-based intervention meets the needs of modern people for health.

Unlike the intervention effect of an app, SMS text messaging intervention has no significant improvement effect on PA. However, 2 systematic reviews are inconsistent with the results of this review. Ludwig et al [48] performed a systematic review of the efficacy of the intervention that uses SMS text messaging to improve PA and found that some studies have potential effects on improving PA in adolescents. Similarly, Feter et al [21] found that SMS text messaging intervention can significantly improve PA in adults. However, interventions in some studies included in these 2 reviews are SMS text messaging plus other components, so it is difficult to discern whether the actual effect comes from SMS text messaging or other interventions. Unfortunately, there are no controlled trials on separate interventions for SMS text messaging–only and SMS text messaging plus other components, which is also an issue that researchers need to study further.

Effects of the Smartphone on Different Age and Study Duration

Our subgroup analyses found that smartphone intervention has a significant effect on improving PA of children. In the studies in this review where the participants are children, the implementation of interventions requires parental assistance. A previous study found that parents play an important role in supporting and managing child-related health behaviors (eg, PA, sedentary behavior) [49]. The assistance of parents is conducive to the implementation of the intervention, which may lead to a positive effect on increasing PA. For adolescents, smartphone intervention has played a role in the intervention to a certain extent. However, adolescence is a transition period from the growth of children to adults, and it is also a stage of emotional fluctuation and frequent physiological changes. Rebellious emotions in the adolescent stage may resist and not cooperate with the implementer, which affects the effectiveness of the intervention and the compliance with the research.

The short-term (≤8 weeks) intervention effects may be attributed to the curiosity of the participants in the early stages of the intervention, and that they are willing to participate in the implementation. Over time, the decline in the interest and compliance of the participants led to the intervention effect not being maintained. A 4-week game app intervention found that the first-week intervention significantly improved PA in children, but the second-week and the fourth-week follow-up had no significant effect [22]. When all the games are unlocked or participants are familiar with the game, the participants are no longer interested in continuing, and the intervention effect of PA cannot be maintained. Therefore, considering the interest and passion of children and adolescents, we should strive to propose a novel strategy along with the design for a long-duration intervention.

Strengths and Limitations

This review has several strengths. First, scientifically rigorous RCT studies were included in the meta-analysis. Second, the included studies are smartphone-alone intervention, excluding studies with other intervention content, so the results can better reflect the intervention effect of smartphone. Lastly, this review conducted a subgroup analysis to explore the potential modifying effect of different factors thoroughly.

Despite these strengths, the review has several limitations. First, there are not enough studies to examine potential modifying effects of LPA, economic levels, and demographic characteristics (eg, gender, body mass index, economic status). Second, the different characteristics of the included studies lead to high heterogeneity. However, we have included the latest Chinese and English literature and conducted a subgroup analysis based on literature characteristics.

Conclusions

The findings of this meta-analysis indicated that interventions based on smartphone may be a promising strategy to increase the number of steps and TPA of children and adolescents, but...
the effect of intervention on MVPA remains to be studied. Currently, app intervention may be a more effective strategy among smartphone intervention technologies. To extend the promise of smartphone intervention, the future needs to design comparative trials among different smartphone technologies (ie, app vs SMS text messaging, app vs app + SMS text messaging, SMS text messaging vs app + SMS text messaging). Moreover, additional studies are needed to determine the effects on different participants, such as for children who are overweight and obese and low-income people.

Acknowledgments
This work was supported by grants from the MOE (Ministry of Education in China) Project of Humanities and Social Sciences (18YJC890060), National Natural Science Foundation of China (81703252), the Innovation Fund Designated for Graduate Students of Jiangxi Province, China (YC2019-S012), and the Zhejiang Provincial Medical and Health Science Technology Plan, China (2019KY217).

Conflicts of Interest
None declared.

Multimedia Appendix 1
Search strategy.
[DOCX File, 13 KB - mhealth_v9i2e22601_app1.docx]

Multimedia Appendix 2
Characteristics of Included Studies.
[DOCX File, 15 KB - mhealth_v9i2e22601_app2.docx]

Multimedia Appendix 3
Robustness of the Results.
[PDF File (Adobe PDF File), 148 KB - mhealth_v9i2e22601_app3.pdf]

References
3. World Health Organization. Physical Activity and Young People: Recommended Levels of Physical Activity for Children Aged 5-17 Years. URL: https://www.who.int/news-room/fact-sheets/detail/physical-activity [accessed 2021-01-19]


TPA: total physical activity
WHO: World Health Organization
WMD: weighted mean difference

Edited by L Buis; submitted 18.07.20; peer-reviewed by J Li, N Mohammad Gholi Mezerji; comments to author 11.08.20; revised version received 28.08.20; accepted 12.01.21; published 01.02.21.

Please cite as:
He Z, Wu H, Yu F, Fu J, Sun S, Huang T, Wang R, Chen D, Zhao G, Quan M
Effects of Smartphone-Based Interventions on Physical Activity in Children and Adolescents: Systematic Review and Meta-analysis
JMIR Mhealth Uhealth 2021;9(2):e22601
URL: https://mhealth.jmir.org/2021/2/e22601
doi:10.2196/22601
PMID:33522980

©Zihao He, Hua Wu, Fengyu Yu, Jinmei Fu, Shunli Sun, Ting Huang, Runze Wang, Delong Chen, Guanggao Zhao, Minghui Quan. Originally published in JMIR mHealth and uHealth (http://mhealth.jmir.org), 01.02.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on http://mhealth.jmir.org/, as well as this copyright and license information must be included.
Addressing Implementation Challenges to Digital Care Delivery for Adults With Multiple Chronic Conditions: Stakeholder Feedback in a Randomized Controlled Trial

Kelly Williams1*, MPH, PhD; Sarah Markwardt1*, MID; Shannon M Kearney1*, CPH, MPH, DrPH; Jordan F Karp2*, MD; Kevin L Kraemer2*, MD; Margaret J Park3*, MDiv; Paul Freund4*, MEd; Andrew Watson5*, FACS, M. Litt, MD; James Schuster1*, MBA, MD; Ellen Beckjord1*, PhD, MPH

1UPMC Center for High-Value Health Care, Insurance Services Division, UPMC, Pittsburgh, PA, United States
2Department of Psychiatry, College of Medicine-Tucson, University of Arizona, Tuscon, AZ, United States
3Community Wellness Consultancy, Pittsburgh, PA, United States
4Consumer Action Response Team of Allegheny County, NAMI Keystone Pennsylvania, Pittsburgh, PA, United States
5Department of Surgery, UPMC, Pittsburgh, PA, United States
*all authors contributed equally

Corresponding Author:
Kelly Williams, MPH, PhD
UPMC Center for High-Value Health Care
Insurance Services Division
UPMC
600 Grant St
Floor 40, Center for High-Value Health Care
Pittsburgh, PA, 15219
United States
Phone: 1 412 454 1198
Email: williamsk17@upmc.edu

Abstract

Background: Digital tools accessed via smartphones can promote chronic condition management, reduce disparities in health care and hospital readmissions, and improve quality of life. However, whether digital care strategies can be implemented successfully on a large scale with traditionally underserved populations remains uncertain.

Objective: As part of a randomized trial comparing care delivery strategies for Medicaid and Medicare-Medicaid beneficiaries with multiple chronic conditions, our stakeholders identified implementation challenges, and we developed stakeholder-driven adaptions to improve a digitally delivered care management strategy (high-tech care).

Methods: We used 4 mechanisms (study support log, Patient Partners Work Group log, case interview log, and implementation meeting minutes) to capture stakeholder feedback about technology-related challenges and solutions from 9 patient partners, 129 participants, and 32 care managers and used these data to develop and implement solutions. To assess the impact, we analyzed high-tech care exit surveys and intervention engagement outcomes (video visits and condition-specific text message check-ins sent at varying intervals) before and after each solution was implemented.

Results: Challenges centered around 2 themes: difficulty using both smartphones and high-tech care components and difficulty using high-tech care components due to connectivity issues. To respond to the first theme’s challenges, we devised 3 solutions: tech visits (eg, in-person technology support visits), tech packet (eg, participant-facing technology user guide), and tailored condition-specific text message check-ins. During the first 20 months of implementation, 73 participants received at least one tech visit. We observed a 15% increase in video call completion for participants with data before and after the tech visit (n=25) and a 7% increase in check-in completion for participants with data before and after the tech visit (n=59). Of the 379 participants given a tech packet, 179 completed care during this timeframe and were eligible for an exit survey. Of the survey respondents, 76% (73/96) found the tech packet helpful and 64% (62/96) actively used it during care. To support condition-specific text message check-in completion, we allowed for adaption of day and/or time of the text message with 31 participants changing the time they received check-ins and change in standard biometric settings with 13 physicians requesting personalized settings for participants.
To respond to the second theme’s challenges, tech visits or phone calls were made to demonstrate how to use a smartphone to connect or disconnect from the internet, to schedule video calls, or for condition-specific text message check-ins in a location with broadband/internet.

Conclusions: Having structured stakeholder feedback mechanisms is key to identify challenges and solutions to digital care engagement. Creating flexible and scalable solutions to technology-related challenges will increase equity in accessing digital care and support more effective engagement of chronically ill populations in the use of these digital care tools.

Trial Registration: ClinicalTrials.gov NCT03451630; https://clinicaltrials.gov/ct2/show/NCT03451630.

JMIR Mhealth Uhealth 2021;9(2):e23498 doi:10.2196/23498

KEYWORDS
telehealth tools; smartphones; remote patient monitoring; care management; mobile phone

Introduction

Over the last 10 years, ownership and use of smartphones has more than doubled in the United States, from 35% to 81% of the population [1]. One potential benefit of increased access to smartphones is the reduction in health disparities. As smartphone ownership becomes more equitable across socioeconomic categories [1], the use of smartphones provides an opportunity for traditionally underserved or isolated populations to remain connected to health care professionals despite geographic distance or mobility limitations, to quickly receive up-to-date and accurate health education information, and to monitor changes in health conditions using digital health care strategies [2]. In the context of care management teams, remote monitoring platforms provide the opportunity to scale programs, allowing teams of health care professionals and social workers to reach a higher number of individuals living in medically underserved areas [3].

Digital care tools have a growing evidence base, including evidence supporting the effectiveness of such technologies for patients managing chronic conditions. For example, these tools can support individuals with diabetes in lowering hemoglobin A1c levels, improve quality of life and lower number of hospital readmissions for individuals with heart failure [4], improve symptoms and outcomes for individuals with respiratory conditions [5], support better blood pressure control for individuals with hypertension [6], and reduce symptoms of depression [7]. Increased access to and use of digital care strategies has the potential to increase health systems’ ability to achieve the quadruple aim: improving population health, enhancing both patient and provider experience, and reducing costs [8].

Despite increased smartphone use and evidence supporting the benefits of digital care strategies such as remote patient monitoring, biosensors, and wearable devices, barriers to care are ever-present when implementing digital tools on a large scale. Notably, there is a lack of research on the challenges that occur when implementing digital care with traditionally underserved populations and those with high-burden, high-cost medical conditions [9]. A pervasive barrier to the success of digital tools is that individuals may lack confidence in their ability to learn how to use these new tools, which may impact their readiness to engage with such tools [10]. According to recent research on mobile devices and health, over half of Americans are considered to have low digital literacy skills when it comes to using mobile devices [11]. A recent survey of Americans aged above 65 years indicated that although respondents generally had a positive view of technology, they doubted their capacity to learn to use new technology without extra help [12]. In addition, although access to smartphones is becoming more prevalent among all socioeconomic groups, a digital divide still exists in the United States between high- and low-income Americans [13,14], with some research showing that both access and ability may be contributing reasons for why low-income adults may use online health resources less [15,16]. Finally, research has historically focused on remote patient monitoring care effectiveness with older Americans [17] and may not address common challenges faced by younger and other diverse populations.

Overcoming these challenges to realize the full potential of digital care to support the management of chronic conditions and reduce health disparities requires an iterative development approach that includes ongoing consumer and community stakeholder input. We are conducting a large-scale, randomized controlled trial comparing the effectiveness of 3 care management strategies (ie, high-touch, high-tech, and usual care) delivered by a commercial insurance organization for adult Medicaid and Medicare-Medicaid beneficiaries living with multiple chronic conditions. To address the unique needs and challenges experienced by this population and to ensure that our digital interventions are patient-centered and pragmatic, we describe early implementation challenges and our stakeholder-driven process adaptions specific to the digitally delivered chronic disease care management strategy (high-tech care).

Methods

Randomized Controlled Trial Overview

The 3 study comparators are approaches to care delivered by a chronic disease care management program and incorporate fundamental, evidence-based components of integrated care models including interdisciplinary care management [18,19], individualized care plans [18,19], chronic disease self-management education or self-care support [8,18,20-24], and linkages to medical/behavioral health and social services [18,25,26]. High-touch is delivered primarily face-to-face, with telephonic support as needed. High-tech is delivered via a remote care management platform. Both high-touch and

high-tech participants receive care management for at least 4 months and can continue care for up to 1 year, based on need. Usual care consists of an initial visit and care management for 14 days, which includes connections to condition management support and resources. Care managers receive a weekly worklist denoting individuals who are eligible to be offered participation in the study. To be eligible, individuals must be 21 years or older, have Medicaid or Medicare-Medicaid insurance, have at least 2 chronic conditions, including 1 physical health condition, and have been discharged from a hospital within the past 30 days.

**Introducing Participants to the High-tech Care Strategy**

All study participants work, one on one, with a care manager (ie, nurse, social worker, and licensed professional counselor) to create individualized care plans [18,19] centered around chronic disease self-management education, self-care support [8,18,20-24], and to form linkages to medical, behavioral health, and social services [18,25,26]. Participants in the high-tech care management strategy have an initial face-to-face appointment with their care manager and are provided with a preconfigured iPhone that allows for care to continue digitally via a remote monitoring platform. We provide iPhones to participants to ensure access to smartphones, and the cost of cellular data is not a barrier to participation. At the initial appointment, the care manager explains 2 key components of the remote monitoring platform that the participant will use on a regular basis: video visits (eg, video conferencing between patients and care managers) and condition-specific text message check-ins. The remote monitoring platform facilitates video conferencing (eg, video visits) between participants and their care manager. Moreover, as our study population has multiple physical and behavioral health conditions, condition-specific check-in questions are sent via text messages to each participant at varying intervals (eg, daily, weekly, or biweekly) based on their condition(s). Check-ins allow care managers to monitor diverse participant needs, symptoms, or condition exacerbations, including specific biometric readings such as pulse, blood glucose level, weight, and blood pressure.

**Stakeholder Engagement and Feedback Processes**

At the onset of implementation, the study team was acutely aware of the need for continued stakeholder engagement and feedback to promote effective high-tech care implementation. On the basis of the Patient-Centered Outcome Engagement principles [27], we developed 4 mechanisms (Textbox 1) to capture feedback and input from various stakeholder groups. Patients, care managers, and clinical leadership all provided key insights and observations related to technology challenges that care managers or participants experienced during high-tech care implementation. This stakeholder input was collated from the study support log, the case interview log, and the implementation meeting minute log (Textbox 1). Feedback was iteratively reviewed by the study and the clinical team. Stakeholder input was organized by topic and content to understand early stage implementation challenges. Topics were then reviewed by the study team and organized into 2 major thematic categories; themes were reviewed with key stakeholders for validation. Using information from these 3 feedback mechanisms, we developed solutions to the identified technology-related challenges. Solutions were discussed, refined, and implemented with input from the study team, care managers, and clinical leadership. Solutions were also vetted through the Patient Partners Work Group. A work group of patient partners, who have similar characteristics and lived experiences similar to those experienced by our study population, was established through a collaboration with the National Alliance on Mental Illness Southwestern Pennsylvania’s Consumer Action Response Team. Patient partner feedback was tracked in the Patient Partners Work Group log (Textbox 1).

**Textbox 1. Mechanisms to capture stakeholder feedback on challenges and/or solutions.**

Data source and information collected and provided:

- Study support log
  - Study team created a study-specific, toll-free, hotline staffed during office hours
  - Hotline supports care managers and participants with study-related questions or challenges
- Implementation meeting minutes
  - Study team meets with clinical leadership weekly and meets monthly with all care managers
  - Meetings provide a time and space for care managers and their clinical leadership to voice implementation challenges and to strategize potential solutions
- Case interview log
  - Semistructured interviews were conducted with care managers to identify technology-related challenges, participants experience, and workflow impacts
- Patient Partners Work Group log
  - Study team meets regularly with the work group to discuss high-tech care implementation
  - The work group provides feedback on materials and solutions supporting high-tech care engagement/implementation
Understanding Process Modifications: Sources of Information

In order to assess a change in participants’ abilities to overcome technology-related challenges, we analyzed intervention engagement outcomes that may have been impacted by stakeholder-driven implementation solutions. We reviewed the following 3 sources of engagement data pertaining to care activity from April 23, 2018, to December 31, 2019: (1) the participants’ ability to complete a video visit as defined by answering the video call from their care manager, (2) the participants’ ability to answer condition-specific text message check-ins, as defined by receiving a check-in via text message and submitting all answers to condition-specific questions, and (3) participant responses to exit survey questions sent via the remote monitoring platform. All pre- and postdata presented are based on the first in-person technology support visit completed by the participant.

Results

Overview

Stakeholders reported challenges centered around 2 major themes: (1) difficulties using basic functionalities of the smartphone and high-tech care components and (2) difficulties using high-tech care components due to cellular reception and internet connectivity issues. Approximately 500 study hotline phone calls, about technology-specific challenges, were made by 129 participants and 32 care managers to the study team, and the calls were tracked in the study support log. Feedback was also provided during clinical leadership meetings (n=82), monthly care management staff meetings (n=16, tracked via implementation meeting minutes), and semistructured interviews with care managers (n=4, tracked via case interview log). For each thematic challenge, we present: (1) specific stakeholder feedback that leads to solution development, (2) the description of the stakeholder-driven solutions as they are a direct result of stakeholder feedback, and (3) changes in participant engagement data after solution implementation.

Theme 1 Challenge: Smartphone and High-tech Care Digital Component Use

With support from care managers and clinical leadership, the study team focused on common functionality challenges experienced by high-tech care participants and devised 3 main solutions: (1) tech visits (technology support visits), (2) a tech packet (participant-facing technology user guide), and (3) tailored condition-specific text message check-in. Table 1 displays stakeholder feedback regarding the participants’ experiences when using the smartphone and high-tech care components.

<table>
<thead>
<tr>
<th>Data source</th>
<th>Information provided</th>
<th>Informed solution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case interview log and study support log</td>
<td>1. Care managers concerned about time spent teaching participants basic smartphone functionalities 2. Specific technology challenges faced by participants include screen pressure difficulties, home screen navigation, phone charging, text message access, including opening condition-specific check-ins, and phone volume manipulation</td>
<td>1. Tech visits and tech packet 2. Tech packet</td>
</tr>
<tr>
<td>Implementation meeting minutes</td>
<td>1. Clinical leadership interpreted technology-education time concerns as a workflow issue in which care managers had to make up time to ensure participant clinical care needs were met</td>
<td>1. Tech visits and tech packet</td>
</tr>
<tr>
<td>Study support log</td>
<td>1. Smartphone factory resets and reconfiguration were time consuming for care managers 2. Smartphone volume manipulation and battery power were 2 participant challenges that resulted in missed high-tech care video visits and condition-specific check-ins 3. Check-in assignments were sometimes automatically scheduled at inconvenient times for participants (ie, work, school or sleeping hours) 4. Biometric check-in settings were automatically standardized for each participant</td>
<td>1. Tech visits and tech packet 2. Tech packet 3. Tailored condition-specific check-ins 4. Tailored condition-specific check-ins</td>
</tr>
</tbody>
</table>

Technology Support Visit (Tech Visit) Solution for Theme 1: Description and Changes in Engagement

Tech visits are structured to allow a study team member to assist participants with time-consuming digital literacy challenges, either at the participant’s home or at a community location. Tech visits do not replace initial face-to-face training that care managers provide to high-tech care participants; rather, it is a form of supplemental training to ensure that participants are able to use their smartphone to receive care and to reduce the time care managers spend on high-tech care training during the initial appointment. Participants are selected for tech visits if (1) they have called the study hotline multiple times with issues that could not be completely resolved, (2) a care manager is unable to connect with the member due to technology challenges, and (3) clinical supervisors believe a participant’s level of digital literacy requires a substantial amount of care manager’s time.

During the first 20 months of implementation, 73 participants received at least one tech visit. Before the tech visit, 23% (17/73) of the participants completed a video call with their care
Within 30 days of a tech visit, the average rate for video call completion increased to 51% (n=33). In total, 21 members, who had never been able to connect with their care manager via video calls, before the tech visit completed a video call after receiving the tech support. Of the 73 participants, 25 (34%) had video call data for both before and after the tech visit. For these participants, the average rate of completed calls increased by 15% after the completion of a tech visit (Figure 1).

**Figure 1.** Video call data for participants with completed pre- and postdata. CM: care manager.

<table>
<thead>
<tr>
<th>Before tech visit</th>
<th>After tech visit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean number of video calls attempted by CM</td>
<td>▲ 2.44</td>
</tr>
<tr>
<td>1.96 SD= 1.57</td>
<td>4.40 SD= 4.25</td>
</tr>
<tr>
<td>Mean number of video calls answered per participant</td>
<td>▲ 1.36</td>
</tr>
<tr>
<td>0.36 SD= 0.57</td>
<td>1.72 SD= 2.84</td>
</tr>
<tr>
<td>Mean number of video calls completed with a participant</td>
<td>▲ 8</td>
</tr>
<tr>
<td>8</td>
<td>16</td>
</tr>
<tr>
<td>Average rate of video calls completed with a participant</td>
<td>▲ 15%</td>
</tr>
<tr>
<td>22% SD= 0.37</td>
<td>37% SD= 0.39</td>
</tr>
</tbody>
</table>

Data for participants with whom CM attempted calls both before and after tech visit. (N=25)

Specific to condition-specific check-ins sent via text message, most participants completed at least one check-in on their own before their first tech visit (60/73, 82%), with an average rate of 36% (n=59). Within 30 days of a tech visit, the average rate for engagement with check-ins increased to 43% (n=59). Two participants completed check-ins after receiving support, who had never completed a check-in prior. Conversely, 15 participants completed check-ins before receiving support but...
never completed a check-in after the visit. Of the 73 participants, 59 (81%) had check-in data both before and after the tech visit. For these participants, the average rate of completed check-ins increased by 7% after the completion of a tech visit (Figure 2).

**Figure 2.** Check-in data for participants with completed pre- and postdata.

<table>
<thead>
<tr>
<th>Before tech visit</th>
<th>After tech visit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total number of participants who completed check-ins</td>
<td>▼ 12</td>
</tr>
<tr>
<td>58</td>
<td>46</td>
</tr>
<tr>
<td>Average rate of completed check-ins among all participants</td>
<td>▲ 7%</td>
</tr>
<tr>
<td>36% SD= 0.33</td>
<td>43% SD= 0.40</td>
</tr>
</tbody>
</table>

Data for participants who received check-in messages *both before and after tech visit* (N=59)

Note: ▲ Indicates increase    ▼ Indicates decrease
SD indicates standard deviation

**Participant-Facing Technology User Guide (Tech Packet) Solution for Theme 1: Description and Changes in Engagement**

The participant-facing technology user guide (tech packet) is an educational resource for participants learning how to use the smartphone and high-tech care digital components. The tech packet outlines key smartphone functions (ie, how to answer a phone call, how to navigate to the *home screen*, how to open a text message, etc) and high-tech digital care components (ie, how to answer a video call and how to respond to a condition-specific check-in). Care managers give the tech packet to participants during the initial appointment when the participant is randomized into high-tech care; care managers explain the tech packet and have participants practice key functions that they will use throughout their care. Figure 3 shows 2 pages from the tech packet that were developed in response to specific challenges reported in the study support log (Table 1). Feedback from our patient partners, members of the Patient Partners Work Group, was collected over 2-hour-long meetings. Table 2 displays the Patient Partners Work Group feedback on the solution and details on how their feedback was incorporated, and Figure 4 presents a visual example of feedback incorporation. The tech packet is updated regularly based on participant, patient partner, care manager, and clinical leadership feedback.
Figure 3. Tech packet additions based on study support log.

Table 2. Patient Partners Work Group technology user guide feedback.

<table>
<thead>
<tr>
<th>Work group feedback</th>
<th>Examples of feedback</th>
<th>Feedback incorporated</th>
</tr>
</thead>
<tbody>
<tr>
<td>Use less abbreviations/jargon as these are difficult to follow</td>
<td>1. App has many meanings to different participants</td>
<td>1. Addition of definitions page and spelling out abbreviations: app to application</td>
</tr>
<tr>
<td></td>
<td>2. Remote monitoring may have a negative connotation for participants</td>
<td>2. Replacement of technical jargon: Check-ins with your care manager instead of remote monitoring</td>
</tr>
<tr>
<td>Provide an easy start point for each section; assume the lowest level of digital literacy when creating the instructions</td>
<td>1. Each section begins on the smartphone home screen, which assumes participants can find the home screen</td>
<td>1. Addition of instructions on how to navigate to home screen at the beginning and end of each section</td>
</tr>
<tr>
<td>Dexterity and pressure difficulties may be a concern</td>
<td>1. Press, click, touch, and open were used interchangeably</td>
<td>1. Used touch for screen actions and press for the home screen button to distinguish amount of pressure to be applied</td>
</tr>
<tr>
<td>Highlight important contact information associated with care management and study activities</td>
<td>1. Participants may not be able to distinguish who is sending a text message, and some may be concerned about the legitimacy of messages</td>
<td>1. Stated explicitly messages from the remote monitoring platform come from the same phone number each time and make the number visible on all sections</td>
</tr>
<tr>
<td>Participants can be difficult to reach</td>
<td>1. Participants may not be reached during business hours</td>
<td>1. Displayed the study hotline number frequently</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1. Included a section on how to set up and check a voicemail box</td>
</tr>
</tbody>
</table>
We distributed the tech packets to 379 participants. Of the individuals who received a packet, 179 participants were eligible (eg, completed high-tech care before January 1, 2020) to receive the high-tech care exit survey. Of the 179 participants, 96 responded to the survey; 73 respondents strongly agreed or agreed that the tech packet was useful and 62 actively used the guide at least 1 to 2 times a week during care. See Figure 5 for detailed exit survey results.
Figure 5. High-tech care exit survey (96/179, 53.6%).

Check-in Tailoring Solution for Theme 1: Description and Impact

Not all challenges presented to the study team were solved through tech visits and the tech packet. To support participants in completing their condition-specific check-ins, the study team adapted their workflow to allow for a modification of when check-ins are assigned (ie, day of the week and time) based on participant preference. In total, 31 participants requested to change the day/time when check-in text messages were to be received. Second, as all biometric check-in settings (eg, normal boundary parameters for pulse, blood glucose level, weight, and blood pressure) were standardized across participants by default, the study team allowed individual check-in settings to be modified based on the agreement of a participant’s primary medical provider. For 13 participants, the primary medical provider requested biometric setting changes to reflect the participant’s normal range; this modification allows care managers to better track if biometric readings fall outside the participant’s expected range. Textbox 2 describes specific cases in which check-in assignments or biometric settings were modified.

Textbox 2. Condition-specific check-in tailoring examples.

- Check-in assignments modified based on participant preferences
  - Participant rescheduled their diabetes check-in for the morning, based on when their medical provider had instructed them to check their glucose
  - Participant rescheduled their check-in assignment to the day they are off work
- Biometric settings modified based on provider preferences
  - Primary medical provider verified that the participant takes glucose readings before taking insulin and requested setting alerts to be set at 270 or above
  - Primary medical provider approved to change a participant’s blood pressure settings; allowing notification to only send to the care manager when the participant is out of their expected range >170/100 or <90/60

Theme 2 Challenge: Limited Cellular Reception and Internet Connectivity

It was reported, by both care managers and participants themselves, that participants were having difficulties using high-tech care digital components (eg, video visits and condition-specific check-ins) due to limited cellular reception and internet connectivity. Depending on the circumstances, solutions include (1) participant education on how to connect...
a smartphone to the internet, (2) schedule video visits or condition-specific check-ins when the participant is in a location with cellular service or internet, and (3) participant education on how to disconnect from the internet. Table 3 displays the sources of information that led the study team to become aware of the challenge participants experienced with connectivity.

### Table 3. Sources of information and solutions: understanding limited cellular reception and intervention connectivity.

<table>
<thead>
<tr>
<th>Data source</th>
<th>Information provided</th>
<th>Informed solution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study support log and implementation meeting minutes</td>
<td>• Server error messages reported by participants preventing check-in messages from opening</td>
<td>Education on connecting a smartphone to the internet</td>
</tr>
<tr>
<td>Study support log and implementation meeting minutes</td>
<td>• No internet at the home preventing digital tool use</td>
<td>Scheduling video visits/check-ins when participant has access to cellular service or internet</td>
</tr>
<tr>
<td>Study support log and implementation meeting minutes</td>
<td>• Poor internet service at home chosen as default connection method prevented digital tool use</td>
<td>Education on disconnecting from the internet</td>
</tr>
<tr>
<td>Study support log and implementation meeting minutes</td>
<td>• Poor internet service in community chosen as default connection method prevented digital tool use</td>
<td></td>
</tr>
</tbody>
</table>

### Solutions for Theme 2: Description and Changes in Engagement

Care managers communicate with each participant regarding the best way to provide care, given limited cellular reception and internet connectivity. All 3 connectivity solutions are addressed via a tech visit or phone call. Providing education on connecting a smartphone to the internet is often a solution for participants receiving a server error message when attempting to open condition-specific check-ins. Participants were advised to connect to the home internet in case of a bandwidth issue. Scheduling video visits and condition-specific check-ins at specific times affords participants the knowledge that they will have access to cellular service or internet and has been another viable solution. For example, one participant rescheduled their condition-specific check-ins to days of the week when a routine visit was set with family members who have internet access or to business hours as their employer offers internet access. Finally, providing education to participants on how to disconnect from the internet can support video visits or check-in completion. For example, participants are instructed to disconnect weak internet connections, such as home connections or public connections, when they have strong cellular service.

### Discussion

#### Principal Findings

This paper highlights the impact of stakeholder-driven solutions on early implementation challenges specific to a digital care strategy. To support participants in engaging with their smartphone and high-tech care digital components (theme 1 challenge), 3 main solutions were implemented (eg, in-person technology support visits, participant-facing technology user guides, and tailored condition-specific check-ins). For participants who received an in-person technology support visit, we saw an overall increase in engagement with video calls and condition-specific check-ins. For participants who received the participant-facing technology user guide and completed both high-tech care and the exit survey, we found that most used the tech packet while receiving care and or believed it was useful. Finally, condition-specific check-ins were tailored for participants to support engagement and meet their primary medical provider’s care goals. To support participants experiencing difficulties engaging in high-tech care due to limited cellular reception or internet connectivity (theme 2 challenges), 3 solutions were devised and implemented as needed to support engagement in the program (eg, education on both connecting and disconnecting a smartphone to the internet, scheduling high-tech care video visits or condition-specific check-ins at times when the participant is in a physical location that allows connectivity to occur).

Our findings suggest that concurrent stakeholder feedback has the potential to increase implementation success; therefore, it is pivotal to provide stakeholders with multiple and continuous avenues for communicating challenges to the study team. Furthermore, the results stress the importance of working collaboratively with stakeholders early in the implementation of digital interventions to design scalable solutions such as educational materials (tech packet) and activities (tech visits and telephonic support) and refine condition-specific check-ins that suit the specific needs of the patient and their primary medical provider. Moreover, although measuring the success of solutions created during implementation is not always preplanned, early results indicate positive changes in participant technology engagement after tech visits are implemented. Our positive trends in engagement highlight the need for earlier identification of patients who require tech visits to promote early engagement, reduce demoralization, and potentially achieve earlier clinical benefits. Understanding the nuanced challenges of delivering interventions and engaging patients—as well as how to create effective solutions—will advance the efficiency and reach of digital care.

#### Comparison With Prior Work

Current digital care literature focuses on either how tool engagement impacts desired health outcome(s) [6] or defining tool use metrics [28,29]. Processes and solutions to overcome tool utilization barriers are underdeveloped topics in the field that has implications for replication and scaling. One systematic review of digital mental health interventions targeting college students found that of the 89 studies, 45 reported outcomes focused on usability and acceptability (many with low rates of response) and only 2 studies reported on feasibility [30]. It is
critical to expand knowledge centered on how to design and adapt implementation processes in order for digital care teams to be equipped with the right knowledge and resources to best overcome challenges to digital care provision for chronically ill and low-income populations. Providers delivering care digitally must understand and be able to adequately address patient-specific barriers to using digital care tools before patients can engage in the evidence-based tool functions and work toward improved health outcomes.

Our work is an important addition to the discussion on digital care provision as we provide a systematic framework for how digital care providers can work with stakeholder groups to identify and address care delivery implementation challenges. Although most research on remote patient monitoring focuses on single-diagnosis care for older Americans, our intervention targets adults aged 21 years and older who are managing multiple chronic physical and/or behavioral health conditions [6]. Our findings expand knowledge beyond the traditional populations included in digital health research.

Finally, as participants in this research are exclusively eligible for Medicaid, our work promises to reduce health disparities by improving access to digitally delivered evidence-based care management for low-income patients. Supporting consumer adoption of digital health tools is one way to both support patients in their management of chronic conditions and the ethical imperative of reducing health disparities. However, realizing the full potential of digital tools to positively impact health disparities requires continued work on understanding the ways to best support traditionally underserved populations to use digital tools (and how to design those tools and their implementation protocols to meet diverse consumer needs).

Limitations

Although this study contributes novel stakeholder-driven solutions to stakeholder-reported implementation challenges that affect participants’ engagement in a digital care intervention, there are several limitations. Data collected via the high-tech care exit survey may be limited due to a nonresponse bias, as participants who completed the survey had to be able to access the check-in and be willing to complete the survey after completing all care goals. However, our current response rate of 54% indicates that this bias may be less salient [31]. In addition, we were not able to control for additional factors such as in-home caregiver support, improvement in health conditions, or time participating in care that may have also impacted a participant’s ability to overcome the specific challenges that our solutions were designed to target. However, as our solutions were codesigned with key stakeholders and our ability to review pre- and postsolution data, it is reasonable to assume that our solutions influenced positive trends in high-tech care engagement. We also acknowledge that our provision of a smartphone to all study participants may be perceived as a potential barrier to scalability. However, it is important to understand engagement-related challenges for individuals with varying levels of experience with such technology and provide the same phone to all study participants allows us to understand how heterogeneity in technology comfort/experience manifests over the course of the intervention. For future efforts, care managers can and do support individuals with the procurement of a government-issued smartphone that has similar functionality to the phone provided for this study; thus, our findings related to technology engagement can likely be generalized beyond the scope of this study.

Conclusions

To better understand digital care barriers specific to a patient population or care program, it is critical to develop and employ methods for obtaining feedback from key stakeholders before and during implementation. Key stakeholders may include care providers and implementation teams, digital tool creators, and community organizations that represent the population of interest and patients. Our stakeholder-informed solutions include in-person tech visits, a tech packet detailing how to access and use key technology components, along with the tailoring of digital care components to meet both patient and provider needs.

Using high-tech care exit survey responses and remote monitoring engagement data, we measured the impact of our solutions and continued to improve high-tech care delivery. As solutions to challenges develop, detailed tracking of their implementation may positively impact patient engagement with digital tools and ultimately show increased participation in care resulting in improved health outcomes and reductions in health outcome disparities.

Acknowledgments

The authors thank all care managers, clinical leadership, Patient Partners Work Group, Stakeholder Advisory Board, and our study participants for their collaboration and input into each stage of the study thus far and for their continued commitment to this research. The authors also thank the study team for their dedication and preserving adaptability. ClinicalTrials.gov identifier: NCT03451630. This study was partially funded by the Patient-Centered Outcomes Research Institute (PCORI). All statements in this report, including its findings and conclusions, are solely those of the authors and do not necessarily represent the views of PCORI, its Board of Governors or Methodology Committee. The participation of JS in the research project occurred before JS’s appointment to the Board of Governors of the PCORI (R).

Conflicts of Interest

JK has received medication supplies for investigator-initiated trials from Pfizer and Indivior. He received compensation for an educational webinar from Otsuka (condition-specific, not product-focused) within the past 12 months. He receives ongoing compensation for editorial work from the American Association for Geriatric Psychiatry and Physicians Postgraduate Press.
References


Abbreviations

- **PCORI**: Patient-Centered Outcomes Research Institute
Examining an Integrative Cognitive Model of Predicting Health App Use: Longitudinal Observational Study

Kwanho Kim¹*, MA; Chul-Joo Lee²*, PhD

¹Annenberg School for Communication, University of Pennsylvania, Philadelphia, PA, United States
²Department of Communication, College of Social Science, Seoul National University, Seoul, Republic of Korea

*all authors contributed equally

Corresponding Author:
Kwanho Kim, MA
Annenberg School for Communication
University of Pennsylvania
3620 Walnut Street
Philadelphia, PA, 19104
United States
Phone: 1 215 898 7041
Email: kwanho.kim@asc.upenn.edu

Abstract

Background: Specifying the determinants of using health apps has been an important research topic for health scholars as health apps have proliferated during the past decade. Socioeconomic status (SES) has been revealed as a significant determinant of using health apps, but the cognitive mechanisms underlying the relationship between SES and health app use are unknown.

Objective: This study aims to examine the cognitive mechanisms underlying the relationships between SES and use of health apps, applying the integrative model of behavioral prediction (IM). The model hypothesizes the indirect influences of SES on intentions to use health apps, which in turn predict actual use of health apps. The relationships between SES and intentions to use health apps were assumed to be mediated by proximal variables (attitudes, perceived behavioral control [PBC], injunctive norms, and descriptive norms).

Methods: We conducted path analyses using data from a two-wave opt-in panel survey of Korean adults who knew about health apps. The number of respondents was 605 at baseline and 440 at follow-up. We compared our model with two alternative theoretical models based on modified IM to further clarify the roles of determinants of health app use.

Results: Attitudes (β=.220, P<.001), PBC (β=.461, P<.001), and injunctive norms (β=.186, P<.001) were positively associated with intentions to use health apps, which, in turn, were positively related to actual use of health apps (β=.106, P=.03). Income was positively associated with intentions to use health apps, and this relationship was mediated by attitudes (B=0.012, 95% CI 0.001-0.023) and PBC (B=0.026, 95% CI 0.004-0.048). Education was positively associated with descriptive norms (β=.078, P=.03), but descriptive norms were not significantly related to intentions to use health apps. We also found that PBC interacted with attitudes (B=0.043, SE 0.022, P=.046) and jointly influenced intentions to use health apps, whereas the results did not support direct influences of education, income, and PBC on health app use.

Conclusions: We found that PBC over using health apps may be the most important factor in predicting health app use. This suggests the necessity of designing and promoting health apps in a user-friendly way. Our findings also imply that socioeconomic inequalities in using health apps may be reduced by increasing positive attitudes toward, and boosting PBC over, health app use among individuals with low income.

(JMIR Mhealth Uhealth 2021;9(2):e24539) doi:10.2196/24539

KEYWORDS
mHealth; health apps; digital divide; integrative model of behavioral prediction; path analysis
**Introduction**

**Overview**

Health-related apps (health apps) are software on mobile devices providing various health care services [1,2]. Health apps have been considered new communication technologies that may substantially affect public health [1]. As of 2019, it has been estimated that 54.2% of US adults have health apps [3]; use of health apps can promote prohealth behaviors such as healthy eating [4-6] and weight loss [7,8], though effectiveness of each app may vary [9]. To assess the public health impacts of health apps, scholars have explored predictors of health app use.

Several pioneering studies have reported that demographic factors, including education and income, which are widely used indicators of socioeconomic status (SES), are positively associated with use of health apps [10-13]. Furthermore, some studies examined the roles of SES and cognitive factors as potential determinants of health app use [11,14]. Nevertheless, they did not theorize how SES and cognitive factors are related to each other in predicting health app use. As a result, it is still largely unknown why people with higher SES are more apt to use health apps than those with lower SES.

To address this issue, we propose a comprehensive model of predicting health app use that utilizes the integrative model of behavioral prediction (IM). This theory has been frequently adopted by health researchers to explain the cognitive mechanisms underlying people’s health-related behaviors (eg, safe sex, cancer screening, quitting smoking) [15,16]. We test our model with data from a two-wave panel survey of South Korean adults. Last but not least, to further investigate the relationships between determinants of behaviors in IM, we compare our model with other competing models based on modified IM.

**Applying IM to the Context of Health App Use**

IM succeeds the theory of reasoned action [17] and the theory of planned behavior [18]; all three theories posit that behavioral intention is the primary determinant of a behavior [15,16]. Then, IM theorizes the roles of two different types of variables in predicting behaviors: proximal and distal variables. Only proximal variables directly affect intentions; the influences of distal variables on intentions are mediated by proximal variables.

IM claims that intentions can be sufficiently explained with three proximal variables: (1) attitudes (overall favorableness) toward a behavior, (2) subjective norms regarding a behavior, consisting of injunctive norms (perceptions of what is approved or disapproved by close others) and descriptive norms (perceptions about prevalence of a behavior among close others) on a behavior, and (3) perceived behavioral control (PBC) over a behavior (self-evaluated capability in performing a behavior) [16]. Adopting this argument, we posit hypotheses as follows:

H1-H3: Attitudes toward (H1), subjective norms regarding (H2), and PBC over (H3) health app use will be positively associated with intentions to use health apps, which, in turn, will be associated with increased health app use.

However, resources available for those who attempt to change people’s health-related behaviors are limited. Specifying the proximal variable that most strongly influences a target behavior is necessary to find the most efficient way of affecting it [16]. When it comes to health apps, app developers can devise better promotion strategies and improve the design of their apps more efficiently than they do without such knowledge. For instance, if app developers know that PBC is the strongest determinant of adopting and using health apps, and if the developers have a tight budget, they will want to focus on making apps easy to use. This would be the most cost-efficient way of developing the apps. Hence, we pose the following research question:

RQ1: Which proximal variable will most strongly predict intentions to use health apps?

**Applying IM to Investigate Digital Divide in Using Health Apps**

Next, IM categorizes all potential determinants of behaviors other than the proximal variables as distal variables. The relationships between distal variables and intentions are hypothesized to be fully mediated by the proximal variables [16]. Thus, SES is conceptualized as a distal variable in IM. Why, then, does this study concentrate on the relationship between SES and health app use?

Investigating whether and how SES relates to health app use is important in order to know how to reduce the digital divide in using the apps. The digital divide refers to inequalities in accessing and utilizing information and communication technologies (ICT) across different social groups [19-22]; this has been revealed as a substantial cause of health disparities [23-25]. Evidence has supported the digital divide in using health apps due to SES [10-13].

Given that cognitive approaches have contributed to understanding the diffusion of ICT [26,27], theorizing the cognitive mechanisms behind the digital divide in using health apps is important to find effective ways of diminishing it.

Nevertheless, former studies have not asked how SES is associated with cognitive factors in predicting health app use. For example, Chae (2018) juxtaposed education and income with cognitive factors in predicting health app use but did not theorize the relationship between SES and cognitive factors [11]. Mackert et al (2016) controlled for demographics when testing the potential connection between health information literacy (ie, cognitive capacity for processing health information) and use of health ICT including health apps [14]. In sum, we propose the following hypothesis:

H4: Individuals’ SES (education and income) will be positively associated with intentions to use health apps through the mediation of proximal variables.

Moreover, in the following research question, we specify the proximal variable that most strongly mediates the influences of SES on behavioral intentions. This will show what will be the most efficient way of decreasing the gaps in use of health apps across people with different SES:

RQ2: Which proximal variable will most strongly mediate the effects of SES on intentions to use health apps?
Revisiting the Roles of Distal Variables and PBC in IM

Though IM is considered well-established, there are three ongoing controversies regarding the roles of distal variables and PBC in the model [16]. This study will test those competing arguments in the context of health app use.

First, a handful of health studies have found evidence supporting significant direct influences of distal variables on behavioral intentions and actual behaviors [28-30]. Those findings confront two fundamental assumptions of IM: (1) indirect relationships between distal variables and intentions and (2) behavioral intention as the primary determinant of behavior. Given those prior findings, we revisit the role of distal variables in the context of health app use as follows:

RQ3: Will SES be directly associated with intentions to use health apps or actual use of health apps?

Next, the original IM argues that the influence of PBC on behaviors is fully mediated by intentions, and it has been consistently supported by evidence [16,31]. However, a few researchers have suggested that PBC may be directly related to behaviors, bypassing the mediation of intentions, to the extent that PBC may reflect actual control over behaviors [31]. Some health studies have reported evidence supporting this competing argument [32-35]. To examine these potential direct influences of PBC on behaviors in the context of health apps, we propose the following research question:

RQ4: Will PBC be directly associated with actual use of health apps?

Lastly, it has been proposed that PBC may moderate the attitudes-intentions and subjective norms–intentions relationships [31,36-38]. The logic of this hypothesis is that positive attitudes and subjective norms may not translate into a behavior when people do not feel that they have sufficient control over (not) conducting the behavior [31,38]. This issue has not been addressed in the context of health apps, and findings from health studies have been mixed. For instance, PBC significantly moderated only attitudes-intentions relationships in the context of prostate-specific antigen testing, whereas only norms-intentions relationships were significantly moderated by PBC in the context of performing regular exercise [37]. Given these mixed findings and the lack of studies addressing this issue in the health app context, we ask the following question:

RQ5: Will PBC moderate attitudes-intentions and subjective norms–intentions relationships in the context of health app use?

Methods

Survey Data

This study is a part of a larger health communication research project conducted in South Korea. A two-wave opt-in panel survey of Korean adults was collected by a survey company (Embrain). 1718 respondents participated in the baseline survey in February 2016 (completion rate=1718/2415=71.1%). 1304 of those respondents cooperated with a follow-up survey in April 2016 (attrition rate=414/1718=24.1%). The final sample size decreased to 605 at baseline and 440 at follow-up because we included only those who answered “yes” for the following filter question: “Do you know about health apps? Health app refers to health-related software installed on a smartphone or tablet PC to help users to manage their health behaviors.”

If we had provided a brief explanation about health apps, we could have measured proximal variables and intentions even among the respondents who did not know about the apps. However, measures based on very little knowledge have been considered “tentative”; those should be distinguished from “real” views based on good knowledge about the topic [39]. We adopted the filtered sample that allowed us to focus on those real views about health apps; this increased the validity of our measures and findings [39-41].

When it comes to demographic characteristics, the full (N=1718) and filtered (N=605) samples significantly differed only in years of education (full sample mean 14.97 years, 95% CI 14.86-15.08; filtered sample mean 15.30 years, 95% CI 15.13-15.47). The selected baseline (N=605) and follow-up (N=440) respondents were not significantly different with regard to demographic characteristics. For descriptive statistics, see Table 1.
Table 1. Descriptive statistics of variables.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Baseline (n=605)</th>
<th>Follow-up (n=440)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years), mean (SD)</td>
<td>39.00 (10.94)</td>
<td>40.42 (10.81)</td>
</tr>
<tr>
<td>Male sex, n (%)</td>
<td>292 (48.3)</td>
<td>210 (47.7)</td>
</tr>
<tr>
<td>Employment status (employed), n (%)</td>
<td>431 (71.2)</td>
<td>326 (74.1)</td>
</tr>
<tr>
<td>Marital status (married), n (%)</td>
<td>365 (60.3)</td>
<td>284 (64.5)</td>
</tr>
<tr>
<td>Years of education, mean (SD)</td>
<td>15.30 (2.14)</td>
<td>15.32 (2.22)</td>
</tr>
<tr>
<td>Monthly household income (US $)^a, mean (SD)</td>
<td>3910.59 (1627.45)</td>
<td>3920.45 (1600.84)</td>
</tr>
<tr>
<td>Body mass index</td>
<td>23.10 (3.15)</td>
<td>23.16 (3.02)</td>
</tr>
<tr>
<td>Perceived health status^b, mean (SD)</td>
<td>3.41 (0.75)</td>
<td>3.40 (0.75)</td>
</tr>
<tr>
<td>Other source use^c, mean (SD)</td>
<td>2.64 (0.50)</td>
<td>2.66 (0.50)</td>
</tr>
<tr>
<td>Attitudes^d, mean (SD)</td>
<td>5.01 (0.98)</td>
<td>—</td>
</tr>
<tr>
<td>Injunctive norms^f, mean (SD)</td>
<td>2.69 (0.65)</td>
<td>—</td>
</tr>
<tr>
<td>Descriptive norms^g, mean (SD)</td>
<td>2.01 (0.70)</td>
<td>—</td>
</tr>
<tr>
<td>Perceived behavioral control^h, mean (SD)</td>
<td>3.41 (1.01)</td>
<td>—</td>
</tr>
<tr>
<td>Intentions to use health apps^i, mean (SD)</td>
<td>3.32 (1.07)</td>
<td>—</td>
</tr>
<tr>
<td>Types of health apps in use^j, mean (SD)</td>
<td>2.01 (1.25)</td>
<td>1.80 (1.57)</td>
</tr>
<tr>
<td>Frequency of health app use^k, mean (SD)</td>
<td>2.83 (2.47)</td>
<td>2.50 (2.54)</td>
</tr>
<tr>
<td>Health app use (composite measure)^l, mean (SD)</td>
<td>0.00 (1.56)</td>
<td>0.00 (1.65)</td>
</tr>
</tbody>
</table>

---

^aIncome was measured on an 8-point scale: 1=US $990 or lower to 8=US $7000 or higher. We averaged household income after recoding the response options into a ratio variable (eg, 2=US $1000 to $1990 was recoded as US $1500).

^b1=very bad to 5=very good.

^cMean of seven items tapping use of health information sources other than health apps. 1=never, 4=frequently.

^dMean of two attitudes items. 1=very negative, 7=very positive.

^eNot available.

^f1=strongly disagree (low norms), 4=strongly agree (high norms).

^g1=none of them (low norms), 4=everyone (high norms).

^h1=no confidence, 5=complete confidence.

^i1=extremely low, 5=extremely high.

^jSum of 14 items of a certain type of health app use. 0=no (no use), 1=yes (use).

^k0=never to 7=everyday.

^lSum of the standardized values of frequency of health app use and the number of types of health apps in use.

Measures

We created survey items capturing IM-related variables in the context of health apps, following the guidelines from Fishbein and Ajzen (2010) [16]. Moreover, we developed a composite measure of health app use following prior studies’ measure creation procedures [42,43].

Distal Variables (Baseline)

We adopted education and income as proxies of SES. For education, we asked respondents their highest level of schooling completed (1=elementary school not completed to 8=doctorate). The response options were recoded into years usually required to finish a given type of education in the nation. Income was captured employing an 8-point scale (1=US $990 or lower to 8=US $7000 or higher) and then recoded using the midpoint of each option (eg, 2=US $1000 to $1990 was recoded as US $1500) [44-47].

Proximal Variables (Baseline)

To measure attitudes toward using health apps, we averaged participants’ answers to the following two 7-point semantic differential scale items (r=0.81): “Using mobile health apps in the next two months would be...” (1=very bad to 7=very good; 1=very unenjoyable to 7=very enjoyable).

We measured injunctive norms by asking respondents about their agreement with the following sentence: “Most people important to me think I should use health apps in the next two months (1=strongly disagree to 4=strongly agree).” Descriptive norms regarding health app use were captured by asking for respondents’ perceptions of how many people important to them had employed health apps in the past two months (1=none of
them to 4=everyone). As the correlation between the two norms was only 0.28, they were treated as separate variables in the analyses.

PBC was measured by asking how sure respondents were that they could use health apps on most days in the next two months if they wanted to (1=very unsure to 5=very sure).

**Intentions to Use Health Apps (Baseline)**
To capture intentions to use health apps, we asked participants to report their likelihood of using health apps in the next two months (1=very unlikely to 5=very likely).

**Health App Use (Baseline & Follow-up)**
A composite scale of health app use was constructed by summing the standardized values of the two variables: the frequency of health app use and the number of types of health apps in use (r=0.21 at baseline and 0.37 at follow-up). Each reflects the depth and breadth of health app use. The frequency was captured by asking participants about how often they used any health app (0=not at all to 7=all 7 days a week). The other was measured with an additive index of 14 dichotomous items (0=no, 1=yes; the number of users at baseline in parentheses): (a) exercise & fitness (539); (b) healthy diet (145); (c) weight control (141); (d) blood pressure (66); (e) blood sugar (31); (f) menstruation (132); (g) pregnancy (17); (h) baby care (30); (i) medication (12); (j) health information & news (93); (k) mental health (35); (l) sleep (69); (m) quit smoking (20); and (n) beauty (28).

**Control Variables (Baseline)**
We measured demographic variables (age, sex, marital status, employment status, education, and monthly household income) and health-related variables (body mass index, perceived health status). Additionally, we captured use of health information sources other than health apps (hereafter, other source use) by averaging how often respondents obtained health information from the following seven sources (1=never to 4=frequently): printed media, TV, social media, health websites, general websites, friends, and health professionals.

**Analysis Strategy**
We performed path analyses via Mplus 8.3 (Muthén & Muthén). Throughout all analyses, we controlled for all potential confounders described above and health app use at baseline (ie, past behavior). To evaluate model fit, we used a root mean square error of approximation (RMSEA), a comparative fit index (CFI), and a standardized root mean square residual (SRMR); RMSEA≤0.05, CFI≥0.95, and SRMR≤0.08 indicated a well-fitting model [48]. We employed the maximum likelihood with robust standard errors (MLR) method to test direct and interaction effects and bootstrapping (5000 samples) to examine indirect effects. Indirect relationships were considered significant when the bias-corrected 95% CI of unstandardized coefficients did not contain 0 [48]. We adopted the full information maximum likelihood method to handle missing values (aka, FIML and Direct ML). Figure 1 is a graphical illustration of the analysis strategy.

**Figure 1.** Conceptual framework of the analytic procedure. Bolded solid lines represent paths included in the basic model (Model A). Dashed lines represent paths added to the basic model in each competing model. None of the additional paths were adopted in the final model (Model A). Thin solid lines represent interaction relations added to each of Models D, E1, and E2. SES: socioeconomic status.
The Original IM-based Model

We began by fitting an original IM-based model (hereafter, Model A). In Model A, health app use at follow-up was directly predicted only by intentions at baseline; the associations between intentions and the distal variables (education and income) were fully mediated by proximal variables. We estimated direct and indirect path coefficients to test original IM-based hypotheses (H1 to H4); to compare the relative importance of proximal variables (RQ1 and RQ2), the differences between certain pairs of coefficients were examined with the $\chi^2$ difference test. For bivariate correlations of variables in Model A, see Table 2.

Table 2. Bivariate correlations of variables in Model A (best-fitting model).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Education</th>
<th>Income</th>
<th>Attitudes</th>
<th>IN(^a)</th>
<th>DN(^b)</th>
<th>PBC(^c)</th>
<th>Intentions</th>
<th>App use (B)(^d)</th>
<th>App use (F)(^e)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Education</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$r$</td>
<td>1.000</td>
<td>0.157</td>
<td>0.083</td>
<td>0.052</td>
<td>0.094</td>
<td>0.021</td>
<td>0.086</td>
<td>0.078</td>
<td>0.090</td>
</tr>
<tr>
<td>$P$ value</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td><strong>Income</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$r$</td>
<td>0.157</td>
<td>1.000</td>
<td>0.151</td>
<td>0.087</td>
<td>0.078</td>
<td>0.137</td>
<td>0.097</td>
<td>0.122</td>
<td>0.097</td>
</tr>
<tr>
<td>$P$ value</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td><strong>Attitudes</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$r$</td>
<td>0.083</td>
<td>0.151</td>
<td>1.000</td>
<td>0.482</td>
<td>0.227</td>
<td>0.582</td>
<td>0.584</td>
<td>0.396</td>
<td>0.268</td>
</tr>
<tr>
<td>$P$ value</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td><strong>IN</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$r$</td>
<td>0.052</td>
<td>0.087</td>
<td>0.482</td>
<td>1.000</td>
<td>0.275</td>
<td>0.460</td>
<td>0.511</td>
<td>0.321</td>
<td>0.290</td>
</tr>
<tr>
<td>$P$ value</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td><strong>DN</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$r$</td>
<td>0.094</td>
<td>0.078</td>
<td>0.227</td>
<td>0.275</td>
<td>1.000</td>
<td>0.208</td>
<td>0.224</td>
<td>0.184</td>
<td>0.137</td>
</tr>
<tr>
<td>$P$ value</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td><strong>PBC</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$r$</td>
<td>0.021</td>
<td>0.137</td>
<td>0.582</td>
<td>0.460</td>
<td>0.208</td>
<td>1.000</td>
<td>0.680</td>
<td>0.500</td>
<td>0.336</td>
</tr>
<tr>
<td>$P$ value</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td><strong>Intentions</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$r$</td>
<td>0.086</td>
<td>0.097</td>
<td>0.584</td>
<td>0.511</td>
<td>0.224</td>
<td>0.680</td>
<td>1.000</td>
<td>0.581</td>
<td>0.368</td>
</tr>
<tr>
<td>$P$ value</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td><strong>App use (B)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$r$</td>
<td>0.078</td>
<td>0.122</td>
<td>0.396</td>
<td>0.321</td>
<td>0.184</td>
<td>0.500</td>
<td>0.581</td>
<td>1.000</td>
<td>0.514</td>
</tr>
<tr>
<td>$P$ value</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td><strong>App use (F)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$r$</td>
<td>0.090</td>
<td>0.097</td>
<td>0.268</td>
<td>0.290</td>
<td>0.137</td>
<td>0.336</td>
<td>0.368</td>
<td>0.514</td>
<td>1.000</td>
</tr>
<tr>
<td>$P$ value</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
</tbody>
</table>

\(^a\)IN: injunctive norms.  
\(^b\)DN: descriptive norms.  
\(^c\)PBC: perceived behavioral control.  
\(^d\)B: baseline.  
\(^e\)F: follow-up.  
\(^f\)Not applicable.

Model Comparisons

To address the theoretical controversies about IM, we first created Model B by modifying Model A to include direct links of distal variables (ie, education and income) with intentions and health app use. We compared Models A and B via the $\chi^2$ difference test (RQ3). Notably, since we used $\chi^2$ values from MLR estimations, the values were first adjusted using the scaling correction factors and then employed for the difference tests. The model fitting the data better was selected and then compared...
with Model C, constructed by allowing the selected model (Model A or B) to have a direct association between PBC and health app use (RQ4). The better-fitting model in the last comparison was chosen as the final model.

**Moderating Roles of PBC**
To assess the potential moderating roles of PBC (RQ5), we added three mean-centered interaction terms to the final model one at a time ("Attitudes × PBC," “Injunctive norms × PBC,” and "Descriptive norms × PBC”; Models D, E1, and E2, respectively). We checked the significance of the interaction in each model.

**Results**

**The Original IM-based Model**
Model A fit the data well (Table 3). We found that intentions at baseline predicted health app use at follow-up (B=0.164, SE 0.075, β=.106, P=.03). The effects of attitudes (B=0.241, SE 0.039, β=.220, P<.001) (H1), injunctive norms (B=0.307, SE 0.027, β=.209, P<.001) (H2), and PBC (B=0.491, SE 0.046, β=.461, P<.001) (H3) on intentions were significant, whereas descriptive norms showed no significant association with intentions (B=0.041, SE 0.046, β=.027, P=.38) (H2). Accordingly, the indirect effects of attitudes (B=0.040, 95% CI 0.002-0.077) (H1), injunctive norms (B=0.050, 95% CI 0.001-0.102) (H2), and PBC (B=0.081, 95% CI 0.007-0.154) (H3) on follow-up health app use were significant, while the indirect effects of descriptive norms on health app use were not (B=0.007, 95% CI −0.010 to 0.023) (H2). PBC was more strongly related to intentions than were attitudes (β=0.053, SE 0.022, P<.001) and injunctive norms (β=0.046, SE 0.027, P=.04). However, the attitudes-intentions relationship was not significantly different from the injunctive norms–intentions relationship (χ² diff,1=6.1, P=.19). That is, direct effects of distal variables on health app use were not supported (RQ3). The winning model, Model A, was further compared with Model C. Still, Model C did not fit the data significantly better than Model A (χ² diff,1=1.4, P=.24). In other words, there was no evidence for direct associations of PBC with health app use (RQ4).

**Conceptual Models**

<table>
<thead>
<tr>
<th>Modelsa</th>
<th>Chi-square (df)b</th>
<th>RMSEA² (90% CI)</th>
<th>CFI²</th>
<th>SRMR²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model A</td>
<td>69.2 (35)</td>
<td>0.040 (0.026-0.054)</td>
<td>0.975</td>
<td>0.027</td>
</tr>
<tr>
<td>Model B</td>
<td>63.1 (31)</td>
<td>0.041 (0.027-0.056)</td>
<td>0.976</td>
<td>0.025</td>
</tr>
<tr>
<td>Model C</td>
<td>67.7 (34)</td>
<td>0.040 (0.026-0.055)</td>
<td>0.975</td>
<td>0.026</td>
</tr>
<tr>
<td>Model D</td>
<td>74.7 (37)</td>
<td>0.041 (0.027-0.054)</td>
<td>0.988</td>
<td>0.027</td>
</tr>
<tr>
<td>Model E1</td>
<td>70.4 (37)</td>
<td>0.039 (0.024-0.052)</td>
<td>0.975</td>
<td>0.025</td>
</tr>
<tr>
<td>Model E2</td>
<td>70.3 (37)</td>
<td>0.039 (0.024-0.052)</td>
<td>0.975</td>
<td>0.025</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Modelsa</th>
<th>Chi-square (df)b</th>
<th>RMSEA² (90% CI)</th>
<th>CFI²</th>
<th>SRMR²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model A</td>
<td>69.2 (35)</td>
<td>0.040 (0.026-0.054)</td>
<td>0.975</td>
<td>0.027</td>
</tr>
<tr>
<td>Model B</td>
<td>63.1 (31)</td>
<td>0.041 (0.027-0.056)</td>
<td>0.976</td>
<td>0.025</td>
</tr>
<tr>
<td>Model C</td>
<td>67.7 (34)</td>
<td>0.040 (0.026-0.055)</td>
<td>0.975</td>
<td>0.026</td>
</tr>
<tr>
<td>Model D</td>
<td>74.7 (37)</td>
<td>0.041 (0.027-0.054)</td>
<td>0.988</td>
<td>0.027</td>
</tr>
<tr>
<td>Model E1</td>
<td>70.4 (37)</td>
<td>0.039 (0.024-0.052)</td>
<td>0.975</td>
<td>0.025</td>
</tr>
<tr>
<td>Model E2</td>
<td>70.3 (37)</td>
<td>0.039 (0.024-0.052)</td>
<td>0.975</td>
<td>0.025</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Modelsa</th>
<th>Chi-square (df)b</th>
<th>RMSEA² (90% CI)</th>
<th>CFI²</th>
<th>SRMR²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model A</td>
<td>69.2 (35)</td>
<td>0.040 (0.026-0.054)</td>
<td>0.975</td>
<td>0.027</td>
</tr>
<tr>
<td>Model B</td>
<td>63.1 (31)</td>
<td>0.041 (0.027-0.056)</td>
<td>0.976</td>
<td>0.025</td>
</tr>
<tr>
<td>Model C</td>
<td>67.7 (34)</td>
<td>0.040 (0.026-0.055)</td>
<td>0.975</td>
<td>0.026</td>
</tr>
<tr>
<td>Model D</td>
<td>74.7 (37)</td>
<td>0.041 (0.027-0.054)</td>
<td>0.988</td>
<td>0.027</td>
</tr>
<tr>
<td>Model E1</td>
<td>70.4 (37)</td>
<td>0.039 (0.024-0.052)</td>
<td>0.975</td>
<td>0.025</td>
</tr>
<tr>
<td>Model E2</td>
<td>70.3 (37)</td>
<td>0.039 (0.024-0.052)</td>
<td>0.975</td>
<td>0.025</td>
</tr>
</tbody>
</table>

aModel A: the original IM model; Model B: Model A with direct paths of distal variables on intentions and behavior; Model C: Model A with a direct path of perceived behavioral control (PBC) on behavior; Model D: Model A with an interaction term “Attitudes × PBC”; Model E1: Model A with an interaction term “Injunctive norms × PBC”; Model E2: Model A with an interaction term “Descriptive norms × PBC”. Models D, E1, and E2 are employed only for interaction tests.

bAll χ² test results were significant (P<.001).

RMSEA: root mean square error of approximation.

CFI: comparative fit index.

SRMR: standardized root mean square residual.

Education was positively associated with descriptive norms (B=0.026, SE 0.012, β=0.078, P=.03), while income was positively related to attitudes (B=0.050, SE 0.022, β=0.096, P=.02) and PBC (B=0.053, SE 0.022, β=0.100, P=.02). The indirect associations between income and intentions through the mediation of attitudes (B=0.012, 95% CI 0.001-0.023) and PBC (B=0.026, 95% CI 0.004-0.048) were significant. However, the indirect relationships between education and intentions through the mediation of descriptive norms were not significant (B=0.001, 95% CI −0.002 to 0.004) (H4). There was no significant difference between the income-attitudes relationship and the income-PBC relationship (χ² diff,1=0.0, P=.88) (RQ2). Overall, we found some evidence supporting H4; the importance of proximal variables in predicting intentions to use health apps was not statistically different.

**Model Comparisons**
Model A turned out to be the best-fitting model, although Models B and C also fit the data well (Table 3). In the first round of comparison (Model A vs Model B), we found that Model B did not explain the data significantly better than Model A (χ² diff,1=6.1, P=.19). That is, direct effects of distal variables on health app use were not supported (RQ3). The winning model, Model A, was further compared with Model C. Still, Model C did not fit the data significantly better than Model A (χ² diff,1=1.4, P=.24). In other words, there was no evidence for direct associations of PBC with health app use (RQ4).

**Moderating Roles of PBC**
All models (Models D, E1, and E2) fit the data well (Table 3). The interaction between attitudes and PBC (B=0.043, SE 0.022, P=.046) was significant, while neither injunctive norms (B=0.012, SE 0.036, P=.75) nor descriptive norms (B=−0.032, P=.186) showed significant association with intentions (B=0.041, SE 0.046, β=.027, P=.38) (H2). Accordingly, the indirect effects of attitudes (B=0.040, 95% CI 0.002-0.077) (H1), injunctive norms (B=0.050, 95% CI 0.001-0.102) (H2), and PBC (B=0.081, 95% CI 0.007-0.154) (H3) on follow-up health app use were significant, while the indirect effects of descriptive norms on health app use were not (B=0.007, 95% CI −0.010 to 0.023) (H2). PBC was more strongly related to intentions than were attitudes (β=0.053, SE 0.022, P<.001) and injunctive norms (β=0.046, SE 0.027, P=.04). However, the attitudes-intentions relationship was not significantly different from the injunctive norms–intentions relationship (χ² diff,1=6.1, P=.19). That is, direct effects of distal variables on health app use were not supported (RQ3). The winning model, Model A, was further compared with Model C. Still, Model C did not fit the data significantly better than Model A (χ² diff,1=1.4, P=.24). In other words, there was no evidence for direct associations of PBC with health app use (RQ4).
SE 0.047, \( P = .49 \) significantly interacted with PBC in predicting intentions. The results from the Johnson-Neyman technique [49] showed that the 95% CI of the conditional effect of attitudes on intentions was always above 0 (Figure 2). That is, at any range of PBC, the influence of attitudes on intentions was significantly larger for people with higher PBC than for those with lower PBC (RQ5).

**Figure 2.** Conditional effect of attitudes on intentions to use health apps as a function of perceived behavioral control from the Johnson-Neyman technique. The 95% CI of the conditional effect of attitudes on intentions to use health apps is always above 0, which means that the effect of attitudes is significantly positive for any value of PBC. PBC: perceived behavioral control.

### Discussion

As expected, attitudes, PBC, and injunctive norms were associated with intentions, which, in turn, were related to health app use. In contrast, descriptive norms were not significantly related to intentions; thus, they did not affect health app use. PBC positively interacted with attitudes and jointly influenced intention. The association between income and intention was mediated by attitudes and PBC; education was associated with descriptive norms, but the indirect relationship between education and intention was not significant.

Several limitations of this study should be discussed. First, as our data do not represent the Korean adult population, the generalizability of our findings may be restricted. Second, we used single-item questions to measure norms and PBC; future studies should consider employing multiple-item measures. Third, we cannot eliminate the concern of reverse causality because distal and proximal variables and intentions were measured at baseline. Fourth, our health app use measure cannot distinguish people using one app frequently from those who use many apps, but less frequently. However, our measure may better capture the actual pattern of health app use than binary measures (ie, use or no use) adopted in prior studies [11-14,50]. Lastly, future studies may need to control for factors possibly related to both a distal variable and health app use, such as health literacy, which may correlate to SES and health app use.

The theoretical implications of the findings should be highlighted. First, PBC was most strongly associated with intentions. This finding is inconsistent with a well-known argument that subjective norms are the most powerful predictors of behavioral intentions in collectivist cultures, including Korean, while attitudes are key determinants of intentions in individualistic cultures [51]. The relatively low penetration rate of health apps in Korea may explain this discrepancy [52,53]. PBC positively moderated the effects of attitudes on intentions; this has been repeatedly reported in other contexts [36,37,54]. In contrast, the effects of subjective norms on intentions were not moderated by PBC. This finding is not consistent with the prior evidence from Western countries [36,37,54,55]. Perhaps, as subjective norms are more robustly related to behavioral intentions in collectivist cultures than they are in individualistic...
that PBC was the strongest determinant of intentions to use health apps and moderated the influences of attitudes on intentions. To boost PBC, an app should be designed and promoted in a user-friendly way (eg, using plain and easy-to-read language; providing easy-to-follow guidelines) so that potential users will not experience difficulties in using the app. Second, this study suggests that, to reduce the digital divide in health app use, public health professionals should instill in low-income individuals beliefs about expected positive outcomes from, and confidence in, using health apps. This strategy would thereby form favorable attitudes toward and greater PBC over health app use. Health apps are frequently monetized; thus, they are designed to target people with high SES to maximize their developers’ profits [56,57]; given our findings, this trend is particularly worrisome because it can maintain or even worsen inequalities in public health outcomes.

Acknowledgments
This work was supported by the National Research Foundation of Korea (NRF-2018S1A5B8070398 to CJL) and the Institute of Communication Research at Seoul National University (to CJL).

Conflicts of Interest
None declared.

References


45. Hout M. Getting the most out of the GSS income measures.: University of California, Berkeley; 2018. URL: https://www2.isr.umich.edu/src/~/Publications/technicalReports/2018/2018TR015.pdf


51. Kelly BJ, Leader AE, Mittermaier DJ, Hornik RC, Cappella JN. The HPV vaccine and the media: how has the topic been covered and what are the effects on knowledge about the virus and cervical cancer? Patient Educ Couns 2009 Nov;77(2):308-313 [FREE Full text] [doi: 10.1016/j.pec.2009.03.018] [Medline: 19395221]


Abbreviations

- CFI: comparative fit index
- ICT: information and communication technology
- IM: integrative model of behavioral prediction
- MLR: maximum likelihood with robust standard errors
- PBC: perceived behavioral control
- RMSEA: root mean square error of approximation
- SES: socioeconomic status
- SRMR: standardized root mean square residual
Digital Medicine Community Perspectives and Challenges: Survey Study

Brinnae Bent¹, MSc; Ida Sim², PhD, MD; Jessilyn P Dunn¹,³, PhD

¹Department of Biomedical Engineering, Duke University, Durham, NC, United States
²Department of Medicine, University of California, San Francisco, San Francisco, CA, United States
³Department of Biostatistics & Bioinformatics, Duke University Medical Center, Durham, NC, United States

Corresponding Author:
Jessilyn P Dunn, PhD
Department of Biostatistics & Bioinformatics
Duke University Medical Center
2424 Erwin Road
Durham, NC, 27705
United States
Phone: 1 9196689798
Email: jessilyn.dunn@duke.edu

Abstract

Background: The field of digital medicine has seen rapid growth over the past decade. With this unfettered growth, challenges surrounding interoperability have emerged as a critical barrier to translating digital medicine into practice. In order to understand how to mitigate challenges in digital medicine research and practice, this community must understand the landscape of digital medicine professionals, which digital medicine tools are being used and how, and user perspectives on current challenges in the field of digital medicine.

Objective: The primary objective of this study is to provide information to the digital medicine community that is working to establish frameworks and best practices for interoperability in digital medicine. We sought to learn about the background of digital medicine professionals and determine which sensors and file types are being used most commonly in digital medicine research. We also sought to understand perspectives on digital medicine interoperability.

Methods: We used a web-based survey to query a total of 56 digital medicine professionals from May 1, 2020, to July 10, 2020, on their educational and work experience, the sensors, file types, and toolkits they use professionally, and their perspectives on interoperability in digital medicine.

Results: We determined that the digital medicine community comes from diverse educational backgrounds and uses a variety of sensors and file types. Sensors measuring physical activity and the cardiovascular system are the most frequently used, and smartphones continue to be the dominant source of digital health information collection in the digital medicine community. We show that there is not a general consensus on file types in digital medicine, and data are currently handled in multiple ways. There is consensus that interoperability is a critical impediment in digital medicine, with 93% (52) of survey respondents in agreement. However, only 36% (20) of respondents currently use tools for interoperability in digital medicine. We identified three key interoperability needs to be met: integration with electronic health records, implementation of standard data schemas, and standard and verifiable methods for digital medicine research. We show that digital medicine professionals are eager to adopt new tools to solve interoperability problems, and we suggest tools to support digital medicine interoperability.

Conclusions: Understanding the digital medicine community, the sensors and file types they use, and their perspectives on interoperability will enable the development and implementation of solutions that fill critical interoperability gaps in digital medicine. The challenges to interoperability outlined by this study will drive the next steps in creating an interoperable digital medicine community. Establishing best practices to address these challenges and employing platforms for digital medicine interoperability will be essential to furthering the field of digital medicine.

(JMIR Mhealth Uhealth 2021;9(2):e24570) doi:10.2196/24570

KEYWORDS
digital medicine; digital health; interoperability; mHealth; wearables; sensors
Introduction

Digital medicine is defined as the use of technologies as tools for measurement and intervention in the service of human health [1]. Here we will focus on the use of mobile health (mHealth) and wearable sensors for digital medicine applications, which are growing rapidly in importance for health care applications, particularly during the COVID-19 pandemic. The field of digital medicine has seen rapid growth over the last decade [2]. This growth has resulted from a combination of health care costs and utilization at an all-time high [3] and the consistent improvements in mHealth and wearable technology that have resulted in their wide prevalence and accessibility [4-6].

While the field of digital medicine has seen rapid growth, many challenges remain: standards, best practices, and oversight methodology are still under development [7], sensors and devices used in digital medicine are constantly evolving and are often not validated [8], and a lack of interoperability results in a fragmented, inconvenient, and sometimes impossible adoption of digital medicine into medical practice [9-11]. In order to understand how to mitigate these challenges, it is critical to understand who is using and developing digital medicine tools, which tools are most utilized, and what common perspectives are on current challenges in the field.

Few data are available regarding the landscape of the digital medicine community and the challenges researchers in digital medicine are currently facing. In this study, we surveyed 56 digital medicine professionals to understand the topography of digital medicine, including the background and perspectives of digital medicine professionals, which sensors and file types they use, and perspectives on interoperability challenges. We asked additional questions on background and perspectives on interoperability challenges. If participants checked that they were involved in digital medicine research, research and development (R&D), or both, we asked additional questions on sensors and file types they use (40 participants). Participants were able to review their survey answers prior to submission.

Methods

An open, web-based survey was conducted from May 1, 2020, to July 10, 2020. The survey was conducted using Google Forms, and usability was tested internally prior to survey deployment. The survey consisted of 19 questions over 4 pages. The survey (Multimedia Appendix 1) was disseminated via monthly email newsletter and Slack to the Digital Medicine professional society (DiMe; approximate reach of 1250 digital medicine professionals), via email to 10 subject matter experts who disseminated the survey in their networks, and via social media (Multimedia Appendix 2). A total of 22 tweets and retweets were tweeted on Twitter, deploying the survey to networks totaling more than 1500 individuals. This survey protocol was approved by the Duke University Campus Institutional Review Board (#2020-0450). All participants provided consent and were provided the survey duration and purpose of the study prior to the survey. No personal information was collected or stored. There were no incentives offered to complete the survey. Survey completion after consent was 100%. All respondents (56 participants) were asked questions on background and perspectives on interoperability challenges. If participants checked that they were involved in digital medicine research, research and development (R&D), or both, we asked additional questions on sensors and file types they use (40 participants). Participants were able to review their survey answers prior to submission.

PubMed literature review was conducted on July 14, 2020, with keywords listed in Multimedia Appendix 3. Results were limited to the time span of 2010-2020.

Results

Who Makes Up the Field of Digital Medicine?

We found that our sample of the field of digital medicine is made up of people with diverse backgrounds (Figure 1). The most common educational backgrounds and current roles among the 56 survey respondents include data science/analytics/machine learning (18), business/entrepreneurship (17), and medicine (practitioners) (16). However, backgrounds were diverse and included nutrition, psychology, economics, design, marketing, and theatre. The sector breakdown for digital medicine professionals in this survey was industry (25, 45%), academia (17, 30%), medical institution (5, 9%), startup/freelance (4, 7%), government (3, 5%), and nonprofit (2, 4%) (Figure 1A). Just over half (52%, 29) of survey respondents hold a doctorate as their terminal degree, and 39% (22) hold a master’s degree as their terminal degree (Figure 1B).
Which Sensors Do Digital Medicine Researchers Use?

Respondents who were active participants in digital medicine research and R&D at the time of the survey (n=40) answered with a wide variety of responses to the question “Which sensors and devices do you regularly work with?” A total of 153 sensors were reported to be used by these 40 researchers (mean 3.6, median 3 sensors or devices per respondent; Figure 2). Of these 153 sensors, 143 sensors were associated with a particular measurement modality: 39.2% (56) monitored the cardiovascular system, 35.7% (51) measured physical activity, 8.4% (12) measured physiological temperature, 4.9% (7) measured electrodermal activity, 4.9% (7) monitored behavior, adherence, or location, 4.2% (6) monitored brain activity, 2.1% (3) monitored respiration or oxygen consumption, and 0.7% (1) of sensors were reported as proprietary (Figure 2B). The top three reported devices or sensors used by survey respondents include smartphone (iPhone or Android) (8), Apple Watch (7), and Fitbit (7).
Figure 2. Sensors used by researchers in digital medicine. (A) Types of devices and sensors reported. (B) Measurement modalities of reported devices and sensors. ECG: electrocardiogram; EEG: electroencephalogram; HR: heart rate; PPG: photoplethysmogram.

The results of the survey are consistent with the literature: studies indexed in PubMed in the last decade include electrocardiogram (number of articles \( n = 59,114 \)), photoplethysmogram (\( n = 2819 \)), accelerometer (\( n = 11,430 \)), electrodermal activity (\( n = 729 \)), temperature sensor (\( n = 11,980 \)), gyroscope (\( n = 1417 \)), and pulse oximetry (\( n = 7638 \)). Smartphones (\( n = 12,485 \)) are also popular tools for medical research. Fitbit was found to be the most common smartwatch cited in PubMed-indexed research (\( n = 624 \)), followed by Garmin (\( n = 141 \)) and Apple Watch (\( n = 136 \)).

**Which Data Formats Are Most Commonly Used in Digital Medicine Research?**

The most commonly used data formats among survey respondents include comma-separated values (.csv), JavaScript Object Notation (JSON), and Microsoft XML spreadsheets (.xls/.xlsx) (Figure 3). The most popular file type for both raw files (data sourced directly from the device or company database) and processed files was .csv. While JSON was used more frequently in raw file types, .xls/.xlsx was more frequently the file type researchers reported to use for analysis (Figure 3). Interestingly, there was more diversity among raw file formats (\( n = 114 \)) than file types that researchers map to for analysis (\( n = 79 \)). For raw file formats, respondents listed a mean of 2.75 file types and a median of 2.5 file types. For analysis file formats, respondents listed a mean of 2 file types and median of 2 file types.
Interoperability in Digital Medicine

Of the 56 survey respondents, 93% (52) agreed that interoperability is a problem in digital medicine (Figure 4A).

Figure 4. Interoperability in digital medicine. (A) Challenges in digital medicine interoperability. (B) Tools used for interoperability. API: application programming interface; FHIR: Fast Healthcare Interoperability Resources; HL7: Health Level Seven.

While nearly all respondents (52, 93%) believe interoperability is a problem in digital medicine, only 36% (20) currently use tools for interoperability. Those tools for interoperability include Health Level Seven (HL7) Fast Healthcare Interoperability Resources (FHIR), Apple Health, general application programming interfaces (APIs), the Medisafe platform, Epic, Protege, and in-house solutions (Figure 4B).

When asked if they would utilize a platform for standardizing and validating digital medicine algorithms, methodologies, and analyses, 100% (53) of respondents said “Yes” (27, 51%) or “Maybe” (26, 49%). Reasons for use, as described by survey respondents, included that open science increases efficiency and improves reproducibility, work quality, and readability, and that such a platform would allow for direct comparisons of
analytical results. Considerations for using this type of platform included understanding data security and potential risks.

When asked whether they would use a platform mapping raw data files to a standard format, 84% (47) of respondents said “Yes” (13, 23%) or “Maybe” (34, 61%). Considerations for using this platform included understanding compatibility with electronic health record systems, whether this platform would save time, and whether the resulting file types aligned with the desired file types for analysis.

**Discussion**

**Overview**

In a survey of 56 digital medicine professionals, we sought to better understand the digital medicine community. Key challenges in the field of digital medicine include the following: standards, best practices, and oversight methodology are still under development [7]; sensors and devices used in digital medicine are constantly evolving and are often not validated [8]; and a lack of interoperability discourages adoption of digital medicine into medical practice [9-11]. To understand how these challenges can be mitigated, we explored which sensors are most commonly used in digital medicine research, which file types are most commonly used by the digital medicine community, and perspectives in the field regarding challenges in interoperability.

**Diversity of the Digital Medicine Community Sets Us Apart**

We showed great diversity of backgrounds and current roles in the digital medicine community. We hope that this diversity will be encouraging to those looking to join the digital medicine community who may come from “nontraditional” backgrounds. The diversity of backgrounds in the digital medicine community is one of our most important assets for developing a common language, strong frameworks, and best practices. This diversity in thought and experience has resulted in a uniquely heterogeneous set of voices and perspectives contributing to community standards and best practices in research and application [1,7,12-14].

**Digital Medicine Research Is Largely Multimodal**

Of the 40 digital medicine researchers in the survey, they regularly work with a total of 153 sensors, indicating that many researchers are using a number of sensors in their research. These sensors are largely used to monitor the cardiovascular system (electrocardiogram, photoplethysmogram, blood pressure) and participant activity (accelerometry, gyroscope, gait map). Literature review of PubMed-indexed studies showed similar results: researchers are primarily employing sensors measuring physical activity and the cardiovascular system in their work. Many are also using temperature, electrodermal activity, or electroencephalogram sensors in their research. Overall, smartphones continue to be the dominant source of digital health information collection in the digital medicine community.

**There Is Not a Consensus on File Types in Digital Medicine**

In order to inform teams working to establish common data schemas and file types, we examined file types that are commonly used in digital medicine. There are 15 unique file types used by digital medicine researchers, either as raw files or as file types mapped to for analysis. While the .csv file type is the most commonly used, there are other commonly used file types, including JSON and .xlsx/.xlst. There was more diversity among raw file formats (n=114) than file types researchers map data to for analysis (n=79), and respondents are averaging 2.75 raw file types versus a mean of 2 file types for analysis, indicating that while researchers may receive raw data in a number of file formats, they are mapping them to a smaller subset of file types for analysis. We show that there is not a general consensus on file types in digital medicine and that data is currently handled in multiple ways (both as raw files and as files mapped for analysis). Noteworthy is the low number of proprietary file types, indicating that the digital medicine community is largely using accessible file types that could be mapped to a standard, interoperable format.

**Interoperability Remains a Critical Challenge in Digital Medicine**

Nearly all digital medicine professionals surveyed (52/56, 93%) agree that interoperability is a problem in digital medicine. Literature points to this lack of interoperability being a critical barrier to using digital medicine in clinical practice, causing fragmented, inconvenient, and sometimes impossible clinical adoption of digital medicine [9-11]. While many specific challenges to digital medicine interoperability were revealed in this study, the most cited challenges include integration with electronic health records, lack of standard data schemas, and a lack of standard and verifiable methods for digital medicine research. These three areas should be the focus of future directions in developing standards, frameworks, and best practices for interoperability in digital medicine.

Despite agreeing that interoperability is a problem facing the digital medicine community, only 36% (20) of respondents currently use existing tools for interoperability. Generally employed tools used in the community include HL7 FHIR, Apple Health, unspecified APIs, and in-house solutions. A large proportion of those declaring that they use interoperability tools use in-house solutions, which addresses a key problem in the current field of digital medicine—siloed solutions that are not generalizable and are adopted by only a small number of digital medicine professionals.

One of the critical interoperability needs identified is the development of standard data schemas. When asked whether they would use a platform mapping raw data files to a standard format, respondents identified considerations that would have to be made to use this platform: they would need to understand compatibility with electronic health record systems and whether this platform would save time. The most cited consideration for using a platform mapping raw data files is the standard format that data would be mapped to. We identified that there is a strong preference for .csv and .xlsx filetypes for data analysis among respondents. When developing standard data schemas
for the field of digital medicine, it is important to consider the most commonly used file formats and how standard formats could map from and between these popular file formats. Open mHealth, currently the leading mobile health data interoperability standard, maps data to a common JSON data structure [15].

Other critical interoperability needs included standard and verifiable methods for digital medicine research, including preprocessing and postprocessing, algorithms, models, and analyses. When asked if they would use a platform for standardizing and validating digital medicine algorithms, methodologies, and analyses, respondents identified considerations for using this platform, which include understanding data security, potential risks, and extensibility. MD2K Cerebral Cortex provides a complete software platform that allows for data collection and analysis with interactive web dashboards [16]. Recently, we identified a need for an open-source, crowdsourced software platform where digital medicine researchers could share and compare methods, algorithms, and processing methods. To address this need, we have developed the Digital Biomarker Discovery Pipeline (DBDP) [3]. Future directions for the digital medicine community include addressing and implementing solutions to the critical interoperability needs in digital medicine: integration with electronic health records, standard data schemas, and standard and verifiable methods for digital medicine research.

This study was limited in the small sample size and the short time frame of the survey; thus, extending this work will be necessary to further understanding of the challenges facing the digital medicine community as research in mHealth and wearables expands.

In conclusion, understanding the digital medicine community, the sensors and file types commonly used, and perspectives on interoperability will enable the development and implementation of solutions to the critical interoperability needs in digital medicine. As the digital medicine community builds tools, platforms, and resources for mobile health and wearable sensor data, this study can be leveraged to meet real needs and address existing technology gaps. The challenges to interoperability outlined by this study will drive the next steps in creating an interoperable digital medicine community. Establishing best practices to address these challenges and employing platforms for digital medicine interoperability will be essential to furthering the field of digital medicine.

Acknowledgments

All authors are funded through the Chan-Zuckerberg Donor-Advised Fund, an advised fund of the Silicon Valley Community Foundation (grant number 2020-218599). BB is funded through a Duke Forge predoctoral fellowship. JPD is funded through the Whitehead Scholars program. IS funded by a National Institutes of Health/National Institute of Biomedical Imaging and Bioengineering grant (#R24EB025845).

The authors would like to acknowledge Jennifer Goldsack and the Digital Medicine (DiMe) Society for their support and distribution of this survey. They would like to acknowledge the Banff International Research Station Workshop on Implantable and Wearable Sensors 2020 for providing an environment where the idea for this survey was conceived.

Authors’ Contributions

BB was involved in concept development, survey development and deployment, analysis, manuscript preparation, and figure development. JPD was involved in concept development, survey deployment, and manuscript preparation. IS was involved with manuscript editing.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Web-based survey.
[PDF File (Adobe PDF File), 466 KB - mhealth_v9i2e24570_app1.pdf]

Multimedia Appendix 2

Recruitment blurb for distribution of survey via email, Slack, and social media.
[PDF File (Adobe PDF File), 75 KB - mhealth_v9i2e24570_app2.pdf]

Multimedia Appendix 3

PubMed literature review was conducted on July 14, 2020, with the following keywords. Results were limited to the time span 2010-2020 to account for newer technologies.
[DOCX File , 14 KB - mhealth_v9i2e24570_app3.docx]

References


15. mHealth Data Interoperability | Open mHealth. 2020. URL: https://www.openmhealth.org/ [accessed 2020-12-17]

16. MD2K GitHub.: GitHub; 2020. URL: https://github.com/MD2Korg/ [accessed 2020-12-17]

Abbreviations

API: application programming interface
.csv: comma-separated values
DiMe: Digital Medicine professional society
FHIR: Fast Healthcare Interoperability Resources
HL7: Health Level Seven
JSON: JavaScript Object Notation
mHealth: mobile health
R&D: research and development
.xls/.xlsx: Microsoft XML spreadsheets

Edited by L Buis; submitted 24.09.20; peer-reviewed by M Holko, F Abujarad; comments to author 03.11.20; revised version received 17.11.20; accepted 17.12.20; published 03.02.21.

Please cite as:
Bent B, Sim I, Dunn JP
Digital Medicine Community Perspectives and Challenges: Survey Study
JMIR Mhealth Uhealth 2021;9(2):e24570
URL: http://mhealth.jmir.org/2021/2/e24570/
doi:10.2196/24570
PMID:33533721

http://mhealth.jmir.org/2021/2/e24570/
WeChat as a Platform for Baduanjin Intervention in Patients With Stable Chronic Obstructive Pulmonary Disease in China: Retrospective Randomized Controlled Trial

Junjie Bi*, MM; Wei Yang*, MB; Ping Hao*, MB; Yongmei Zhao, MM; Dan Wei*, MB; Yipeng Sun, PhD; Yuhua Lin*, MM; Meng Sun*, MM; Xuan Chen*, MD; Xuming Luo*, MD, PhD; Wei Zhang*, MM; Xiongbiao Wang*, MD, PhD

1Department of Respiratory Medicine, Putuo Hospital, Shanghai University of Traditional Chinese Medicine, Shanghai, China
2Department of Nursing, Putuo Hospital, Shanghai University Of Traditional Chinese Medicine, Shanghai, China
3Department of Respiratory Medicine, Shuguang Hospital, Shanghai University of Traditional Chinese Medicine, Shanghai, China
4Department of Pulmonary Medicine, Zhongshan Hospital, Fudan University, Shanghai, China

* these authors contributed equally

Corresponding Author:
Xiongbiao Wang, MD, PhD
Department of Respiratory Medicine
Putuo Hospital
Shanghai University of Traditional Chinese Medicine
No.164, LanXi Road
Shanghai, 200062
China
Phone: 86 13818666735
Email: xiongbiao6@yahoo.com

Abstract

Background: Pulmonary rehabilitation is a crucial part of the nonpharmacological treatment of stable chronic obstructive pulmonary disease (COPD), but management remains problematic. WeChat could serve as a useful tool in patient management. Baduanjin is a popular exercise in China that is usually applied in pulmonary rehabilitation, which has been confirmed to be effective in improving lung function and life quality.

Objective: This study aimed to explore the efficiency of WeChat in the management of Baduanjin exercise in COPD patients.

Methods: A total of 200 patients from the respiratory department of Putuo Hospital participated in the Baduanjin rehabilitation project from September 2018 to October 2019, and were randomly assigned to the WeChat and control groups and followed up using the WeChat platform or telephone for 12 weeks. The frequency of Baduanjin exercise, lung function (percentage of forced expiratory volume in 1 second predicted, FEV1% predicted), and COPD assessment test (CAT) scores were collected and compared between the two groups. The number of message exchanges and a satisfaction survey on the WeChat platform were used to assess the feasibility of WeChat management outside the hospital.

Results: The Baduanjin exercise frequency significantly differed between the control group and WeChat group (F=33.82, P<.001) and across various time points (F=214.87, P<.001). After the follow-up on WeChat, there were fewer patients not performing Baduanjin exercise. The FEV1% predicted value significantly differed before and after Baduanjin exercise in the control group (Z=-3.686, P<.001) and the WeChat group (Z=-6.985, P<.001). A significant difference in the FEV1% predicted value was observed after Baduanjin exercise between the two groups (Z=-3.679, P<.001). The CAT score significantly differed before and after Baduanjin exercise in the control group (Z=-4.937, P<.001) and the WeChat group (Z=-5.246, P<.001). A significant difference in the CAT score was observed after Baduanjin exercise between the two groups (Z=-5.246, P<.001). The number of completed Baduanjin exercises, lung function, and CAT scores in active patients were higher than those in nonactive patients. All satisfaction survey items were scored with more than 4 points. Among the items, the highest score (mean 4.54, SD 0.77) was for continued WeChat management, followed by the effective management of Baduanjin exercise (mean 4.46, SD 0.87). The patients in the WeChat group showed much higher enthusiasm for and compliance with Baduanjin exercise, resulting in improved lung function and quality of life.

Conclusion: WeChat could be a promising tool in the management of Baduanjin exercise in COPD patients.
in better life quality and lung function. The patients were very satisfied with the WeChat management because of the obvious curative effect and home feeling.

**Conclusions:** The WeChat platform provided a feasible, effective, and sustainable management plan for Baduanjin rehabilitation.

**Trial Registration:** Chinese Clinical Trial Registry ChiCTR1900028248; http://www.chictr.org.cn/showprojen.aspx?proj=46995

**KEYWORDS**
WeChat management; chronic obstructive pulmonary disease; Baduanjin rehabilitation

**Introduction**
Chronic obstructive pulmonary disease (COPD) is a chronic disease characterized by persistent respiratory symptoms and airflow limitation. COPD develops progressively and is related to an abnormal inflammatory response to harmful gases or particles [1]. COPD, ischemic heart disease, cerebrovascular disease, and cancer are the four major causes of human death. The prevention and treatment of COPD have introduced substantial economic burdens for families and countries [2]. The currently used drugs cannot prevent the progression of COPD. The incidence is approximately 13.6% among persons older than 40 years in China [3], and patients usually have a low quality of life with a high disability rate. As a chronic disease, COPD is prone to acute exacerbation [4]. The goals of COPD treatment are to reduce the symptoms, reduce the frequency and severity of exacerbations, and improve the health status and exercise tolerance [5]. Except for hospitalization during acute exacerbation, most patients need to perform self-care at home for many years. Therefore, the management of stable patients is especially important [6]. However, most COPD patients lack disease-related knowledge. Although medical staff members provide disease health knowledge education using different modes, such as education during hospitalization, owing to differences in patient awareness of the disease, education level, compliance, and other factors, many patients cannot effectively manage themselves [7]. It is difficult to meet the needs of the increasing morbidity of COPD and satisfy the increasing expectation of a quality life with the existing medical and health service model in China. Therefore, it is necessary to explore a new management model for stable COPD patients [8].

Pulmonary rehabilitation for COPD has been given more attention in recent years. Pulmonary rehabilitation has been demonstrated to improve the symptoms of dyspnea, improve exercise endurance, reduce the number of in-hospital days, and reduce the frequency of acute exacerbation [9]. Baduanjin, as a traditional aerobics exercise, is listed as the 97th sports item officially launched by the State General Administration of Sport in 2003 and has been widely promoted in China. Baduanjin involves the following eight components: (1) pushing up the heavens, (2) drawing the bow, (3) separating heaven and earth, (4) shaking the heavenly pillar, (5) punching with an angry gaze, (6) shaking the head and swaying the buttocks to extinguish fire in the heart, (7) touching the toes and bending back, and (8) bouncing on the toes. Baduanjin has been applied in lung rehabilitation for many years in China. Previous studies have confirmed that Baduanjin has a strong clinical effect on lung rehabilitation in COPD patients [10]. Thus, Baduanjin has become an exercise prescription for COPD [11]. However, owing to the lack of effective management, pulmonary rehabilitation, including Baduanjin, has not been practiced widely and continuously, and the benefit for patients is limited. In addition, there is a lack of consistency in the length of the research period among many published studies [12].

With the key position of home-based physical activity, long-term management becomes critical. With the rapid development of computer technology, network technology, and multimedia technology, new media have become convenient, fast, interactive, and wide-spread platforms for information dissemination. WeChat is the mainstream instant communication platform in China [13]. The “2018 WeChat Annual Data Report” reported that 1.01 billion users logged on to WeChat daily in 2018. The function of the platform is quite strong but is very easy to operate. Various message forms, such as text, picture, voice, and video, could be presented in communication. WeChat has rapidly developed into a comprehensive information platform integrating communication, information, entertainment, search, e-commerce, office collaboration, and corporate customer service. It is reasonable to apply new technology in medical areas. The “Internet Medical” model is very suitable for China’s current situation, that is, a large population with a shortage of medical resources [14]. WeChat has been gradually used in medical education and the follow-up of patients, and it has produced successful outcomes [15,16].

Because the benefits disappear over time if activity and other good behaviors are not continued, COPD patients should be offered consistent guidance. The telephone has been used for follow-up in early studies but can no longer meet the requirements [17]. New technologies, such as WeChat, could be applied for more effective management. Thus, we established a new system for continuous management from inside the hospital to outside the hospital under the WeChat platform. In the hospital, patients are provided disease knowledge and educated regarding the skills of pulmonary rehabilitation, such as Baduanjin, diet nutrition, correct drug use, etc [18,19]. After discharge, all patients are placed under management using the WeChat platform.

As Baduanjin requires long-term persistence by COPD patients, consistent encouragement from medical staff is essential. In this study, the established WeChat platform was applied to manage Baduanjin exercise in patients with stable COPD, and the results are encouraging.
Methods

Research Design and Flow Chart
This study was a parallel controlled study conducted from September 2018 to October 2019 in Shanghai, China. Two hundred stable COPD patients were included, and the participants were randomly divided into the following two groups (1:1 ratio): the WeChat intervention group and routine nursing control group (Figure 1). The research plan was approved by the ethics committee of Putuo Hospital (grant number: PTEC-A-2018-25-1).

Figure 1. Flowchart of the research design. CAT: chronic obstructive pulmonary disease assessment test; COPD: chronic obstructive pulmonary disease.

Procedures of Baduanjin Exercise
The rehabilitation therapist instructed the patients to perform Baduanjin exercise until they could accomplish proficiency according to the Baduanjin video (produced by the State Sports General Administration in 2003). All patients were asked to perform the exercise two to four times a day for no less than 5 days per week.

Establishment of the WeChat Platform
A WeChat platform team with six members was established. The head nurse of the ward served as the team leader and was responsible for the operation and guidance of the entire project. Two nurse supervisors were responsible for teaching the patients and their families how to use the platform, providing all educational documents to the patients at the appropriate time, maintaining contact with the patients, and urging the patients to perform Baduanjin exercise. Two physicians were responsible for evaluating the patients' condition and devising the medical plan. One technician in the pulmonary function room was responsible for evaluating the lung function of all patients, collecting the data, and performing statistical analyses of the data. The team members cooperated very well, and the platform was operated effectively. The patients very actively participated in the communication. Responses to all questions were obtained from the medical staff in a timely manner.

Participants
All patients with stable COPD were discharged from the respiratory medicine ward of a large general hospital. All candidates first completed a brief screening questionnaire. Patients who met the inclusion criteria were invited to participate in the study and received more detailed information regarding the study. After providing written informed consent, all qualified patients were divided into the WeChat group and control group.

The inclusion criteria were as follows: (1) 50 to 80 years of age regardless of gender; (2) confirmed clinical diagnosis of stable COPD according to the standard of GOLD 2018 [20]; (3) informed consent (patients and families); (4) ability to use WeChat proficiently (patients and primary caregivers); and (5) ability to perform Baduanjin independently.

The exclusion criteria were as follows: (1) severe heart, liver, and kidney diseases, tumors, or other conditions that may affect the observation; (2) life expectancy less than 1 year; and (3) history of conducting physical exercise for a long time (≥3 times/week, ≥20 minutes/time, persisting for more than 12 months) [21].

Pulmonary function (percentage forced expiratory volume in 1 second [FEV1%]) was the outcome index. Based on previous studies showing FEV1% improvement in clinical trials [22], the mean and SD of FEV1% in the control group were 57.09 and 22.53, respectively, while the mean and SD of FEV1% in
the experimental group were 60.24 and 20.15, respectively. The alpha level was set to .05, the power was 80%, and the participant dropout rate was 20%. A sample size of 192 patients (96 per group) was required for the primary analysis. Thus, we recruited 200 patients in the trial (100 in the WeChat group and 100 in the control group).

Randomization and Masking
Using the method of block randomization, 200 research patients were randomly assigned to the WeChat and control groups. The block size was defined as four, and there were six sequential arrangements and combinations. Excel (Microsoft Corp) randomly generated 50 (1/50) numbers that did not repeat. Then, the numbers were divided by six to obtain the remainder, and six combinations were matched according to the remainder. A random block group table was then completed. After the research patients were included, the random block table was assigned to a group. By the nature of the trial design, neither the research staff nor the participants were blinded to the intervention.

Intervention Method

Control Group
Handbooks were distributed to the participants, and they included the following: (1) a Baduanjin video (produced by the State Sports General Administration in 2003); (2) Baduanjin notes; (3) information regarding COPD disease; (4) the importance of pulmonary rehabilitation; (5) nutrition and diet suggestions; and (6) prevention of acute exacerbation of COPD. The patients were asked to comply with medical advice, including performing Baduanjin exercise at home and being followed up by telephone once a week. The chronic obstructive pulmonary disease assessment test (CAT) questionnaire and lung function evaluation were performed before and after the study using a spirometer.

WeChat Group
The participants also received the necessary education on WeChat. All education materials, including text, pictures, videos, and voice messages, were sent to the patients as an electric document via WeChat and individuals as needed. The patients were supervised while performing Baduanjin and maintaining healthy behaviors. A WeChat voice conference was held every 4 weeks for 40 minutes. The medical staff were required to be online, and they held a seminar regarding common problems that the patients usually discussed and answered questions. The feedback information was as follows: (1) dyspnea status; (2) discomfort status; (3) appetite status; (4) number of Baduanjin exercises completed every day (the patients in the WeChat group submitted their information online every day); (5) completion of the satisfaction survey by all patients at the end of the 12th week; and (6) the CAT questionnaire and lung function evaluation before and after the study.

Quality Control
A designated attending physician supervised the project every month and evaluated the progress.

Outcome Measures

Baseline Assessment
The general demographic data included the patients’ gender, age, smoking status, disease course, lung function, classification of airflow limitation severity, comorbidities, and combined COPD assessment according to GOLD 2018. Data Collection and Evaluation at the Endpoint
The Baduanjin exercise frequency was collected each week in WeChat and once a week by telephone. At the end of the study, the lung function evaluation, quality of life evaluation (CAT), personal activity evaluation, and satisfaction survey were completed by all participants.

Lung Function and CAT
All patients underwent a spirometric analysis before and after the trial. Given the noninvasive and simple characteristics of lung function tests, they are generally adopted in clinical practice for early diagnosis and prognostic evaluations. Lung function tests can effectively reflect the pathological changes in the airway and their degree, and provide reliable results of the presence and severity of spirometric abnormalities. The CAT score is a reliable and effective indicator used to assess the living conditions and quality of life of COPD patients. The CAT questionnaire is very simple, can be completed by patients in 2 to 3 minutes, and includes eight items, including daily living ability and physical health. Compared with the St. George’s Respiratory Questionnaire (SGRQ), the content is greatly reduced, but the required indicators are comprehensive and can more intuitively reflect changes in patient health. Related studies have shown that the lung function of patients with COPD is correlated with the CAT score [23] and that the CAT score has good repeatability [24]. Thus, the CAT score is a good evaluation of stable COPD.

Activation of the Platform
We recorded the number of messages exchanged on the WeChat platform. The communication included text/graphics, voice, pictures, videos, etc. The activity on the platform and the degree of concern regarding the disease were evaluated.

Assessment of Satisfaction With the WeChat Platform
A self-designed questionnaire was used to evaluate the satisfaction of the participants at the end of the trial. In total, the following eight items were assessed: effective management of Baduanjin exercise, practical information, convenient communication and interaction, smooth platform operation, simple operation process, continued WeChat management, effective COPD rehabilitation management, and service attitude of the medical staff. A 5-point Likert scale was used with scores ranging from 1 to 5 points (1, strongly disagree; 2, disagree; 3, uncertain; 4, agree; and 5, strongly agree) [25]. A higher score was associated with better COPD patient experience and service evaluation of the WeChat platform and better effectiveness of the WeChat management of Baduanjin exercise.
**Statistical Analysis**

The baseline data were based on continuous measurement data and categorized count data. The measurement data are expressed as mean (SD) or median (IQR). The count data are characterized as frequencies. Two-tailed $t$ tests (mean and SD) or chi-squared tests were used to analyze the normally distributed data. A nonparametric test was used to analyze the nonnormally distributed data (median and IQR).

To analyze the results, repeated measures analysis of variance, a simple effect model, and two-sample $t$ tests were used. To determine the within-group differences and between-group differences, the 95% CIs were calculated for the continuous measures. The statistical analysis was performed using SPSS 22.0 software (IBM Corp), and $P<.05$ was considered statistically significant.

**Results**

**Sample Characteristics**

Based on the inclusion and exclusion criteria, we screened 200 patients and completed baseline assessments before randomization to the trial of 100 patients in each group. Of the 200 patients, 67 were female and 133 were male. The ratio of male and female patients assigned to the two groups was not different. The mean age was 68.16 years (WeChat group, 69.43 years; control group, 68.87 years), and the mean course of COPD was 4.46 years (WeChat group, 4.50 years; control group, 4.31 years). Most patients had severe COPD, accounting for 73% (146/200) of cases (WeChat group, 72% [72/100]; control group, 74% [74/100]). Patients with moderate and very severe COPD accounted for 11.5% (23/200) of cases (WeChat group, 12% [12/100]; control group, 11% [11/100]) and 15.5% (31/200) of cases (WeChat group, 16% [16/100]; control group, 15% [15/100]), respectively. We also assessed the number of acute exacerbations and current smokers in both groups. Patients who had acute exacerbation at least once in the past year accounted for 31% (31/100) of cases in the control group and 35% (35/100) of cases in the WeChat group. The proportion of current smokers in the control group (34% [34/100]) was comparable to that in the WeChat group (37% [37/100]). According to statistical analysis, the baseline characteristics were balanced in both groups (Table 1).

**Table 1.** Baseline characteristics of the participants in the WeChat and control groups.

<table>
<thead>
<tr>
<th>Baseline characteristics</th>
<th>Control group (n=100)</th>
<th>WeChat group (n=100)</th>
<th>$P$ value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender (female/male), n</td>
<td>32/68</td>
<td>35/65</td>
<td>.76</td>
</tr>
<tr>
<td>Age (years), mean (SD)</td>
<td>68.87 (5.33)</td>
<td>69.43 (7.38)</td>
<td>.54</td>
</tr>
<tr>
<td>Course (years), mean (SD)</td>
<td>4.31 (2.07)</td>
<td>4.50 (2.46)</td>
<td>.56</td>
</tr>
<tr>
<td>Acute exacerbation*, n (%)</td>
<td>31 (31%)</td>
<td>35 (35%)</td>
<td>.65</td>
</tr>
<tr>
<td>Current smoker, n (%)</td>
<td>34 (34%)</td>
<td>37 (37%)</td>
<td>.77</td>
</tr>
<tr>
<td><strong>Severity, n</strong></td>
<td></td>
<td></td>
<td>.95</td>
</tr>
<tr>
<td>Moderate</td>
<td>11</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>Severe</td>
<td>74</td>
<td>72</td>
<td></td>
</tr>
<tr>
<td>Very severe</td>
<td>15</td>
<td>16</td>
<td></td>
</tr>
</tbody>
</table>

*Acute exacerbation: at least once in the past year.

Further analysis by a simple effects model showed a difference within and between the groups. The Baduanjin exercise frequencies in the control and WeChat groups (1 week to 4 weeks) increased from week 1 to week 4 and then maintained a plateau until the end of the project. In the within-group comparisons, there were statistically significant differences at each time point in weeks 1 to 3 compared with week 4, but no difference was observed between weeks 5 and 12 and week 4. According to the between-group comparison results, significant differences were found between the WeChat and control groups at the same time points (all $P<.001$). We found that WeChat management was more effective for performing Baduanjin exercise (Multimedia Appendix 1, Figure 2).
Lung Function Assessment

The FEV1% predicted value significantly differed before and after the trial in both the control group ($Z=-3.686, P<.001$) and WeChat group ($Z=-6.985, P<.001$), and the FEV1% predicted value in both groups improved after Baduanjin exercise. A significant difference was found in the FEV1% predicted values after Baduanjin exercise between the two groups ($Z=-3.679, P<.001$), and no difference was observed before the trial (Multimedia Appendix 2, Figure 3). The results demonstrate that Baduanjin exercise was more effective in the WeChat group.

CAT Score Evaluation

All patients completed the questionnaire independently. We found that the CAT score significantly differed before and after the exercise in the control group ($Z=-4.937, P<.001$) and the WeChat group ($Z=-5.246, P<.001$). The life quality of the patients in both groups improved after Baduanjin exercise (Figure 4). No statistically significant difference was found in the CAT score between the two groups before the study ($Z=-1.407, P=.30$), and a significant difference was found after the exercise ($Z=-5.246, P<.001$). These results demonstrate greater improvement in the life quality of patients in the WeChat group after the exercise (Multimedia Appendix 3, Figure 4).
Activity and Interaction on the WeChat Platform
Communication in the WeChat group was mainly conducted via text/graphics, with 13,911 items, followed by voice messages, with a total of 1317 items. Text/graphics and voice accounted for the main methods of message exchange. Pictures and videos were also used; however, the frequency was relatively low (Table 2).

Table 2. Distribution of different message forms present on the WeChat platform.

<table>
<thead>
<tr>
<th>Message form</th>
<th>1-4 weeks, n</th>
<th>5-8 weeks, n</th>
<th>9-12 weeks, n</th>
<th>Total (N=16,100), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Text/graphics</td>
<td>4680</td>
<td>4601</td>
<td>4630</td>
<td>13,911 (86.40%)</td>
</tr>
<tr>
<td>Voice</td>
<td>388</td>
<td>428</td>
<td>501</td>
<td>1317 (8.18%)</td>
</tr>
<tr>
<td>Picture</td>
<td>173</td>
<td>200</td>
<td>289</td>
<td>662 (4.11%)</td>
</tr>
<tr>
<td>Video</td>
<td>52</td>
<td>54</td>
<td>77</td>
<td>183 (1.14%)</td>
</tr>
<tr>
<td>Other files</td>
<td>10</td>
<td>7</td>
<td>10</td>
<td>27 (0.17%)</td>
</tr>
</tbody>
</table>

All messages could be divided into the following three groups: information provided by the medical team, medical staff-patient interaction communication, and patient-patient interaction communication. The frequency of the provided information was relatively fixed. Medical staff-patient interaction was relatively frequent during the first 2 weeks but was less frequent thereafter. However, patient-patient interaction gradually increased, indicating that a COPD family was established. Interestingly, we found that the patients interacted and encouraged each other. The platform became a common home for all WeChat participants.

We divided the WeChat group into active patients and nonactive patients based on the number of messages exchanged. Those with more than two messages per day were considered active patients, and those with one or fewer messages were considered inactive patients. Eighty-six patients were more active in the group. The number of completed Baduanjin exercises, lung function, and CAT scores of these active patients were higher than those of the nonactive patients (Figure 5). The amount of information exchanged by the WeChat group reflected the activity of the WeChat management platform, indicating that COPD patients actively participated in the management of the disease and received better outcomes.
Figure 5. Comparison of outcomes between active patients and nonactive patients. (A) Daily exercise frequency; (B) Difference in lung function before and after Baduanjin exercise; and (C) Difference in the chronic obstructive pulmonary disease assessment test (CAT) score before and after Baduanjin exercise. FEV1% pred: percentage of forced expiratory volume in 1 second predicted. **P<0.001.

Evaluation of the Satisfaction With the WeChat Platform

According to the score of the 5-point Likert scale, items that received more than 4 points were as follows: continued WeChat management, effective Baduanjin exercise management, practical information content, convenient communication and interaction, smooth platform operation, simple operation process, service attitude of the medical staff, and COPD rehabilitation. The results indicated that the COPD patients were satisfied with the management of the WeChat platform. Among these items, the highest score was for continued WeChat management (mean 4.54, SD 0.77), followed by effective management of Baduanjin exercise (mean 4.46, SD 0.87), indicating the feasibility, sustainability, and effectiveness of disease management with the WeChat platform (Table 3).

Table 3. Satisfaction scores of the WeChat platform.

<table>
<thead>
<tr>
<th>Contents</th>
<th>Agreement (N=50)</th>
<th>Score, mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Strongly agree, n (%)</td>
<td>Agree, n (%)</td>
</tr>
<tr>
<td>Continued WeChat management</td>
<td>28 (56%)</td>
<td>21 (42%)</td>
</tr>
<tr>
<td>Effective Baduanjin exercise management</td>
<td>26 (52%)</td>
<td>21 (42%)</td>
</tr>
<tr>
<td>Practical information content</td>
<td>22 (44%)</td>
<td>28 (56%)</td>
</tr>
<tr>
<td>Convenient communication and interaction</td>
<td>24 (48%)</td>
<td>23 (46%)</td>
</tr>
<tr>
<td>Smooth platform operation</td>
<td>21 (42%)</td>
<td>28 (56%)</td>
</tr>
<tr>
<td>Simple operation process</td>
<td>21 (42%)</td>
<td>28 (56%)</td>
</tr>
<tr>
<td>Service attitude of the medical staff</td>
<td>24 (48%)</td>
<td>20 (40%)</td>
</tr>
<tr>
<td>COPD rehabilitation</td>
<td>22 (44%)</td>
<td>22 (44%)</td>
</tr>
</tbody>
</table>

aStrongly agree, 5 points; agree, 4 points; uncertain, 3 points; disagree, 2 points; strongly disagree, 1 point.
bCOPD: chronic obstructive pulmonary disease.

Discussion

Principal Findings

The main purpose of this study was to test the feasibility and efficiency of WeChat in the management of Baduanjin exercise in COPD patients. The results of this study suggest that the management of Baduanjin exercise with the WeChat platform improves the enthusiasm and compliance of patients. The frequency of Baduanjin exercise using the WeChat-based intervention was 0 to 21 times (mean 18.84) to 0 to 13 times (mean 10.43) per person per week compared with traditional interventions at 12 weeks. The Baduanjin exercise frequency in the WeChat group was markedly higher than that in the control group. Under the management of WeChat, some patients did not perform Baduanjin exercise. However, after the follow-up on WeChat, the number of patients with a Baduanjin exercise frequency of 0 decreased (Figure 6). Therefore, Baduanjin exercise management based on the WeChat platform improved the enthusiasm and compliance of patients. These results are consistent with the beneficial effect of compliance using WeChat management in other clinical populations [26].
Figure 6. Number of patients with a Baduanjin exercise frequency of 0 (per week) in the WeChat group and control group.

Based on the WeChat platform, communication between medical staff and patients increased, and the effectiveness of Baduanjin lung rehabilitation improved. Studies have shown that the frequency of the use of the platform is crucial for the effectiveness of the intervention [27]. Our results indicate that the total number of messages exchanged gradually increased and reached a peak at approximately 4 weeks; thereafter, the number maintained a plateau. Interaction with text/graphics exhibited the most obvious increase (Figure 7). We also found that active participants had better outcomes. The interaction between patients increased, and mutual exchange, sharing, and supervision promoted more enthusiasm among the patients. The role of the team was always favorable.

Figure 7. Trend of messages exchanged on WeChat.

All participants completed the satisfaction survey, and a high score was obtained in the WeChat group. Based on the questionnaire responses, we found that continuing WeChat management and Baduanjin exercise management were the top two ranked items. The other items also achieved relatively high scores. Thus, the patients expected to continue WeChat management, and the WeChat management of Baduanjin exercise is feasible.

Early hospital follow-up, coaching, COPD action plans and management programs, telemedicine, pulmonary rehabilitation, and other interventions require more research. Improvement in management is an effective strategy for improving patient recovery [28]. A recent study also noted that reducing readmission could improve prognosis [29]. Although the management of COPD patients has been studied since the 1960s [30], the situation of management is not optimistic. Many problems are associated with COPD management, and a key problem is lung rehabilitation [31]. Currently, published data lack sufficient evidence for the identification of an effective pathway to improve the symptoms of COPD by management. Therefore, better ways to achieve management must be explored [32].

With the development of the internet, major communication platforms have begun to be applied in health care. Social media platforms, such as Twitter and Facebook, have been steadily applied in medical education [33,34]. Facebook has been used to study many aspects of health care, and it provides a convenient and easily accessible way to collect unsolicited and
observational patient data [35]. An analysis of the content of COPD group chats on Facebook showed that COPD Facebook group members share specific disease-related experiences and request information regarding select self-management topics. This information can be used to improve the quality of self-management support provided to members of popular COPD Facebook groups [36]. WeChat is also gradually being used in medical teaching and the clinic, but the clinical application needs improvement [37]. Clinical studies based on the WeChat platform, such as cough variant asthma and promoting weight loss, have achieved good results [38]. The internet could also help patients with COPD understand the disease and manage the condition by themselves [39].

Limitations
This study has some limitations. First, the study period was quite short (approximately 12 weeks), which may have affected the judgement of compliance, but our study is ongoing, and more data will be reported in the future. Second, the number of participants in this study was relatively small, although the sample size was calculated statistically. However, because of the complexity of clinical trials and the uncontrollability of patients, a larger number of patients could supply more intact information. Third, as the original data were obtained from patient reports, inaccurate reports could not be fully excluded. Thus, there is a problem of credibility in patient reports using WeChat management. Finally, our study only explored the effectiveness of Baduanjin rehabilitation training under WeChat platform management. More projects in self-management supported by the WeChat platform could be fully studied in the future.

Conclusions
In this study, a new management platform involving WeChat from the in-hospital period to the out-of-hospital period was established for patients. Pulmonary rehabilitation using Baduanjin exercise for COPD patients was applied using this platform. This online rehabilitation training model met the patients’ requirements as this platform was convenient, easy to understand, and time saving. Information is shared in a timely manner, and seamless communication is maintained with patients. Additionally, the compliance and enthusiasm of the patients were much higher. Baduanjin exercise based on WeChat platform management was effective in improving the life quality and lung function of COPD patients. Patient management based on WeChat is effective, feasible, and sustainable.

Acknowledgments
This work was supported by the National Key Research and Development Program of China (number: 2018YFC1313600), the Shanghai Science and Technology Committee (number: 18140904000, 17401970900), Department of Respiratory Medicine Development Fund of Putuo District (2016PTZK03), the Scientific Innovation Foundation of Putuo District (ptkwws201714), Shanghai Municipal Commission of Health and Family Planning (20174Y0239), the Peiying Program of Putuo Hospital (2017206A), and the scientific project of Shanghai University of Traditional Chinese Medicine (2019LK096) and Specialty Construction of Respiratory and Critical Care Medicine (2020tszk02).

Conflicts of Interest
None declared.

Editorial Notice
This randomized study was only retrospectively registered by the authors due to technical issues and the limitation of time. The editor granted an exception from ICMJE rules mandating prospective registration of randomized trials because the risk of bias appears low and the study was considered formative, guiding the development of the application. However, readers are advised to carefully assess the validity of any potential explicit or implicit claims related to primary outcomes or effectiveness, as retrospective registration does not prevent authors from changing their outcome measures retrospectively.

Multimedia Appendix 1
Effects of the WeChat intervention on Baduanjin exercise times.
[DOCX File, 15 KB - mhealth_v9i2e23548_app1.docx ]

Multimedia Appendix 2
Lung function in the WeChat group and control group.
[DOCX File, 15 KB - mhealth_v9i2e23548_app2.docx ]

Multimedia Appendix 3
Chronic obstructive pulmonary disease assessment test score in the WeChat group and control group.
[DOCX File, 15 KB - mhealth_v9i2e23548_app3.docx ]

Multimedia Appendix 4
CONSORT-eHEALTH checklist.


35. Bi et al. JMIR Mhealth and Uhealth 2021;9(2):e23548

Abbreviations

CAT: chronic obstructive pulmonary disease assessment test
COPD: chronic obstructive pulmonary disease
FEV1%: percentage of forced expiratory volume in 1 second
Heart Rate Variability and Firstbeat Method for Detecting Sleep Stages in Healthy Young Adults: Feasibility Study

Liisa Kuula, PhD; Anu-Katriina Pesonen, PhD
SleepWell Research Program, University of Helsinki, Helsinki, Finland

Abstract

Background: Polysomnography (PSG) is considered the only reliable way to distinguish between different sleep stages. Wearable devices provide objective markers of sleep; however, these devices often rely only on accelerometer data, which do not enable reliable sleep stage detection. The alteration between sleep stages correlates with changes in physiological measures such as heart rate variability (HRV). Utilizing HRV measures may thus increase accuracy in wearable algorithms.

Objective: We examined the validity of the Firstbeat sleep analysis method, which is based on HRV and accelerometer measurements. The Firstbeat method was compared against PSG in a sample of healthy adults. Our aim was to evaluate how well Firstbeat distinguishes sleep stages, and which stages are most accurately detected with this method.

Methods: Twenty healthy adults (mean age 24.5 years, SD 3.5, range 20-37 years; 50% women) wore a Firstbeat Bodyguard 2 measurement device and a Geneactiv actigraph, along with taking ambulatory SomnoMedics PSG measurements for two consecutive nights, resulting in 40 nights of sleep comparisons. We compared the measures of sleep onset, wake, combined stage 1 and stage 2 (light sleep), stage 3 (slow wave sleep), and rapid eye movement (REM) sleep between Firstbeat and PSG. We calculated the sensitivity, specificity, and accuracy from the 30-second epoch-by-epoch data.

Results: In detecting wake, Firstbeat yielded good specificity (0.77), and excellent sensitivity (0.95) and accuracy (0.93) against PSG. Light sleep was detected with 0.69 specificity, 0.67 sensitivity, and 0.69 accuracy. Slow wave sleep was detected with 0.91 specificity, 0.72 sensitivity, and 0.87 accuracy. REM sleep was detected with 0.92 specificity, 0.60 sensitivity, and 0.84 accuracy. There were two measures that differed significantly between Firstbeat and PSG: Firstbeat underestimated REM sleep (mean 18 minutes, P=0.03) and overestimated wake time (mean 14 minutes, P<0.001).

Conclusions: This study supports utilizing HRV alongside an accelerometer as a means for distinguishing sleep from wake and for identifying sleep stages. The Firstbeat method was able to detect light sleep and slow wave sleep with no statistically significant difference to PSG. Firstbeat underestimated REM sleep and overestimated wake time. This study suggests that Firstbeat is a feasible method with sufficient validity to measure nocturnal sleep stage variation.

KEYWORDS
electroencephalogram; actigraphy; polysomnography; sleep; heart rate; rapid eye movements

Introduction

Sleep stages alternate throughout the typical nighttime sleep period. After the initial sleep onset, nonrapid eye movement (NREM) sleep stages 1 (N1), 2 (N2), and 3 (N3) emerge alongside rapid eye movement (REM) sleep [1]. Together, both NREM and REM sleep stages form sleep cycles, which, in healthy adults, rotate approximately four or five times over the course of a single night [2]. This alteration between stages correlates with changes in physiological measures such as muscle tone [3,4], blood pressure [5-7], temperature regulation [8,9], as well as heart rate and heart rate variability (HRV) [10,11].
More specifically, NREM sleep stages are related to stability in the cardiovascular system and stronger parasympathetic cardiac modulation. This, in turn, is reflected in REM sleep so that the heart rate increases and becomes less stable [5,7]. Within NREM sleep stages, the differences between deep sleep (N3, or slow wave sleep [SWS]) and lighter sleep stages (N1 and N2) also have some physiological differences, but these are less pronounced than those between REM and NREM [12]. Specifically, the deeper the sleep, the stronger the parasympathetic cardiac modulation (i.e., SWS is associated with a lower heart rate compared to N2 and N1) [13,14].

Polysomnography (PSG) is considered the gold-standard means for measuring sleep stages, as the combination of electromyography (EMG) and electroencephalograph (EEG) is, by definition, the only way to distinguish between the different sleep stages [1]. Although PSG provides reliable data on sleep, other less laborious methods are needed as the increasing prevalence of sleep disorder diagnoses [15] has highlighted an urgent gap to be filled in the development of reliable, cost-efficient sleep study tools for both clinical and consumer use [16].

Some recent studies suggest that HRV might provide a noninvasive marker for detecting sleep behavior such as differentiating between sleep stages [10,11]. HRV has also been widely utilized for assessing phenomena such as stress and recovery [17], physical activity [18], and oxygen consumption [19]. Recently, a sleep analysis method was developed based on HRV and acceleration data (Firstbeat Technologies Oy, Jyväskylä, Finland) for providing personalized feedback and guidance regarding the quantity and quality of sleep. HRV as measured by a single-lead ECG device (Firstbeat Bodyguard 2, Firstbeat Technologies Oy) can estimate atrial fibrillation accurately [20], making it a reliable measurement device regarding HRV-related phenomena.

Based on the need to evaluate the validity of commonly available and easy-to-administer sleep measurement solutions, we investigated how PSG and the Firstbeat sleep analysis algorithm correlate in detecting sleep stages. From analog measurements, we estimated the sensitivity, specificity, and accuracy of the Firstbeat method in relation to PSG measurement over two nights.

**Methods**

**Participants**

The study protocol has been described in detail in a previous publication [21]. We recruited 20 voluntary participants (mean age 24.5 years, SD 3.5, range 20-37 years; 10 [50%] women) by word of mouth in Helsinki, Finland. Participants were recruited from the research team’s circle of acquaintances: if the acquaintance showed initial interest in participating, they received a detailed description of the procedure via email. After reading the description, if the potential participant was still interested in taking part in the sleep study, they were screened for suitability. Their sleep was then measured for two consecutive nights using PSG, chest-worn Firstbeat Bodyguard 2, and a wrist-worn Geneactiv actigraph (Activinsights Ltd, Kimbolton, United Kingdom). The inclusion criteria were as follows: aged between 20 and 45 years, and a relatively stable sleep schedule (e.g., no shift work or jet lag). Exclusion criteria were any diagnosed sleep disorder, the use of any medication that could affect sleep, acute sickness (e.g., the flu), and gold allergy (as the electrodes used for the PSG recording were gold-plated). Each participant received a compensation of 100 euros (US $115) and structured feedback on their sleep stages. Written informed consent was obtained from all participants. The study was approved by the Ethical Committee of the Helsinki University Central Hospital. All procedures followed were in accordance with the Helsinki Declaration and its later amendments.

In our previous study, we investigated two different intervention groups within this setting, and demonstrated that these groups did not differ significantly from each other [21]. The Pittsburgh Sleep Quality Index (PSQI) scores of the participants, ranging from 2 to 12 points, indicate some variation in sleep quality. For the purpose of this study, all nights from all participants were pooled together for comparisons of PSG and Firstbeat sleep metrics.

**Procedure**

A research assistant visited participants at their homes on two consecutive nights. Participants had been asked not to consume alcohol or caffeine after 4 PM on the measurement nights. The evening visit started between 6 and 10 PM, depending on the participant’s current sleeping schedule. The research assistant attached the measurement electrodes to the participant during the house call and began the recording. Before the research assistant left, participants were given instructions for the night: the participants were instructed to spend the evening as usual but to refrain from vigorous activities. They were also asked to keep their phones and any other electrical devices with transmitters at least 2 meters away from the bed so they would not interfere with the PSG recording. Participants were instructed to sleep normally, and the visit for the following morning was scheduled according to the participant’s expected wake-up time. The research assistant arrived the following morning approximately 0 to 30 minutes after the wake-up time.

**Physical Measurements**

We measured HRV with Firstbeat Bodyguard 2, including two chest electrodes and 3-axis acceleration data obtained from the wrist with a Geneactiv actigraph. The Firstbeat sleep analysis method evaluates the physiological state of the person as being awake or asleep based on HRV and acceleration data, and scores sleep as N1+N2 (light), N3, or REM. The method uses a neural network–based algorithm with HRV, HRV-derived respiration rate, movement, and time of day data for sleep and wake detection and for sleep classification. To align the measurement modes, we combined PSG-measured sleep stages N1 and N2 to correspond to “light” sleep of the Firstbeat method.

We used overnight PSG to measure sleep at home (SOMNOScreen plus, SOMNomedics GmbH, Germany) with the following recorded parameters: EEG (left and right for F, C, O), left and right electrooculogram (EOG), left and right EMG, and ECG. The setup for the PSG and the removal in the
morning measurements were carried out by a trained research assistant. EEG measurements were recorded with gold cup electrodes at 6 EEG locations (F3, F4, C3, C4, O1, and O2) and 2 channels for the mastoids (A1, A2) according to the standardized 10/20 system. The ECG, EOG, and EMG were measured using disposable adhesive electrodes (Ambu Cardiology Blue Sensor M; Ambu Neuroline 715, Ambu A/S, Denmark) with two locations for ECG and EOG, and three locations for EMG. In addition, an online reference Cz and a ground electrode in the middle of the forehead were used. The sampling rate was 256 Hz. All signals were filtered with a pass band of 0.5-40 Hz (Hamming windowed sinc zero-phase FIR filter, cutoff [−6 dB] 0.25 Hz and 44.3 Hz, respectively) and rereferenced to the average signal of A1 and A2 electrodes. Sleep stages from PSG data were scored manually with the DOMINO program version 2.7 (SOMNOmedics GmbH, Germany) by two experienced researchers in 30-second epochs. The scoring was completed with both researchers visually inspecting the data together and agreeing over each epoch. The scoring was paused if any disagreement emerged and continued after agreement was found based on careful inspection of all channels, in accordance with the rules published by the American Academy of Sleep Medicine (AASM) [1].

Statistical Analyses

Following standard sleep score practices in the AASM manual [1], we used 30-second epochs for sleep stage comparisons. The entire data were compared side by side after lights off; following AASM scoring rules, sleep onset was defined as the first epoch of any sleep stage as detected by the PSG measurement. We compared how Firstbeat was able to detect the actual sleep onset by calculating the difference between the two time points, which were statistically evaluated using a paired t test. All comparisons of sleep staging between PSG and Firstbeat were performed from the PSG-measured actual sleep onset onward.

First, we used paired-sample t tests to compare sleep metrics for evaluating differences between PSG and Firstbeat in sleep onset, and minutes spent in wake, light sleep, SWS, and REM sleep. Second, we conducted epoch-by-epoch comparisons between Firstbeat and PSG to calculate the sensitivity (ability of Firstbeat to detect true sleep), specificity (ability of Firstbeat to detect true wake), and accuracy (ability of Firstbeat to detect both sleep and wake). This comparison was performed across all sleep stages, as well as for overall sleep-wake comparisons between PSG and Firstbeat.

Third, we used a confusion matrix to compare epoch-by-epoch measures of true positives, true negatives, false positives, and false negatives between Firstbeat and PSG across all measured nights for sleep versus wake as well as for light sleep, SWS, and REM sleep stages. True positives arise when both the PSG and Firstbeat score the 30-second epoch as sleep. True negatives arise when both the PSG and Firstbeat score the epoch as awake. False positives arise when the PSG scores the epoch as sleep but Firstbeat scores it as wake. False negatives arise when the PSG scores the epoch as wake but Firstbeat scores it as sleep.

Fourth, we evaluated the differences between the amount of sleep scored as wake, light sleep, SWS, or REM sleep when comparing Firstbeat and PSG using minute-based Bland-Altman plots, and visually demonstrate how many observations remained within a 30-minute window of the PSG measure.

Finally, we used t tests to compare whether specificity, sensitivity, and accuracy differed based on sex, measurement night, or the intervention we reported previously [21].

Results

The 40 nights from 20 participants measured with both PSG and Firstbeat were included in all analyses with no exclusions. Table 1 shows the participants’ characteristics as well as their mean sleep measures.
Table 1. Characteristics of the sample (N=20).

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years), mean (SD)</td>
<td>24.50(3.50)</td>
</tr>
<tr>
<td>Sex (female), n (%)</td>
<td>10 (50)</td>
</tr>
<tr>
<td>BMI (kg/m²), mean (SD)</td>
<td>23.64(3.10)</td>
</tr>
<tr>
<td>PSQI score, mean (SD)</td>
<td>5.40(2.35)</td>
</tr>
<tr>
<td>Poor sleep quality (PSQI score&gt;5), n (%)</td>
<td>5 (25)</td>
</tr>
</tbody>
</table>

Polysomnography-measured sleep, mean (SD)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sleep onset (hour:minute)</td>
<td>23:44(1:05)</td>
</tr>
<tr>
<td>TSTb (hour:minute)</td>
<td>7:28(0:49)</td>
</tr>
<tr>
<td>Sleep efficiency (%)</td>
<td>91.55(5.82)</td>
</tr>
<tr>
<td>N1c of TST (%)</td>
<td>4.68(2.71)</td>
</tr>
<tr>
<td>N2d of TST (%)</td>
<td>47.46(6.14)</td>
</tr>
<tr>
<td>N3e of TST (%)</td>
<td>22.18(7.11)</td>
</tr>
<tr>
<td>REMf of TST (%)</td>
<td>25.68(5.48)</td>
</tr>
</tbody>
</table>

PSQI: Pittsburgh Sleep Quality Index.
TST: total sleep time.
N1: sleep stage 1 (light sleep).
N2: sleep stage 2 (light sleep).
N3: sleep stage 3 (slow wave sleep).
REM: rapid eye movement.

Table 2 shows paired t test comparisons and the mean differences between Firstbeat and PSG sleep stage scores. Sleep onset did not differ significantly between Firstbeat and PSG (mean difference 0, SD 9 minutes, SE 1 minute; t(39)=0.578, P=.57). Three nights’ sleep onset was detected accurately, whereas for 12 nights, the Firstbeat method assumed earlier sleep onset than PSG. To detect the true difference in detecting sleep onset, we calculated the absolute difference between Firstbeat onset to PSG onset, and found a mean difference of 7.06 minutes (SD 6.64 minutes).

Table 2. Paired comparisons and mean differences of sleep parameters recorded by Firstbeat and polysomnography.

<table>
<thead>
<tr>
<th>Parameter (minutes)</th>
<th>Mean differencea (SD)</th>
<th>SE</th>
<th>95% CI</th>
<th>t(df=39)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wake</td>
<td>14.03 (16.65)</td>
<td>2.63</td>
<td>8.70 to 19.35</td>
<td>5.327</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Light sleep</td>
<td>-0.80 (42.25)</td>
<td>6.68</td>
<td>-14.31 to 12.72</td>
<td>-0.120</td>
<td>.91</td>
</tr>
<tr>
<td>Slow wave sleep</td>
<td>4.68 (46.79)</td>
<td>7.40</td>
<td>-10.29 to 19.64</td>
<td>0.632</td>
<td>.53</td>
</tr>
<tr>
<td>REMf sleep</td>
<td>-17.90 (50.44)</td>
<td>7.98</td>
<td>-34.03 to -1.77</td>
<td>-2.244</td>
<td>.03</td>
</tr>
</tbody>
</table>

Mean differences calculated as Firstbeat – polysomnography.
REM: rapid eye movement.

Table 2 shows paired t test comparisons and the mean differences between Firstbeat and PSG sleep stage scores. Sleep onset did not differ significantly between Firstbeat and PSG (mean difference 0, SD 9 minutes, SE 1 minute; t(39)=0.578, P=.57). Three nights’ sleep onset was detected accurately, whereas for 12 nights, the Firstbeat method assumed earlier sleep onset than PSG. To detect the true difference in detecting sleep onset, we calculated the absolute difference between Firstbeat onset to PSG onset, and found a mean difference of 7.06 minutes (SD 6.64 minutes).

Table 3 shows the confusion matrix [22] regarding the two measurement devices and their differences.
Table 3. Confusion matrix of the Firstbeat method and polysomnography sleep stage epoch comparisons.

<table>
<thead>
<tr>
<th>Firstbeat (N)</th>
<th>Polysomnography (N)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Light</td>
</tr>
<tr>
<td>Light</td>
<td>12,737(^c)</td>
</tr>
<tr>
<td>SWS</td>
<td>2683</td>
</tr>
<tr>
<td>REM</td>
<td>2072</td>
</tr>
<tr>
<td>Wake</td>
<td>1435</td>
</tr>
<tr>
<td>Total</td>
<td>18,927</td>
</tr>
<tr>
<td>Correct stage classification (%)</td>
<td>67.3</td>
</tr>
</tbody>
</table>

\(^a\)SWS: slow wave sleep.  
\(^b\)REM: rapid eye movement.  
\(^c\)Diagonals indicate the number of correctly categorized epochs in the respective sleep stage.  
\(^d\)N/A: not applicable.

Figure 1 shows the Bland-Altman mean difference plots, which illustrate the share of observations that were within 30 minutes from each other in wake state, or in different sleep stages as measured with different devices.

Discussion

As a further sensitivity check, we compared the means of specificity, sensitivity, and accuracy to detect possible differences based on sex, measurement night, or the intervention reported previously [21]. Compared to females, there was a better specificity in REM sleep in male participants (0.95 vs 0.89, \(P=0.004\)), but there were no other differences between sexes \((P>.06)\). There was no first- or second-night effect in the specificity, sensitivity, and accuracy \((P>.38)\), nor regarding the presence of the previously reported music or slow-breathing intervention \((P>.13)\).

Principal Findings

Wearable devices have gained a significant share of the health and well-being consumer market, and new wearable devices and algorithms emerge frequently. Although a great majority of this research aims to detect sleep quality and duration based on data derived from accelerometer sensors [23], other measures such as respiratory signals have also been utilized [24]. Several reviews have evaluated the accuracy of accelerometer-based sleep wearables [23,25,26], and a recent review summarized an
overall evaluation of wearables utilizing other sensors [27]. They concluded that detecting sleep from wake is relatively successful in many devices, but when wearables aim to classify sleep stages as opposed to simply distinguish between sleep and wake, there is a challenge in distinguishing four choices (wake, light, deep, and REM sleep) [27], which makes the result more inaccurate.

Commercial accelerometers typically yield accuracy between 0.81 and 0.91, sensitivity values between 0.87 and 0.99, and specificity values between 0.10 and 0.52 in distinguishing sleep from wake [26]. However, when attempting to detect sleep stages, the results are less consistent. A recent review focusing on commercial accelerometers identifying sleep stages found great variation in accuracy depending on the study [26]. For instance, accuracy in detecting light sleep varied between 69% and 81%, accuracy in detecting SWS was between 36% and 89%, and that for REM sleep ranged between 62% and 89%. Such variation suggests that acceleration itself may not be sufficient in reliably identifying sleep stages.

Previous studies have implied that HRV may be a useful marker for detecting sleep stages [10,11]. One study reported an accuracy of up to 89% in detecting SWS, but their method included respiratory signals alongside HRV [28]. When detecting sleep stages by utilizing both HRV and accelerometer data, one study managed to identify 75% of SWS correctly [29]. In that study, REM sleep was identified correctly in over 70% of epochs, whereas light sleep detection was the weakest with correct identification varying between 42% and 52%. Our findings are of similar accuracy, which further supports the notion of combining accelerometer and HRV-based measures for reproducible sleep staging.

This study was performed to evaluate the ability of HRV- and acceleration-based Firstbeat sleep analysis methods to detect sleep and different sleep stages. In pairwise comparisons, the Firstbeat method detected light sleep and SWS with no statistically significant difference to the gold-standard PSG method. There were two measures that differed significantly between the Firstbeat method and PSG: Firstbeat underestimated REM sleep (mean 18 minutes) and overestimated wake (mean 14 minutes). Considering the number of minutes in the context of a typical night’s sleep, the differences are not alarmingly high in practice, especially when measuring sleep over repeated nights. Sleep onset detection was very accurate, which is in accordance with a review published earlier this year [30]. Sleep stages can only be detected using PSG, as the stages are, by definition, separated by different patterns in ECG, EOG, and ECM. REM sleep is particularly difficult to detect without measuring activity from EOG and EMG channels. Thus, relying on other physiological measures as a means for separating sleep stages is always based on secondhand information. Although HRV has both previously [10,11] and in this study reflected sleep stages relatively well, it cannot detect the immediate changes in EEG, EOG, and EMG. However, this study suggests that HRV-assisted sleep stage detection can serve as a good estimate of sleep architecture despite being less accurate in detecting specific sleep stages.

When observing the comparisons in more analytical detail, we found that comparing the Firstbeat method against PSG yielded good specificity, and excellent sensitivity and accuracy in detecting wake. Regarding light sleep, the measures of specificity, sensitivity, and accuracy were less convincing. SWS detection had excellent specificity, adequate sensitivity, and good accuracy, while REM sleep was detected with similarly excellent specificity, adequate sensitivity, and good accuracy. These results suggest that the Firstbeat method is best at detecting sleep stages that have strong parasympathetic cardiac markers; however, light sleep is typically not significantly differentiated based its physiological fingerprint [12,14].

Strengths and Limitations
Our study was fully balanced in sex distribution and we were able to evaluate the Firstbeat method across two different nights in two different settings in the participants’ own homes. Thus, the ecological validity in this study can be considered excellent. As a limitation, even though our sample had some variation in PSQI-measured sleep quality, this study did not include any participants with diagnosed sleep disorders. Our study included only healthy participants, and the results are likely to be different if any health issues, particularly cardiovascular, or any sleep disorders are present. This is a question to solve before utilizing the Firstbeat method in clinical contexts.

Conclusion
Combining HRV with accelerometer measurements can be considered a feasible method with sufficient validity to measure nocturnal sleep stage variation. We found that the specificity, sensitivity, and accuracy were the weakest in detecting light sleep. Nevertheless, considering its availability, affordability, and ease of administration, Firstbeat may be a useful tool in various contexts, particularly in consumer-based sleep-measuring environments to produce an overview of sleep structures.

Acknowledgments
We thank Kristiina Kajanto and Miina Peltonen for their assistance in data collection.

Conflicts of Interest
None declared.

References

http://mhealth.jmir.org/2021/2/e24704/


Abbreviations

  AASM: American Academy of Sleep Medicine
  EEG: electroencephalography
  EMG: electromyography
  EOG: electrooculogram
  HRV: heart rate variation
  N1: stage 1 (nonrapid eye movement) sleep
  N2: stage 2 (nonrapid eye movement) sleep
  N3: stage 3 (nonrapid eye movement) sleep (slow wave sleep)
  NREM: nonrapid eye movement
  PSG: polysomnography
  PSQI: Pittsburgh Sleep Quality Index
  REM: rapid eye movement
  SWS: slow wave sleep

Edited by L Buis; submitted 01.10.20; peer-reviewed by S Myllymaa, O Chelidoni; comments to author 23.10.20; revised version received 07.12.20; accepted 23.12.20; published 03.02.21.

Please cite as:

Kuula L, Pesonen AK
Heart Rate Variability and Firstbeat Method for Detecting Sleep Stages in Healthy Young Adults: Feasibility Study
JMIR Mhealth Uhealth 2021;9(2):e24704
URL: http://mhealth.jmir.org/2021/2/e24704/
doi:10.2196/24704
PMID:33533726

©Liisa Kuula, Anu-Katriina Pesonen. Originally published in JMIR mHealth and uHealth (http://mhealth.jmir.org), 03.02.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on http://mhealth.jmir.org/, as well as this copyright and license information must be included.