Mobile and tablet apps, ubiquitous and pervasive computing, wearable computing and domotics for health
Volume 9 (2021), Issue 9  ISSN: 2291-5222

Contents

Original Papers

Exploring the Shift in International Trends in Mobile Health Research From 2000 to 2020: Bibliometric Analysis (e31097)
Jianfei Cao, Yeongjoo Lim, Shintaro Sengoku, Xitong Guo, Kota Kodama. .......................................................... 2

Toward a Better Understanding of the Intention to Use mHealth Apps: Exploratory Study (e27021)
Pedro Palos-Sanchez, Jose Saura, Miguel Rios Martin, Mariano Aguayo-Camacho. .................................................. 19

Nursing Interns’ Attitudes Toward, Preferences for, and Use of Diabetes Virtual Simulation Teaching Applications in China: National Web-Based Survey (e29498)
Fang Liu, Huiting Weng, Rong Xu, Xia Li, Zhe Zhang, Kuaile Zhao, Zhiguang Zhou, Qin Wang. ........................................... 38

Upper-Limb Motion Recognition Based on Hybrid Feature Selection: Algorithm Development and Validation (e24402)
Qiaoqin Li, Yongguo Liu, Jiajing Zhu, Zhi Chen, Lang Liu, Shangming Yang, Guanyi Zhu, Bin Zhu, Juan Li, Rongjiang Jin, Jing Tao, Lidian Chen. .......................................................... 51

User Experiences of the NZ COVID Tracer App in New Zealand: Thematic Analysis of Interviews (e26318)
Alexei Tretiakov, Inga Hunter. .......................................................................................................................... 67

Diagnostic Accuracy of Smartphone-Based Audiometry for Hearing Loss Detection: Meta-analysis (e28378)
Chih-Hao Chen, Heng-Yu Lin, Mao-Che Wang, Yuan-Chia Chu, Chun-Yu Chang, Chii-Yuan Huang, Yen-Fu Cheng. ........................................................................................................... 82

Corrigenda and Addenda

Correction: Effect of Physician-Pharmacist Participation in the Management of Ambulatory Cancer Pain Through a Digital Health Platform: Randomized Controlled Trial (e33223)
Lu Zhang, Howard McLeod, Ke-Ke Liu, Wen-Hui Liu, Hong-Xing Huang, Ya-Min Huang, Shu-Sen Sun, Xiao-Ping Chen, Yao Chen, Fang-Zhou Liu, Jian Xiao. .......................................................... 65
Original Paper

Exploring the Shift in International Trends in Mobile Health Research From 2000 to 2020: Bibliometric Analysis

Jianfei Cao¹, MSc; Yeongjoo Lim², PhD; Shintaro Sengoku³, PhD; Xitong Guo⁴, PhD; Kota Kodama¹, PhD

¹Graduate School of Technology Management, Ritsumeikan University, Ibaraki, Japan
²Department of Business Administration, Ritsumeikan University, Ibaraki, Japan
³Department of Innovation Science, Tokyo Institute of Technology, Tokyo, Japan
⁴Institute at School of Management, Harbin Institute of Technology, Harbin, China

Corresponding Author:
Kota Kodama, PhD
Graduate School of Technology Management
Ritsumeikan University
Ibaraki
Japan
Phone: 81 0726652448
Email: kkodama@fc.ritsumei.ac.jp

Abstract

Background: Smartphones have become an integral part of our lives with unprecedented popularity and a diverse selection of apps. The continuous upgrading of information technology has also enabled smartphones to display great potential in the field of health care.

Objective: We aimed to determine the future research direction of mobile health (mHealth) by analyzing its research trends and latest research hotspots.

Methods: This study collected mHealth-related literature published between 2000 and 2020 from the Web of Science database. Descriptive statistics of publication trends of mHealth research were determined by analyzing the annual number of publications in the literature and annual number of publications by country. We constructed visualization network maps of country (or regional) collaborations and author-provided keyword co-occurrences, as well as overlay visualization maps of the average publication year of author-provided keywords to analyze the hotspots and research trends in mHealth research.

Results: In total, 12,593 mHealth-related research papers published between 2000 and 2020 were found. The results showed an exponential growth trend in the number of annual publications in mHealth literature. JMIR mHealth and uHealth, the Journal of Medical Internet Research, and JMIR Research Protocols were the 3 top journals with respect to number of publications. The United States remained the leading contributor to the literature in this area (5294/12,593, 42.0%), well ahead of other countries and regions. Other countries and regions also showed a clear trend of annual increases in the number of mHealth publications. The 4 countries with the largest number of publications—the United States, the United Kingdom, Canada, and Australia—were found to cooperate more closely. The rest of the countries and regions showed a clear geographic pattern of cooperation. The keyword co-occurrence analysis of the top 100 authors demonstrated 5 clusters, namely, development of mHealth medical technology and its application to various diseases, use of mHealth technology to improve basic public health and health policy, mHealth self-health testing and management in daily life, adolescent use of mHealth, and mHealth in mental health. The research trends revealed a gradual shift in mHealth research from health policy and improving public health care to the development and social application of mHealth technologies.

Conclusions: To the best of our knowledge, the most current bibliometric analysis dates back to 2016. However, the number of mHealth research published between 2017 and 2020 exceeds the previous total. The results of this study shed light on the latest hotspots and trends in mHealth research. These findings provide a useful overview of the development of the field; they may also serve as a valuable reference and provide guidance for researchers in the digital health field.

(JMIR Mhealth Uhealth 2021;9(9):e31097) doi:10.2196/31097

https://mhealth.jmir.org/2021/9/e31097
Introduction

In recent years, smartphones have become popular in many countries; especially in high-income countries such as the United Kingdom and the United States—as of September 2019, the smartphone penetration rate is as high as 80% [1]. With the popularity of smartphones, the richness of smartphone app functions and the anytime-anywhere operability provide more opportunities for health promotion, especially in the medical field [2,3]. Services for medical and public health supported by mobile devices is defined as mobile health (mHealth). The outbreak of coronavirus disease 2019 (COVID-19) in 2020 has exposed the lack of medical resources in many countries [4-7]. In this context, mHealth apps can monitor body information, such as heart rate, as well as behavioral information, such as real-time acceleration, through smartphones, smartwatches, and other mobile devices. This can enable people to check their health status at any time and provide medical staff with more reference data [8-10]. Therefore, the development of mHealth can alleviate the shortage of medical resources to a certain extent [11,12]. The great potential shown by mHealth in the medical field has received attention from researchers in many countries [13]. A focus for an increasing number of researchers is to determine how further developments in the mHealth field can reasonably create more social value; therefore, it is necessary to have an in-depth understanding of current research trends and hot spots in mHealth.

Bibliometrics can quantify comprehensive textual information to provide numerical statistics on the development process of a particular topic [14]. The quantified numerical information can also help scholars identify the future trends of a subject [15]. Bibliometrics is widely used in academics, specifically for the in-depth analysis of journal papers [16,17]. Recently, researchers have developed many tools that meet the needs of bibliographic analysis and enrich the bibliographic treatment, such as for the analysis of co-authors’ countries (or regions) and research institutions to elucidate the collaboration between different regions or research institutions [18-20], the extraction of keywords for co-occurrence analysis to identify research hotspots [21,22], and keyword clustering to identify the main research directions in a field [23,24]. Thus, bibliometrics plays an important role, both in providing an overview of the past and to provide predictive information.

Currently, there are only a few papers on bibliometric analyses of mHealth literature. Sweileh et al [13] searched Scopus for mHealth papers between 2006 and 2016 and found that most keywords were related to diabetes, medication adherence, and obesity. This study also found an exponential growth in mHealth literature.

Shen et al [25] collected 2704 papers related to mHealth from the Web of Science database as of 2016. Although different from the database searched by Sweileh et al [13], the results of the 2 studies were similar in that both found the United States to be the most active country in mHealth research worldwide; they also showed an exponential growth trend in publications on mHealth in the Web of Science. By identifying the keywords, Shen et al [25] classified the research hotspots in mHealth research into the following 4 main areas: (1) patient engagement and patient intervention, (2) health monitoring and self-care, (3) mobile device and mobile computing, and (4) security and privacy.

Another bibliometric analysis [26] of mHealth literature, published in 2020, focused on papers related to mHealth apps. A total of 2802 papers published between 2000 and 2019 were collected from the Web of Science. The current state of research, research trends, hotspots, and coauthorship networks showed that the United States, England, Australia, and Canada were the most productive countries for mHealth apps research and the hot topics of mHealth apps research formed 5 clusters: (1) technology and system development of mobile health apps, (2) mobile health apps used in mental health, (3) mobile health apps used as mobile health tools in telemedicine, chronic disease, and medication adherence management, (4) mobile health apps used in health behavior and health promotion, and (5) mobile health apps used in disease prevention via the internet.

However, a gap—between 2017 and 2020—in bibliometric analysis of mHealth research remains. Both Sweileh et al [13] and Shen et al [25] found that there was an exponential growth trend up until 2016; therefore, it can be expected that the number would have grown substantially from 2017 to 2020. In fact, the number of publications in the mid-2017 to 2020 period surpasses the previous total. Therefore, a renewed bibliometric analysis of mHealth research from 2000 to 2020 was necessary. The period 2000 to 2020, instead of only 2017 to 2020, was chosen to facilitate the calculation of logical growth curves for publications and the visualization of trends in research hotspots.

Methods

Data Collection

We collected metadata (paper title, abstract, author keywords, author information, country, and references [27]) on papers related to mHealth published between 2000 and 2020 from the Web of Science database. The Web of Science database was chosen because it covers a wide range of fields of study and includes 21,000 peer reviewed and high-quality journals. In addition, the Web of Science includes 6 high-impact citation databases, the Science Citation Index extension, Social Science Citation Index, and many regional databases [28] in its core collection. Thus, the Web of Science database was considered to be appropriate for the bibliometric analysis.

We conducted searches using mHealth and its synonyms as search-topic keywords (in titles, abstracts, and author-provided keywords) to find potential publications related to mHealth;
however, this simple approach has a major limitation. As Sweileh et al. [13] suggested, many researchers might not identify their papers as focusing on mHealth though the papers are mHealth-related. Therefore, a second search strategy was also used. Given that mHealth depends on mobile devices, we searched for author-provided keywords related to both mobile devices and mHealth (smartphone, mobile phone, etc) and general health (health, health care, etc). Author-provided keyword searches were used instead of topic searches because the latter may have led to the inclusion of papers that did not emphasize the study of mobile devices and health, whereas the former represents keywords chosen by authors to highlight the contents of their papers. Thus, we determined that searching for author-provided keywords would be more appropriate to collect articles related to mHealth. Both search strategies were conducted for the period from 2000 to 2020, and only papers published in English were retrieved (Figure 1). We implemented the search on March 2, 2021. The results from both strategies were aggregated, and duplicates were removed.

Figure 1. Data collection strategy for mHealth research bibliometric analysis. AK: author-provided keywords; TS: topic search.

Data Analysis

We used VOSviewer (version 1.6.15) for data analysis. In bibliometric analysis, mapping and clustering techniques can provide insight into network structure and are usually used together [28,29]; however, these 2 techniques were developed independently and rely upon different ideas and assumptions. Waltman et al. [30] proposed a unified mapping approach and clustering, which is used in VOSviewer [31]. This tool has been used in bibliometric analyses in many fields [32,33].

The annual number of publications, the annual growth rate, $AGR$; relative growth rate, $RGR$; doubling time, $DT$; and the growth curve of publications were calculated to observe publication trends in mHealth literature using Excel (version 2013; Microsoft Inc). In the growth curve, $x$ is the number of years of growth since 2000, and $y$ is the cumulative number of publications. We examined the coefficient of determination ($R^2$) to confirm the explanatory power of the growth curve. $AGR$ was defined as the percentage change in the number of publications per year and is calculated with the following formula:

$$AGR = \left(\frac{N_2 - N_1}{N_1}\right) \times 100$$

where $N$ is the annual number of publications.

$RGR$ was defined as the growth rate of the cumulative number of publications per unit of time and was calculated with the following formula [13,34]:

$$RGR = \left(\frac{\ln(TN_2) - \ln(TN_1)}{T_2 - T_1}\right) \times 100$$

where $T$ is the year and $TN$ is the cumulative number of publications.

$DT$ was defined as the number of times the number of publications double in 1 year and was calculated with the following formula [13,34]:

$$DT = \frac{0.693}{RGR}$$

In addition, we analyzed the publication trends by country (or region) and the distribution of publications by journal. Using...
VOSviewer, we created bibliometric maps for social networks, based on countries and regions, to identify international partnerships in the mHealth field.

In this study, we did a co-occurrence analysis using author-provided keywords in VOSviewer to elucidate research hotspots in the mHealth field. We set the minimum number of co-occurrences to 50. The keywords mHealth and smartphone (as well as keywords with a similar meaning) appeared more frequently because of the search strategy and took up a large weight in the co-occurrence network graph. Such keywords were considered to influence the distribution of the remaining keywords; hence, we removed the keywords used in the search strategy that appeared in the results, to focus the results on valuable research-topic buzzwords. We then extracted the top 100 keywords and mapped them into a keyword co-occurrence network. The top 100 author-provided keywords were superimposed and visualized according to the average publication year to determine the changes in research hotspots of mHealth over time. The node size indicates the number of times the author’s keyword appeared, and the color of the node changes gradually, according to the average publication year.

Results

mHealth Research Publications

Through the first search strategy, 6604 search results were obtained, and through the second strategy, 7037 search results were obtained. After removing 1048 documents; there were 12,593 remaining (Figure 1). The number of publications related to mHealth has been increasing since 2004 (Table 1, Figure 2, and Figure 3) and has demonstrated an approximately exponential growth trend. By fitting an exponential function equation, the growth curve can be represented by $y = 37e^{0.3062x}$, with $R^2 = 0.9935$. Specifically, the year 2015 was a flashpoint. The number of documents published in 2015 increased by 366 compared to 2014, and the annual growth rate reached 61%, becoming the highest annual growth rate in 20 years. $RGR$ dropped from 58% in 2001 to 30% in 2003 and then stabilized at 28% (SD 5%). $DT$ increased from 1.2 in 2001 to 2.3 in 2003 and then stabilized at 2.6 (SD 0.5). The stability of $RGR$ and $DT$ demonstrates the exponential growth trend [13,34] of the number of publications and confirms that the curve in Figure 3 is exponential, which indicates that the field of mHealth is increasingly receiving attention from scholars.
Table 1. Descriptive statistics of the collected mHealth literature.

<table>
<thead>
<tr>
<th>Year</th>
<th>Publications, n</th>
<th>Annual growth, n</th>
<th>$AGR^a$ (%)</th>
<th>$RGR^b$ (%)</th>
<th>$DT^c$</th>
<th>Cumulative total, n</th>
</tr>
</thead>
<tbody>
<tr>
<td>2000</td>
<td>37</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>37</td>
</tr>
<tr>
<td>2001</td>
<td>29</td>
<td>–8</td>
<td>–22</td>
<td>58</td>
<td>1.2</td>
<td>66</td>
</tr>
<tr>
<td>2002</td>
<td>41</td>
<td>12</td>
<td>41</td>
<td>48</td>
<td>1.4</td>
<td>107</td>
</tr>
<tr>
<td>2003</td>
<td>37</td>
<td>–4</td>
<td>–10</td>
<td>30</td>
<td>2.3</td>
<td>144</td>
</tr>
<tr>
<td>2004</td>
<td>50</td>
<td>13</td>
<td>35</td>
<td>30</td>
<td>2.3</td>
<td>194</td>
</tr>
<tr>
<td>2005</td>
<td>77</td>
<td>27</td>
<td>54</td>
<td>33</td>
<td>2.1</td>
<td>271</td>
</tr>
<tr>
<td>2006</td>
<td>86</td>
<td>9</td>
<td>12</td>
<td>28</td>
<td>2.5</td>
<td>357</td>
</tr>
<tr>
<td>2007</td>
<td>96</td>
<td>10</td>
<td>12</td>
<td>24</td>
<td>2.9</td>
<td>453</td>
</tr>
<tr>
<td>2008</td>
<td>123</td>
<td>27</td>
<td>28</td>
<td>24</td>
<td>2.9</td>
<td>576</td>
</tr>
<tr>
<td>2009</td>
<td>158</td>
<td>35</td>
<td>28</td>
<td>24</td>
<td>2.9</td>
<td>734</td>
</tr>
<tr>
<td>2010</td>
<td>184</td>
<td>26</td>
<td>16</td>
<td>22</td>
<td>3.1</td>
<td>918</td>
</tr>
<tr>
<td>2011</td>
<td>236</td>
<td>52</td>
<td>28</td>
<td>23</td>
<td>3.0</td>
<td>1154</td>
</tr>
<tr>
<td>2012</td>
<td>302</td>
<td>66</td>
<td>28</td>
<td>23</td>
<td>3.0</td>
<td>1456</td>
</tr>
<tr>
<td>2013</td>
<td>437</td>
<td>135</td>
<td>45</td>
<td>26</td>
<td>2.6</td>
<td>1893</td>
</tr>
<tr>
<td>2014</td>
<td>603</td>
<td>166</td>
<td>38</td>
<td>28</td>
<td>2.5</td>
<td>2496</td>
</tr>
<tr>
<td>2015</td>
<td>970</td>
<td>367</td>
<td>61</td>
<td>33</td>
<td>2.1</td>
<td>3466</td>
</tr>
<tr>
<td>2016</td>
<td>1206</td>
<td>236</td>
<td>24</td>
<td>30</td>
<td>2.3</td>
<td>4672</td>
</tr>
<tr>
<td>2017</td>
<td>1383</td>
<td>177</td>
<td>15</td>
<td>26</td>
<td>2.7</td>
<td>6055</td>
</tr>
<tr>
<td>2018</td>
<td>1725</td>
<td>342</td>
<td>25</td>
<td>25</td>
<td>2.8</td>
<td>7780</td>
</tr>
<tr>
<td>2019</td>
<td>2132</td>
<td>407</td>
<td>24</td>
<td>24</td>
<td>2.9</td>
<td>9912</td>
</tr>
<tr>
<td>2020</td>
<td>2681</td>
<td>549</td>
<td>26</td>
<td>24</td>
<td>2.9</td>
<td>12,593</td>
</tr>
</tbody>
</table>

$^aAGR$: annual growth rate.

$^bRGR$: relative growth rate.

$^cDT$: doubling time.

$^dN/A$: not applicable.
Publishing Trends and Cooperation Among Countries and Regions

We found that scholars from 166 countries and regions contributed to publications on mHealth (Multimedia Appendix 1). The United States had the largest number of publications (5294/12,593, 42.0%), followed by the United Kingdom (1372/12,593, 10.9%), and then Australia (979/12,593, 7.8%), China (842/12,593, 6.7%), and Canada (828/12,593, 6.6%) (Table 2). Compared with that of other countries, the growth curve of the United States shows explosive growth (Figure 4); mHealth received more attention, early on, from scholars in the United States which continued throughout the period. All countries and regions show growth, though not as high as that of the United States.
Table 2. Top 10 contributing countries in mHealth literature between 2000 and 2020.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Country and territory</th>
<th>Publications (n=12,593), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>United States</td>
<td>5294 (42.0)</td>
</tr>
<tr>
<td>2</td>
<td>United Kingdom</td>
<td>1372 (10.9)</td>
</tr>
<tr>
<td>3</td>
<td>Australia</td>
<td>979 (7.8)</td>
</tr>
<tr>
<td>4</td>
<td>China</td>
<td>842 (6.7)</td>
</tr>
<tr>
<td>5</td>
<td>Canada</td>
<td>828 (6.6)</td>
</tr>
<tr>
<td>6</td>
<td>Germany</td>
<td>583 (4.6)</td>
</tr>
<tr>
<td>7</td>
<td>The Netherlands</td>
<td>526 (4.2)</td>
</tr>
<tr>
<td>8</td>
<td>Spain</td>
<td>445 (3.5)</td>
</tr>
<tr>
<td>9</td>
<td>Italy</td>
<td>426 (3.4)</td>
</tr>
<tr>
<td>10</td>
<td>India</td>
<td>424 (3)</td>
</tr>
</tbody>
</table>

*Due to research cooperation between scholars of different nationalities, some papers have been counted more than once.

Figure 4. Comparison of the growth trends of mHealth-related research publications in various countries between 2000 and 2020. Due to research cooperation between scholars of different nationalities, some papers have been counted more than once.

![Comparison of the growth trends of mHealth-related research publications in various countries between 2000 and 2020.](image)

Usually, the closer the two circles, the thicker the links and the stronger the relationship (between the countries). Different colors indicate different clusters, and circles belonging to the same cluster usually have similar properties or characteristics [31]. All countries had a cooperative relationship with the United States (Figure 5). Of the top 5 countries, in terms of the number of publications, United States, the United Kingdom, Canada, and Australia occupy the center of the network diagram with similar distances between the nodes; these 4 productive countries have strong collaborative relationships. Furthermore, it is evident from the location of the countries’ nodes that the cooperation between countries and regions have geographic characteristics.
Figure 5. Visual network diagram of cooperation between countries or regions. The size of the circles indicates the number of publications. The larger the circle, the greater the number of publications. The length and thickness of the links between the circles indicate the strength of partnerships between countries. Asian countries and regions represented by the red cluster and the European countries and regions represented by the green cluster.

Journal Distribution

Literature related to mHealth was distributed among 3268 journals (Table 3). The Canadian Journal of Medical Internet Research and its sister journals JMIR mHealth and uHealth, JMIR Research Protocols, and JMIR Mental Health were in the top 10 journals, with respect to number of publications, and together represented 14% of all publications (1763/12,593). In addition, all of the top 10 journals, except JMIR Research Protocols, have an impact factor above 2.

Table 3. Top 10 journals, in terms of the number of mHealth publications, between 2000 and 2020.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Journal</th>
<th>Country</th>
<th>2-year impact factor (in 2019)</th>
<th>Publications (n=12,593), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>JMIR mHealth and uHealth</td>
<td>Canada</td>
<td>4.31</td>
<td>956 (7.6)</td>
</tr>
<tr>
<td>2</td>
<td>Journal of Medical Internet Research</td>
<td>Canada</td>
<td>5.03</td>
<td>463 (3.7)</td>
</tr>
<tr>
<td>3</td>
<td>JMIR Research Protocols</td>
<td>Canada</td>
<td>___a</td>
<td>235 (1.9)</td>
</tr>
<tr>
<td>4</td>
<td>Telemedicine and Health</td>
<td>The United States</td>
<td>2.841</td>
<td>202 (1.6)</td>
</tr>
<tr>
<td>5</td>
<td>International Journal of Environmental Research and Public Health</td>
<td>Switzerland</td>
<td>2.849</td>
<td>145 (1.2)</td>
</tr>
<tr>
<td>6</td>
<td>BMC Public Health</td>
<td>The United Kingdom</td>
<td>2.69</td>
<td>139 (1.1)</td>
</tr>
<tr>
<td>7</td>
<td>JMIR Mental Health</td>
<td>Canada</td>
<td>3.54</td>
<td>109 (0.87)</td>
</tr>
<tr>
<td>8</td>
<td>International Journal of Medical Informatics</td>
<td>Ireland</td>
<td>3.025</td>
<td>106 (0.84)</td>
</tr>
<tr>
<td>9</td>
<td>BMC Medical Informatics and Decision Making</td>
<td>The United Kingdom</td>
<td>2.317</td>
<td>101 (0.80)</td>
</tr>
<tr>
<td>10</td>
<td>Sensors</td>
<td>Switzerland</td>
<td>3.275</td>
<td>99 (0.79)</td>
</tr>
</tbody>
</table>

aNot available.
Author Keywords Co-occurrence Analysis

The top 100 keywords (Multimedia Appendix 2) were classified into 5 clusters using keyword clustering analysis (Figure 6), and the top 10 keywords by co-occurrence frequency are shown (Table 4). The average year of publication for the keywords shown in Table 4 ranged from 2015.26 to 2017.90, and the average number of citations ranged from 10.75 to 17.98. The most frequently occurring keyword was mental health, with a co-occurrence frequency of 449, followed by physical activity, with a co-occurrence frequency of 285.

Figure 6. Co-occurrence network diagram of the top 100 author keywords in mHealth research between 2000 and 2020.

Table 4. Top 10 author-provided keywords of mHealth research between 2000 and 2020.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Keyword</th>
<th>Cluster</th>
<th>Occurrences, n</th>
<th>Average year of publication</th>
<th>Average number of citations</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>mental health</td>
<td>Purple</td>
<td>449</td>
<td>2017.30</td>
<td>12.48</td>
</tr>
<tr>
<td>2</td>
<td>physical activity</td>
<td>Blue</td>
<td>285</td>
<td>2017.46</td>
<td>14.00</td>
</tr>
<tr>
<td>3</td>
<td>health promotion</td>
<td>Green</td>
<td>243</td>
<td>2015.26</td>
<td>14.97</td>
</tr>
<tr>
<td>4</td>
<td>self-management</td>
<td>Red</td>
<td>234</td>
<td>2017.90</td>
<td>10.75</td>
</tr>
<tr>
<td>5</td>
<td>public health</td>
<td>Red</td>
<td>232</td>
<td>2016.29</td>
<td>13.41</td>
</tr>
<tr>
<td>6</td>
<td>depression</td>
<td>Purple</td>
<td>227</td>
<td>2017.51</td>
<td>17.98</td>
</tr>
<tr>
<td>7</td>
<td>HIV</td>
<td>Yellow</td>
<td>208</td>
<td>2017.57</td>
<td>11.37</td>
</tr>
<tr>
<td>8</td>
<td>text messaging</td>
<td>Yellow</td>
<td>207</td>
<td>2016.90</td>
<td>13.22</td>
</tr>
<tr>
<td>9</td>
<td>obesity</td>
<td>Blue</td>
<td>173</td>
<td>2016.65</td>
<td>13.81</td>
</tr>
<tr>
<td>10</td>
<td>adherence</td>
<td>Yellow</td>
<td>157</td>
<td>2017.48</td>
<td>13.85</td>
</tr>
</tbody>
</table>
The average publication year range of the top 15 author-provided keywords was 2017.98 to 2020.05 (Table 5), and the occurrence range was 41 to 135 (Figure 7). Among the top 15 keywords, 8 belonged to cluster red, 5 belonged to cluster purple, 1 belonged to cluster yellow, and 1 belonged to cluster green. The average publication year range of the bottom 15 author-provided keywords was 2015.26 to 2016.19, and the occurrence range was 46 to 243. Among the bottom 15 keywords, 10 belonged to cluster green, 2 belonged to cluster red, 2 belonged to cluster yellow, and 1 belonged to cluster purple.

Table 5. Comparison of the top 15 and bottom 15 author-provided keywords.

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Occurrences, n</th>
<th>Average publication year</th>
<th>Keyword</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Top 15</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Red</td>
<td>86</td>
<td>2020.05</td>
<td>covid-19</td>
</tr>
<tr>
<td>Red</td>
<td>41</td>
<td>2019.05</td>
<td>artificial intelligence</td>
</tr>
<tr>
<td>Red</td>
<td>43</td>
<td>2018.79</td>
<td>wearables</td>
</tr>
<tr>
<td>Red</td>
<td>85</td>
<td>2018.55</td>
<td>machine learning</td>
</tr>
<tr>
<td>Purple</td>
<td>42</td>
<td>2018.54</td>
<td>gamification</td>
</tr>
<tr>
<td>Red</td>
<td>47</td>
<td>2018.48</td>
<td>feasibility</td>
</tr>
<tr>
<td>Red</td>
<td>87</td>
<td>2018.31</td>
<td>wearable devices</td>
</tr>
<tr>
<td>Purple</td>
<td>67</td>
<td>2018.26</td>
<td>ecological momentary assessment</td>
</tr>
<tr>
<td>Yellow</td>
<td>135</td>
<td>2018.24</td>
<td>randomized controlled trial</td>
</tr>
<tr>
<td>Red</td>
<td>69</td>
<td>2018.22</td>
<td>internet of things</td>
</tr>
<tr>
<td>Purple</td>
<td>54</td>
<td>2018.16</td>
<td>mindfulness</td>
</tr>
<tr>
<td>Purple</td>
<td>45</td>
<td>2018.09</td>
<td>sleep</td>
</tr>
<tr>
<td>Purple</td>
<td>93</td>
<td>2018.05</td>
<td>anxiety</td>
</tr>
<tr>
<td>Green</td>
<td>59</td>
<td>2018.04</td>
<td>qualitative</td>
</tr>
<tr>
<td>Red</td>
<td>55</td>
<td>2017.98</td>
<td>schizophrenia</td>
</tr>
<tr>
<td><strong>Bottom 15</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Green</td>
<td>243</td>
<td>2015.26</td>
<td>health promotion</td>
</tr>
<tr>
<td>Green</td>
<td>95</td>
<td>2015.35</td>
<td>primary health care</td>
</tr>
<tr>
<td>Green</td>
<td>81</td>
<td>2015.40</td>
<td>health policy</td>
</tr>
<tr>
<td>Green</td>
<td>55</td>
<td>2015.45</td>
<td>evaluation</td>
</tr>
<tr>
<td>Yellow</td>
<td>80</td>
<td>2015.53</td>
<td>children</td>
</tr>
<tr>
<td>Red</td>
<td>103</td>
<td>2015.58</td>
<td>medical devices</td>
</tr>
<tr>
<td>Yellow</td>
<td>117</td>
<td>2015.79</td>
<td>prevention</td>
</tr>
<tr>
<td>Green</td>
<td>97</td>
<td>2015.82</td>
<td>health disparities</td>
</tr>
<tr>
<td>Green</td>
<td>152</td>
<td>2015.84</td>
<td>internet</td>
</tr>
<tr>
<td>Green</td>
<td>50</td>
<td>2015.94</td>
<td>focus groups</td>
</tr>
<tr>
<td>Green</td>
<td>87</td>
<td>2015.97</td>
<td>primary care</td>
</tr>
<tr>
<td>Green</td>
<td>67</td>
<td>2016.00</td>
<td>developing countries</td>
</tr>
<tr>
<td>Purple</td>
<td>46</td>
<td>2016.07</td>
<td>well-being</td>
</tr>
<tr>
<td>Red</td>
<td>59</td>
<td>2016.15</td>
<td>health informatics</td>
</tr>
<tr>
<td>Green</td>
<td>84</td>
<td>2016.19</td>
<td>health education</td>
</tr>
</tbody>
</table>
Figure 7. Overlay visualization maps of the average publication year of the top 100 author keywords. The more the node displays a yellow gradient, the later the average publication year of the keyword.

Discussion

Principal Results

Publishing Trends of mHealth Literature

The emergence of mHealth is a great innovation in the rapid development of information technology. It has circumvented the obstacles of location and medical resources of traditional health care, making health care more accessible to a wider range of people. The growth trend for mHealth literature published between 2000 and 2020 was exponential, which suggests that, when mHealth first started, acceptance was low, the number of users was small, and research on mHealth progressed relatively slowly. As the number of users of mHealth gradually increased, more and more researchers focused on this area, and the number of mHealth publications showed an increasing trend. Based on the theory of diffusion of innovation [35], the growth curve (Figure 3) coincides with the early part of the diffusion model of innovations; we can surmise that the development of mHealth technology is currently in the early stages of rapid growth.

International Trends

A comparison with the bibliometric analysis [25] of mHealth research up to 2016 shows that the United States remains the most productive country in this field. The number of annual publications in the United States continues to show a steady growth trend. This is followed by the United Kingdom, China, Australia, and Canada, which are also experiencing rapid growth in their publication trends.

The 4 most productive countries—the United States, the United Kingdom, Canada, and Australia—had close cooperative relationships with each other. In contrast, the rest of the countries and regions showed a clear geographic pattern. Cluster red contains mainly of Asian countries such as Japan, South Korea, Russia, Malaysia, Thailand, and Singapore (Figure 5), and cluster green is composed mainly of European countries such as France, Netherlands, Germany, Spain, and Italy. It is not difficult to speculate that the specificity of the EU has led to closer research cooperation among EU countries. Cluster blue comprises African countries such as Kenya, South Africa, Ghana, Nigeria, Tanzania and 3 of the most productive countries—the United States, Canada, and United Kingdom. It
can be presumed that African countries establish cooperation based on geography and have a major cooperation relationship with these 3 countries. In addition, China and Taiwan may be grouped in cluster purple because of the same language. Australia and New Zealand belong to cluster yellow, due to their close geographic locations. Therefore, we conjecture that international partnerships may be influenced by geography, regional characteristics, language, international relations, political, and economic alliances.

Research Hotspots

Cluster red contains the most author-provided keywords, comprising 29 keywords such as artificial intelligence, electronic health records, global health, health informatics, health information technology, machine learning, medical devices, self-care, and wearable devices. The keywords breast cancer, cancer, COVID-19, heart failure, and other diseases also appeared in the list. This cluster focuses on the development of mHealth technologies and their application to various diseases. Globally, health issues such as aging populations and cancer pose a serious challenge to health care providers [36-38]. Researchers are increasingly trying to address many health issues with the use of mHealth technologies. COVID-19 also appears among the high-frequency keywords. Importantly, the COVID-19 pandemic exposed the shortage of health care resources in several countries. The demand for telemedicine, including mHealth, has also been indirectly increased by countries promoting policies to prevent their population from going outside under social isolation measures adopted to tackle COVID-19 [39]. It is also worth noting that the keyword privacy appears in this cluster. Patient privacy, security in data transmission, and privacy-related health policy issues remain major barriers to the development of mHealth in both high-income and low- to middle-income countries [40].

Cluster green focuses on the use of mHealth technologies to improve basic public health and health policies. Some of the 25 keywords under this cluster include health promotion, primary health care, health education, health policy, health communication, and community health workers. Health care is one of the largest industries in the world. According to the World Health Organization, global health expenditure in 2017 was US $7.8 trillion, or approximately 10% of the total gross domestic product [41]. Compared to traditional health services, mHealth, which relies upon mobile devices such as smartphones, provides timely health information and fast, inexpensive access to primary care. As of 2017, mobile phone apps related to mHealth exceeded 325,000 [42]. Therefore, it is necessary to formulate corresponding health policies to ensure that mHealth technology can serve society more effectively and to provide direction for future health initiatives. The author-provided keyword developing countries appeared in this cluster. The development of mHealth in low- and middle-income countries faces more serious challenges than those faced in high-income countries. Although smartphones have become commonplace globally, challenges exist in terms of the cost of owning and using smartphones in low- and middle-income countries. For example, resource scarcity and other issues have forced low- and middle-income countries to reduce the budget for building mHealth and related infrastructure to allocate resources to other necessities such as potable water and food. The shortage of trained medical professionals and technical skills in low- and middle-income countries has also made the development of mHealth difficult [38]. Therefore, research focused on low- and middle-income countries remains a key research priority for the future of the field.

Cluster blue focuses on self-health testing and management in daily life. This cluster comprises 18 keywords such as behavior change, diet, exercise, health behavior, lifestyle, and self-monitoring. An increasing number of people are using emerging mHealth apps to improve their lifestyles and manage their health; these apps have a variety of functions. For example, people can control their daily calorie intake by recording their diet [43] or detect changes in their health by recording their weight, heart rate, and breathing rate [44,45]. In fact, the emergence of such apps has played a positive role in the popularization of mHealth. For example, mobile phone apps related to physical exercise have been combined with users’ social networks. People are more willing to use the tracking function of such apps to record their physical changes and share their exercise status with others, thereby increasing their social contacts’ motivation to exercise [46,47].

Cluster yellow focuses on the use of mHealth among adolescents. This cluster contains 16 keywords such as adolescent, adherence, children, HIV, intervention, sexual health, social media, social support, and youth. Research shows that the youth are the most prone to smartphone addiction [48,49]. There has been considerable research on the negative effects of smartphone addiction on health [50-52]. Excessive smartphone use affects sleep quality, and thus, other daily activities [48,53]. Adolescents are also a priority group for HIV prevention. mHealth apps that use social media technology make it easier for health workers to spread sexual health information more effectively, and thus, reduce the risk of HIV infection among adolescents [54]. Therefore, mHealth research focusing on adolescents is essential.

Cluster purple focuses on the use of mHealth in the context of mental health. It contains 12 keywords, including mental health, anxiety, mindfulness, stress, and well-being. The keyword mental health has the most frequent co-occurrence. Therefore, it can be assumed that this topic is the primary focus of researchers. Various factors influence mental health, such as past experiences [55], social stress [56], and interpersonal relationships [57]. People with mental health problems often resist talking to others [58], and even those who have undergone psychotherapy and have recovered are at high risk of reoccurrence [59]. Mental illness is a severe social problem, especially in high-income countries. For example, in Japan, the suicide rate due to depression has been high, and it has been increasing among youth in recent years [60,61]. For a country with a serious aging problem, an increase in the suicide rate among young people can incur a huge cost to the national economy. Moreover, people with depression can have poor physical health compared with that of individuals in the general population [62]. Timely intelligence technology that captures body information provided by mHealth can provide psychologists with more reference data to detect physical changes in patients through ecological
momentary assessment, thus providing more guidance to patients.

**Research Trends**

Based on the clusters to which these keywords belong, we can speculate that mHealth research hotspots have gradually shifted from research on mHealth policy and the improvement of public health care to the development of mHealth technology and social apps (cluster green to cluster red and cluster purple). Thus, we find that the development of mHealth requires appropriate health policy as a cornerstone. However, individual governments usually develop health policies, leading to national and regional limitations in the scope of policy application. In contrast, the scope of web-based mHealth services can be global. This may also make it more difficult to regulate mHealth services; therefore, it is still necessary to continue to explore how to establish regulations for cross-border telehealth in the future.

Furthermore, we note that in high-income countries, especially in the health care field, government regulatory formation is critical to the growth of the mHealth market [63]. Governmental oversight measures often limit the development of mHealth technologies and services [64]. Although the United States is absolutely central to mHealth research, health care regulations in the country may be more conservative and less susceptible to change due to the huge health care infrastructure. Conversely, mHealth policy reforms are likely to be smoother in low- and middle-income countries because they are met with less opposition and fewer infrastructural barriers [65]. Therefore, effective strategies are needed to advance regulatory reforms related to mHealth.

**Limitations**

To the best of our knowledge, the results obtained in this study are the most recent available for mHealth bibliometric analysis; however, this study has some limitations. First, we developed a search strategy that included as many mHealth-related studies as possible, but we still could not guarantee the inclusion of all mHealth-related studies. Second, our search strategy collected only English-language literature, which narrowed the scope. Hence, the data results are not representative of papers and conference papers published in other languages. Finally, the data used in this study were extracted only from the Web of Science and did not include other search engines such as Scopus and PubMed. Although the Web of Science has a large enough database to ensure the accuracy of the data to a certain extent, there are still many papers that are included only in the other databases, which may have impacted the study results. For example, our finding suggest that only 175 mHealth papers were in Japan (Multimedia Appendix 1); however, many mHealth papers published in Japanese are included in the CiNii database maintained by the National Institute of Informatics in Japan. The Chinese Science Citation Database in China also contains many papers published in Chinese; therefore, future studies can include more databases and languages to make the research results more accurate and rigorous.

**Conclusions**

This study reveals the latest research trends and hotspots and the current state of international collaboration in mHealth research. As previously suggested, mHealth has shown great potential in recent years for use in all aspects of our lives; however, the development of mHealth faces challenges from regulatory policies, national economies, and personal privacy. Therefore, we advise researchers in this field to work on these issues to further develop the mHealth field. We also hope that the results of this study provide valuable guidance for future mHealth research.

**Acknowledgments**

We would like to express our gratitude to Ritsumeikan University for providing access to the Web of Science database. We would also like to thank all study participants for their constructive advice and guidance for this research. This work was supported by the Foundation France-Japon/Air Liquide Fellowship. The authors gratefully acknowledge the generous support and assistance of the Fondation France-Japon de École des Hautes Études en Sciences Sociales and Air Liquide.

**Conflicts of Interest**

None declared.

**Multimedia Appendix 1**

List of countries and regions that have contributed to publications on mHealth.

[XLS File (Microsoft Excel File), 41 KB - mhealth_v9i9e31097_app1.xls ]

**Multimedia Appendix 2**

Details of the top 100 author keywords in mHealth research between 2000 and 2020.

[XLS File (Microsoft Excel File), 42 KB - mhealth_v9i9e31097_app2.xls ]

**References**


60. Tanaka T, Okamoto S. Increase in suicide following an initial decline during the COVID-19 pandemic in Japan. Nat Hum Behav 2021 Feb;5(2):229-238. [doi: 10.1038/s41562-020-01042-z] [Medline: 33453498]


Abbreviations

COVID-19: coronavirus disease 2019
HIV: human immunodeficiency virus
mHealth: mobile health

©Jianfei Cao, Yeongjoo Lim, Shintaro Sengoku, Xitong Guo, Kota Kodama. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 08.09.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License.
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 https://mhealth.jmir.org/, as well as this copyright and license information must be included.
Toward a Better Understanding of the Intention to Use mHealth Apps: Exploratory Study

Pedro R Palos-Sanchez1*, PhD; Jose Ramon Saura2*, PhD; Miguel Ángel Rios Martín1*, PhD; Mariano Aguayo-Camacho1**, PhD

1Department of Financial Economy and Operations Management, Faculty of Economic and Business Sciences, University of Seville, Seville, Spain
2Department of Business Economics, Rey Juan Carlos University, Madrid, Spain
*all authors contributed equally

Corresponding Author:
Pedro R Palos-Sanchez, PhD
Department of Financial Economy and Operations Management
Faculty of Economic and Business Sciences
University of Seville
Av. Ramon y Cajal, 1
Seville, 41017
Spain
Phone: 34 954557515
Email: ppalos@us.es

Abstract

Background: An increasing number of mobile health (mHealth) apps are becoming available for download and use on mobile devices. Even with the increase in availability and use of mHealth apps, there has still not been a lot of research into understanding the intention to use this kind of apps.

Objective: The purpose of this study was to investigate a technology acceptance model (TAM) that has been specially designed for primary health care applications.

Methods: The proposed model is an extension of the TAM, and was empirically tested using data obtained from a survey of mHealth app users (n=310). The research analyzed 2 additional external factors: promotion of health and health benefits. Data were analyzed with a PLS–SEM software and confirmed that gender moderates the adoption of mHealth apps in Spain. The explanatory capacity ($R^2$ for behavioral intention to use) of the proposed model was 76.4%. Likewise, the relationships of the external constructs of the extended TAM were found to be significant.

Results: The results show the importance of healthy habits developed by using mHealth apps. In addition, communication campaigns for these apps should be aimed at transferring the usefulness of eHealth as an agent for transforming attitudes; additionally, as more health benefits are obtained, ease of use becomes greater. Perceived usefulness (PU; $\beta$=.415, $t_{0.001;499}$=3.442, $P$=.001), attitude toward using ($\beta$=.301, $t_{0.01;499}$=2.299, $P$=.02), and promotion of health ($\beta$=.210, $t_{0.05;499}$=2.108, $P$=.03) were found to have a statistically significant impact on behavior intention to use eHealth apps ($R^2$=76.4%). Perceived ease of use (PEOU; $\beta$=.179, $t_{0.01;499}$=2.623, $P$=.009) and PU ($\beta$=.755, $t_{0.001;499}$=12.888, $P$<.001) were found to have a statistically significant impact on attitude toward using ($R^2$=78.2%). Furthermore, PEOU ($\beta$=.203, $t_{0.01;499}$=2.810, $P$=.005), health benefits ($\beta$=.448, $t_{0.01;499}$=4.010, $P$<.001), and promotion of health ($\beta$=.281, $t_{0.01;499}$=2.393, $P$=.01) exerted a significant impact on PU ($R^2$=72.7%). Finally, health benefits ($\beta$=.640, $t_{0.001;499}$=14.948, $P$<.001) had a statistically significant impact on PEOU ($R^2$=40.9%), while promotion of health ($\beta$=.865, $t_{0.001;499}$=29.943, $P$<.001) significantly influenced health benefits ($R^2$=74.7%).

Conclusions: mHealth apps could be used to predict the behavior of patients in the face of recommendations to prevent pandemics, such as COVID-19 or SARS, and to track users’ symptoms while they stay at home. Gender is a determining factor that influences the intention to use mHealth apps, so perhaps different interfaces and utilities could be designed according to gender.

(JMIR Mhealth Uhealth 2021;9(9):e27021) doi:10.2196/27021
KEYWORDS
mHealth apps; mobile apps; eHealth; promotion of health; TAM; PLS–SEM; COVID-19

Introduction

Overview
The use of mobile health (mHealth) apps increased during the first decade of the 21st century [1] and this has led to an increase in the amount of time that users devote to improve their health using mHealth app(s). New ways of monitoring and controlling health indicators and daily activities using new technologies and improvements on the internet have now become available [2].

The increasing use of technology and the internet has forced companies to adapt their marketing strategies to this digital ecosystem. This growth has led to an increase in the use of smartphones around the world [3,4].

For this reason, user behavior and consumption habits with mobile apps have become important fields of research [3,5]. Alharbi et al [6] reported that one type of app which has been increasingly used in recent years is mHealth apps. Support for patients has become more widespread due to the use of these apps. However, users sometimes stop using these apps because they perceive that their usefulness may not cover health quality standards or because the service is not of the same quality as, for example, a visit to the doctor offline [2].

Telemedicine and eHealth have duly become important factors for the analysis, study, improvement, and development of patients’ medical and health care. Electronic health or eHealth was defined by Eysenbach [7] as “health services and information provided by the Internet and related technologies.”

Many mHealth apps provide direct communication links between patients and health care professionals, health education, health portals, wellness management for measuring calories and following a diet, management of diseases such as diabetes and asthma, self-diagnosis to identify symptoms and early diagnosis, medication reminders, and rehabilitation processes and therapies. Therefore, this kind of app could be used to predict what the behavior of patients would be in the face of recommendations to prevent pandemics, such as COVID-19 or SARS, and to track users’ symptoms while they stay at home and follow doctors’ recommendations [8].

The term “application” or “app” refers to a self-contained program or piece of software that is designed to fulfill a particular purpose, and is usually optimized to run on mobile devices, such as smartphones, tablet computers, and wearable devices such as smart watches [3].

Therefore, mHealth apps can improve users’ health by monitoring risks, symptoms, and health care programs. Consumer interest in mHealth apps has increased at the same rate as new technology use in the health care sector. Taking the characteristics shown by mHealth apps into consideration, the technology acceptance model (TAM) was chosen for this study [9].

TAM is a computational system, presented by Davis [10], which analyses users’ decision-making processes when adopting a new technology. The TAM was used in this research paper to investigate the adoption of mHealth apps. External factors that help describe the user adoption of mHealth apps were incorporated into the TAM.

This research therefore fills a gap in the information currently available because it incorporates innovative factors for the adoption of mHealth apps that creators and developers should take into account for successful acceptance and adoption of new mHealth apps. This information duly adds to the existing literature that can be consulted by professionals and researchers. Therefore, this study addresses the following research question: What factors, including the innovative TAM variables such as promotion of health and health benefits, determine the acceptance of mHealth apps?

This paper is divided into 5 sections. First, the theoretical framework for adoption of mHealth apps is explained. TAM is analyzed and the hypotheses to be studied are formulated. The next section explains the methodology used in the study. The characteristics of the chosen research technique, a survey, are given. This section covers all aspects of questionnaire design and data collection.

Finally, the results of PLS–SEM analysis of the hypotheses and relationships are presented. This section also includes the interpretation, discussion, and implications of the results obtained. The conclusions of the study and the main theoretical and practical implications of the results are also presented.

Theoretical Background
As stated above, in recent years, researchers have become interested in the adoption of mHealth apps. Research by Housman [11] investigated health information on social media by studying how mHealth apps share results on social media platforms. The increase in use of social networks and the factors that affect the relationship and use of mHealth app were also studied by investigating the social acceptance of mHealth apps by internet user communities [3].

Likewise, Li et al [12] studied emotional bonding of patients with mHealth apps. They showed that users accept this type of apps from an emotional perspective, keeping the disease more in mind and, therefore, applying better monitoring protocols.

Handel [13] studied the use of mobile apps for health and wellness and identified the uses of mHealth apps for health, weight loss, consumption of healthy diet and food, monitoring glucose levels and diabetes, calculating calories consumed, disease diagnosis, meditation, yoga, monitoring sleep quality, and tracking sports activities [14,15]. Therefore, these categories of health care have already been accepted as interesting topics for scientific research in the area of mHealth apps.

Atienza and Patrick [16] studied the acceptance of mHealth apps for the care industry. Furthermore, Grundy et al [17] studied the use of high-quality mHealth apps with...
innovation-based systems and systematically described the characteristics of recent apps.

Following this line of research, Mueller [18] studied the types of mHealth apps recommended by doctors to their patients, concluding that this type of app is a valid technological support for disease monitoring and treatment.

Likewise, Bloomfield et al [19] studied the influence of SMART goals on the behavior of mHealth app users. Cho [20] investigated the impact of postadoption sentiments on mHealth app use with the postacceptance model and the TAM to find the users’ continued intention to use health apps.

Bort-Roig et al [21] investigated how mHealth apps could improve employees’ sedentary lifestyles while at work and studied the users’ acceptance and continued use of mHealth app. In a similar way, Ashurst and Jones [22] studied the acceptance of mHealth apps among people with diabetes who used one to check and control their condition. It can be seen that the diagnosis and control of medical conditions with technology is an accepted area of scientific research.

Accordingly, Gorkem et al [23] investigated what factors may influence users’ behavioral intentions to adopt and use mHealth apps. To this end, the authors extended the TAM with external factors such as price value, trust factors, and perceived risk and evaluated users’ technology acceptance. The results of this study showed that the first 2 presented a statistical significance with intention to use.

Deng et al [24] studied which determinants influence the adoption of mHealth services among Chinese patients using the TAM extended with trust, perceived risks, and patients’ age and chronic diseases. All external variables were found to be positively correlated with mHealth service adoption.

The study carried out by Mao et al [25] highlighted the importance of studying the recommendations made by patients who have used this type of app to predict what the behavior of patients would be in the face of a change in medical treatment.

In this context, aiming to understand the main advances of mHealth apps, this study takes as a reference the apps regulated by the Food and Drug Administration (FDA). As noted by Humphries et al [2], the FDA is a leading international institution in the regulation of new health products and services and serves as a guide and institutional leader for all other regulatory institutions in the health field around the world, including Spain. Table 1 shows the main mHealth app categories related to this study’s objectives.

### Table 1. mHealth App categories regulated by the Food and Drug Administration.

<table>
<thead>
<tr>
<th>Mobile health app (category)</th>
<th>Description</th>
<th>Functions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Slendertone Connect (health and wellness)</td>
<td>Allows users to measure physical exercise intensity by connecting an intelligent device.</td>
<td>Measures patients’ resistance, tracks distance traveled, and allows patients to monitor calories.</td>
</tr>
<tr>
<td>Kardia (medicine)</td>
<td>Allows patients to measure blood glucose levels to detect possible risks and evaluate the patient’s condition.</td>
<td>Tracks patients’ heartbeat, measures the glucose and oxygen level in blood, and shares the patient’s data with the doctor.</td>
</tr>
<tr>
<td>Diasend (health and fitness)</td>
<td>Measures the patient’s diabetes constants.</td>
<td>Shares data in real time with other users of the app, allows patients to track exercise and calories, and connects patients’ data with other health and medicine apps.</td>
</tr>
<tr>
<td>LibreLink (medicine)</td>
<td>Checks blood glucose without extracting blood from the finger using a small external device that connects to the app.</td>
<td>Can add notes about food, insulin, and exercise; gives blood glucose readings; and shares information with family, friends, and doctors in real time.</td>
</tr>
<tr>
<td>Qardio heart health (health and fitness)</td>
<td>Allows patients to control blood pressure, heart rate, and weight. A small external device is used to send the data.</td>
<td>Measurement of blood pressure and patient’s weight, monitoring of heart rate and prediction of heart attacks, and helps share patient information in real time with family and friends.</td>
</tr>
</tbody>
</table>

### Conceptual Framework and Hypothesis Elaboration

The TAM was used to explain the relationship between the acceptance and adoption of technology and the users’ intention to use it [26]. Au and Zafar [27] and Chen and Tan [28] used TAM to demonstrate that perceived usefulness (PU) and perceived ease of use (PEOU) are the most critical factors in the process of adoption and use of new technology. In the TAM, PU and PEOU are considered beliefs and evaluations, respectively, given by users, which influence their attitude toward and intention to use the product (in this case, an app) [29], and finally result in behavior change [30,31].

In the study by Davis [10], TAM was used to explain and predict the use of information systems; in other words, TAM was used to understand the influence of the variables PU and PEOU on the use of technology. The PU is the belief that a certain technology can improve users’ performance while using it. The PEOU is defined as the degree to which a person believes that technology can improve users’ performance while using it. The TAM consistently explains a large part of the variance, 40% according to many authors such as Venkatesh and Davis [32], in the intention to use different information and communication technologies by users in different environments and countries [27,33,34]. Since its appearance, the TAM has been widely analyzed and expanded in different ways [35].

The most important evolutions of TAM have been the TAM2 model by Venkatesh and Davis [32], the Unified Theory of the Acceptance and the Use of Technology by Venkatesh et al [36].
the model for the acceptance of technology and user satisfaction by Wixom and Todd [37], and the TAM3 model.

The reasons for choosing the TAM are its tremendous popularity and besides many studies have used the model. The TAM is often considered a common and robust model to address consumer acceptance of an innovative technology [38]. Scherer et al [39] confirmed that the TAM successfully predicts user behavior and can thus be of interest to all potential users of a new technology [29,40,41]. TAM is widely used, with its application extending to a multitude of technologies, especially websites and apps [38].

The TAM has found relevant support in the literature: there are more than 14,870 citations regarding this model within the core collection of the Web of Science database, and more than 51,495 citations have been retrieved from Google Scholar for the article by Davis [10] as of June 2020, 30 years after his first theory.

Therefore, TAM has established itself well as a robust, powerful, and parsimonious model for predicting user acceptance. However, it has been modified through different extensions.

The first of the TAM extensions, the so-called TAM2 [42], is based on the expansion of the PU background. Subsequently, with the same intention as in TAM2, but to complete the model by incorporating the background of the original TAM, Venkatesh and Bala [35] developed the TAM3. More specifically, while TAM2 added the history of PU, TAM3 was expanded into the constructions that precede the PEOU and that were already established in [43,44].

Legris et al [45] made an important critical review of the model and concluded that TAM is useful, but it has to be integrated into a broader one that includes variables related to social and human processes of change.

Similarly, Tang and Chen [46] concluded that current studies on TAM and its extended models have made great progress and recommended paying more attention for future research on new variables that come from other theories or topics that must be introduced in the new model to make it easier to interpret.

Thus, in many health care studies where TAM was applied, the authors have added variables to extend the original TAM to better adapt it to the context of health care [47].

We can find research studies that have used the TAM, such as [48], in which the authors evaluated the acceptance of home telemedicine services by elderly patients. Within the health care domain, the TAM has been used to examine the determinants of adoption or the intention to adopt health technologies [49,50] and to know the effects of cognitive and contingent factors on the health adoption of smartphone apps [51].

The use of computing in the health care field is increasing, but adoption remains a challenge. To understand and introduce the health information technology, a series of behavioral models and innovation acceptance models have been studied and specifically applied the TAM to understand the acceptance of technology [52].

In addition, as in our work, TAM was developed with a focus on technology that can be used voluntarily without the assistance of professional health staff [47].

Furthermore, a recent study in the field of mHealth, in which extended TAM was used [53], indicated that the findings in the literature are contradictory regarding the adoption of mHealth self-monitoring tools, thereby suggesting a gap in the literature that must be covered.

Besides, Thies et al [54] justified that the lack of adoption of a mobile app to support patients in self-management of chronic diseases was mainly due to problems related to the usability of the app and that patients are not comfortable with the technology.

Likewise, Paré et al [55] indicated that people who declare themselves ill are less likely to use digital or traditional tools to monitor their well-being/health than people in good health. Therefore, it is especially important to investigate the adoption of these instruments by consumers considering the characteristics of both the technology and the individuals (users), especially those related to their health [53], as well as the reliability of the model. Extended TAM is decisive in using unused constructs to cover this gap identified in the literature.

The hypotheses below were chosen after reviewing research studies on mHealth apps by Cho [20], Kim and Park [56], and Jeon and Park [57].

Cho [20] and Jeon and Park [57] demonstrated the influence of PEOU on the use of mHealth apps and its effect on PU. Veer et al [58] explained how the intention to use mHealth app influences PU in communities of older people. The following hypothesis was therefore proposed:

\[ H1: \text{Perceived ease of use has a positive influence on perceived usefulness} \]

Veer et al [58], Hu and Bentler [59], and Deng [60] explored the influence and effect of PEOU on attitude toward using. Thompson et al [61] studied the effect of attitude toward using a technology on the intention to use it. Based on their study, the following hypothesis was proposed:

\[ H2: \text{Perceived ease of use has a positive influence on attitude toward use} \]

With the emergence of mHealth, some studies [49,62] confirm the influence of the PU of patients’ intention to adopt a mHealth management service in other cultural contexts [53].

Chauhan and Jaiswal [63] showed that PU influences attitude toward using an mHealth app. The influence of different variables for using different types of mHealth app was also reported. PU demonstrates how a user feels that a particular technology can have a positive effect on his/her life. This influences the user’s attitude toward using the technology [64]. Consequently, the following hypothesis was proposed:

\[ H3: \text{Perceived usefulness has a positive influence on attitude toward use} \]

To investigate PU, Chang et al [65] analyzed the acceptance of a hospital-based eHealth service. The influence of PU on the behavioral intention to use this service by hospital users was
also found. Likewise, Klein [66] concluded that PU has a positive effect on behavioral intention to use in his research on patient psychology and the use of eHealth services. Therefore, the following hypothesis was proposed:

**H4: Perceived usefulness has a positive influence on behavioral intention to use**

Moores [67] concluded that the attitude toward use variable influences the adoption of technological health care services. In addition, Mun et al [68] concluded that behavioral intention to use has a positive effect on the PU of technology by eHealth professionals. From these investigations, the following hypothesis was proposed:

**H5: Attitude toward use has a positive influence on behavioral intention to use**

Lin and Yang [69] and Buntin et al [70] examined the health benefits of mHealth apps and reported on the main positive health benefits of mHealth apps by applying the TAM for the PEOU construct. Beldad and Hegner [51] studied health benefits with the “health valuation” construct after users tried a fitness app. The confidence that users have in the app was found by extending the TAM with trust, social influence, and health valuation variables. Consequently, the following hypothesis was proposed:

**H6: Health benefits have a positive influence on perceived ease of use**

Jeon and Park [57] investigated the factors that affect the acceptance of mHealth apps for obesity and found the influence and effect of health benefits on PU. They suggested that more studies should be carried out with the TAM to find out how mHealth apps can help manage and reduce problems with health and chronic diseases [67]. Based on this, the following hypothesis was proposed:

**H7: Health benefits have a positive influence on perceived usefulness**

Kim and Park [56] improved the TAM with the promotion of health external variable to apply it for evaluating health information technology. Melzner et al [71] studied the influence of mHealth apps on the promotion of health at the workplace and also the attitude of employees toward using an mHealth app. The effects on productivity and health benefits at work upon using an mHealth app were studied by Kelly et al [72]. Ramtohul [73] performed a comprehensive analysis of the decision to adopt eHealth services from the user’s perspective. Therefore, the following hypothesis was proposed:

**H8: Promotion of health has a positive effect on the health benefits of mHealth apps**

Bert et al [74] studied the influence of mobile phones on promotion of health and concluded that some mHealth apps can help prevent diseases and also influence changes in the users’ health behavior. Ramtohul [73] investigated promotion of health with a construct called “Health Needs,” which expresses the benefits for mHealth app users. Consequently, the following hypothesis was proposed:

**H9: Promotion of health has a positive effect on behavioral intention to use an mHealth app**

Ramtohul [73] also analyzed the influence of promotion of health on PU in a study on psychological variables. Cho et al [51] analyzed the influence of PU on health benefits for workers who use smartwatch apps [75]. Moores [67] linked the PU of an mHealth app with the promotion of health of app users [8]. Therefore, the following hypothesis was proposed:

**H10: Promotion of health has a positive effect on perceived usefulness of an mHealth app**

Venkatesh et al [36] pointed out that men and women have different perceptions of usefulness when deciding on technology acceptance. Shabani [76] studied the importance of gender as a moderating variable for adolescents’ emotional health. Bidmon et al [77] indicated that although both men and women use mHealth app, men tend to use it more on mobile devices. Dyck et al [78] studied the moderating effects of age, gender, and education variables and the influence of these on patients’ physical activity. Based on the studies by Venkatesh et al [36] and Shabani [76], the following hypothesis was proposed:

**H11: Gender and age moderate all the relationships of constructs in the research model**

The research model in Figure 1 was formulated to explore the influence of health benefits and promotion of health on the mHealth app adoption model.
Figure 1. Research model to explore the influence of health benefits and promotion of health on the mHealth app adoption model. TAM: technology acceptance model.

Methods

Measurement

A questionnaire was created with 24 questions on attitudes and behavior and 5 questions for group classification. The classification questions were for gender, age, job, residence, and education level. The questionnaire was divided into 3 sections. The first section dealt with questions on the users’ behavior, beliefs, and attitudes toward an mHealth app. Before answering this section and the next one, users could watch a video on different mHealth apps and try them out. A total of 12 different FDA-approved mHealth apps were suggested for trial purposes. The apps can be found in Google Play or Apple Store by searching their names: my mhealth, Mhealth Medical App, MHealth, Babylon, HealthForYou, Medipal mHealth app, Walking: Pedometer, Medical ID: ICE, Symptom Tracker, ContinuousCare, Medical Record, and ManageMyHealth. All sample members were selected because they indicated that they had previously used mHealth apps and were aware of their functionality and traceability. They were informed about the other apps so that they could take into account additional features of the apps.

The first section of the questionnaire contained 15 questions on PU (n=5), PEOU (n=3), attitude toward using (n=3), and behavioral intention to use (n=4).

The second section consisted of a block of questions on health and disease prevention. These were grouped into health benefits (n=4) and promotion of health (n=4). The last section consisted of 5 questions on the demographic profile of the sample.

Adapted items were used to measure the variables in the TAM [10]. The behavioral items for health were adapted from the studies by Lin and Yang [69] and Jeon and Park [57]. Lin and Yang [69] studied the influence of mHealth app on patients with asthma problems and Jeon and Park [57] studied the influence of mHealth apps on patients with obesity problems. Altogether, there were 24 items in the questionnaire.

All the items, except the demographic profile, were measured using a 5-point Likert scale that ranged from total disagreement (1) to total agreement (5).

A pilot survey was conducted to find the pilot sample’s opinions about the content and structure of the questionnaire, so that the questions could be refined if needed. The pilot survey was conducted on a subsample of 31 individuals whose answers were not added to the final sample.

The subsample followed all the instructions and answered all the questions. Participants were asked to provide comments and suggestions to improve both the instructions and the questions in the questionnaire.

The most important comments were made regarding the items with unclear wording, which were not easily understood, which could cause confusion about the question, or with possible ambiguity in the answers. The wording of these erroneous items was later modified or changed.

The psychometric properties of the proposed scale were then evaluated, along with its ability to identify theoretical concepts and constructs from the data extracted from the questionnaire. The criteria, procedures, and validation techniques for scales proposed by Mackenzie et al [79] were used to create the validation process for the scale used. The measurement model gave satisfactory results.

Recruitment

The questionnaires were distributed in Spain, both in Madrid and in towns and cities in nearby regions. The prerequisite for the sample was that the user had 4G or Wi-Fi connectivity to the internet. In total, 442 valid questionnaires were collected from the interviewees between January and February 2020.

The sampling was nonprobabilistic and convenient. Google Forms (Google LLC/Alphabet Inc.) was used to prepare the
questionnaire, which was then distributed on different social networks, especially LinkedIn (Microsoft).

The SPSS 24 statistical software (IBM) was used to calculate the frequency tables and statistics generated by the sample.

Demographic Information
The results from the questionnaires showed that 242/442 members (54.8%) were men, 195/442 (44.1%) were women, and 5/442 were others (1.1%).

Of these, 186 participants live in small populations of less than 5000 inhabitants (42.1%), which makes the sample interesting, as getting to hospitals and health centers may be difficult for them. Furthermore, 336 participants were aged between 18 and 30 years (76.0%) and 291 had studied at a university (65.9%); 64.9% (n=287) of the sample were students.

Statistical Analysis
Data analysis and hypothesis testing were carried out using structural equation modeling (SEM) with variance, which allowed for a statistical examination of the interrelated dependency relationships between the latent variables and the indicator variables of the research model by directly measuring observable variables [80].

SEM was used together with partial least squares (PLS). PLS trajectory modeling can be understood as a complete SEM method to study composite factor models by measuring constructs, estimating structural models, and performing model fitting tests [81].

The PLS–SEM statistical analysis technique, based on the structural equation model, was used, as it is especially recommended for exploratory research. It allows the modeling of latent constructs with both formative and reflective indicators to analyze the collected data [82]. In addition, PLS is appropriate for the prediction and analysis of relatively new phenomena [83]. The SmartPLS 3 software (SmartPLS GmbH) was used in this study [84].

Reinartz et al [85] investigated the conditions under which PLS–SEM should be used in research analysis, and concluded that the technique can be applied for a relatively new object of research with a model that is not fully consolidated. As these were the conditions in this research, we chose to use PLS–SEM. Besides, ours is an exploratory approach [86] for which this type of data analysis is highly recommended [87].

The PLS–SEM technique was also used because one of the aims of this research was to check whether the model was predictive. Chin and Newsted [83], Fornell and Larcker [88], and Hair et al [89] had already shown that PLS–SEM can be used for this purpose.

Fornell and Bookstein [90] state that PLS explicitly defines the latent variables, constructs, or combinations, which can easily be measured. The use of these factors is another point that justifies the use of SEM, as shown in similar studies by Sarstedt et al [80], Henseler [91], and Rigdon et al [92].

Based on the research studies by Sarstedt et al [80], Hair et al [89], and Cepeda-Carrion et al [93], the choice of the best SEM approach depends on the type of latent variables being measured, with the aforementioned studies recommending PLS for reflective or common factor constructs. The information required to analyze these factors was found from other related variables, which is another condition for which PLS–SEM is recommended [80].

Investigation and adoption of mHealth apps is a recent area of research. Because this study is exploratory, PLS–SEM is recommended.

The Harman single-factor test was used as an indicator in the subsequent common method bias test [94,95]. Using this test, no single factor was detected that could explain most of the total variance, which suggests that it is very unlikely that any selection bias exists.

Results
Measurement Model
The measurement model was tested for internal reliability, convergent validity, and discriminant validity. The internal reliability was evaluated using Cronbach α which needs a value of at least .70 for acceptable internal consistency [96]. Causality was analyzed using indicator loadings. Composite reliability was also used to investigate causality [97]. All the constructs had internal consistency, as their Cronbach α values were higher than .7 [86,88,98]. To assess convergent validity, Fornell and Larcker [88] used the average variance extracted (AVE) method and stated that an acceptable value for this factor is 0.50 or more.

The structural model was then analyzed using a bootstrapping technique configured to readjust 5000 subsamples to estimate the statistical significance of the path coefficients [99].

Table 2 shows the element loads, Cronbach α, and AVE which were found for the constructs. Cronbach α values ranged from .899 to .789, which is higher than the recommended level of .70, and therefore indicates strong internal reliability for the constructs. The composite reliability ranged between 0.930 and 0.877 and the AVE between 0.651 and 0.783, which are higher than the recommended levels. The conditions for convergent validity were therefore met. The discriminant validity was calculated with the square root of the AVE and the cross-loading matrix. For satisfactory discriminant validity, the square root of the AVE of a construct should be greater than the correlation with other constructs [88].

These researchers carried out simulation studies to demonstrate that a lack of discriminant validity is better detected by means of another technique, the heterotrait-monotrait ratio, which they had discovered earlier. All the heterotrait-monotrait ratios for each pair of factors was less than 0.90.

https://mhealth.jmir.org/2021/9/e27021
Table 2. Reliability, validity of the constructs, Fornell–Larcker criterion, and HTMT.

<table>
<thead>
<tr>
<th>Construct</th>
<th>Cronbach’s α</th>
<th>CR&lt;sup&gt;a&lt;/sup&gt;</th>
<th>AVE&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Fornell-Larcker Criterion</th>
<th>HTMT&lt;sup&gt;c&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATU</td>
<td>.861</td>
<td>0.915</td>
<td>0.783</td>
<td>0.898</td>
<td></td>
</tr>
<tr>
<td>HB</td>
<td>.899</td>
<td>0.930</td>
<td>0.768</td>
<td>0.703</td>
<td>0.867</td>
</tr>
<tr>
<td>BIU</td>
<td>.883</td>
<td>0.920</td>
<td>0.742</td>
<td>0.743</td>
<td>0.711</td>
</tr>
<tr>
<td>PEOU</td>
<td>.789</td>
<td>0.877</td>
<td>0.703</td>
<td>0.596</td>
<td>0.555</td>
</tr>
<tr>
<td>POH</td>
<td>.844</td>
<td>0.906</td>
<td>0.762</td>
<td>0.722</td>
<td>0.814</td>
</tr>
<tr>
<td>PU</td>
<td>.866</td>
<td>0.903</td>
<td>0.651</td>
<td>0.771</td>
<td>0.740</td>
</tr>
</tbody>
</table>

<sup>a</sup> CR: composite reliability.
<sup>b</sup> AVE: average variance extracted.
<sup>c</sup> HTMT: heterotrait-monotrait.
<sup>d</sup> ATU: attitude toward using.
<sup>e</sup> HB: health benefits.
<sup>f</sup> BIU: behavioral intention to use.
<sup>g</sup> PEOU: perceived ease of use.
<sup>h</sup> POH: promotion of health.
<sup>i</sup> PU: perceived usefulness.

Structural Model

In this next stage, the proposed model was analyzed in detail. The structural model was built up from the different relationships between the constructs. The hypotheses for the study were tested by analyzing the relationships between the different constructs in the model to see if they were supported [83,85,100].

The assessment of the significance of structural model is usually preceded by performing an analysis of the indicator reliability and the internal consistency reliability to prove the lack of multicollinearity. The variance inflation factor values obtained were less than 5 and ranged from 1.603 (PEOU3) to 3.496 (behavioral intention to use 3).

The variance is found from the values for the reflective indicators given by the constructs [101,102]. This was found numerically by calculating the $R^2$ values, which are a measure of the amount of variance for the construct in the model. The bootstrap method was used to test the hypotheses. The detailed results (path coefficient, $β$, and $t$ statistic) are summarized in Table 3 and Figure 2.

PEOU is positively associated with PU ($β=0.203$, $t_{0.01,499}=2.810$, $P=0.005$) and attitude toward using ($β=0.179$, $t_{0.01,499}=2.623$, $P=0.009$), and therefore, H1 and H2 were compatible with the proposed model with a 99% level of confidence.

Likewise, PU, another relationship established in the TAM, positively influenced the variable attitude toward using. This relationship was therefore confirmed and was compatible with the proposed model ($β=0.755$, $t_{0.01,499}=12.888$, $P<0.001$) with a high level of confidence (99.9%).

The TAM constructs that influence behavioral intention to use, such as PU ($β=0.415$, $t_{0.01,499}=3.442$, $P=0.001$), have a significant influence on the intention to use an mHealth app. Therefore, H4 was supported for the proposed model with a confidence level of 99.9%.

The results also indicated that the research model explains 76.4% of the variance of the intention to use an mHealth app ($R^2$ for behavioral intention to use=76.4%, $R^2$ values for attitude toward using, health benefits, PEOU, and PU are 78.2%, 74.7%, 40.9%, and 72.7%, respectively). The result of a single linear regression from attitude toward using mHealth apps and behavioral intention to use confirmed that attitude toward using is positively associated with behavioral intention to use an mHealth app ($β=0.301$, $t_{0.01,499}=2.299$, $P=0.02$). This means that H5 was supported (99%).

The hypotheses for the external variable health benefits of the original TAM were all supported with the same level of confidence (99.9%). Therefore, the health benefits variable was shown to have a significant influence on PU ($β=0.448$, $t_{0.001,499}=4.010$, $P<0.001$) and therefore H6 was supported.

Likewise, health benefits also positively influenced PEOU ($β=0.640$, $t_{0.001,499}=14.948$, $P<0.001$), which shows that H7 was supported. The other external variable (ie, promotion of health) was found to significantly influence health benefits ($β=0.865$, $t_{0.001,499}=29.943$, $P<0.001$), which means that H8 is supported with the highest values in this research model (99.9%). H7 and H8 had the highest $t$ statistic value of all the studied hypotheses (Table 3).

H9 and H10 studied the association of promotion of health with behavioral intention to use ($β=0.210$, $t_{0.05,499}=2.108$, $P=0.03$) and PU ($β=0.281$, $t_{0.01,499}=2.393$, $P=0.01$) with a 99% level of confidence. H9 had the lower $t$ statistic value of all the studied hypotheses (95%).
Table 3. Results of hypothesis: path coefficients and statistical significance (n=5000 subsamples).\textsuperscript{a}

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>$\beta$ (coefficient path)</th>
<th>$t$ statistic</th>
<th>$P$ value</th>
<th>Supported</th>
</tr>
</thead>
<tbody>
<tr>
<td>H1: Perceived ease of use $\rightarrow$ Perceived usefulness</td>
<td>.203</td>
<td>2.810</td>
<td>.005</td>
<td>Yes\textsuperscript{b}</td>
</tr>
<tr>
<td>H2: Perceived ease of use $\rightarrow$ Attitude toward using</td>
<td>.179</td>
<td>2.623</td>
<td>.009</td>
<td>Yes\textsuperscript{b}</td>
</tr>
<tr>
<td>H3: Perceived usefulness $\rightarrow$ Attitude toward using</td>
<td>.755</td>
<td>12.888</td>
<td>&lt;.001</td>
<td>Yes\textsuperscript{c}</td>
</tr>
<tr>
<td>H4: Perceived usefulness $\rightarrow$ Behavioral intention to use</td>
<td>.415</td>
<td>3.442</td>
<td>.001</td>
<td>Yes\textsuperscript{c}</td>
</tr>
<tr>
<td>H5: Attitude toward using $\rightarrow$ Behavioral intention to use</td>
<td>.301</td>
<td>2.299</td>
<td>.02</td>
<td>Yes\textsuperscript{b}</td>
</tr>
<tr>
<td>H6: Health benefits $\rightarrow$ Perceived usefulness</td>
<td>.448</td>
<td>4.010</td>
<td>&lt;.001</td>
<td>Yes\textsuperscript{c}</td>
</tr>
<tr>
<td>H7: Health benefits $\rightarrow$ Perceived ease of use</td>
<td>.640</td>
<td>14.948</td>
<td>&lt;.001</td>
<td>Yes\textsuperscript{c}</td>
</tr>
<tr>
<td>H8: Promotion of health $\rightarrow$ Health benefits</td>
<td>.865</td>
<td>29.943</td>
<td>&lt;.001</td>
<td>Yes\textsuperscript{c}</td>
</tr>
<tr>
<td>H9: Promotion of health $\rightarrow$ Behavioral intention to use</td>
<td>.210</td>
<td>2.108</td>
<td>.03</td>
<td>Yes\textsuperscript{d}</td>
</tr>
<tr>
<td>H10: Promotion Of Health $\rightarrow$ Perceived usefulness</td>
<td>.281</td>
<td>2.393</td>
<td>.01</td>
<td>Yes\textsuperscript{b}</td>
</tr>
</tbody>
</table>

\textsuperscript{a}For 5000 subsamples, we used a $t$ distribution (4999) of students in single queue.
\textsuperscript{b}$P<.01$ ($t_{0.01;499}$=2.33843952).
\textsuperscript{c}$P<.001$ ($t_{0.001;499}$=3.106644601).
\textsuperscript{d}$P<.05$ ($t_{0.05;499}$=1.64791345).

Figure 2. Analysis results (path coefficient, $\beta$, and $t$ statistic are presented). TAM: technology acceptance model.

The measurements for approximate adjustments of the model [81,91] are given by the standardized root mean square residual (SRMR) value [103], which measures the difference between the observed correlation matrix and the implied correlation matrix of the model. SRMR shows the average magnitude of these differences.

A low value of SRMR means that the fit is better. In our case SRMR=0.023, which was within the recommendations for a model with a good fit. A good fit is considered to be shown with an SRMR value of less than 0.08 [103].

Regarding the evaluation of the overall fit of the model, Benitez et al [104] recommend evaluating a saturated structural model by investigating discrepancy between empirical and model-implied indicator variance–covariance matrix. Bootstrapping results show that the SRMR sample mean for the saturated model (0.023) is below the 95% mark of its corresponding reference distribution (0.027).

The blindfolding procedure omits part of the data for a given construct during the estimation of parameters. Estimated parameters are then used to try to recreate the omitted data [101]. It is possible to study the predictive relevance of the model in this way using the Stone–Geisser ($Q^2$) test [105,106]. This test revealed that the model has predictive capability. As can be seen in Table 4, all endogenous constructs fulfill $Q^2 > 0$. Values
of 0.02, 0.15, and 0.35 for $Q^2$ in the Stone–Geisser test indicate small, medium, and great predictive relevance [107].

As per the $R^2$ (see Table 4 and Figure 2) values reported by Chin [101], we conclude the following: If $R^2 = 0.67$, the result is considered substantial; 0.33, the result is considered moderate, and 0.19, the result is considered weak. The $R^2$ obtained for the main dependent variable of the model, behavioral intention to use, was 76.4%.

This value shows that this model is “substantially” applicable for the adoption of an mHealth app. The variables that are not endogenous do not have a value for $R^2$.

The blindfolding technique consists in omitting part of the data for a given construct during the estimation of parameters, and then trying to estimate what was omitted from the estimated parameters [83].

In this way the predictive relevance of the model was studied and using the Stone–Geisser ($Q^2$) test the model was shown to have predictive capacity [105].

Therefore, all constructs, except PEOU, in the studied model have great predictive relevance, as the values of $Q^2$ are greater than 0.35 (Table 4). The proposed research model thus has good predictive power when explaining behavioral intention to use an mHealth app.

Effect size shows the strength of the relationship between 2 variables in the research model on a numeric scale. The effect size ($f^2$) shows how much an exogenous latent variable contributes to the $R^2$ value of an endogenous latent variable. The $f^2$ values 0.02, 0.15, and 0.35 indicate small, medium, and large effect size [100]. Cohen's tables [107] showed that for 95.2% statistical power and an average effect size of $f^2=0.15$, a minimum of 107 questionnaires would be needed. In our case the number of samples was 442, showing that this research has adequate statistical power.

Table 4. $R^2$ and $Q^2$ results.

<table>
<thead>
<tr>
<th>Construct</th>
<th>$Q^2$</th>
<th>$R^2$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attitude toward using</td>
<td>0.478</td>
<td>78.2</td>
</tr>
<tr>
<td>Health benefits</td>
<td>0.465</td>
<td>74.7</td>
</tr>
<tr>
<td>Behavioral intention to use</td>
<td>0.491</td>
<td>76.4</td>
</tr>
<tr>
<td>Perceived ease to use</td>
<td>0.229</td>
<td>40.9</td>
</tr>
<tr>
<td>Promotion of health</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Perceived usefulness</td>
<td>0.381</td>
<td>72.7</td>
</tr>
</tbody>
</table>

$^aN/A$: not applicable.

PLS–SEM Results With Moderator (Gender and Age)

In order to check H11 and measure the potential moderating influence of gender and age, we performed a multigroup analysis [108].

First, the sample was divided by gender into men and women. The following process was then repeated, dividing members of the sample into old and young people.

However, before doing this test it is necessary to analyze the measurement invariance of the composite models (MICOM) technique [80]. This test will ensure that the effect of gender is restricted to the trajectory coefficients of the structural model and not to the parameters of the measurement model [109]. As described in Tables 5 and 6, we find the invariance of the measurement in the case of gender, but not in the case of age (Table 6) for the variables attitude toward using, health benefits, behavioral intention to use, perceived ease to use (PEOU), promotion of health, and PU.
### Table 5. Results of the measurement invariance of composite models (MICOM) procedure (gender).

<table>
<thead>
<tr>
<th>Construct</th>
<th>Step 1</th>
<th>Step 2</th>
<th>Step 3a</th>
<th>Step 3b</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Configural invariance</td>
<td>Compositional invariance</td>
<td>Equal variances</td>
<td>Mean original difference (men–women)</td>
</tr>
<tr>
<td></td>
<td>Original correlation 5%</td>
<td>Partial measurement invariance established</td>
<td>Variance original difference (men–women) 2.5% 97.5% Equal</td>
<td>2.5% 97.5% Equal</td>
</tr>
<tr>
<td>ATU(^a)</td>
<td>Yes 1.000 1.000</td>
<td>Yes 0.209 −0.182 0.182 No −0.070 −0.308 0.303 Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BIU(^b)</td>
<td>Yes 1.000 1.000</td>
<td>Yes −0.011 −0.211 0.176 Yes 0.234 −0.278 0.265 Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HB(^c)</td>
<td>Yes 1.000 1.000</td>
<td>Yes 0.061 −0.197 0.185 Yes 0.146 −0.277 0.282 Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PEOU(^d)</td>
<td>Yes 1.000 0.998</td>
<td>Yes −0.054 −0.202 0.187 Yes 0.095 −0.267 0.270 Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>POH(^e)</td>
<td>Yes 1.000 1.000</td>
<td>Yes 0.092 −0.189 0.167 Yes 0.242 −0.288 0.279 Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PU(^f)</td>
<td>Yes 0.999 0.999</td>
<td>No 0.152 −0.198 0.174 Yes −0.123 −0.287 0.265 Yes</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^a\)ATU: attitude toward using.
\(^b\)BIU: behavioral intention to use.
\(^c\)HB: health benefits.
\(^d\)PEOU: perceived ease of use.
\(^e\)POH: promotion of health.
\(^f\)PU: perceived usefulness.

### Table 6. Results of the measurement invariance of composite models (MICOM) procedure (age).

<table>
<thead>
<tr>
<th>Construct</th>
<th>Step 1</th>
<th>Step 2</th>
<th>Step 3a</th>
<th>Step 3b</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Configural invariance</td>
<td>Compositional invariance</td>
<td>Equal variances</td>
<td>Mean original difference (young people–old people)</td>
</tr>
<tr>
<td></td>
<td>Original correlation 5%</td>
<td>Partial measurement invariance established</td>
<td>Variance original difference (young people–old people) 2.5% 97.5% Equal</td>
<td>2.5% 97.5% Equal</td>
</tr>
<tr>
<td>ATU(^a)</td>
<td>Yes 1.000 0.999</td>
<td>Yes −0.399 −0.298 0.307 No −0.665 −0.398 0.537 No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BIU(^b)</td>
<td>Yes 1.000 0.999</td>
<td>Yes −0.520 −0.298 0.298 No −0.590 −0.423 0.538 No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HB(^c)</td>
<td>Yes 1.000 0.998</td>
<td>Yes −0.461 −0.300 0.304 No −0.624 −0.410 0.538 No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PEOU(^d)</td>
<td>Yes 0.999 0.993</td>
<td>Yes −0.324 −0.315 0.298 No −0.432 −0.400 0.512 Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>POH(^e)</td>
<td>Yes 1.000 0.998</td>
<td>Yes −0.470 −0.299 0.298 No −0.239 −0.419 0.539 Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PU(^f)</td>
<td>Yes 1.000 0.997</td>
<td>Yes −0.606 −0.312 0.293 No −0.360 −0.401 0.500 Yes</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^a\)ATU: attitude toward using.
\(^b\)BIU: behavioral intention to use.
\(^c\)HB: health benefits.
\(^d\)PEOU: perceived ease of use.
\(^e\)POH: promotion of health.
\(^f\)PU: perceived usefulness.
Discussion

Principal Findings

The results of this study confirmed that the variable that has the strongest impact on the behavioral intention to use of mHealth apps in Spain is PU. This variable also has a very high predictive capacity as its determination coefficient is high \( [81, 108] \). The next most important variable in the model is health benefits.

The results of this research could be applicable to other EU countries with similar levels of internet access. However, it must be taken into account that most of the participants lived in areas with less than 5000 inhabitants (186/442 participants, 42.1%), where acceptance of mHealth apps is also determined by the close social environment. In this type of environment, users of mHealth apps can offer an effective short-term consultation for families and acquaintances before they make a decision to visit hospitals or health clinics.

Comparison With Prior Work

These findings are consistent with previous studies on PU for the acceptance of medical information systems \([57, 69, 101]\). These studies also found that PU significantly influences the adoption of medical information systems.

Promotion of health was also found to have a significant effect on health benefits of using mHealth apps in this study, as mHealth apps positively promote and improve the health of mHealth app users in Spain. This relationship was the strongest among all the relationships studied in this research and shows the usefulness of mHealth apps for improving health.

This is important when promoting the idea of preventing diseases and other ailments with mHealth apps, such as controlling continued physical exercise, consumption of certain foods, monitoring the evolution of potential and current patients, and using smartphones or tablet PCs to help prevent health problems. These results are consistent with the findings from a previous study \([110]\).

H8 has been revealed as the relationship with the greatest burden and confirms the extraordinary influence it has on health benefits (β=.865, \( t_{0.001,499}=29.943, P<.001 \)). This means that eHealth apps that take care of nutrition, improve sports activity, or make mealtimes more respectful are perceived by respondents as favoring aspects related to blood pressure, weight loss, blood sugar levels, or mood. In other words, users consider that apps related to healthy habits should be developed. This means that H8 is the most reliable and significant relationship among all.

The second hypothesis with the greatest burden and influence was H7. The relationship between health habits and PEOU of apps indicates that the more beneficial the eHealth app is, the easier it should be to use. Furthermore, the third hypothesis with the greatest intensity is H3, which shows that the perception of usefulness of an eHealth app has an extraordinary influence on the attitude of use. This means that the selling strategy of these apps must be aimed at transferring 2 very important aspects to the user: on the one hand, the usefulness of eHealth as an attitude transforming agent, and on the other, the more health benefits are obtained, the easier it is to use. In addition, these 3 relationships (ie, H3, H7, and H8) were very significant (99.9%).

The TAM is applicable to the use of eHealth apps as was the case with other studies, but with the influence of the “health promotion” and “health benefits” constructs. In addition, health promotion is directly related to the main dependent variable in the behavioral intention to use model. Therefore, health promotion is a construct that should be considered in future research, as it is also directly related to the final construct of the behavioral intention to use as well as indirectly to the PU.

In this study it was demonstrated that mHealth apps were easy to use and that users were familiar with the basic functions and applications of the internet. This is justified by the fact that health benefits had a very significant influence on the perceived usability (PEOU). This is an important point to highlight when explaining mHealth apps, as this can help ensure that mHealth apps are used as often as necessary to achieve effective results. However, the influence of PEOU on PU is the relationship with the lowest load among all (β=.179, \( t_{0.01,499}=2.623, P=.009 \)) and a 99% confidence level. Likewise, the PEOU has a moderate variance (\( R^2=40.9% \)), which is why a moderately atypical result was obtained in this research. PEOU has a positive relationship with PU, which suggests that users will not need to learn new skills to use mHealth apps. The sample in this study, however, did not consider it an important factor in this model. In all probability, the advancement of usability of smartphone interfaces reduces the influence of PEOU, so people might need to use smartphones to be able to use these types of apps \([111]\). These results could be explained by the fact that the Spanish population is already familiar with health promotion and also that current mHealth apps are easy to use and accessible.

The remaining endogenous variables had a very high explanatory capacity (>70%). This gives the model a great capacity to explain the reality of the users’ behavior before using eHealth apps, as in the case of behavioral intention to use it was 76.4%.

The results obtained for the relationship between the PEOU and the attitude toward use predict a smooth learning curve. This suggests that the adoption of mHealth app will be permanent and stable in the future. The use of mHealth apps will not present any significant difficulties that may cause users to abandon it.

Our study also confirmed that health promotion has a positive influence on behavioral intention to use and perception of usefulness (PU). In both cases, the level of trust is high, which shows that health promotion is an important factor in this model. Health promotion was also found to have an indirect influence on health benefits. This result supports the previously reported finding that app titles influence behavioral intention to use \([112]\). Specifically, we found that apps with titles related to symptoms have a significantly lower number of installs as compared with those whose titles are not related to symptoms.

Finally, a moderating capacity was found with a 95% confidence level regarding gender. We found that the 2 relationships with the lowest level of confidence in the model (Table 7), H1 or the relationship of the perception of ease of use with PU (β=.422, \( P=.015 \)) and H9 or the relationship of health promotion with...
behavioral intention of use ($\beta = -0.239, P = .04$), show significant differences between men and women. Furthermore, gender-moderated behaviors were found in H10, indicating that health promotion also influences the perception of usefulness differently according to gender ($\beta = 0.178, P = .01$).

### Table 7. PLS$^a$–SEM$^b$ results with moderator (gender).

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>$\beta$ (Coefficient path)</th>
<th>$P$ value</th>
<th>Support</th>
</tr>
</thead>
<tbody>
<tr>
<td>H1: Perceived ease of use $\rightarrow$ Perceived usefulness</td>
<td>$-0.422$</td>
<td>$0.01$</td>
<td>Yes$^c$</td>
</tr>
<tr>
<td>H2: Perceived ease of use $\rightarrow$ Attitude toward using</td>
<td>$-0.100$</td>
<td>$0.24$</td>
<td>No</td>
</tr>
<tr>
<td>H3: Perceived usefulness $\rightarrow$ Attitude toward using</td>
<td>$0.318$</td>
<td>$0.18$</td>
<td>No</td>
</tr>
<tr>
<td>H4: Perceived usefulness $\rightarrow$ Behavioral intention to use</td>
<td>$0.166$</td>
<td>$0.21$</td>
<td>No</td>
</tr>
<tr>
<td>H5: Attitude toward using $\rightarrow$ Behavioral intention to use</td>
<td>$-0.107$</td>
<td>$0.49$</td>
<td>No</td>
</tr>
<tr>
<td>H6: Health benefits $\rightarrow$ Perceived usefulness</td>
<td>$0.266$</td>
<td>$0.11$</td>
<td>No</td>
</tr>
<tr>
<td>H7: Health benefits $\rightarrow$ Perceived ease of use</td>
<td>$0.003$</td>
<td>$0.94$</td>
<td>No</td>
</tr>
<tr>
<td>H8: Promotion of health $\rightarrow$ Health benefits</td>
<td>$-0.318$</td>
<td>$0.22$</td>
<td>No</td>
</tr>
<tr>
<td>H9: Promotion of health $\rightarrow$ Behavioral intention to use</td>
<td>$-0.239$</td>
<td>$0.04$</td>
<td>Yes$^c$</td>
</tr>
<tr>
<td>H10: Promotion of health $\rightarrow$ Perceived usefulness</td>
<td>$0.178$</td>
<td>$0.01$</td>
<td>Yes$^c$</td>
</tr>
</tbody>
</table>

$^a$PLS: partial least squares.

$^b$SEM: structural equation modeling.

$^c$For 500 subsamples, we used a $t$ distribution ($4999$) of students in a single queue: $P < 0.05$ ($t_{0.05,4999} = 1.64791345$).

The other moderating variable (ie, age) was not supported, coinciding with the results of similar studies [113]. Therefore, mHealth app is an effective way to promote good health and habits in the population. Participants in the study believed that mHealth apps could help them improve their health, maintain a meal schedule, take part in more sporting activities, or improve the hours slept at night. Thus, mHealth apps can promote healthy habits and improve the users’ quality of life.

### Conclusions

#### Theoretical Implications

As has often been addressed in previous mHealth studies [114,115], health apps on smartphones can serve as very realistic health care alternatives, helping people save on medical expenses and being more effective in managing their personal health. Therefore, we agree with a previous work [20] that the potential advantages of using health apps (mHealth) in terms of improving overall health can be harmed without the use of apps.

The extended TAM adoption model was found to be fully valid for the study of mHealth app use and acceptance in Spain. This result could be extrapolated to other EU countries with similar levels of internet accessibility and sociodemographic characteristics.

This study identified the variables that influence people’s intention to use mHealth apps. Using an extended TAM, PU was found to be the most significant variable influencing adoption of mHealth apps in Spain. This means that the most important factor for users are the ways in which mHealth apps can help them. This result is important because users of this type of apps must first understand the utility of the use of these apps, so that they can become cognizant about how they can improve treatment of their diseases and their control.

#### Practical Implications

Other external variables, such as promotion of health, have a significant effect on the health benefits of mHealth app use. This result showed that users consider maintenance or improvement of health as an additional health benefit provided by these apps.

The predictive capacity of the model and the predictive capacity can be very useful in preventing diseases that need controlled habits. Examples are indulging in regular physical exercise; consumption of certain foods; monitoring the evolution of current and potential patients; and using smartphones, tablets, and other medical devices to prevent health problems. Besides, care centers should have Wi-Fi access so that patients can carry out real-time diagnostic tests.

The results of this research show that gender is neither completely decisive nor moderating in the behavioral intention to use mHealth apps. This means that adoption of mHealth apps for promotion of health was moderated only by gender. Another important factor influencing mHealth app use is PEOU.

Therefore, user-friendliness and health promotion should be gender sensitive when applying utilities to apps. Accordingly, app developers should take into account users’ gender and introduce some changes in usage and health promotion levels.

The results obtained using the extended TAM show that promotion of health and health benefits are important variables for mHealth apps users because they indirectly influence the adoption of the technology. This means that mHealth apps could be an alternative way to promote and improve health and could
become a service that minimizes primary care consultations for simple cases.

This is because PU and PEOU are not the only mediators for the final intention to use. Promotion of health is directly related to behavioral intention to use. This was a highly significant relationship and means that users prefer mHealth apps that promote health. This recommendation is important for designers, developers, and start-ups creating new mHealth apps. Therefore, we could start thinking that barriers such as standards, security, and interoperability [116] could be overcome by the activities derived from promotion of health.

The significance of the association between PU and behavioral intention to use explains the importance of mHealth apps for the users. This could explain the evolution of mHealth apps that offer an increasing number of benefits to the user.

An example is that the users’ health information can now be transmitted online. This could help health centers have real-time information and minimize visits to health centers for primary care. To increase the adoption and use of mHealth apps, there should be an approved catalog of health service providers and an adoption strategy for citizens.

Based on our study results, the authorities could take the following as indicators for the use of mHealth apps: connectivity of the mHealth app, interaction between the patient and the health professional via the app, the need to prescribe additional quality hardware that allows measurements and analyses, and the personalized and nonautomated accessibility of these apps to the use and analysis of patient data remotely. These tools could be key indicators to measure the quality of this type of apps by health authorities.

In conclusion, gender is a determining factor that influences the intention to use eHealth apps, and therefore, different interfaces and utilities could be designed according to gender.

The findings of this study are beneficial for organizations, governments, and policymakers to provide strategies and policies to improve mHealth app in different hospitals and Spanish primary health care centers.

Limitations
The limitations of the research are those related to the analysis technique used, the country under study, and the size of the sample.

Conflicts of Interest
None declared.

References


URL: https://tinyurl.com/ndxhelps [accessed 2021-06-01]


73. Ramtohul I. The adoption of e-health services: Comprehensive analysis of the adoption setting from the user s perspective. Health Policy and Technology 2015 Sep;4(3):286-293. [doi: 10.1016/j.hpt.2015.04.007]
76. Shabani J, Damavandi AJ. The Importance of Gender as a Moderator for the Relationship between Emotional Intelligence and Mental Health of Adolescents. ASS 2011 Aug 31;7(9):142-148. [doi: 10.5539/ass.v7n9p142]


Abbreviations

- **PLS**: partial least squares
- **SEM**: structural equation modeling
- **TAM**: technology acceptance model
Nursing Interns’ Attitudes Toward, Preferences for, and Use of Diabetes Virtual Simulation Teaching Applications in China: National Web-Based Survey

Fang Liu¹ ², RN, MD; Huiting Weng¹ ², RN, MNurs; Rong Xu³, RN, MNurs; Xia Li¹, MD; Zhe Zhang¹ ², RN; Kuaile Zhao¹ ², RN; Zhiguang Zhou¹, MD; Qin Wang² ³, RN, MNurs

¹National Clinical Research Center for Metabolic Diseases, Department of Metabolism and Endocrinology, The Second Xiangya Hospital, Central South University, Changsha, China
²Clinical Nursing Teaching and Research Section, The Second Xiangya Hospital, Central South University, Changsha, China
³Department of Ophthalmology, The Second Xiangya Hospital, Central South University, Changsha, China

Corresponding Author:
Qin Wang, RN, MNurs
Clinical Nursing Teaching and Research Section
The Second Xiangya Hospital
Central South University
139 Middle Renmin Road
Changsha, 410011
China
Phone: 86 18774806226
Email: wangqin3421@csu.edu.cn

Abstract

**Background:** Diabetes has placed heavy social and economic burdens on society and families worldwide. Insufficient knowledge and training of frontline medical staff, such as nurses, interns, and residents, may lead to an increase in acute and chronic complications among patients with diabetes. However, interns have insufficient knowledge about diabetes management. The factors that affect interns’ current level of diabetes-related knowledge are still unclear. Therefore, understanding the behavioral intentions of interns is essential to supporting the development and promotion of the use of virtual simulation teaching applications.

**Objective:** This study aimed to identify the determinants of nursing interns’ intentions to use simulation-based education applications.

**Methods:** From December 1, 2020, to February 28, 2021, the web-based survey tool Sojump (Changsha Xingxin Information Technology Co) was used to survey nursing interns in hospitals across China. Two survey links were sent to 37 partner schools in 23 major cities in China, and they were disseminated through participants’ WeChat networks. Multiple regression analysis was used to determine the association between demographic information and basic disease information and the use of the application for treating adult patients.

**Results:** Overall, 883 nursing interns from 23 provinces in China responded to the survey. Among them, the virtual simulation utilization rate was 35.6% (314/883) and the awareness rate was 10.2% (90/883). In addition, among the interns, only 10.2% (90/883) correctly understood the concept of virtual simulation, and most of them (793/883, 89.8%) believed that scenario-simulation training or the use of models for teaching are all the same. Multiple regression analysis showed that the educational level, independent learning ability, and professional identity of the interns were related to use of the application \((P<.05)\). Skills and knowledge that the interns most wanted to acquire included the treatment of hypoglycemia (626/883, 70.9%), functional test simulation (610/883, 69.1%), and blood glucose monitoring technology (485/883, 54.9%). A total of 60.5% (534/883) of the interns wanted to acquire clinical thinking skills, while 16.0% (141/883) wanted to acquire operational skills. Nursing trainees believed that the greatest obstacles to virtual simulation included limited time (280/883, 31.7%), the degree of satisfaction (129/883, 14.6%), the demand for satisfaction (108/883, 12.2%), and test scores (66/883, 7.5%).

**Conclusions:** The understanding and usage rate of diabetes virtual simulation teaching applications by Chinese nursing interns is very low. However, they have high requirements regarding this teaching method. Conducting high-quality randomized controlled trials and designing applications that are suitable for the needs of different nurse trainees will increase students’ interest in learning and help improve diabetes knowledge among nursing interns.
Introduction

Background

With the rapid development of the social economy, continuous changes in modern people’s behaviors and lifestyles, and the aging of the population, the incidence of diabetes is also increasing rapidly in all parts of the world. According to forecasts, from 1995 to 2030, the number of patients with diabetes worldwide will increase from 135 million to 472 million, among which more than 75% are in developing countries [1]. According to the World Health Organization [2], as of 2015, the prevalence of diabetes in China had reached 10%, and the estimated prevalence of diabetes in adults over 18 years old in China was 11.6%. The number of adult patients with diabetes in China has reached 92.4 million [3]; China is now the country with the largest number of patients with diabetes.

Effective diabetes education for patients is indispensable; it is a necessary means to ensure that patients receive effective therapies. Nurses are the most important providers of diabetes education in China. Due to China’s national conditions, there are no specialized diabetes educators that provide health guidance and dietary education to patients with diabetes. Such work is often undertaken by clinical staff (doctors, nurses, interns, etc.). Nurses and interns have the most contact with patients with diabetes and are most likely to provide patients with diabetes-related knowledge [4]. In nursing programs, Chinese students often need to go to designated hospitals for 8- to 12-month internships in their senior year. During these internships, they learn basic knowledge about diseases, professional skills, and communication skills at the hospital. At the end of the internship, they need to pass a unified examination jointly organized by the hospital and the school before they can graduate. The clinical internship is a critical period for nursing students to transition from student to nurse. At this stage, interns often cannot perform nursing activities alone, and all of these require the demonstration and guidance of clinical teachers. In most clinical teaching sessions, the teacher and the interns are apprentices. In addition to regular and unified theoretical training and operational training in the nursing department, teachers often use a one-to-one teaching mode. The importance of nursing interns in disease prevention is enormous, as they are the members of health care teams who spend the most time with the patients [5]. They also serve as resources for patients with diabetes seeking information about the early detection of diabetes complications [6]. The knowledge and practice they acquire during their studies and internships play important roles in providing accurate and up-to-date information to improve the health behaviors and outcomes of patients with diabetes. Nursing interns must possess the necessary knowledge to enable them to care for patients with diabetes, helping them to achieve a high quality of life devoid of complications [7]. In addition, an intern can detect hypoglycemia for the first time when measuring a patient’s blood sugar. If nursing interns know how to deal with hypoglycemic events, they can instruct patients to eat the correct glucose-increasing food immediately, thereby reducing the time interval from discovery to treatment of hypoglycemia and reducing the occurrence of adverse events [8].

The cultivation of self-learning ability by interns is inseparable from the application of self-learning methods and tools [9]. With the goal of improving the self-learning ability of nursing students, many scholars at home and abroad have carried out various studies in this field. Some studies have shown that project-based learning methods, preceptorship programs, and reflective diaries have improved students’ abilities for critical thinking, clinical decision making, and humanistic care [10-13]. However, these methods focus on cultivating students’ information-seeking and cooperation abilities in order to enhance the autonomous learning ability of interns, and the effect is not lasting. Without the supervision of teachers, the internal motivation of students to learn independently is still insufficient [14].

Interns’ apprenticeships have always constituted a challenge faced by the government, health educators, health managers, and the students themselves to ensure the quality and safety of learning and clinical practice [15]. Students in the 21st century are using information and communications technology (ICT) every day [16,17]. The use of ICT has led to different learning processes and information structure processes. The development of digital and virtual technology has simplified the ability to reconstruct reality using virtual patients depicted on a computer touch screen (ie, virtual simulation) [18].

A virtual simulation is a real-life reproduction depicted on a computer screen, and it involves a real person operating the simulation system. This type of simulation puts people at the center of a situation by exercising decision making, motor control, and communication skills [19]. Virtual simulation uses virtual patients in dynamic and immersive clinical environments, ranging from prehospital to community environments [20]. The latest technological advances in virtual simulation have improved their authenticity and dynamic interaction, and it is possible to display thousands of clinical situations on a touch screen or on the web [21,22]. However, little is known about their effect on students’ learning satisfaction, self-efficacy, knowledge retention, and clinical reasoning, especially when using the latest developments in virtual simulation [21].

This study aimed to assess the knowledge needs of nursing students for managing diabetes mellitus. By evaluating the self-learning ability of nursing students and the degree of demand for diabetes-related knowledge, the demand for virtual simulation teaching applications for nursing students was explored. Therefore, the purpose of this study was to evaluate the level of understanding of diabetes specialist knowledge and...
the demand for virtual simulation teaching among nursing students in China.

Objectives
We aimed to investigate the use of virtual simulation teaching applications by nursing interns as well as their perspectives, attitudes, and associated factors regarding these teaching applications. We also aimed to investigate interns’ needs for these applications in order to provide information for the design of virtual simulation teaching applications and to learn how best to promote their use, which will help teachers to further improve their teaching methods and strengthen the willingness of nursing students to learn independently.

Methods
Questionnaire Design
An expert group consisting of five nursing educators and five clinical nursing staff members searched for applications on the national, virtual simulation, education platform; they then designed a questionnaire based on the current diabetes guidelines and the problems encountered in clinical practice. These questions were presented in a selective format. If the respondent disagreed with the listed options, they could select “other options” and write their answer in the “remarks” column. The questionnaire collected information about respondents’ demographics and their views, attitudes, and needs for virtual simulation education applications.

To determine the validity of the questionnaire content, a total of 15 experts, consisting of 12 nursing education experts and three diabetes education nurses with at least 5 years of experience, rated the relevance and clarity of the items on a 4-point scale ranging from 1 (irrelevant) to 4 (highly relevant), with a content validity index of 0.91. Before administering the questionnaire survey, we conducted a pilot test on 18 interns at Xiangya Second Hospital in China. The Cronbach α value of the questionnaire was .83.

Survey Platform and Methods
WeChat has become one of the largest mobile traffic platforms in China. It provides many services, including messaging, free phone calls, browsing and publishing for instant sharing of information, and mobile payments [23]. It has been installed on more than 90% of mobile phones and has become part of the daily lives of most people [24]. As of 2019, the number of monthly active accounts on WeChat reached 1.15 billion, and the number of daily active accounts of mini programs exceeded 300 million [25]. As the most commonly used social media tool in China, WeChat has an expansive network of contacts. The network makes it possible for administrators to manage questionnaires through WeChat.

From December 1, 2020, to February 28, 2021, we used Sojump (Changsha Xingxin Information Technology Co), a web-based survey tool, to conduct snowball sampling through the WeChat contact network and to conduct convenience sampling through WeChat public accounts to recruit interns. The survey link was initially sent to 35 universities in 23 representative major cities in China. We asked the teachers at these universities to post the survey link on their WeChat account to reach their network contacts.

Survey respondents were all nursing trainees in China. Other nursing students who did not take part in internships at hospitals were excluded from our survey. Before administering the survey, we introduced the purpose of the survey, and the questionnaire was filled out by respondents voluntarily without any compensation.

Ethical Approval
This study was approved by the ethics committee of the Second Xiangya Hospital, Central South University, China (ID: 2020-S790).

Statistical Data
The data were analyzed using SPSS, version 23.0 (IBM Corp). Quantile-quantile (Q-Q) charts were used to check the normality of all continuous variables and express them as the mean (SD) or median (IQR) where appropriate. Categorical variables were expressed as frequencies and percentages. The chi-square test was used to assess the differences between groups. The generalized logic model was used to obtain the odds ratio (OR) and its 95% CI at the same time. First, we conducted a univariate analysis to analyze the OR of the potential correlation between demographic factors and autonomous learning ability. Then, we inputted all important factors into the multivariate analysis to obtain the multivariate adjusted OR. Questionnaires with missing values were excluded from the multivariate analysis. Statistical significance was defined as \( P < .05 \).

Results
Sample Characteristics
A total of 883 interns distributed among 26 provinces in China (Figure 1) responded to the patient survey. The respondents’ characteristics are shown in Table 1. Among the respondents, 10.1% (89/883) were male, and respondents had a mean age of 20.64 (SD 2.1) years. Overall, 56.9% (502/883) had a bachelor’s degree. A total of 83.0% (733/883) of the respondents had been an intern for more than 8 months, and 46.5% (411/883) did not know their reason for choosing to study nursing (Table 1).
Figure 1. Distribution of the nursing intern sample in China by province. The numbers represent how many questionnaires were collected in each corresponding province.
Table 1. Characteristics of nursing interns.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Respondents (N=883)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gender, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>89 (10.1)</td>
</tr>
<tr>
<td>Female</td>
<td>794 (89.9)</td>
</tr>
<tr>
<td><strong>Age (years), mean (SD)</strong></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>20.64 (2.1)</td>
</tr>
<tr>
<td><strong>Educational level, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Middle school</td>
<td>26 (2.9)</td>
</tr>
<tr>
<td>High school</td>
<td>102 (11.6)</td>
</tr>
<tr>
<td>Technical college</td>
<td>502 (56.9)</td>
</tr>
<tr>
<td>Bachelor’s degree</td>
<td>242 (27.4)</td>
</tr>
<tr>
<td>Master’s degree or higher</td>
<td>11 (1.2)</td>
</tr>
<tr>
<td><strong>Internship time (months), mean (SD)</strong></td>
<td></td>
</tr>
<tr>
<td>Internship</td>
<td>6.02 (1.6)</td>
</tr>
<tr>
<td><strong>Reason for choosing nursing, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>I like nursing</td>
<td>287 (32.5)</td>
</tr>
<tr>
<td>Parents’ suggestion</td>
<td>247 (28.0)</td>
</tr>
<tr>
<td>Acquaintances’ recommendation</td>
<td>55 (6.2)</td>
</tr>
<tr>
<td>The school transferred me</td>
<td>99 (11.2)</td>
</tr>
<tr>
<td>Good employment</td>
<td>195 (22.1)</td>
</tr>
<tr>
<td><strong>Feelings about nursing, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>I love the nursing career</td>
<td>410 (46.4)</td>
</tr>
<tr>
<td>Not sure</td>
<td>197 (22.3)</td>
</tr>
<tr>
<td>I can accept as a job, but not as a career</td>
<td>260 (29.4)</td>
</tr>
<tr>
<td>I don’t like nursing</td>
<td>16 (1.8)</td>
</tr>
<tr>
<td><strong>Employment intention, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Nurse</td>
<td>747 (84.6)</td>
</tr>
<tr>
<td>Nursing-related industries</td>
<td>115 (13.0)</td>
</tr>
<tr>
<td>Others</td>
<td>21 (2.4)</td>
</tr>
</tbody>
</table>

**Assessment of the Self-Learning Ability of Interns**

All of the interns (N=883) were able to fill out the self-learning ability scale. The Q-Q normality was the sum of the total scores of the self-learning ability of interns (Figures 2 and 3). The total score is represented by the diagonal line, so it is considered that the total score of the autonomous learning ability of nursing students conforms to the normal distribution. The data were analyzed using the Pearson correlation coefficient. Age, gender, educational level, and length of internship were not related to the self-learning ability of interns ($P>.05$). The correlation coefficient between the “reason for choosing nursing” and the “autonomous learning ability scale score” was 0.993; the correlation between them was statistically significant ($P<.001$). This correlation was also reflected with “feelings about nursing” ($P=.02$), which showed that interns who love nursing had stronger self-learning ability. In addition, the correlation between “feelings about nursing” and the “score of the learning strategy scale” was statistically significant ($P=.001$). This indicates that the more positive feelings the nursing student interns had toward the nursing profession, the higher their scores were on the learning strategy scale (Table 2).
Figure 2. The normal quantile-quantile (Q-Q) chart for the score of autonomous learning ability.
Table 2. Correlation analysis of self-learning ability of interns (N=883).

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Autonomous learning ability</th>
<th>Score of learning strategy scale</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>r</td>
<td>P value</td>
</tr>
<tr>
<td>Age</td>
<td>-0.020</td>
<td>.56</td>
</tr>
<tr>
<td>Gender</td>
<td>-0.101</td>
<td>.10</td>
</tr>
<tr>
<td>Educational level</td>
<td>0.008</td>
<td>.81</td>
</tr>
<tr>
<td>Internship time</td>
<td>0.410</td>
<td>.22</td>
</tr>
<tr>
<td>Reason for choosing nursing</td>
<td>0.993</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Feelings about nursing</td>
<td>0.595</td>
<td>.02</td>
</tr>
<tr>
<td>Employment intention</td>
<td>0.011</td>
<td>.75</td>
</tr>
</tbody>
</table>

Interns’ Needs and Expectations of Diabetes Virtual Simulation Applications

Nursing trainees believed that important functions of a diabetes virtual simulation application are to help them treat patients with hypoglycemia and the simulation of functional tests. Almost all respondents believed the listed functions were important or very important. However, most interns believed that oral administration, venofusion, and intramuscular injection were important (Figure 4). When comparing teaching methods with the expectations of nurse interns, PowerPoint presentations (222/883, 25.1%) and face-to-face teaching (219/883, 24.8%) were the most-used teaching methods, while students expected to use more virtual simulations (204/883, 23.1%) and to reduce the use of PowerPoint presentations (148/883, 16.8%) (Figure 5).
Figure 4. Importance of different simulation scenes on a diabetes virtual simulation application as reported by interns.

Figure 5. Comparison of teaching methods with interns’ expectations. PPT: PowerPoint.

In this study, out of 883 interns, 569 (64.4%) had never participated in virtual simulation teaching and 793 (89.8%) had not heard of the concept of virtual simulation before this survey. Table 3 shows that through the virtual simulation application, what interns most want to improve is their clinical thinking ability (534/883, 60.5%), followed by their comprehension ability (156/883, 17.7%).
Table 3. Nurse interns’ usage and preferences of a diabetes virtual simulation application.

<table>
<thead>
<tr>
<th>Question</th>
<th>Respondents (N=883), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Have you participated in virtual simulation teaching?</strong></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>569 (64.4)</td>
</tr>
<tr>
<td>Yes</td>
<td>231 (26.2)</td>
</tr>
<tr>
<td><strong>Have you participated in similar activities? (yes)</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td>83 (9.4)</td>
</tr>
<tr>
<td><strong>Do you know about virtual simulation teaching?</strong></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>793 (89.8)</td>
</tr>
<tr>
<td>Yes</td>
<td>90 (10.2)</td>
</tr>
<tr>
<td><strong>What do you think of virtual simulation teaching?</strong></td>
<td></td>
</tr>
<tr>
<td>Very good</td>
<td>138 (15.6)</td>
</tr>
<tr>
<td>Good</td>
<td>251 (28.4)</td>
</tr>
<tr>
<td>Neutral</td>
<td>76 (8.6)</td>
</tr>
<tr>
<td>Bad</td>
<td>7 (0.8)</td>
</tr>
<tr>
<td>Very bad</td>
<td>2 (0.2)</td>
</tr>
<tr>
<td>Do not know</td>
<td>409 (46.3)</td>
</tr>
<tr>
<td><strong>What is your acceptance level of virtual simulation teaching?</strong></td>
<td></td>
</tr>
<tr>
<td>Very good</td>
<td>354 (40.1)</td>
</tr>
<tr>
<td>Good</td>
<td>393 (44.5)</td>
</tr>
<tr>
<td>Neutral</td>
<td>126 (14.3)</td>
</tr>
<tr>
<td>Bad</td>
<td>7 (0.8)</td>
</tr>
<tr>
<td>Very bad</td>
<td>3 (0.3)</td>
</tr>
<tr>
<td><strong>Which ability do you most want to improve in virtual simulation teaching?</strong></td>
<td></td>
</tr>
<tr>
<td>Comprehension skills</td>
<td>156 (17.7)</td>
</tr>
<tr>
<td>Analytical skills</td>
<td>141 (16.0)</td>
</tr>
<tr>
<td>Judgment skills</td>
<td>48 (5.4)</td>
</tr>
<tr>
<td>Clinical thinking ability</td>
<td>534 (60.5)</td>
</tr>
<tr>
<td>Others</td>
<td>4 (0.5)</td>
</tr>
<tr>
<td><strong>What do you think is appropriate for the average duration of each session? (minutes)</strong></td>
<td></td>
</tr>
<tr>
<td>0-10</td>
<td>209 (23.7)</td>
</tr>
<tr>
<td>11-30</td>
<td>483 (54.7)</td>
</tr>
<tr>
<td>31-60</td>
<td>169 (19.1)</td>
</tr>
<tr>
<td>61-90</td>
<td>22 (2.5)</td>
</tr>
</tbody>
</table>

**Discussion**

**Principal Findings**

**The Use of a Virtual Simulation Teaching Application and its Influencing Factors Among Interns**

Among the interns, 26.2% (231/883) had participated in virtual simulation education, and 9.4% (83/883) had participated in similar activities. These rates are comparable to results from surveys conducted in New York [17] and Florida [16], and higher than the rate (7%) found in a 2011 survey in Canada [18]. In China, more nursing interns in Southern China (87.3%) participated in virtual simulation teaching than in Northern China (12.7%). One possible reason is that China’s economic development is uneven, and medical resources are unevenly distributed. These resources are more highly concentrated in economically developed areas. In these developed areas, the economy is developing well, and the government and families attach great importance to education [26]. In addition, nursing students who participated in virtual simulation teaching preferred it (271/883, 30.7% vs 43/883, 4.9%), which is consistent with previous studies [27,28]. This could be the case because virtual simulation teaching caters to the thinking skills of young people more than traditional teaching. Through the use of virtual simulations, nursing trainees have the opportunity to practice skills and deal with difficult situations. Virtual simulation teaching allows greater access rights, and it allows interns to...
appear “virtually” only as participants. In addition, the virtual environment provides a safe environment for practicing nontechnical skills such as teamwork.

Suggestions to Promote the Use of Virtual Simulation Teaching Applications

The utilization rate of virtual simulation teaching applications in China is low because of the low awareness of this teaching method among interns. Only 10.2% (90/883) of interns had heard about virtual simulation teaching. In 2008, Tsinghua University launched a medical-related virtual simulation project for the first time to help doctors complete neurosurgery operations [29]. In 2018, China established a national virtual simulation experiment teaching platform, and virtual simulation teaching began to develop [30].

Through virtual simulation, clinical thinking was the ability that interns wanted to acquire the most (534/883, 60.5%); the second most desired ability was analytical skills (141/883, 16.0%). This result is consistent with a study in Canada [31]. This indicates that virtual simulation is a supplementary teaching strategy that provides opportunities to improve students’ clinical reasoning ability through exposure to a large number of clinical situations. The use of clinical virtual simulation as a teaching strategy should be integrated and coordinated with other teaching strategies in the classroom and other resources (eg, the high-, medium-, and low-tech simulators used in our simulation laboratory) to maximize the development of students’ cognitive, emotional, and psychomotor skills [32,33].

Nursing trainees believed that the scenarios that should be included in the virtual simulation of diabetes care are the treatment of patients with hypoglycemia (626/883, 70.9%), functional test simulation (610/883, 69.1%), and blood glucose monitoring technology (485/883, 54.9%).

Several studies have also shown that nursing interns lack the knowledge to properly handle patients with hypoglycemia, especially elderly patients with diabetes, which could increase the risk of acute complications in these patients [34,35]. This reminds us that virtual simulation is an interactive learning strategy that can increase students’ intrinsic motivation and satisfaction. It focuses on the application of basic knowledge to clinical learning challenges that reproduce the clinical scenarios that students will face in the future. It allows for competency-based education and assessment to enable deeper learning and the development of clinical expertise. Virtual simulation can help reduce clinical errors and improve the safety and quality of health care. When designing diabetes virtual simulations, we should focus on the design of scenarios for patients with hypoglycemia.

Comparison With Previous Work

To the best of our knowledge, no large-scale survey on the use and demand of virtual simulation has been previously conducted among Chinese nursing interns. An Indian survey showed that the need for diabetes knowledge by interns is urgent, consistent with our research, but that study did not identify what kinds of teaching tools the interns wanted. The survey only investigated the needs of first-year nursing students in one city in regard to virtual simulation [36], while our research collected information about the understanding of virtual simulation among interns in various provinces of China. Our research found that students who received virtual simulation teaching tended to be younger, more educated, and have a stronger autonomous learning ability, which is consistent with a survey conducted in Canada [37].

Strengths and Limitations

A strength of our research is that the initial survey links for patients and diabetes experts were sent to 37 partner schools in 23 representative major cities in China, and these were disseminated through their WeChat contact networks. In addition to this snowball-sampling method, the survey was also carried out through three convenience-sampling methods on WeChat Moments.

Our research also has some limitations. First, the sample of 883 nurse interns could not fully represent all interns in China. Our sample came from 23 provinces in China; thus, not all provinces were represented. Second, our sampling was not stratified by geographic area, urban or rural area, school level, or hospital level where internships were based. Certain selection biases were inevitable. Finally, our sampling was based on the WeChat network. Although WeChat has 1.04 billion monthly active users [38], some people rarely use WeChat or surf the internet. Our research methods included a cross-sectional survey. Although the views and attitudes of interns are very important in developing teaching methods for them, people’s attitudes toward the usefulness of simulations and their possible effects depend to a large extent on their current technological development and implementation methods. Therefore, with the development of technology and changes in people’s perceptions, these findings must be updated over time. In addition, many factors affect the use of teaching methods. Although we adjusted for some factors in the multivariate analysis, other potential confounding factors still exist.

Conclusions

Chinese nursing interns’ awareness and usage of diabetes virtual simulation teaching methods are low. However, interns desire the knowledge they would gain by using these methods. Designing virtual simulations of diabetes that are suitable for the needs of different nurse trainees will increase students’ interest in learning and help improve diabetes knowledge among nursing interns. High-quality randomized controlled trials can be conducted to improve the effectiveness of virtual simulation teaching of diabetes, provide evidence for teachers to choose suitable teaching tools, and help with the promotion of the correct management of diabetes. China should improve people’s understanding of virtual simulation teaching in universities, and relevant policies and regulations should be published to support teachers in using virtual simulation teaching tools in schools or hospitals. Virtual simulation is a potentially effective supplement for teaching. It can be used anywhere and at any time to improve the self-learning methods of Chinese nursing interns.
Acknowledgments
This work was supported by the Education and Teaching Reform Research Project of Central South University (2020y166-11) and the Clinical Nursing Research Fund Project of the Second Xiangya Hospital, Central South University (2019-HLKY-25), the People's Republic of China.

Conflicts of Interest
None declared.

References


Upper-Limb Motion Recognition Based on Hybrid Feature Selection: Algorithm Development and Validation

Qiaoqin Li¹, PhD; Yongguo Liu¹, PhD; Jiajing Zhu¹, PhD; Zhi Chen¹, PhD; Ling Liu¹, MD; Shangming Yang¹, PhD; Guanyi Zhu², BA; Bin Zhu², BSc; Juan Li³, PhD; Rongjiang Jin¹, PhD; Jing Tao⁵, PhD; Lidian Chen⁵, PhD

¹Knowledge and Data Engineering Laboratory of Chinese Medicine, School of Information and Software Engineering, University of Electronic Science and Technology of China, Chengdu, China
²College of Electrical and Information Engineering, Hunan University, Changsha, China
³Chengdu Chronic Diseases Hospital, Chengdu, China
⁴College of Health Preservation and Rehabilitation, Chengdu University of Traditional Chinese Medicine, Chengdu, China
⁵College of Rehabilitation Medicine, Fujian University of Traditional Chinese Medicine, Fuzhou, China

Corresponding Author:
Yongguo Liu, PhD
Knowledge and Data Engineering Laboratory of Chinese Medicine
School of Information and Software Engineering
University of Electronic Science and Technology of China
No 4, Section 2
Jianshe North Road
Chengdu, 610054
China
Phone: 86 13980786625
Email: liuyg@uestc.edu.cn

Abstract

Background: For rehabilitation training systems, it is essential to automatically record and recognize exercises, especially when more than one type of exercise is performed without a predefined sequence. Most motion recognition methods are based on feature engineering and machine learning algorithms. Time-domain and frequency-domain features are extracted from original time series data collected by sensor nodes. For high-dimensional data, feature selection plays an important role in improving the performance of motion recognition. Existing feature selection methods can be categorized into filter and wrapper methods. Wrapper methods usually achieve better performance than filter methods; however, in most cases, they are computationally intensive, and the feature subset obtained is usually optimized only for the specific learning algorithm.

Objective: This study aimed to provide a feature selection method for motion recognition of upper-limb exercises and improve the recognition performance.

Methods: Motion data from 5 types of upper-limb exercises performed by 21 participants were collected by a customized inertial measurement unit (IMU) node. A total of 60 time-domain and frequency-domain features were extracted from the original sensor data. A hybrid feature selection method by combining filter and wrapper methods (FESCOM) was proposed to eliminate irrelevant features for motion recognition of upper-limb exercises. In the filter stage, candidate features were first selected from the original feature set according to the significance for motion recognition. In the wrapper stage, k-nearest neighbors (kNN), Naïve Bayes (NB), and random forest (RF) were evaluated as the wrapping components to further refine the features from the candidate feature set. The performance of the proposed FESCOM method was verified using experiments on motion recognition of upper-limb exercises and compared with the traditional wrapper method.

Results: Using kNN, NB, and RF as the wrapping components, the classification error rates of the proposed FESCOM method were 1.7%, 8.9%, and 7.4%, respectively, and the feature selection time in each iteration was 13 seconds, 71 seconds, and 541 seconds, respectively.

Conclusions: The experimental results demonstrated that, in the case of 5 motion types performed by 21 healthy participants, the proposed FESCOM method using kNN and NB as the wrapping components achieved better recognition performance than the traditional wrapper method. The FESCOM method dramatically reduces the search time in the feature selection process. The results also demonstrated that the optimal number of features depends on the classifier. This approach serves to improve feature recognition performance.
selection and classification algorithm selection for upper-limb motion recognition based on wearable sensor data, which can be extended to motion recognition of more motion types and participants.

(JMIR Mhealth Uhealth 2021;9(9):e24402) doi:10.2196/24402

KEYWORDS
feature selection; inertial measurement unit; motion recognition; rehabilitation exercises; machine learning

Introduction

Background

The combination of wearable devices and wireless network technologies enables modern healthcare service providers to ubiquitously monitor patients out of hospital who require long-term exercise [1-3]. Motion recognition plays an important role in maintaining the intensity and quality of autonomous training with no or reduced supervision [4]. O’Brien et al [5] investigated the performance of action recognition based on signals collected by accelerometer, gyroscope, and barometer sensors in a mobile phone in a home setting for stroke patients. Zhang et al [6] proposed a fuzzy kernel motion classifier to address the overlapping motion class issue caused by irregular motion samples performed by patients with different functional impairments. Cui et al [7] developed an automatic gait analysis system for stroke patients based on multimodal fusion architecture. Cai et al [8] investigated the feasibility of a support vector machine (SVM) classifier for motion recognition of the upper-limb exercises via surface electromyogram (sEMG) signals [8]. Huang et al [9] proposed a knowledge-driven multimodal activity recognition framework that exploits external knowledge to fuse multimodal data.

Feature Selection in Motion Recognition

Most motion recognition methods are based on feature extraction and machine learning algorithms [10]. Time-domain and frequency-domain features are extracted from original time series data [11,12]. Castiblanco et al [13] exploited myoelectric signals (EMG) to identify finger and hand motions through pattern recognition techniques. Several methods for feature extraction, ranking, and classification from EMG signals were implemented, and the performance of motion identification was compared. Shawen et al [14] developed 4 classifiers that use accelerometer and gyroscope data collected by mobile phone from able-bodied individuals to detect falls in individuals with a lower limb amputation. A set of 40 features was computed from the original sensor data, and classifiers were trained to detect falls. Lin et al [15] used 2 sensors on the arm and wrist to collect acceleration and angular velocity of 6 types of upper-limb exercises performed by 13 volunteers. Motor features were used to train a back-propagation neural network (BPNN) algorithm for motion recognition. Wu et al [16] developed a method to identify upper-limb motion for community rehabilitation. The feature vector space was established by variance, mean absolute value, the fourth-order autoregressive, zero crossings, and root mean square. Various feature sets were extracted for classification.

Feature selection is an essential step to eliminate redundant or irrelevant features for specific classification task so as to deal with high-dimensional data [17,18]. Its task is to find the most representative feature subset from the original feature set. Ramezani et al [19] analyzed physical activity sensor features and activities with regard to indoor localization. Random forest (RF) was used to build a predictive model based on the most significant features. The study demonstrated that a subset of features can better distinguish between at-risk patients that can gain independence versus patients that will be rehospitalized. Wang et al [20] proposed 2 feature selection methods to improve activity recognition. Experimental results showed that the proposed methods reduce the dimensionality of the original feature space and contribute to the enhancement of overall recognition accuracy. Fang et al [21] compared feature selection methods based on interclass distance for human activity recognition in smart home environments. The experimental results showed that activity recognition accuracy is related to the feature set selected and an unsuitable feature set increases computational complexity and degrades activity recognition accuracy. Zhou et al [22] proposed a feature selection method for human motion recognition based on open human motion data. The experimental results showed that their feature selection method yields better recognition accuracy than non-feature selection models.

Feature selection methods can be categorized into filter and wrapper methods [23]. For filter methods, the selection of the feature subset is independent of the classification algorithm. The feature fitness is evaluated via the statistical characteristics of the dataset, and the features with top ranking fitness are selected [24,25]. Banos et al [26] proposed a feature selection method for physical activity recognition using a feature quality group ranking via statistical criteria based on discrimination and robustness. Satisfactory results were achieved in both laboratory and seminaturalistic activity living datasets for real problems using several classification models. Hong et al [27] proposed a motion gesture recognition system via accelerometer (MGRA) implemented on mobile devices. The best feature vector including 27 items was selected using the minimal-redundancy-maximal-relevance criterion taking both static and mobile scenarios into consideration. The experimental results confirmed that MGRA can accommodate a broad set of gesture variations within each class and achieve higher accuracy than previous methods [28]. As for wrapper methods, the feature subset is selected simultaneously with the estimation of its goodness in a specific classification task [29,30]. Camargo and Young [31] implemented motion classification from sEMG signals for prosthetic control by exploiting Chow-Liu trees for selecting features and evaluating 6 different classification algorithms as the wrapping component [32]. The results demonstrated that feature selection is critical for improving classification accuracy. Xue et al [33] presented a novel wrapper feature selection algorithm that utilizes a generic algorithm to
wrap an extreme learning machine to search for the optimum feature subset. Experiments were conducted on benchmark datasets and compared with 4 filter methods and 2 hybrid wrapper methods. The results revealed that the presented wrapper method is useful for feature selection problems and outperforms other algorithms in comparison. Chen and Chen [34] introduced a wrapper method to eliminate irrelevant features during classifier construction by introducing the cosine distance into SVM. The feature selection method has been applied to fault diagnosis of rolling element bearings and diagnosis of mild cognitive impairment. The results showed that the proposed method has great capacity for feature selection and pattern recognition.

The Hybrid Feature Selection Method

Filter methods are often time-efficient, but the results are not always satisfactory. On the other hand, wrapper methods usually achieve better performance, but could be computationally intensive and the obtained feature subset optimized only for the specific learning algorithm [35]. As a result, hybrid feature selection methods take advantages of both filter and wrapper methods. Manbari et al [36] presented a hybrid feature selection algorithm based on the combination of clustering and the modified binary ant system to overcome the search space and high-dimensional data processing challenges. A damped mutation strategy was introduced to avoid local optima, and a new redundancy reduction policy was adopted to estimate the correlation between the selected features so as to further improve the algorithm.

Existing feature selection methods usually select feature sets that are relevant for specific classification tasks. To select the most representative features for motion recognition of upper-limb exercises, we propose a hybrid feature selection method combining the filter and wrapper methods called FESCOM in this paper. In the filter stage, candidate features are selected by ranking the feature significance index, which reflects the importance of each feature for motion recognition. In the wrapper stage, a classifier-specific feature selection algorithm is applied to further refine the candidate features. Classifiers including kNN, NB, and RF are constructed as the wrapping components. To the best of our knowledge, FESCOM is the first method that exploits hybrid feature selection for motion recognition of upper-limb exercises.

Methods

Workflow

The general workflow of this work is illustrated in Figure 1. An inertial measurement unit (IMU) node was customized for motion data collection. Motion data including acceleration and angular velocity from 5 types of upper-limb exercises were collected. Original data were preprocessed by applying a median filter to remove outliers. Time-domain and frequency-domain features were extracted from the preprocessed acceleration and angular velocity data on each axis. The feature selection method was built to select the most representative features for motion recognition. Then, motion recognition was implemented using the optimal feature set and corresponding classifier.

Figure 1. General workflow. FESCOM: hybrid feature selection method by combining filter and wrapper methods; IMU: inertial measurement unit.
Construction of the IMU

The IMU module consists of 1 inertial sensor (MPU9250), 1 8-bit low power consumption micro-controller (ATmega328P on board nano V3.0), 1 Bluetooth wireless transmitter (HC-06), and 1 battery, shown in Figure 2. The inertial sensor MPU9250 is comprised of a 3-axis accelerometer, gyroscope, and magnetometer. The built-in digital motion processing engine in the MPU9250 can reduce the complex computation and load of the microcontroller. The measurement ranges of the accelerometer and gyroscope in the MPU9250 are ±16 g and ±2000 °/s, respectively, where g represents gravitational acceleration. These specifications meet the needs of upper-limb exercises. Sampled motion data can be transmitted to a PC station by Bluetooth in real time. The battery is 3.7 V and 200 mAh. No recharge module is used. It is convenient for wearable devices. The scale ranges of the accelerometer and the gyroscope can be adjusted using the programming interface of the inertial sensor and were set at ±2 g and ±250 °/s, respectively, in this study. The sampling frequency was set at 20 Hz, which is suitable for upper-limb exercises by patients with motion functionality impairment. The baud rate of Bluetooth was set at 19200 bps. Angular velocity was computed based on the gyroscope data. Magnetometer data were not used in this study. These components were connected and embedded into a 58 mm x 32 mm x 19 mm box. The IMU node was attached to the outside of the right upper limb of the participant with a stretchable 350 mm x 38 mm rubber belt, shown in Figure 2. The positive direction of the y-axis points to the wrist.

Figure 2. (A) inertial measurement unit (IMU) components, (B) box, and (C) belt.

Experimental Protocol

In this study, upper-limb exercises for post-stroke rehabilitation training were considered. From the clinical point of view, a subset of the training items can represent the 33 upper limb–related training items in the Fugl-Meyer Assessment (FMA) scale [37]. In this experiment, 5 representative upper-limb exercises based on the FMA scale were selected:

1. Forearm pronation and supination: Raise the right arm to the horizontal position in the sagittal plane. Then, carry out forearm pronation and supination.
2. Lumbar touch: The right arm hangs naturally. Move the right arm back to touch the back of the waist with the hand. Then, slowly move back to the initial position.
3. Shoulder touch: The right arm hangs naturally. Raise the right arm to the horizontal position in the sagittal plane. Then, carry out an elbow adduction motion and rotate the wrist to touch the opposite shoulder with the hand. Finally, put the arm down to the initial position.
4. Shoulder flexion: The right arm hangs naturally. Raise the right arm in the sagittal plane as high as possible. Then, hold for 3 seconds and move back to the initial position.
5. Shoulder extension: The right arm hangs naturally. Raise the right arm in the coronal plane as high as possible. Then, hold for 3 seconds and move back to the initial position.

Figure 3 includes 5 photos of each exercise taken during the execution process.

Motion data were collected from 21 healthy participants (15 men, 6 women; age, mean 33.2, SD 12.7 years; height, mean 172.5, SD 7.1 cm; weight, mean 62.8, SD 17.5 kg) instead of actual patients who are post-stroke. The study was approved by the institutional review board of the Eighth People’s Hospital of Chengdu. Written informed consent was obtained from all participants. In the sampling experiment, participants were first asked to rest for a while. Before the sampling began, they were invited to perform each exercise several times with the guidance of a guiding video until they performed the motions fluidly. Then, they were required to complete 3 valid repetitions of each exercise independently. Each repetition followed an interval of about 3 seconds. A valid repetition was a coherent movement, and each repetition was completed in 1-4 seconds.
In this study, 10 types of time-domain and frequency-domain features were extracted from the motion data from the upper-limb exercises. The time-domain features included the mean, standard deviation, maximum, and minimum values of the signal as well as the kurtosis, skewness, and interquartile range of the signal, which may reflect the exercise frequency, regularity, and symmetry, respectively. The frequency-domain features included average power, average frequency, and median frequency of the signal. As each sample included acceleration and angular velocity data in 3 axes, the dimension of the original feature vector was 60.

The original feature set contains not only the features that are relevant for classification but also some redundancy features, which decrease the computational efficiency and classification accuracy. We propose a hybrid feature selection method, called FESCOM, to remove redundant features so as to improve the computational efficiency and classification accuracy. Figure 4 shows the procedure of the FESCOM method. In the filter stage, the statistical $t$ test method was adopted to compute the statistical significance value ($P$ value) of each feature, reflecting the capability of motion recognition [38,39]. For samples $x$ and $y$, a two-sample $t$ test was considered for analysis, which is defined as:

$$
t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_x^2}{n} + \frac{s_y^2}{m}}} \quad (1)
$$

where $\bar{x}$ and $\bar{y}$ are the sample means, $s_x$ and $s_y$ are the sample standard deviations, and $n$ and $m$ are the sample size. As there are 5 types of motion, the $t$ test method was applied to each class pair. Let $P_k(i,j)$ represent the $P$ value of feature $k$ on class $i$ and $j$, the average $P$ value of feature $k$ on all class pairs is computed as:

$$
avgP_k = \frac{\sum_{i=1}^{C} \sum_{j=i+1}^{C} P_k(i,j)}{C(C-1)}
$$

where $i=1,...,C$, $j=i+1,...,C$, and $C$ is the number of motion classes. The standard deviation of the $P$ value of feature $k$ on all class pairs is:
Then, the significance index of feature $k$ is computed as:

$$s_k$$

A smaller $s$ value means stronger classification capacity. The features with $s$ smaller than a threshold were selected for the candidate feature set, ranked in ascending order. The threshold value was a compromise between time efficiency and the classification accuracy of FESCOM.

**Figure 4.** The hybrid feature selection method by combining filter and wrapper methods (FESCOM).

In the wrapper stage, a sequential feature selection (SFS) method was used to refine the features from the candidate feature set obtained in the filter stage. SFS includes a search algorithm and an objective function, also called criterion [40]. In this study, the search algorithm was sequential forward selection, and the criterion was the classification error rate. Starting from an empty feature set, SFS selects a subset of features from the candidate feature set by sequentially selecting features using the abovementioned criterion until there is no improvement in classification performance, evaluated by each classification algorithm (ie, the wrapping component). The procedure of SFS is illustrated as Figure 5.
Classification Algorithms

In the experiment, kNN, NB, and RF were adopted as classification algorithms in the wrapper stage of FESCOM.

Classifier kNN is a simple classification algorithm based on the calculation of the distance (usually the Euclidean distance) between the new sample to be classified and the closest samples in the training set. The training samples are sorted in descending order according to their distance from the new object. Then, the new sample is assigned to the class that most of its k-nearest neighbors belong to [41].

NB is a probabilistic classifier based on the assumption that all features are independent of each other, given the category variable [42]. For discrete features, multinomial or Bernoulli distributions are popular. Despite apparently over-simplifier assumptions, NB classifier works quite well in many complex real-world applications, such as medical diagnosis, key phrase extraction, and text classification. The NB classifier is particularly useful for handling incomplete data and could yield good predictions even with a small data size.

RF is a type of ensemble learning method that is formed through the combination of multiple decision trees trained on the training dataset. When applied to the test dataset, the predictions of individual tree models within the RF are combined into an overall classification decision through means of a majority vote or the application of weights. The RF model can avoid overfitting and provide robust classification performances [43]. The number of decision trees in this experiment was set at 20.

Results

Overview

In this study, MATLAB 2016a was used to develop the proposed FESCOM method for motion recognition of upper-limb exercises. The original dataset was randomly partitioned into a training set and testing set. The training set was applied to train each classifier by using five-fold cross validation. For each iteration, one of the partitions was held back as the validation set, whereas the other partitions were used to train the classification model. The model was then validated by the validation set. This process was repeated 5 times, so that each subset was used as a validation set once. The results were averaged over all rounds. Finally, the performance of each classifier was evaluated on the testing set.

The performance of the proposed algorithm was evaluated using the metrics of classification error rate, computed as the ratio of number of instances classified incorrectly to the total number of instances.

Experimental Data

Acceleration and angular velocity in 3 axes of 5 exercises performed by 1 female participant are illustrated in Figure 6 and Figure 7, respectively. Exe1, Exe2, Exe3, Exe4, and Exe5 in Figures 6 and 7 represent the 5 types of motion defined in the experimental protocol section. The time-domain waveforms of the 5 exercises showed different characteristics. For example, the acceleration and angular velocity of forearm pronation and supination, lumbar touch, and shoulder touch showed totally different trends. Although the time-domain features of the
acceleration in the y-axis of shoulder flexion were similar with those of shoulder extension, the acceleration in the x-axis of shoulder flexion exhibited a higher peak compared with that of shoulder extension. Moreover, angular velocity in both the x-axis and y-axis of shoulder flexion exhibited a smaller peak than that of shoulder extension. These differences between exercises can be used for motion recognition.

**Figure 6.** Acceleration in each axis of the 5 exercises.

![Figure 6](image)

**Figure 7.** Angular velocity in each axis of the 5 exercises.

![Figure 7](image)

### Feature Significance Index

Table 1 shows the top-10 feature significance index value and rank order computed by the statistical t-test method. An extended version of Table 1, including the significance index of all 60 features, is presented in Multimedia Appendix 1. The significance index value was computed on 2 types of signals (ie, acceleration and angular velocity), represented in parentheses following the feature name, with a suffix representing on which axis it is. The significance index value of the minimum angular_velocity_y ranks the highest, whereas the average acceleration_z ranks the lowest of all 60 features. A smaller significance index means stronger classification capacity.

**Table 1.** Top 10 feature significance index.

<table>
<thead>
<tr>
<th>Feature name</th>
<th>Significance index</th>
<th>Rank order</th>
</tr>
</thead>
<tbody>
<tr>
<td>minimum (angular_velocity_y)</td>
<td>0.00030</td>
<td>1</td>
</tr>
<tr>
<td>average power (angular_velocity_x)</td>
<td>0.00035</td>
<td>2</td>
</tr>
<tr>
<td>average power (acceleration_x)</td>
<td>0.00049</td>
<td>3</td>
</tr>
<tr>
<td>standard deviation (angular_velocity_z)</td>
<td>0.00058</td>
<td>4</td>
</tr>
<tr>
<td>skewness (acceleration_z)</td>
<td>0.00130</td>
<td>5</td>
</tr>
<tr>
<td>average power (acceleration_y)</td>
<td>0.00132</td>
<td>6</td>
</tr>
<tr>
<td>median frequency (acceleration_y)</td>
<td>0.00155</td>
<td>7</td>
</tr>
<tr>
<td>median frequency (angular_velocity_y)</td>
<td>0.00174</td>
<td>8</td>
</tr>
<tr>
<td>maximum (angular_velocity_x)</td>
<td>0.00240</td>
<td>9</td>
</tr>
<tr>
<td>standard deviation (angular_velocity_y)</td>
<td>0.00283</td>
<td>10</td>
</tr>
</tbody>
</table>

### Experimental Results

To analyze the impact of the feature number on the performance of motion recognition of upper-limb exercises, experiments were conducted on the training set including a different number of features. The results of the classification error rate are shown in Figure 8. The general trends for the 3 classifiers are similar. With an increase in feature number, the classification error rate decreases rapidly. With a further increase in feature number, the classification error rate shows an increasing trend. The trend for kNN is not as stable as that of NB and RF. For kNN, there are several local optima with an increase in feature number.
Figure 8. Classification error rate vs feature number. kNN: k-nearest neighbor; NB: Naïve Bayes; RF: random forest.

Table 2 shows the optimal number of features for the different classifiers. There is an obvious distinction between the optimal number of features for the different classifiers. Classifier kNN needs more features to achieve the minimum classification error rate than classifiers NB and RF.

Table 2. Optimal number of features.

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Optimal number of features</th>
</tr>
</thead>
<tbody>
<tr>
<td>kNN\textsuperscript{a}</td>
<td>33</td>
</tr>
<tr>
<td>NB\textsuperscript{b}</td>
<td>13</td>
</tr>
<tr>
<td>RF\textsuperscript{c}</td>
<td>18</td>
</tr>
</tbody>
</table>

\textsuperscript{a}kNN: k-nearest neighbor.
\textsuperscript{b}NB: Naïve Bayes.
\textsuperscript{c}RF: random forest.

As wrapper methods usually achieve better classification performance than filter methods, and until now, there has been no hybrid feature selection method for motion recognition of rehabilitation exercises, we compared the motion recognition performance of the proposed FESCOM method with the traditional wrapper method. Experiments were conducted on the testing set by selecting the optimal feature set. For the traditional wrapper method, SFS was used to search for the optimal feature set from all 60 original features. For FESCOM, SFS was used to refine the features from the candidate feature set, composed of features with a significance index value smaller than 0.05 and ranked in ascending order. The criterion to set the threshold of the significance index was an assumption that the number of candidate features increased 30% (10 out of 33 features) from the highest optimal number of features in Table 2. The classification error rate is shown in Table 3. For both feature selection methods, the classification performance of kNN was better than that of NB and RF. The classification error rate of FESCOM using kNN and NB as the wrapping component was lower than the corresponding wrapper methods.
Table 3. Classification error rate.

<table>
<thead>
<tr>
<th>Feature selection method</th>
<th>Classification error rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>kNN\textsuperscript{a}</td>
</tr>
<tr>
<td>Wrapper</td>
<td>2.2</td>
</tr>
<tr>
<td>FESCOM\textsuperscript{d}</td>
<td>1.7</td>
</tr>
</tbody>
</table>

\textsuperscript{a}kNN: k-nearest neighbor.  
\textsuperscript{b}NB: Naive Bayes.  
\textsuperscript{c}RF: random forest.  
\textsuperscript{d}FESCOM: hybrid feature selection method by combining filter and wrapper methods.

The time consumed on feature selection in each iteration for both feature selection methods is listed in Table 4. As the candidate feature set of FESCOM is smaller than that of the wrapper method, the search time for FESCOM was much less than that of the wrapper method for all classifiers. For the same feature selection method, kNN needs much less search time than NB and RF.

Table 4. Search time for each iteration.

<table>
<thead>
<tr>
<th>Feature selection method</th>
<th>Search time (seconds)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>kNN\textsuperscript{a}</td>
</tr>
<tr>
<td>Wrapper</td>
<td>23</td>
</tr>
<tr>
<td>FESCOM\textsuperscript{d}</td>
<td>13</td>
</tr>
</tbody>
</table>

\textsuperscript{a}kNN: k-nearest neighbor.  
\textsuperscript{b}NB: Naive Bayes.  
\textsuperscript{c}RF: random forest.  
\textsuperscript{d}FESCOM: hybrid feature selection method by combining filter and wrapper methods.

Discussion

Principal Findings

This paper presents a hybrid feature selection method for motion recognition of upper-limb exercises. For motion recognition based on feature extraction and feature selection, the feature set used for the classification algorithm had a direct impact on the performance of classification. The experimental results in this study verified that recognition performance depends on the feature set. For all 3 classifiers in this study, the same trends existed: The classification error rate decreased to an optimum value when the number of features increased and increased with a further increase in feature number due to overfitting. The optimal number of features depended on the classifier. The optimal numbers of features for classifiers kNN, NB, and RF were 33, 13, and 18, respectively.

Each feature contributes differently to the motion classification task. Take the proposed FESCOM method as an example: The frequency-domain features contribute more than other features to recognition performance. When using the classifier kNN as the wrapping component, the top 3 significant features for motion recognition were average power of angular_velocity_x, average acceleration_x, and mean frequency of angular_velocity_x. When using the classifier NB as the wrapping component, the top 3 significant features for motion recognition were average power of acceleration_x, standard deviation of acceleration_y, and kurtosis of angular_velocity_x.

When using the classifier RF as the wrapping component, the top 3 significant features for motion recognition were average power of angular_velocity_x, average power of acceleration_y, and mean frequency of angular_velocity_y.

Motion recognition performance also depends on the classifier. For both feature selection methods, the classification performance of kNN was the best, while NB was the worst classification performance. The proposed FESCOM method reduces the feature space and improves the time efficiency by filtering irrelevant features for motion classification.

Comparison With Previous Works

The common methods for motion recognition combine wearable sensing techniques and machine learning algorithms. Acceleration, angular velocity, or sEMG signals collected by wearable sensors are used to represent the motion characteristics. Cai et al [8] exploited sEMG signals and the SVM classifier for motion recognition of upper-limb exercises; 5 healthy participants participated in the experiments. The average recognition accuracy of 5 motions was 93.34%. Motion recognition of upper-limb exercises in [15] was based on acceleration, angular velocity data, and BPNN algorithm; 13 volunteers participated in the experiments. Five upper-limb exercises involving simple swinging and stretching movements were recognized with an accuracy of 85%-95%, while exercises consisting of spiral rotations were recognized with an accuracy of 60%. The knowledge-driven activity recognition method in [9] focused on egocentric video and accelerometer/gyroscope...
data. Experiments were conducted on 3 public datasets, with a best recognition accuracy of 76.1%.

Using kNN, NB, and RF as the wrapping components, the recognition performance of FESCOM in this study achieved 98.3%, 91.1%, and 92.6%, respectively. Compared with previous studies on upper-limb motion recognition, the recognition performance of FESCOM is at the same level or even better than that in previous works. Time efficiency is one of the main concerns especially in real-time applications, such as motion recognition in autonomous rehabilitation systems. However, previous works seldom considered time efficiency. The FESCOM method in this study reduced the feature space and improved the time efficiency by filtering irrelevant features for motion classification. Compared with the search time of the traditional wrapper method, the search time of FESCOM using kNN, NB, and RF classifiers as the wrapping component reduced the search time by 43% (from 23 seconds to 13 seconds), 55% (from 159 seconds to 71 seconds), and 38% (from 876 seconds to 541 seconds), respectively. Hence, this study contributes by evaluating the number and types of features for different classification algorithms that achieve acceptable performance for motion recognition of upper-limb exercises.

**Limitations**

The FESCOM method proposed in this study has some limitations. It was only evaluated based on data from 21 healthy participants, and only 5 types of upper-limb exercises were considered in the experiments. However, the behavior of patients with a central nervous system lesion, such as that caused by stroke, may be very different from that of healthy participants. The number of samples for training and testing is not high enough for machine learning algorithms, which may also affect the reliability. The customized IMU module in this work is just a prototype. The components in the sensor node are connected with cables. This may lead to unreliable connections, especially when used in movement conditions. Another drawback is that the validation of the system did not use real-time exercise examples.

In our future work, to further confirm the feasibility of FESCOM, we plan to extend our experiment considering the following aspects. First, we will evaluate and compare the performance of different methods in the filter and wrapper stage of FESCOM. Second, we will evaluate the performance of FESCOM considering more classifiers as the wrapping component in the wrapper stage, such as SVM and latent Dirichlet allocation. Third, we will evaluate the performance of FESCOM on more datasets, such as public datasets including more motion types and datasets including not only healthy participants but also real patients with different functional impairments in the recovery stage in a clinical situation. Fourth, we plan to improve the IMU node as an embedded system on a circuit board for real-time data collection and validate the whole system by real-time prediction of upper-limb exercises.

**Conclusions**

In this study, a hybrid feature selection method, FESCOM, was proposed for motion recognition of upper-limb exercises and evaluated using 5 types of upper-limb exercises performed by 21 healthy participants. The experimental results demonstrate that FESCOM is feasible for motion recognition of upper-limb exercises performed by healthy participants. FESCOM improves the recognition accuracy when using kNN and NB as the wrapping component and improves the time efficiency in the wrapper stage. The results also demonstrate that, for different classifiers, different feature sets are selected to achieve optimal performance. This work can be extended to provide motion recognition of more motion types and participants including healthy people and actual patients with minor motor damage.

**Acknowledgments**

The authors would like to thank the participants who made data collection possible. This study was supported in part by the National Key R&D Program of China (grant 2019YFC1710300) and the Sichuan Science and Technology Program (grants 2019YFS0019, 2020YFS0283, 2021YJ0184, and 2021YFS0152).

**Conflicts of Interest**

None declared.

**Multimedia Appendix 1**

Extended version of Table 1 with significance index of all 60 features.

[DOCX File, 16 KB - mhealth_v9i9e24402_app1.docx]

**References**


Abbreviations

BPNN: back-propagation neural network
FESCOM: hybrid feature selection method by combining filter and wrapper methods
FMA: Fygl-Meyer Assessment
IMU: inertial measurement unit
kNN: k-nearest neighbor
MGRA: motion gesture recognition system via accelerometer
NB: Naive Bayes
RF: random forest
sEMG: surface electromyogram
SFS: sequential feature selection
SVM: support vector machine
Correction: Effect of Physician-Pharmacist Participation in the Management of Ambulatory Cancer Pain Through a Digital Health Platform: Randomized Controlled Trial

Lu Zhang, Howard L McLeod, Ke-Ke Liu, Wen-Hui Liu, Hang-Xing Huang, Ya-Min Huang, Shu-Sen Sun, Xiao-Ping Chen, Yao Chen, Fang-Zhou Liu, Jian Xiao

*Corresponding Author:
Jian Xiao, Prof Dr
Department of Pharmacy
Xiangya Hospital
Central South University
Room 404
No. 87 Xiangya Road, Kaifu District
Changsha, 410008
China
Phone: 86 13973168895
Fax: 86 073184327460
Email: admanoxas@163.com

Related Article:
Correction of: https://mhealth.jmir.org/2021/8/e24555/

(JMIR Mhealth Uhealth 2021;9(9):e33223) doi:10.2196/33223

In “Effect of Physician-Pharmacist Participation in the Management of Ambulatory Cancer Pain Through a Digital Health Platform: Randomized Controlled Trial” (JMIR Mhealth Uhealth 2021;9(8):e24555), one error was noted.

This has been corrected to:

Lu Zhang, Howard L McLeod, Ke-Ke Liu, Wen-Hui Liu, Hang-Xing Huang, Ya-Min Huang, Shu-Sen Sun, Xiao-Ping Chen, Yao Chen, Fang-Zhou Liu, Jian Xiao

The correction will appear in the online version of the paper on the JMIR Publications website on September 13, 2021, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.
©Lu Zhang, Howard L McLeod, Ke-Ke Liu, Wen-Hui Liu, Hang-Xing Huang, Ya-Min Huang, Shu-Sen Sun, Xiao-Ping Chen, Yao Chen, Fang-Zhou Liu, Jian Xiao. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 13.09.2021. 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 https://mhealth.jmir.org/, as well as this copyright and license information must be included.
User Experiences of the NZ COVID Tracer App in New Zealand: Thematic Analysis of Interviews

Alexei Tretiakov¹, PhD; Inga Hunter¹, MB, MPhil
School of Management, Massey University, Palmerston North, New Zealand

Corresponding Author:
Alexei Tretiakov, PhD
School of Management
Massey University
Tennent Drive
Palmerston North, 4474
New Zealand
Phone: 64 6 951 7905
Email: A.Tretiakov@massey.ac.nz

Abstract

Background: For mobile app–based COVID-19 contact tracing to be fully effective, a large majority of the population needs to be using the app on an ongoing basis. However, there is a paucity of studies of users, as opposed to potential adopters, of mobile contact tracing apps and of their experiences. New Zealand, a high-income country with western political culture, was successful in managing the COVID-19 pandemic, and its experience is valuable for informing policy responses in similar contexts.

Objective: This study asks the following research questions: (1) How do users experience the app in their everyday contexts? and (2) What drives the use of the app?

Methods: Residents of New Zealand’s Auckland region, which encompasses the country’s largest city, were approached via Facebook, and 34 NZ COVID Tracer app users were interviewed. Interview transcripts were analyzed using thematic analysis.

Results: Interviews ranged in duration from 15 to 50 minutes. Participants ranged in age from those in their late teens to those in their early sixties. Even though about half of the participants identified as White New Zealanders of European origin, different ethnicities were represented, including New Zealanders of South Pacific, Indian, Middle Eastern, South American, and Southeast Asian descent. Out of 34 participants, 2 (6%) identified as Māori (Indigenous New Zealanders). A broad range of careers were represented, from top-middle management to health support work and charity work. Likewise, educational backgrounds ranged broadly, from high school completion to master’s degrees. Out of 34 participants, 2 (6%) were unemployed, having recently lost their jobs because of the pandemic. The thematic analysis resulted in five major themes: perceived benefits, patterns of use, privacy, social influence, and need for collective action. Benefits of using the app to society in general were more salient to the participants than immediate health benefits to the individual. Use, however, depended on the alert level and tended to decline for many participants at low alert levels. Privacy considerations played a small role in shaping adoption and use, even though the participants were highly aware of privacy discourse around the app. Participants were aware of the need for high levels of adoption and use of the app to control the pandemic. Attempts to encourage others to use the app were common, although not always successful.

Conclusions: Appeals to civic responsibility are likely to drive the use of a mobile contact tracing app under the conditions of high threat. Under the likely scenario of COVID-19 remaining endemic and requiring ongoing vigilance over the long term, other mechanisms promoting the use of mobile contact tracing apps may be needed, such as offering incentives. As privacy is not an important concern for many users, flexible privacy settings in mobile contact tracing apps allowing users to set their optimal levels of privacy may be appropriate.

(JMIR Mhealth Uhealth 2021;9(9):e26318) doi:10.2196/26318

KEYWORDS
COVID-19; contact tracing; app; New Zealand; adoption; use; civic responsibility; privacy
Introduction

Background

Contact tracing is a nonpharmaceutical intervention commonly used for curbing the spread of COVID-19 [1,2]. Manual contact tracing, conducted by interviewing patients diagnosed with the disease to identify their close contacts, is rather slow, and digital contact tracing involving digitally recording information about individuals’ movements has been suggested to be potentially considerably more effective based on simulation evidence [3]. Mobile apps that perform digital contact tracing have been implemented in many countries, such as Australia [4] and Singapore [5]. Based on cross-country comparison, Urbaczewski and Lee [6] asserted that mobile app–based contact tracing is effective in helping countries to keep COVID-19 under control.

For mobile app–based contact tracing to be fully effective, a large majority of the population needs to be using the app on an ongoing basis [3,7]. The importance of mobile contact tracing app adoption prompted several empirical studies. Trang et al [8] conducted an experiment in Germany where they suggested alternative designs and made different appeals about the benefits offered, asking the respondents to rate their intent to install the app. They found that citizens can be divided into three categories: critics, undecided, and advocates. Critics were more likely to accept an app based on an appeal that by using it they would protect society in general, and they were more likely to accept app designs with strong privacy features. Undecided citizens, similar to critics, responded to societal-benefit appeals, but valued convenience in app use more than they valued privacy. Neither critics nor undecided citizens cared about the app offering health benefits to them as individuals. Finally, advocates responded to both societal-benefit and self-benefit appeals and did not care about privacy or convenience.

In a similar experiment conducted in the United Kingdom, Wiertz et al [9] offered citizens four configurations of an app differing by self-benefits offered, privacy, and the entity overseeing the app. Citizens—treated as a single group—preferred an app offering self-benefits and that was overseen by an independent entity, rather than by the government. Wiertz et al [9] found no evidence suggesting that citizens valued the privacy features of an app. Jonker et al [10] conducted a discrete choice experiment in the Netherlands, allowing citizens to rate a range of possible features of a mobile contact tracing app. Citizens preferred an app that would store data locally and give them control over whether to share it with the authorities. Further, they preferred an app that would offer a small financial reward. Thus, the results regarding the effects of privacy features and of self-benefits were not consistent across studies.

Walrave et al [11] conducted a survey in Belgium to determine factors affecting citizen intention to adopt a mobile contact tracing app. The study used the unified theory of acceptance and use of technology (UTAUT) framework [12]; performance expectancy (ie, benefits offered by the app, conceptualized by Walrave et al [11] as societal benefits), effort expectancy, social influence, and facilitating conditions (ie, having the knowledge and resources necessary to use the app) were considered as potential factors. Further, innovativeness, privacy concerns, and COVID-19–related stress were added to the basic UTAUT model. The most important factor was performance expectancy, followed by facilitating conditions and social influence. Innovativeness and privacy concerns had weaker effects on intention to adopt. The results by Walrave et al [11] are complemented by a survey by Altmann et al [13] that was conducted in France, Germany, Italy, the United Kingdom, and the United States. Regarding reasons to install a mobile contact tracing app, the respondents rated benefits to family and friends higher than benefits to the broader community. In addition, they rated concerns about government surveillance and security highly regarding the main reasons against installation, with concerns about government surveillance rated the highest. Further, greater trust in the government was associated with higher app installation intent.

In all the studies introduced above, the participants had no exposure to a real mobile contact tracing app and answered questions with a hypothetical app in mind.

Research Questions

The results of the studies published so far are not entirely consistent regarding factors driving the adoption of a mobile contact tracing app. Further, studies of users, as opposed to potential adopters, of mobile contact tracing apps and their experiences are not available.

Better understanding of user experiences with the NZ (New Zealand) COVID Tracer app is of practical interest for countries using mobile contact tracing apps to protect their populations from COVID-19, particularly if COVID-19 becomes endemic [14], possibly necessitating the continued use of contact tracing over the long term. Further, understanding user experiences with a mobile contact tracing app is of broader theoretical interest for epidemiology. Therefore, this study asks the following research questions:

1. How do users experience the app in their everyday contexts?
2. What drives the use of the app?

Methods

Overall Approach and Study Setting

Qualitative design was used, as it is particularly suitable for an exploratory study of user experiences [15-17]. Data were collected via semistructured interviews with users of the NZ COVID Tracer app, New Zealand’s official mobile contact tracing app overseen by the Ministry of Health [18].

The study was conducted in the Auckland region; Auckland is the biggest city in New Zealand. New Zealand’s COVID-19 outbreak response, as assessed in October 2020, has been recognized as successful [19]. Thus, the New Zealand experience may be of interest. The Auckland region was chosen because it has experienced more COVID-19–related disruption than the rest of the country, as detailed in the following section.

Interviews were conducted 5 months after the app became available, allowing us to explore user experiences in the context of how the pandemic situation in the Auckland region and the
functionality of the app evolved over time. This context is described in the following section.

NZ COVID Tracer App and COVID-19 Pandemic in New Zealand

The NZ COVID Tracer app was released by New Zealand’s Ministry of Health on May 20, 2020 [20], simultaneously for iOS and for Android platforms. The app was presented as a “digital diary,” allowing the recording of places the users of the app visited by scanning QR (Quick Response) codes. Users could also register their contact details with the app to make it easier for COVID-19 contact tracers to reach them.

Privacy was emphasized in the app design and in the Ministry of Health’s communications about the app: information about places visited by the user was stored locally on the phone and was not shared with contact tracing services automatically; in the initial release of the app, the user had to open the app and read out the information to contact tracers. Further, the information was automatically deleted after 31 days. Moreover, for security and privacy reasons, to use the app users had to log on using a strong (ie, sufficiently long and complex) password. The password had to be re-entered every 30 days, resulting in confusion for some of the users, who would not remember the password and, thus, were locked out of the app, as documented in comments on the Apple App Store [21] and on Google Play [22].

On June 15, 2020, the app was updated to allow users to be notified if they visited a venue around the same time as a known COVID-19 case [23]. This feature was implemented without sending users’ location data to the Ministry of Health. Further, users could now send their location data to contact tracers if they chose to do so. If the initial version of the app solely supported the contact tracing process, thus offering benefits for the community or for the country as a whole, the updated app offered immediate benefits to the users, who, in case of exposure, could be diagnosed earlier and could receive early treatment, thus improving their prognosis [24]. Moreover, users receiving an alert could self-isolate, thus protecting their family, friends, and colleagues.

Benefits to the community, the user’s family, and the user as an individual have been repeatedly highlighted in subsequent communications by the Ministry of Health: “Taking a few seconds to scan in with the app means we can quickly inform you when you may have been exposed to the virus, so you can take steps to protect yourself and your whānau [extended family].” “It also means if you test positive for the virus, you can instantly provide your digital diary to contact tracers to give them a massive head-start,” and “The faster we can contact trace, the quicker we can get ahead of the virus and prevent spread in the community” [25].

Another major update of the app was on July 30, 2020, when the ability to add manual entries to record visits to locations with no QR codes, such as visits to friends and family, was added [26], allowing one “to maintain a complete – and private – record.” Initially, organizations were encouraged but not required to display QR codes compatible with the Ministry of Health’s NZ COVID Tracer app [20]. However, starting from August 19, 2020, displaying QR codes became compulsory for most business premises and for many transport services [27].

Even though location data were held locally on users’ phones, usage data, including the number of app registrations, the number of active devices, the number of QR code scans, and the number of manual entries, were available to the Ministry of Health, and some aggregate data were routinely shared via media releases (eg, Ministry of Health [28,29]). Further, historical data were available for download, and some of them are presented in Figure 1, where they are combined with historical data on the number of active COVID-19 cases in New Zealand on the COVID-19 data portal from Stats NZ, New Zealand’s official data agency [30]. A more detailed graph of the number of active COVID-19 cases in New Zealand, distinguishing import-related and locally acquired cases, can be viewed at the Ministry of Health website [31]. The history of COVID-19 alert levels in the Auckland region—Auckland is the biggest city in New Zealand; the population of the Auckland region centered on Auckland is 1.6 million, about one-third of the total population of New Zealand [32]—is also shown in Figure 1, based on a document released by the New Zealand government [33]; Alert Level 4 corresponds to a lockdown with substantial restrictions on movement, while Alert Level 1 suggests heightened vigilance, but very few restrictions.

As seen in Figure 1, the NZ COVID Tracer app was introduced at the end of the first lockdown, which covered the whole of New Zealand, including Auckland [33], and received very little acceptance over June and July, while the country remained at Alert Level 1. Nonetheless, on August 12, 2020, a COVID-19 case with unknown source was discovered in Auckland, resulting in the alert level being raised to Alert Level 3 in the Auckland region and to Alert Level 2 in the rest of the country. This prompted a steep increase in the use of the NZ COVID Tracer app, with the daily number of QR scans growing by two levels of magnitude. However, the level of use decreased considerably once the country returned to Alert Level 1, although it remained considerably higher than before the second lockdown. Relatively high levels of active cases in October and November were almost exclusively imported cases, reflecting the growth of the pandemic overseas [34], and were not associated with higher use of the app.

New Zealand’s COVID-19 outbreak response, as assessed in October 2020, has been recognized as successful [19]. However, in spite of the growth in adoption over the second lockdown in Auckland, the potential of the NZ COVID Tracer app in contributing to this response was not fully realized. As of November 13, 2020, even though 2.3 million users—almost half of the population of the country—were registered with the app, fewer than 1 in 6 of them were using it daily [25]. In an incident in Auckland involving a COVID-19 case visiting business premises on November 7, 2020, the number of potential contacts who could be traced via the app was very low, prompting the Ministry of Health to issue an appeal to citizens to use the app more [35]. In November 2020, improving user engagement with the NZ COVID Tracer app remained a problem for New Zealand.
Semistructured Interviews

The semistructured interview guide (Multimedia Appendix 1) was based on the UTAUT framework [11,12] and focused on effort expectancy (ie, effort associated with using the app), social influence (ie, the extent to which important others are perceived as encouraging the use of the app), facilitating conditions (ie, help available), and habit. Following Walrave et al [11], privacy concerns also received focus. Further, the interview guide emphasized perceived severity of COVID-19 (ie, the perceived consequences of being infected) and perceived susceptibility to COVID-19 (ie, the perceived likelihood of getting infected), concepts borrowed from the protection motivation theory (PMT) [36,37]. The benefits of using a mobile contact tracer app were explored at several levels, following Altmann et al [13], distinguishing benefits to the individual, the family, and society in general. Further, the self-reported patterns of use and the associated experiences were explored in detail, focusing both on current use and on how the approach to using the app by the respondent has changed over time. Moreover, the respondents were asked to project how they are anticipating using the app in the future. Respondents were allowed to deviate from the framework suggested by the interview guide, for as long as the interview remained overall relevant to the research questions of the study.

Participants were recruited using an advertising campaign on Facebook targeting Auckland region residents aged 18 to 64 years (55.19% of the New Zealand population are Facebook users [38]). The campaign invited users of the NZ COVID Tracer app to contribute to the fight against COVID-19 by granting an interview. Further, the participants were entered into a draw to win a token prize. All individuals meeting the criteria who expressed interest in being interviewed were interviewed until the desired sample size was reached; thus, a nonprobability consecutive sampling strategy was used. Following Braun and Clarke [39], the sample size was based on the sample sizes found to be sufficient to answer research questions in similar studies, such as Wessels et al [15] and Byambasuren et al [40], and on pragmatic considerations, such as the ability of the researchers to analyze the resulting volume of data within a reasonable time. The interviews were conducted by the first author over Zoom in late October and early November 2020. The interviews were transcribed in full for analysis.

Analysis

Thematic analysis of interview transcripts was conducted following Braun and Clarke [39]. Both deductive and inductive approaches were used, with deductive coding drawing from the UTAUT and the PMT. Following Braun and Clarke, concepts drawn from the UTAUT and the PMT, as introduced in the previous section, were used as a sensitizing device that was used to attract analysts’ attention to potentially relevant aspects in the data; the aim was to understand user experiences and drivers of app use, rather than to test the UTAUT or the PMT. NVivo 12 (QSR International) was used for coding.
Both coauthors analyzed the data. Both researchers have higher degrees in information technology–related disciplines, with the first coauthor having a stronger technical background and the second coauthor having a background in medicine. Because of the difference in backgrounds, the researchers provided complementary perspectives. The researchers analyzed the data independently, periodically integrating the findings, and resolved differences via discussion.

Ethics

Following the university’s ethics procedures, a low-risk notification was filed. Participants were informed in writing of their rights, such as the right to withdraw from the study at any point and the right to ask questions about the study. After receiving this information, the participants gave consent in writing.

Results

Participants

Interviews were conducted with 34 residents of the Auckland region, with interview durations ranging from 15 to 50 minutes (mean 23, SD 8.9; median 21.4). Participants (Table 1) ranged in age from those in their late teens to those in their early sixties. Even though about half of the participants identified as White New Zealanders of European origin, different ethnicities were represented, including New Zealanders of South Pacific, Indian, Middle Eastern, South American, and Southeast Asian descent. Out of 34 participants, 2 (6%) identified as Māori (Indigenous New Zealanders) and 1 (3%) was a temporary visitor from Europe stranded in New Zealand because of the COVID-19 pandemic (Participant #4). A broad range of careers was represented, from top-middle management to health support work and charity work. Out of 34 participants, 1 (3%) was a female homemaker and 1 (3%) was retired. Likewise, educational backgrounds ranged broadly, from high school completion to master’s degrees. Out of 34 participants, 2 (6%) were unemployed (Participants #1 and #2), having recently lost their jobs because of the pandemic.
Table 1. Characteristics of the participants.

<table>
<thead>
<tr>
<th>Participant No.</th>
<th>Gender</th>
<th>Age range (years)</th>
<th>Ethnicity</th>
<th>Education</th>
<th>Occupation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Male</td>
<td>18-29</td>
<td>NZ European&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Undergraduate diploma</td>
<td>Sales</td>
</tr>
<tr>
<td>2</td>
<td>Female</td>
<td>40-49</td>
<td>NZ European</td>
<td>Bachelor’s degree</td>
<td>Pilot</td>
</tr>
<tr>
<td>3</td>
<td>Female</td>
<td>40-49</td>
<td>Macedonian</td>
<td>Master of Business Administration</td>
<td>Data scientist</td>
</tr>
<tr>
<td>4</td>
<td>Male</td>
<td>18-29</td>
<td>Caucasian</td>
<td>Master’s degree</td>
<td>Biologist</td>
</tr>
<tr>
<td>5</td>
<td>Male</td>
<td>50-59</td>
<td>NZ European</td>
<td>Master’s degree</td>
<td>Training provider</td>
</tr>
<tr>
<td>6</td>
<td>Female</td>
<td>50-59</td>
<td>Latin American</td>
<td>Master’s degree</td>
<td>Interpreter</td>
</tr>
<tr>
<td>7</td>
<td>Male</td>
<td>30-39</td>
<td>Māori</td>
<td>Bachelor’s degree</td>
<td>Support worker</td>
</tr>
<tr>
<td>8</td>
<td>Female</td>
<td>40-49</td>
<td>NZ European</td>
<td>Nursing degree</td>
<td>Nurse</td>
</tr>
<tr>
<td>9</td>
<td>Male</td>
<td>40-49</td>
<td>Middle Eastern</td>
<td>Master’s degree</td>
<td>Account manager</td>
</tr>
<tr>
<td>10</td>
<td>Female</td>
<td>50-59</td>
<td>NZ European</td>
<td>Postgraduate degree</td>
<td>Teacher</td>
</tr>
<tr>
<td>11</td>
<td>Female</td>
<td>40-49</td>
<td>NZ European</td>
<td>Postgraduate degree</td>
<td>Head of compliance</td>
</tr>
<tr>
<td>12</td>
<td>Female</td>
<td>50-59</td>
<td>NZ European</td>
<td>Master’s degree</td>
<td>Homemaker</td>
</tr>
<tr>
<td>13</td>
<td>Female</td>
<td>60-64</td>
<td>NZ European</td>
<td>Bachelor of Arts</td>
<td>Teacher</td>
</tr>
<tr>
<td>14</td>
<td>Female</td>
<td>40-49</td>
<td>NZ European</td>
<td>Postgraduate diploma</td>
<td>Teacher</td>
</tr>
<tr>
<td>15</td>
<td>Female</td>
<td>40-49</td>
<td>Scottish</td>
<td>Postgraduate degree</td>
<td>Health manager</td>
</tr>
<tr>
<td>16</td>
<td>Female</td>
<td>50-59</td>
<td>NZ European</td>
<td>Bachelor’s degree</td>
<td>Charity worker</td>
</tr>
<tr>
<td>17</td>
<td>Female</td>
<td>50-59</td>
<td>Australian</td>
<td>Undergraduate diploma</td>
<td>Customer service</td>
</tr>
<tr>
<td>18</td>
<td>Female</td>
<td>50-59</td>
<td>NZ European</td>
<td>Postgraduate</td>
<td>Administration</td>
</tr>
<tr>
<td>19</td>
<td>Male</td>
<td>30-39</td>
<td>NZ European</td>
<td>High school diploma</td>
<td>Manager</td>
</tr>
<tr>
<td>20</td>
<td>Female</td>
<td>60-64</td>
<td>NZ European</td>
<td>Undergraduate degree</td>
<td>Retired</td>
</tr>
<tr>
<td>21</td>
<td>Female</td>
<td>50-59</td>
<td>European</td>
<td>Postgraduate</td>
<td>Health consultant</td>
</tr>
<tr>
<td>22</td>
<td>Female</td>
<td>50-59</td>
<td>Indian</td>
<td>Bachelor’s degree</td>
<td>Manager</td>
</tr>
<tr>
<td>23</td>
<td>Female</td>
<td>50-59</td>
<td>NZ European</td>
<td>Master’s degree</td>
<td>Teacher</td>
</tr>
<tr>
<td>24</td>
<td>Female</td>
<td>60-64</td>
<td>NZ European</td>
<td>Bachelor’s degree</td>
<td>Dental receptionist</td>
</tr>
<tr>
<td>25</td>
<td>Male</td>
<td>40-49</td>
<td>Samoan</td>
<td>Trade certificate</td>
<td>Human resources</td>
</tr>
<tr>
<td>26</td>
<td>Male</td>
<td>18-29</td>
<td>Korean</td>
<td>High school diploma</td>
<td>University student</td>
</tr>
<tr>
<td>27</td>
<td>Female</td>
<td>30-39</td>
<td>European</td>
<td>Tertiary certificate</td>
<td>Manager</td>
</tr>
<tr>
<td>28</td>
<td>Male</td>
<td>50-59</td>
<td>Scottish</td>
<td>High school diploma</td>
<td>Business owner</td>
</tr>
<tr>
<td>29</td>
<td>Female</td>
<td>40-49</td>
<td>European</td>
<td>University diploma</td>
<td>Sales</td>
</tr>
<tr>
<td>30</td>
<td>Male</td>
<td>18-29</td>
<td>NZ European</td>
<td>High school diploma</td>
<td>Customer service</td>
</tr>
<tr>
<td>31</td>
<td>Male</td>
<td>40-49</td>
<td>NZ European</td>
<td>Master’s degree</td>
<td>Local government</td>
</tr>
<tr>
<td>32</td>
<td>Female</td>
<td>40-49</td>
<td>NZ European</td>
<td>Postgraduate diploma</td>
<td>Health manager</td>
</tr>
<tr>
<td>33</td>
<td>Male</td>
<td>30-39</td>
<td>Māori</td>
<td>High school diploma</td>
<td>System engineer</td>
</tr>
<tr>
<td>34</td>
<td>Female</td>
<td>50-59</td>
<td>Indian</td>
<td>Postgraduate degree</td>
<td>Travel agent</td>
</tr>
</tbody>
</table>

<sup>a</sup>NZ European: White New Zealanders of European origin.

Themes

The thematic analysis resulted in five major themes: perceived benefits, patterns of use, privacy, social influence, and need for collective action. These themes are depicted along with the underlying subthemes and codes in Table 2. The content of the themes is presented in detail in the following sections.
Table 2. Major themes and the underlying subthemes and codes.

<table>
<thead>
<tr>
<th>Themes and subthemes</th>
<th>Codes</th>
</tr>
</thead>
</table>
| **Perceived benefits** | Contact tracing  
|                      | Location data available to the government  
| Society              | Self-isolate  
| Family               | Does not reduce risk  
| Individual           | Peace of mind from protecting others  
|                      | Reduction in uncertainty  
|                      | Better experience in contact tracing  
|                      | Personal diary  
| **Patterns of use**   | Scanning codes  
|                      | Manual entries during a visit  
|                      | Manual entries after a visit  
| Use by yourself      | Manual entries to record information about others  
|                      | Complaining to location manager if code not available  
|                      | Persistence in the face of technical difficulties  
|                      | Maintaining use irrespective of the level of threat  
|                      | Use driven by the level of threat  
| Involving others     |  
| Persistence and changes |  
|                      |  
| Privacy              | Nothing to hide  
|                      | The information is already out there  
|                      | Others care  
|                      | Benefits outweigh privacy concerns  
|                      | Recording information automatically  
|                      | Using GPS location data  
| Social influence      | Encouraged by others  
|                      | Encouraging others  
|                      | Help from others  
|                      | Helping others  
|                      | Seeking help from the internet  
| Need for collective action |  
| Citizenship          | Trust in New Zealand government  
|                      | Civic responsibility  
| Frustration          | Ignoring conspiracy theorists  
|                      | Frustration by others’ lack of use  

**Perceived Benefits**

For most of the participants, the most prominent benefit of the app was supporting contact tracing in the context of controlling the pandemic in a broad sense:

- *Limits the damage and the spread of the virus drastically.* [Participant #4]
- *Feels like a very easy habit to maintain and a very small price to pay, because I can absolutely see the value of having a very quick easy method of tracking people.* [Participant #11]

*If the app is doing what it says it’s doing, then you know, one click, and everybody knows, and you’ve captured the problem, and you know that much faster, and we don’t have to go into the stress of this lockdown business again.* [Participant #22]

Benefits to the family resulting from the individual being able to self-isolate early if at risk of infection were also mentioned:

*It definitely gives me peace of mind because I have young children; I obviously never want to put them in harm’s way.* [Participant #27]
I better remain cautious because of my wider family. My twin sister...if she caught it she would probably die. [Participant #24]

Specific immediate health benefits to the individual using the app as well as greater likelihood to be diagnosed early and, thus, to receive early treatment, resulting in better prospects for the individual, were often not clear to the participants:

I wouldn’t say I’m protecting myself. Because probably it doesn’t reduce my risk in any way. [Participant #20]

It can’t prevent me from catching COVID. I would say it probably more protects the people around me. [Participant #30]

Some of the participants identified reduction in uncertainty as a benefit to the individual using the app (ie, if you are infected, you are likely to know about it faster if you use the app):

I can see the benefit if it happens that I ended up being in contact with somebody that’s got it. I would rather know quicker. [Participant #27]

I feel like it’s protecting me by keeping me in the know. [Participant #21]

Another individual benefit suggested by the participants was the presumed better experience for the individual that was contract traced in the event contact tracing becomes necessary:

If I get sick, I can concentrate on getting well and I can leave contacting people to the government trackers who are paid to do their job. [Participant #2]

I can instantly track back where I’ve gone, and I can provide that information rather than trying to think back. “Oh where was I, did I do that...?” It’s all there. [Participant #31]

Further, some of the participants found benefits that are not associated with COVID-19 virus control. For them, the app acted as a diary making it easy to recollect where they have been:

And it’s quite good for me too, I find, because sometimes I forget where I’ve been. And I look at my app. [Participant #10]

...helps me remember where I have been. [Participant #26]

Overall, while benefits for contact tracing in the context of protecting society in general arose very naturally in the interviews, benefits to the individual using the app were often mentioned only after specific prompting by the interviewer, and different participants had different views on what they are.

Finally, some of the participants perceived the ability of the government to have access to location data and to conduct research using these data as a benefit, although it was not really a benefit because of the privacy features of the app:

Obviously, the government would know where everyone’s going. So that’s like, you know, we’re helping...It is a good thing the government knows where you’ve been. [Participant #26]

It’s giving them data and it’s giving them a platform to start developing what could be needed if this pandemic continues, or if there’s a future one. [Participant #24]

Patterns of Use

Simply scanning Ministry of Health QR codes displayed by businesses was the most common use of the app:

Just scan it sometimes if I’m not in a hurry. I don’t scan it if I can’t be bothered. I try to do it every time. But I’m not religious about it, you know...I suppose it’s just a habit now. [Participant #1]

I walk into a building, grab my phone out, and swipe the tracer. It doesn’t really change my life, it’s not that difficult. [Participant #2]

I plan out when I’m getting out of the car. I have my phone in my hand. Anyway, and I’ll just open the tracer app. And I just, I just walk by, like, you just hardly even have to stop now. [Participant #21]

Manual entries were used when a QR code was not available or could not be scanned easily (eg, it was in an inconvenient location or was laminated, so that reflected light inhibited scanning):

But if they don’t, for some reason, have the QR code available, I always put a manual entry in. [Participant #23]

I just look for the QR codes and just record visits. Every now and then, it hasn’t worked and I’ve recorded manual visit, but that doesn’t seem to be happening so much. [Participant #14]

Another common use of manual entries was to record visits to locations after the fact, when forgetting to scan the QR code:

I try to not forget, wherever I enter any place. But I’ve also caught myself several times, adding it manually after. [Participant #3]

I have forgotten and I’ve moved away. And I remember like half an hour later, I put a manual entry in. [Participant #31]

When a QR code could not be found, some of the participants complained to the location manager; others just did nothing:

I’ve asked people if you got a poster and gone and found it. [Participant #5]

I just go and tell the management...If you don’t display, I’m not comfortable in coming here...That’s what I do if it’s not available, or not prominently displayed. [Participant #34]

When I’m entering the store or going to a place and I see a poster, that kind of reminds me to use it. I must admit, when I haven’t seen a poster, I haven’t used it. It’s not automatic... [Participant #10]

If there’s nothing I don’t bother asking. [Participant #1]

The app allowed users to enter extra information in addition to scanning a QR code. This was occasionally used to record the
presence of others who did not scan for themselves, such as children. Sometimes the presence of others was recorded, possibly without their explicit approval:

If I’ve been with other people, like, particularly if my family have been with me, but they haven’t had the phone with them, like my teenage sons. I’ll put down that they were with me. [Participant #18]

...she’s trying to download and it doesn’t work on her phone. So when she and I are together, that’s okay, because I know we have a record that we’re out together. And I’m doing it. [Participant #11]

Twice I was with somebody who doesn’t have their phone with them...So I just added their name to mine. [Participant #13]

Patterns of use were impacted by updates rolled out by the app developers. For many of the participants, the app was entirely unusable at the beginning (eg, did not scan QR codes well enough or did not scan them at all). One of the participants reported installing and uninstalling the app multiple times, until a version that worked on her phone had been released:

The first couple of versions it was such crap that each time I would end...I would uninstall it and I scream and shout and say, “I am never putting this back on my phone again.”...Then finally...I try to use it every day, I mean, if I go out. I try to always remember to have my phone. [Participant #17]

Few places I tried to use it, it didn’t work. So I put it away for a bit. I think when we went into the next community transmission and they said, or, you know...They seem to have done some work on it and they were raising again that it was good to use. And then I used it a couple of times and it worked. So I thought, “Okay, if it works, then why not?” [Participant #22]

Some of the participants reported considerable resilience in continuing to use the app even while experiencing difficulties scanning the codes or being forcibly logged out of the app and having to recover the password:

Sometimes it’s a little bit slow. And not just starting up. It’s a little bit slow and sometimes forgets my password, and I have to log in again. [Participant #1]

The only thing that has been a bit annoying would be that sometimes I was logged out...other than that it was a seamless transition. I mean, being logged out means that I need to figure out which was my password and I’m terrible at that. But other than that, I figured it’s quite good. [Participant #3]

Participants could be divided according to how their use of the app related to alert levels. While some of them reported continuing to use the app irrespective of the level of the alert—this behavior tended to be associated with the perception of being highly vulnerable to the virus—others reported less consistent use before and after the second lockdown; this was consistent with the pattern suggested in Figure 1:

I am an asthmatic. I would be considered a high-risk group...As a general rule, every time I see a QR code I scan the app. [Participant #31]

I’ve had pneumonia before and I know it’s worse than that...The codes at places didn’t work properly...Halfway through the first lockdown, I actually deleted it. And then once they said that they done a few of the bug fixes and I’ve downloaded it again and I can definitely say that I’ve used it quite a lot since then. [Participant #27]

I didn’t use it for the first couple of weeks. Because I think when it first came out, we were at Level 1 already...I think I forgot the timing, but it might have been after the second lockdown. I’ve been using it absolutely consistently ever since and continue to do so. [Participant #11]

During the lockdowns I always used the app. I mean, because you know it’s a lockdown...But at the moment, with Level 1, I don’t really use the app. And it’s probably because no one else seems to be, whenever I walk into a place. [Participant #25]

Very strong belief in the benefits and the necessity of the app and active steps to encourage others to use the app did not rule out reducing use once the alert level went down:

I was a true supporter to start off with. I’ve scanned wherever I could. I suggested to businesses to go and get it [the QR code for the app]. Made sure that our business got it right away, being a real supersupporter. And then we got over the first wave, we got back to work, there were no cases, so that it sort of died down, I’ve seen businesses removing the scan codes...there were no cases for quite a long time, and my use of the app changed, actually using it a lot less. [Participant #29]

Privacy

Most of the participants did not worry about the app reducing their privacy. Privacy, however, was very prominent in the interviews, with the participants often discussing it with no prompt from the interviewer. Reasons mentioned for not worrying about privacy included the following: (1) the participant has nothing to hide; (2) the participants already perceive themselves as having no privacy as they are tracked via social media, by mobile phone service providers, via transaction records, or by other means; and (3) the participant relies on the app’s privacy features:

I don’t care that they know where I’m at. I don’t think they’d care that much. [Participant #1]

I don’t have anything to hide. I’m not cheating anybody. [Participant #2]

At the end of the day, it’s like these cameras at your workplace. If you got nothing to hide, you don’t have to worry about it. [Participant #19]

...knowing what Google and the likes of Google can do...If someone wants to get something on you, everything is available. [Participant #3]
All of us like to tap into free Wi-Fi everywhere we go. So really, there’s a lot of data out there about us. But so, no. No, I don’t care. [Participant #22]

I’m not worried about the data that it gathers because that data is mine until it is required. [Participant #31]

On the other hand, the existence of others who do care about privacy was often acknowledged, with their preferences mostly accepted as legitimate:

It [the data being recorded] doesn’t bother me. I have a son who’s a lawyer who refuses to…use the app. But not me. It does not bother me. [Participant #16]

You don’t have to put in your personal details, because there were a few [employees at work] that had privacy concerns with the COVID app, they are worried about people watching them, and we pointed out to them that you do not need to put any personal details into the app. You don’t have to put in your first and your last name, you could call yourself Mister 123 if you really wanted to. [Participant #19]

At the same time, some of the participants expressed negative attitudes toward mainstream and social media discourses overemphasizing privacy issues around the app. One of the participants, when asked about others discouraging her from using the app, pointed at one of the major New Zealand newspapers:

Newspapers, like [name of a major New Zealand newspaper], constantly have articles about how it’s taking away our privacy and stuff like that. [Participant #2]

Of the two participants who expressed concerns about privacy, one reported weighing privacy concerns against the benefits of faster contact tracing and deciding that benefits outweigh the risks. The other participant reiterated privacy concerns throughout the interview, but the concerns were not focused on the app and, rather, were about the overall environment, including social media and mobile phone service providers. At the same time, when asked how the app could be improved, the participant suggested an improvement that would reduce, rather than increase, privacy:

I am slightly a conspiracy theorist, but I thought weighing it all up, I felt it was more wise for me to embrace it. [Participant #24]

This is this app, this is that app, there is a lot, you are controlling my life...It is annoying that you need to be booking everywhere you go, like keeping a diary of everything. It’s a form of controlling, Facebook is a form of controlling, Google knows where you, whatever, is controlling, you have no privacy...Yes, I want to help the government and things like that, but at the same time this is a bit like Animal Farm [a novel by George Orwell]. [Participant #6]

You go to places, you need to park your car. Maybe integrate with your car parking... [Participant #6]

Many participants suggested improvements that would reduce privacy (e.g., recording visits automatically using wireless technology, using GPS to track app user location, or using wireless technology to automatically detect and record the proximity of others). None of the participants suggested changes to the app that would increase privacy.

Social Influence and the Need for Collective Action

Overwhelmingly, the participants expressed high levels of trust in the New Zealand government. Often, the fact that information comes from the government was a criterion of its trustworthiness. Using the app was seen as their civic duty and a way to be a good citizen of New Zealand. Moreover, some of the participants framed patterns of behavior in terms of “good” and “bad.” A phrase introduced by the Prime Minister and repeatedly used in communications about the pandemic by her and other officials was commonly mentioned: “the team of the five million” [41,42]:

I go straight to the New Zealand government COVID-19 website. I try to stay away from the internet. [Participant #28]

I feel like I should do it [use the app]...just to be a good citizen of New Zealand united team of 5 million and all of that. [Participant #1]

I feel like you kind of need to set the example...ultimately, if I get sick I would feel that it is my civic responsibility to make sure that anyone who was near me, came in contact with me, would get the health care that they might need. [Participant #33]

In the beginning there was a lot of talk about it in the news, and people saying, “Oh no, that’s stealing all your personal data.” But I don’t see it this way at all. I think this is good data to be used in a good way. [Participant #25]

However, trust in the government was not a prerequisite for using the app. One of the participants, a manager in an industry that was highly critical in maintaining the functioning of the city during lockdowns, expressed very low levels of trust in the government, even suggesting that the government purposefully distorted some of the information related to the pandemic; at the same time, he reported not only using the app but also ensuring that it was installed on mobile devices used by employees, as well as establishing procedures to ensure that visitors to company premises used the app:

I think there’s a lot covered to avoid panic. So, yeah, yeah, it’s hard to trust. [Participant #19]

The flow of social influence was rather complex, involving multiple actors. Participants reported being encouraged by others to use the app. Further, for some, using the app was a requirement at their workplace. Participants also reported encouraging others, in face-to-face settings and online:

My parents actually...said, “You should probably get it [the app].” And I said, “Yeah, yeah.” [Participant #26]

I installed it I when I still had a job...It was a requirement for me as part of my job to use it. But then when I lost my job, I became more flexible with it. Like, it’s not a requirement. Now for me, it’s just...
something that I grew in, it became a habit. [Participant #1]

My husband, he does use it a lot. Whenever we go out, and if sometimes I rush going somewhere, he just stops me and says, “Just scan it.” He always reminds me and encourages me to be more vigilant...My friends, I just told them to install it, maybe they have, I do not know...When the app was first introduced, I sent messages to a lot of my contacts. [Participant #34]

Businesses displaying the codes occasionally encouraged app use; in their turn, some of the participants actively engaged with businesses when a code could not be found or was unusable, as already highlighted in the Patterns of Use section:

Now [after the second lockdown ended] I’m not really using it, no. I only use it if the shopkeeper asks me to use it and then I will say, “Of course I’ll use it.” [Participant #26]

We do actively, when we have customers coming to pick items, my receptionist will say... “Can you scan in please?” We literally say it to everyone who walks in...To be honest, people don’t scan when they are coming in, the instance you say, “Would you mind scanning,” nobody’s ever thrown anything back at us, they just say, “No worries,” they do it. [Participant #19]

Some of the participants relied on internet-based resources, such as the Ministry of Health website, when having problems using the app. Help was not always readily available:

...just go to the [Ministry of Health] website for help. [Participant #25]

I look at the Ministry website... [Participant #8]

...and I’m going to COVID-19 website and I cannot even find where the bloody test centers are. [Participant #9]

I even tried to contact someone and say, “You know, it’s not working.” And then I figured out, I guess, you know, it’s not working. So they’re getting too much communication...but once it has started working, like once I started using it the second time round, haven’t had any need to contact anyone for help. [Participant #22]

Family members and colleagues may have been a readier source of help:

I’ll ask my 21-year-old son. He is quite tech savvy. So I utilize expertise inside my family. [Participant #28]

I would ask my husband. [Participant #27]

I was coming back by bus...and it was not scanning...somebody was sitting next to me, my colleague...I passed my phone to her, and she scanned it for me, because she was a little bit closer. [Participant #34]

Further, strangers helped each other:

I was in the supermarket and then was having difficulty. And the guy said, “Oh, you know, you need to turn on the camera.” [Participant #6]

I do see a lot of young people helping older people use it. [Participant #8]

I’ve helped a couple of people to download it. [Participant #21]

The existence of “conspiracy theorists” raising, from the perspective of the participants, unreasonable or untrue privacy concerns, was occasionally acknowledged. Although one of the participants described himself as a “conspiracy theorist,” he still used the app, judging that the benefits were greater than the risks, as introduced in the Privacy section:

...some think that the COVID-19 scanning app was taking the wrong information on each individual to hand more information to the government power. I don’t believe that myself. Those conspiracy theorists inside my own family unit, I took it with a grain of salt. So meaning...I believe in the app. [Participant #28]

My mother-in-law, she’s saying, you know, “Don’t use that, don’t install it, because they’re tracking all your data and everything.”...I’m ignoring her for a lot of things. [Participant #18]

Many of the participants were concerned about the behavior of others in using the app and in reducing the risk of COVID-19 spread in other ways. The realization that protecting the country from the pandemic depended on collective action was rather strong. Often, rather than expecting the authorities to improve technological capabilities or ease of use of the app, the participants highlighted the need to encourage its broader use:

I find it quite frustrating going to places and seeing people around me who, you know, are walking on without even bothering to scan. [Participant #23]

I think New Zealand’s getting very complacent. [Participant #2]

My friends or family don’t use it. Full stop. [Participant #19]

My concern is that we have a lot of people just disregarding the impact of COVID. [Participant #8]

I think, some shops are deliberately making it [the QR code] hard to find. [Participant #5]

Discussion

Principal Findings

The main contribution of this study of adoption and use of a mobile contact tracing app is that it is based on data reflecting real user experiences, rather than on perceptions of individuals who are yet to use such an app. Prior studies predicting mobile contact tracing app adoption and use relied on data obtained from nonusers.

The results of this study are consistent with the finding by Trang et al [8] that perceptions of benefits for society as a whole are likely to drive the use of a mobile contact tracing app. However,
the results also suggest that such benefits are mainly relevant when the level of threat to society is high. For many individuals, but not all, the logic of taking individual action to protect the society on which the individuals depend is powerful enough to drive sustained use only when the threat to the society is salient enough.

The results of our study are consistent with Altmann et al [13] in suggesting that trust in the government helps to promote mobile tracing app use. Nonetheless, the finding by Altmann et al [13] that concerns about government surveillance are very important were not confirmed by our study. This may be, in part, because our study covered only the users of the app, who were likely to fall into the “advocates” category following Trang et al [8]. Assuming the participants of this study were “advocates,” the finding that privacy did not matter for them is consistent with the results by Trang et al [8]. The results of our study are consistent with Wiertz et al [9], who also found little evidence that privacy is highly relevant, as well as with the body of literature on the privacy paradox [43], which suggests that in actual use, users are prepared to trade their privacy even for rather small benefits.

The NZ COVID Tracer app was designed in such a way that its use, or nonuse, was highly visible. Thus, social influence, found to have an effect by Walrave et al [11], could be highly influential. Nonetheless, for most of the participants, social influence by peers appeared to play a secondary role in driving their app use. Indeed, some of them continued to use the app while surrounded by nonusers. For them, social influence was coming from the government, not from the peers. At the same time, the results indicate that organizations may be effective in promoting the use of the mobile tracing app by their employees: employees who are not users are likely to comply to become users, rather than resist.

Our study found no indications that an app overseen by an independent entity, rather than by the government, would be better accepted or used more, and in this respect, our results did not confirm the results by Wiertz et al [9]. Indeed, the discourse by the app users around good citizenship and civic duty as reasons for using the app suggested that oversight by the government was a good choice in the New Zealand context. Nonetheless, this conclusion has to be confirmed by a study of nonusers of the app.

The study by Walrave et al [11] did not find effort expectancy to be an important factor. Our results were consistent with this finding. Determined users of the app were prepared to persist in the face of technical difficulties. This is not to suggest that effort expectancy is irrelevant; however, there was little evidence to suggest that, after the initial bugs were fixed, making the app even more effortless to use would result in significantly higher adoption and use.

The implications for practice are that appeals to civic responsibility are likely to drive the use of a mobile tracing app under the conditions of high threat, as citizens “rally around the flag.” Under the likely scenario of COVID-19 remaining endemic and requiring ongoing vigilance over the long term, other mechanisms promoting the use of mobile tracing apps may be needed, such as “nudging” [44] (eg, offering incentives). Further, the results suggest that privacy is not an important concern for many users. Having access to more detailed information faster would benefit contact tracing, enabling faster isolation of probable cases and, thus, better control of the pandemic. Therefore, compared with a mobile tracing app with uniformly restrictive privacy features, an app with flexible privacy settings allowing users to set their optimal levels of privacy—thus allowing users who are less concerned about privacy to opt in to provide more detailed information faster—may be more appropriate.

The value of comparing responses between countries in informing decision making in the COVID-19 pandemic has been highlighted by Pearce et al [45], who characterized health policy responses in different jurisdictions as “numerous natural experiments in progress” (page 1059 in Pearce et al [45]). The case of New Zealand is particularly valuable in this respect because it presents an example of a successful response [46,47] achieved in a country with western political culture [48]. As such, the New Zealand experience in managing the pandemic has received a lot of attention in international literature [49-56]. Our study contributes to this body of literature by focusing on the experiences of the users of the NZ COVID Tracer app. However, the results of this study, as well as of other studies of the New Zealand experience in managing the COVID-19 pandemic, cannot be mechanically applied to other contexts. Rather, as for most qualitative studies, the process of case-to-case transfer [57] should apply: the readers and the consumers of the research should compare their context of interest to the New Zealand context and judge the extent to which the findings apply to their situation (page 1453 in Polit and Beck [57]). A broad description of the New Zealand context as it applies to the management of the COVID-19 pandemic is given by Jefferies et al [54], who assert that the New Zealand response to COVID-19 “has international relevance, particularly for other island nations, high-income and western settings” (page e613 in Jefferies et al [54]). Further, aspects of the context immediately relevant to the research questions of our study, such as the app design and the way it was introduced, relying on persuasion rather than on mandates, are described in the initial sections of this paper.

Conclusions

Appeals to civic responsibility are likely to drive the use of a mobile contact tracing app under the conditions of high threat. Under the likely scenario of COVID-19 remaining endemic and requiring ongoing vigilance over the long term, other mechanisms promoting the use of mobile contact tracing apps may be needed, such as offering incentives. As privacy is not an important concern for many users, flexible privacy settings in mobile contact tracing apps allowing users to set their optimal levels of privacy may be appropriate.
Conflicts of Interest

None declared.

Multimedia Appendix 1

Interview guide.

[DOCX File, 21 KB - mhealth_v9i9e26318_app1.docx ]

References


17. Thomson K, Brouwers C, Damman OC, de Bruijn MC, Timmermans DR, Melles M. How health care professionals evaluate a digital intervention to improve medication adherence: Qualitative exploratory study. JMI Health Humanities 2018 Apr 30;1-19 [FREE Full text] [doi: 10.2196/humanfactors.8948] [Medline: 29463494]


https://mhealth.jmir.org/2021/9/e26318

Tretiakov & Hunter

JMIR MHEALTH AND UHEALTH


Abbreviations

PMT: protection motivation theory
QR: Quick Response
UTAUT: unified theory of acceptance and use of technology

©Alexei Tretiakov, Inga Hunter. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 08.09.2021. 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 https://mhealth.jmir.org/, as well as this copyright and license information must be included.
Diagnostic Accuracy of Smartphone-Based Audiometry for Hearing Loss Detection: Meta-analysis

Chih-Hao Chen¹, MD; Heng-Yu Haley Lin²*, MPH, MD; Mao-Che Wang¹,³,⁴*, MD, PhD; Yuan-Chia Chu⁵,⁶,⁷*, PhD; Chun-Yu Chang⁸*, MD; Chi-Yuan Huang¹,³*, MD, PhD; Yen-Fu Cheng¹,³,⁹,¹⁰*, MD, PhD

¹Department of Otolaryngology-Head and Neck Surgery, Taipei Veterans General Hospital, Taiwan, Taipei City, Taiwan
²Department of Medical Education, Taipei Veterans General Hospital, Taipei City, Taiwan
³Faculty of Medicine, National Yang Ming Chiao Tung University, Taipei City, Taiwan
⁴Institute of Hospital and Health Care Administration, National Yang Ming Chiao Tung University, Taipei City, Taiwan
⁵Information Management Office, Taipei Veterans General Hospital, Taipei City, Taiwan
⁶Medical AI Development Center, Taipei Veterans General Hospital, Taipei City, Taiwan
⁷Department of Information Management, National Taipei University of Nursing and Health Sciences, Taipei City, Taiwan
⁸Department of Anesthesiology, Taipei Tzu Chi Hospital, Buddhist Tzu Chi Medical Foundation, New Taipei City, Taiwan
⁹Department of Medical Research, Taipei Veterans General Hospital, Taipei City, Taiwan
¹⁰Institute of Brain Science, National Yang Ming Chiao Tung University, Taipei City, Taiwan

*these authors contributed equally

Corresponding Author:
Yen-Fu Cheng, MD, PhD
Department of Otolaryngology-Head and Neck Surgery
Taipei Veterans General Hospital
Taiwan
No.201, Sec. 2, Shipai Rd., Beitou District
Taipei City, 11217
Taiwan
Phone: 886 2 2871 2121 ext 1292
Email: yfcheng2@vghtpe.gov.tw

Abstract

Background: Hearing loss is one of the most common disabilities worldwide and affects both individual and public health. Pure tone audiometry (PTA) is the gold standard for hearing assessment, but it is often not available in many settings, given its high cost and demand for human resources. Smartphone-based audiometry may be equally effective and can improve access to adequate hearing evaluations.

Objective: The aim of this systematic review is to synthesize the current evidence of the role of smartphone-based audiometry in hearing assessments and further explore the factors that influence its diagnostic accuracy.

Methods: Five databases—PubMed, Embase, Cochrane Library, Web of Science, and Scopus—were queried to identify original studies that examined the diagnostic accuracy of hearing loss measurement using smartphone-based devices with conventional PTA as a reference test. A bivariate random-effects meta-analysis was performed to estimate the pooled sensitivity and specificity. The factors associated with diagnostic accuracy were identified using a bivariate meta-regression model. Study quality was assessed using the Quality Assessment of Diagnostic Accuracy Studies-2 tool.

Results: In total, 25 studies with a total of 4470 patients were included in the meta-analysis. The overall sensitivity, specificity, and area under the receiver operating characteristic curve for smartphone-based audiometry were 89% (95% CI 83%-93%), 93% (95% CI 87%-97%), and 0.96 (95% CI 0.93-0.97), respectively; the corresponding values for the smartphone-based speech recognition test were 91% (95% CI 86%-94%), 88% (95% CI 75%-94%), and 0.93 (95% CI 0.90-0.95), respectively. Meta-regression analysis revealed that patient age, equipment used, and the presence of soundproof booths were significantly related to diagnostic accuracy.

Conclusions: We have presented comprehensive evidence regarding the effectiveness of smartphone-based tests in diagnosing hearing loss. Smartphone-based audiometry may serve as an accurate and accessible approach to hearing evaluations, especially in settings where conventional PTA is unavailable.
Introduction

Background

Hearing loss is one of the most common disabilities affecting both individual and public health. Hearing loss has been linked to multiple physical [1,2], cognitive [3,4], and psychosocial [5,6] outcomes and is associated with problematic health care use and higher medical expenses [7]. According to previous studies and World Health Organization estimates, more than 5% of the world’s population is affected by hearing impairment, especially older adults aged above 65 years [8-10]. Notably, the prevalence of hearing loss is 50% higher in low-income countries [11]. Within the disease spectrum of hearing impairment, a considerable number of cases, such as those involving idiopathic sudden sensorineural hearing loss (SSNHL) and noise-induced hearing loss, are preventable and can be treated effectively and in a timely manner [12-14].

Pure tone audiometry (PTA) is the gold standard for current hearing assessment batteries [15]. However, this measurement is often unavailable, given its demanding nature with regard to equipment, certified personnel, space, and expenses, particularly in settings such as primary care practices, urgent care, and in low- and middle-income countries [16-18]. As hearing loss has been identified as the single largest potentially modifiable risk factor for dementia in midlife [19] and most patients with hearing impairment can benefit from timely interventions, a more accessible and equally accurate approach to hearing assessment is warranted. Great efforts have been made to create more cost-effective devices and automate audiologic examinations, resulting in the rapid development of smartphone audiometry. Because of the universal availability of mobile technology and cellular networks, smartphone-based hearing tests may provide an adequate assessment of hearing as an alternative to conventional PTA and assist large-scale hearing screening [16,20,21].

Objective

A considerable number of smartphone apps have been introduced for hearing screening [22,23], evaluation [24-26], and even rehabilitation and care [27,28] in recent years, and previous research has compared the performance of these apps with standard audiometry [21,29]. However, these studies were heterogeneous in terms of study design, use of equipment, and baseline characteristics of the participants, which resulted in inconsistent data on the diagnostic performance of smartphone audiometry. The aim of this study is to synthesize the most updated and comprehensive evidence of the diagnostic value of smartphone-based hearing assessments for hearing loss. We performed a meta-analysis with meta-regression to summarize the diagnostic accuracy of smartphone audiometry and investigated the factors affecting the test results. We aim to provide more definitive evidence of the utility of smartphone audiometry in clinical application in the future.

Methods

Study Design

This meta-analysis followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-analyses) Diagnostic Test Accuracy Studies statement [30].

Search Strategy

In all, five databases—PubMed, Embase, Cochrane Library, Web of Science and Scopus—were searched from inception through January 15, 2021, by 2 authors (CHC and HYHL). The Boolean operator OR was used to cover similar concepts, whereas AND was used to intersect different concepts. We used a combination of Medical Subject Headings and text words to create three subsets of citations: the first included studies on hearing loss, the second included studies on smartphones, and the third included studies on the concept of diagnosis, audiometry, and self-examination. The detailed search strategy is presented in Multimedia Appendix 1. The identified citations were imported into the reference software and screened by title, abstract, and keyword. Potentially eligible records were then subjected to a full-text review.

Eligibility Criteria

The included studies were selected based on the following criteria: (1) PTA was used as a reference test, (2) audiometry was used on smart devices (ie, PTA and speech recognition audiometry) as an index test, and (3) adequate information was reported on diagnostic accuracy (ie, prevalence, sensitivity, and specificity) to quantify the effect estimates for meta-analysis. Studies with outcomes that did not relate to the diagnostic accuracy of the index test or did not provide enough information for meta-analysis were excluded. We did not exclude studies based on country, language, or publication date.

Study Selection and Data Extraction

All studies were fully reviewed and selected by 2 authors (CHC and HYHL). If there were any disagreements in the study selection, they were resolved by a third author (YFC) through consensus or discussion. The extracted data included the author’s name, publication year, country, test setting, number of patients, mean age of the study population, operating system of the smart device, equipment used during the examination, and use of a soundproof booth. The disease population was defined as comprising patients with abnormal reference test results in each study. The quantitative data were either extracted directly from raw data or converted from the diagnostic parameters (ie, sensitivity, specificity, and prevalence) in each study to construct...
standard diagnostic test 2×2 tables containing true-positive, false-positive, false-negative, and true-negative samples for the index text.

**Study Quality Assessment**

The quality of the included studies was assessed by 2 authors (CHC and HYHL) using the Quality Assessment of Diagnostic Accuracy Studies-2 tool. A third reviewer (YFC) resolved disagreements regarding the methodological quality through consensus or discussion.

**Statistical Analysis**

**Overview**

Sensitivity and specificity were calculated for each extracted data set. A negative correlation between sensitivity and specificity caused by different thresholds was observed; therefore, we adopted a bivariate random-effects model to estimate the pooled sensitivity and specificity of the index test and to account for the heterogeneity that commonly exists in meta-analyses of diagnostic accuracy tests [31]. The bivariate random-effects model assumes logit-transformed sensitivity and specificity as bivariable distributions, and it also considers the threshold effect, which is an indication of the trade-off phenomenon in most diagnostic accuracy tests because the threshold differs among studies [32]. To investigate the covariate among the index studies, bivariate meta-regression analysis was performed [33], one at a time. For the covariate effect on age, we divided the studies into child, elderly, and adult groups. People aged below 18 years were considered to be in the child group, whereas people aged above 65 years were considered to be in the elderly group based on the World Health Organization criteria [34]. First, we examined whether the covariate caused variance in the sensitivity and specificity measures. The following likelihood-ratio chi-square test was used to determine whether the covariate served as a significant variable by testing the hypothesis that these covariates do not explain variance in the logit-transformed pairs of sensitivity and specificity. To further illustrate the diagnostic accuracy and compare the discriminatory properties, we constructed hierarchical summary receiver operating characteristic curves for the overall result as well as the subgroup results identified by the meta-regression analysis by accounting for the correlation in the data through a hierarchical approach. To deal with zero observations in the 2×2 contingency tables, 0.5 was added to each cell to reduce the influence of small studies. We calculated 95% CIs on the basis of the binominal distribution of the truly positive and truly negative samples. Publication bias was examined using the Deeks funnel plot using the natural logarithm of the diagnostic odds ratio against 1/(effective sample size)½ to plot the asymmetry of the included studies. Effective sample size (ESS) was calculated by the number of examinees who were diseased (n1) and not diseased (n2) as:

\[
    \text{ESS} = \frac{4n_1 n_2}{n_1 + n_2} \quad (1)
\]

ESS considers that unequal numbers of individuals who are diseased and not diseased reduce the precision of test accuracy estimates [31,35]. A \( P<.10 \) for the regression tests suggests significant publication bias. Statistical analyses were conducted using Stata version 15.0 (StataCorp), with the midas and metandi commands. All statistical tests were two-sided, and \( P<.05 \) was considered statistically significant.

**Results**

**Study Identification and Selection**

A total of 1157 studies were identified through the databases. Of the 1157 studies, 648 (56%) remained in the preliminary search after the removal of 509 (44%) duplicates. Of the 648 studies, 584 (90.1%) were excluded after 2 authors (CHC and HYHL) screened the titles and abstracts; a total of 9.9% (64/648) of studies then underwent full-text review. Of the 64 studies, 39 (61%) were excluded because of the following reasons: insufficient data for meta-analysis, index tests not used, inappropriate study design, or unavailability of the full text. As a result, of the 64 studies, 25 (39%) studies with a total of 4470 patients were included in the meta-analysis. The detailed PRISMA flow diagram is presented in Figure 1.
Study Characteristics

Of the 25 studies, 21 were prospective [10,21,22,29,36-52], 1 was retrospective [53], and the remaining 3 studies did not report the study design [23,54,55]. In all, 20 studies used PTA as the index test [10,21,22,29,36-42,44-46,48-52], whereas the remaining 5 studies applied a speech recognition test (SRT) as the index test [23,43,47,53,55]. A total of 4 studies enrolled elderly participants [10,36,37,39], whereas 7 studies included children [21-23,38,41,49,55], and 13 studies enrolled adult participants [29,40,42-44,46,48-54]. The remaining study did not report the age of the study population [45]. In all, 15 studies operated audiometry through an iPhone (Apple Inc) operating system–based app [10,22,29,36,37,39,40,45-48,50,52,54,55], whereas the remaining 10 used an Android (Google LLC) operating system–based audiometry app [21,23,38,41-44,49,51,53]. A total of 15 studies used headphones for testing [10,21,23,38,41-47,49,50,52,54], 9 studies used earphones for the examination [22,29,36,37,39,40,48,53,55], and 1 study did not mention the equipment used [51]. In all, 6 studies conducted the examination in a soundproof booth [10,49-51,53,55], 18 studies did not use a soundproof booth to conduct the examination [21-23,29,36-47,52,54], and 1 study did not report whether the test was conducted in a soundproof booth [48]. A total of 4 studies [45,46,49,50] conducted the index test among different independent populations, yielding a total of 30 study groups for the analysis. Further information regarding the included study populations and statistics is presented in Multimedia Appendices 2 and 3.

Quality and Risk-of-Bias Assessment

Quality Assessment of Diagnostic Accuracy Studies-2 scores were used to evaluate the quality of the included studies. Regarding the evaluation of the risk of bias, all the studies carried out index studies without knowing the results of the reference test in advance and set the threshold before testing. A total of 4 studies did not clearly describe the sequence...
between the index and reference tests [47,53-55]. Regarding the evaluation of applicability, 1 study enrolled patients with underlying otitis media [38], and another 2 studies included patients with SSNHL [29,52]. In all, 5 studies used unmarketed apps as index tests. A detailed assessment and an overall picture of the methodological quality of the included studies are presented in Figure 2.

**Figure 2.** Quality assessment results based on the Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) guidelines.

### Overall Diagnostic Performance

Overall, the studies using a smartphone app with PTA showed a sensitivity of 89% (95% CI 83%-93%) and specificity of 93% (95% CI 87%-97%), whereas studies using an app involving SRT revealed a sensitivity of 91% (95% CI 86%-94%) and specificity of 88% (95% CI 75%-94%). The hierarchical summary receiver operating characteristic curves with summary points for both PTA and SRT are shown in Figures 3 and 4. The predicted values for the area under the receiver operating characteristic curve (AUC) for the PTA and SRT measures were 0.96 (95% CI 0.93-0.97) and 0.93 (95% CI 0.90-0.95), respectively.
Figure 3. The HSROC for pure tone audiometry. HSROC: hierarchical summary receiver operating characteristic.

Figure 4. The HSROC for the speech recognition test. HSROC: hierarchical summary receiver operating characteristic.

Meta-Regression and Subgroup Analysis

The bivariate meta-regression analysis showed a significant influence of the operating system on sensitivity (88% vs 89%). The likelihood-ratio chi-square test revealed that elderly group ($\chi^2=85.9; P<.001$), child group ($\chi^2=62.9; P<.001$), headphone use ($\chi^2=17.8; P<.001$), and soundproof booth use ($\chi^2=19.5; P<.001$) were significant covariates causing variance between paired sensitivity and specificity, whereas the operating system did not reveal such a difference ($\chi^2=0.02; P=.99$). The AUC values for the elderly group versus the adult group were 0.90...
(95% CI 0.87-0.92) versus 0.96 (95% CI 0.94-0.97), respectively, whereas the AUC values for the child group versus the adult group were 0.90 (95% CI 0.88-0.93) versus 0.96 (95% CI 0.94-0.97), respectively. The AUC values for the headphone group versus the earphone group were 0.96 (95% CI 0.94-0.97) versus 0.92 (95% CI 0.89-0.94), respectively. The AUC values for the soundproof booth group versus the non–soundproof booth group were 0.99 (95% CI 0.97-0.99) versus 0.94 (95% CI 0.91-0.96), respectively. The AUC values for the iPhone operating system group versus the Android operating system group were 0.95 (95% CI 0.93-0.97) versus 0.96 (95% CI 0.94-0.97), respectively. The detailed results are presented in Table 1.

Table 1. Results of the bivariate meta-regression analysis (N=25).

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Number</th>
<th>Sensitivity (95% CI)</th>
<th>P value</th>
<th>Specificity (95% CI)</th>
<th>P value</th>
<th>Likelihood-ratio test</th>
<th>Chi-square (df)</th>
<th>Area under the curve (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Elderly [10,36,37,39]</td>
<td>4</td>
<td>0.77 (0.55-0.99)</td>
<td>.04</td>
<td>0.92 (0.80-1.00)</td>
<td>.99</td>
<td>&lt;.001*</td>
<td>85.9 (1)</td>
<td>0.90 (0.87-0.92)</td>
</tr>
<tr>
<td>Child [21,22,38,41,49]</td>
<td>5</td>
<td>0.85 (0.69-1.00)</td>
<td>.10</td>
<td>0.96 (0.89-1.00)</td>
<td>.34</td>
<td>&lt;.001</td>
<td>62.9(1)</td>
<td>0.90 (0.88-0.93)</td>
</tr>
<tr>
<td>Adult [29,40,42,44,46,48-52,54]</td>
<td>14</td>
<td>0.90 (0.85-0.96)</td>
<td>—</td>
<td>0.91 (0.82-1.00)</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>0.96 (0.94-0.97)</td>
</tr>
<tr>
<td><strong>Operating system</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>iPhone operating system</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[10,22,29,36,37,39,40,45,46,48,50,52,54]</td>
<td>17</td>
<td>0.88 (0.82-0.94)</td>
<td>.04</td>
<td>0.93 (0.87-0.99)</td>
<td>.51</td>
<td>.99</td>
<td>0.02(1)</td>
<td>0.95 (0.93-0.97)</td>
</tr>
<tr>
<td>Android [21,38,41,42,44,49,51]</td>
<td>8</td>
<td>0.89 (0.81-0.97)</td>
<td>—</td>
<td>0.93 (0.85-1.00)</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>0.96 (0.94-0.97)</td>
</tr>
<tr>
<td><strong>Equipment</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Headphone [10,21,38,41,42,44-47,49,50,52,54]</td>
<td>17</td>
<td>0.91 (0.87-0.95)</td>
<td>.85</td>
<td>0.89 (0.82-0.97)</td>
<td>.05</td>
<td>&lt;.001</td>
<td>17.8(1)</td>
<td>0.96 (0.94-0.97)</td>
</tr>
<tr>
<td>Earphone [22,29,36,37,39,40,48]</td>
<td>7</td>
<td>0.80 (0.65-0.95)</td>
<td>—</td>
<td>0.97 (0.92-1.00)</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>0.92 (0.89-0.94)</td>
</tr>
<tr>
<td><strong>Soundproof booth</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes [10,49-51]</td>
<td>6</td>
<td>0.95 (0.90-1.00)</td>
<td>.72</td>
<td>0.95 (0.87-1.00)</td>
<td>.83</td>
<td>&lt;.001</td>
<td>19.5(1)</td>
<td>0.99 (0.97-0.99)</td>
</tr>
<tr>
<td>No [21,22,29,36-42,44-46,52,54]</td>
<td>18</td>
<td>0.87 (0.82-0.93)</td>
<td>—</td>
<td>0.91 (0.85-0.98)</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>0.94 (0.91-0.96)</td>
</tr>
</tbody>
</table>

*Significant P<.05.

bReference of likelihood-ratio chi-square test.

Publication Bias

The Deeks funnel plot revealed no asymmetrical distribution for the included studies, and the regression test did not show a significant publication bias (P=.71; Figure 5).
Discussion

Principal Findings

In this study, we performed a meta-analysis to estimate the pooled diagnostic accuracy of smartphone-based hearing tests using conventional PTA as the gold standard. The overall sensitivity of smartphone-based audiometry was 89%, the specificity was 93%, and the AUC was 0.95, which suggested outstanding diagnostic performance for identifying hearing loss using PTA as the gold standard test. When using the SRT as the gold standard test, our results showed a sensitivity of 91%, specificity of 88%, and AUC of 0.93, which also indicated excellent diagnostic accuracy. On the basis of the results of the bivariate meta-regression analysis, we found that participant age, equipment used, and the use of a soundproof booth significantly affected the diagnostic accuracy of smartphone-based audiometry, whereas the operating system of the smartphone did not. To our knowledge, this is the first meta-analysis that provides comprehensive evidence of the diagnostic performance of a smartphone-based approach to detecting hearing loss.

PTA assesses a person’s lowest threshold response to pure tone stimuli at various frequencies [56]. It is still considered the gold standard test for audiologic examinations and provides information regarding the severity and type of hearing loss. According to the American National Standards Institute specifications, there are four types of PTA. Type 1 audiometry (advanced clinical or research) involves a completely equipped audiometer that can conduct both air and bone conduction tests. Type 2 (clinical) fits the same specifications as type 1, except for the requirement of loudspeaker equipment. Portable audiometers without speech-comprehension measurements are classified as type 3 (diagnostic), whereas type 4 (screening) consists of screening audiometers with the basic functions of a hearing test [56]. Although types 1 and 2 are considered the most informative and comprehensive audiometry, they are often not available in many settings, especially in resource-limited areas such as in low- and middle-income countries and rural regions. Even in resource-rich countries, standard PTA is not usually available at primary care practices [17]. Standard PTA tests require certified professionals to administer them, whereas audiologic training is generally lacking in resource-limited countries—there is less than one audiologist for every 1 million people according to previous studies [23,53,54]. Furthermore, the equipment for conventional PTA, including a soundproof booth and a calibrated audiometer, involves both cost and space. The demanding nature of conventional PTA may result in its low accessibility and further affect the generality of hearing screening and quality of hearing care [57,58].

In recent years, mobile health devices have evolved rapidly, as have smartphone-based hearing approaches. Smartphone-based audiometry is a cost-effective, convenient, and reliable tool for screening hearing loss. As smartphones are common in the modern society and the apps are very accessible, given their low cost or no cost, smartphone-based hearing tests could potentially bridge the gap between patients with hearing loss and adequate audiologic assessments and, potentially, hearing care. Previous studies have confirmed that such apps were able to provide basic hearing screening wherever the individual was located as long as the location met the required level of background noise, reducing the need to travel and pay for a hearing examination [59,60]. These smartphone-based hearing tests are usually designed to be user friendly because automated diagnostic audiometry simplifies complex audiologic protocols, allowing their use by nonprofessionals [61,62]. Studies have also described the use of smartphone-based audiometry in settings such as primary care practices and community health clinics for routine hearing screening to identify potentially
handicapping hearing loss [59]. The findings of this study confirm that the diagnostic performance of smartphone-based audiometry aligns perfectly with conventional PTA in identifying hearing loss and adds to previous research with a larger pooled sample size and systematic scope.

Although this study highlights the high diagnostic value of smartphone-based hearing tests and their promising role in hearing screening, we identified several possible variables that may influence diagnostic performance. First, the accuracy of smartphone-based audiometry was lower in elderly individuals and children. This may suggest technical barriers between smart devices and elderly individuals and children. Previous studies have found that factors such as prevalent vision impairments and slower learning curves in managing technological devices because of lack of experience and functional decline may contribute to the higher level of difficulties when using smartphone-based apps among elderly populations [60,63]. At the same time, a previous study also found that children achieved lower accuracy in PTA [64]. Our results also showed that headphone use during the hearing examination may improve the diagnostic accuracy of smartphone-based audiometry. Earphones are a required component in standard audiometry because they prevent the collapse of the external ear canal and reduce the level of ambient noise [65-67]. However, if the participant does not insert the earphone correctly in the automated examination, it could be a problem. A previous study showed that earphone positioning may affect audiologic assessment results and whether the earphone is positioned by a trained examiner or by the examinee may affect audiologic assessment [68]. The negative effects of background noise may further support our finding that examinations conducted in soundproof booths have better diagnostic accuracy. The influence of ambient noise, which results in erroneous test results of smartphone audiometry, has been reported in previous studies [69,70], leading to the conclusion that the use of soundproof booths may increase the diagnostic accuracy of smartphone-based hearing tests [71]. Although some of the included studies reported comparable results of hearing assessments outside of a soundproof booth with passive attenuation and simultaneous ambient noise monitoring [71-73], most of the studies did not provide information regarding the management of ambient noise. The diagnostic value of this subgroup, however, still appeared feasible, because their AUC values exceeded the cutoff point of 0.9 [74]. In summary, our findings suggest that adequate adjustment of the variables that significantly affect the accuracy of smartphone-based audiometry may improve its diagnostic performance in diagnosing hearing loss. Approaches such as adding instructions regarding the examination protocol and correct use of earphones, providing customized audiologist consultations for elderly individuals, improving the app’s function in monitoring environmental noise, and regularly collecting feedback from users could be added to the current implementation methods.

Limitations

This study has several limitations. First, as in most studies of diagnostic test accuracy, different thresholds exist among the studies and may have caused the threshold effect. A priori test calculation of the correlation between sensitivity and specificity revealed a negative result, confirming the threshold effect in this study. Therefore, we adopted the bivariate random-effects model to account for the cross-study threshold difference as suggested by previous studies [31,32]. Second, there was heterogeneity regarding the study designs, test protocols, and reference PTA thresholds for diagnosing hearing loss across the included studies, which may have biased the results when pooling them into the meta-analyses. Future studies with homogenous gold standards and uniform protocols for smartphone-based hearing tests are needed. Third, ambient noise monitoring is a key factor influencing the accuracy of audiometry [75]. Although most of the included studies did monitor noise, no data on the accuracy without ambient noise monitoring were provided. As a result, we were not able to perform the meta-regression analysis according to this factor. Fourth, frequency may act as a confounder, but most of the included studies did not provide diagnostic accuracy for each frequency; therefore, we could evaluate the diagnostic performance of smartphone audiometry only with the average threshold calculated from the frequencies. Fifth, most of the included studies did not describe the masking procedure, possibly because the included studies sampled healthy people, and the threshold difference between bilateral ears could hardly exceed 40 dB. In addition, some smartphone audiometry methods did not provide an automasking procedure during the automated examination. We suggest that future studies describe the masking procedure in detail, regardless of whether it is used. Sixth, of the 25 included studies, most did not describe the calibration method, whereas 9 (36%) used reference equivalent threshold sound pressure levels. A previous study revealed that the differences in hearing thresholds among the device models were significant, which might directly result from the biological calibration method used to determine the reference sound level [75]. Calibration information was lacking, possibly because of the intrinsic lack of a calibration function in the app. We suggest that future studies address this issue. Finally, some included studies enrolled patients with underlying diseases such as otitis media and SSNHL. Although, ideally, subgroup analyses should have been performed for these unique studies for more accurate results, we were not able to implement this investigation because of the scarcity of relevant studies. We look forward to more studies that investigate the value of smartphone audiometry in identifying different types of hearing loss in the future because they can provide more solid and specific evidence for apps in different clinical settings.

Conclusions

In this meta-analysis, we have provided comprehensive evidence regarding the diagnostic performance of smartphone-based audiometry in diagnosing hearing loss. Given the high sensitivity and specificity of smartphone-based audiometry, along with its low cost and high accessibility, smartphone-based hearing assessments may serve as a cost-effective and equally accurate diagnostic tool, in comparison with conventional PTA, for assessing hearing loss, especially in resource-limited settings where conventional PTA is not feasible. Our findings also suggest that future improvements in smartphone-based audiometry should focus on adjusting the potential factors that may affect its diagnostic accuracy.
Acknowledgments

This work was supported by the Taipei Veterans General Hospital (V108C-145, V109C-135, V109E-008-5, and V110E-003-2) and the Ministry of Science and Technology (MOST109-2320-B075-00, MOST109-2314-B075-014-MY2, and MOST109-2320-B-075-006).

Authors’ Contributions

Both YFC and CYH are the corresponding authors of this paper and have contributed equally to this work. CHC and YFC were responsible for data acquisition. CHC, HYHL, YCC, and CYH were responsible for the analysis and interpretation of the data. CHC, CYH, and HYHL drafted the manuscript. CHC, HYHL, and YCC performed the statistical analyses. YFC and CYH obtained the funding, and YCC, CYC, and MCW were responsible for administrative, technical, and material support. YFC and CYH supervised the study. All authors were responsible for the study concept and design, critical revision of the manuscript for important intellectual content, and approval of the final version.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The detailed search strategy.
[DOCX File, 22 KB - mhealth_v9i9e28378_app1.docx ]

Multimedia Appendix 2
Table of study characteristics.
[DOCX File, 22 KB - mhealth_v9i9e28378_app2.docx ]

Multimedia Appendix 3
Table of study diagnostic parameters.
[DOCX File, 22 KB - mhealth_v9i9e28378_app3.docx ]

References


64. Kamil RJ, Genter DJ, Lin FR. Factors associated with the accuracy of subjective assessments of hearing impairment. Ear Hear 2015 Jan;36(1):164-167 [FREE Full text] [Medline: 25158982]


https://mhealth.jmir.org/2021/9/e28378/