

# JMIR mHealth and uHealth

Mobile and tablet apps, ubiquitous and pervasive computing, wearable computing and domotics for health  
Volume 5 (2017), Issue 9 ISSN: 2291-5222

## Contents

### Original Papers

Development of a Whole Slide Imaging System on Smartphones and Evaluation With Frozen Section Samples ( <a href="#">e132</a> ) Hong Yu, Feng Gao, Liren Jiang, Shuoxin Ma. ....	3
Improvements in Patient Acceptance by Hospitals Following the Introduction of a Smartphone App for the Emergency Medical Service System: A Population-Based Before-and-After Observational Study in Osaka City, Japan ( <a href="#">e134</a> ) Yusuke Katayama, Tetsuhisa Kitamura, Kosuke Kiyohara, Taku Iwami, Takashi Kawamura, Junichi Izawa, Koichiro Gibo, Sho Komukai, Sumito Hayashida, Takeyuki Kiguchi, Mitsuo Ohnishi, Hiroshi Ogura, Takeshi Shimazu. ....	22
Atrial Fibrillation Screening in Nonmetropolitan Areas Using a Telehealth Surveillance System With an Embedded Cloud-Computing Algorithm: Prospective Pilot Study ( <a href="#">e135</a> ) Ying-Hsien Chen, Chi-Sheng Hung, Ching-Chang Huang, Yu-Chien Hung, Juey-Jen Hwang, Yi-Lwun Ho. ....	36
Tablet-Based Patient-Centered Decision Support for Minor Head Injury in the Emergency Department: Pilot Study ( <a href="#">e144</a> ) Navdeep Singh, Erik Hess, George Guo, Adam Sharp, Brian Huang, Maggie Breslin, Edward Melnick. ....	48
Effectiveness of SmartMoms, a Novel eHealth Intervention for Management of Gestational Weight Gain: Randomized Controlled Pilot Trial ( <a href="#">e133</a> ) Leanne Redman, L. Gilmore, Jeffrey Breau, Diana Thomas, Karen Elkind-Hirsch, Tiffany Stewart, Daniel Hsia, Jeffrey Burton, John Apolzan, Loren Cain, Abby Altazan, Shelly Ragusa, Heather Brady, Allison Davis, J. Tilford, Elizabeth Sutton, Corby Martin. ....	81
A Smartphone App for Improvement of Colonoscopy Preparation (ColoprAPP): Development and Feasibility Study ( <a href="#">e138</a> ) Benjamin Walter, Roland Schmid, Stefan von Delius. ....	91
A User-Centered Approach: Understanding Client and Caregiver Needs and Preferences in the Development of mHealth Apps for Self-Management ( <a href="#">e141</a> ) Roxanna Bendixen, Andrea Fairman, Meredith Karavolis, Carly Sullivan, Bambang Parmanto. ....	100
Mobile Augmented Reality as a Feature for Self-Oriented, Blended Learning in Medicine: Randomized Controlled Trial ( <a href="#">e139</a> ) Christoph Noll, Ute von Jan, Ulrike Raap, Urs-Vito Albrecht. ....	111
Social Communication Coaching Smartglasses: Well Tolerated in a Diverse Sample of Children and Adults With Autism ( <a href="#">e140</a> ) Neha Keshav, Joseph Salisbury, Arshya Vahabzadeh, Ned Sahin. ....	127

## Review

### Mobile Phone Interventions for Sleep Disorders and Sleep Quality: Systematic Review ([e131](#))

Jong Shin, Julia Kim, Diana Grigsby-Toussaint. . . . . 63

Original Paper

# Development of a Whole Slide Imaging System on Smartphones and Evaluation With Frozen Section Samples

Hong Yu<sup>1\*</sup>, MBBS; Feng Gao<sup>1\*</sup>, MMed; Liren Jiang<sup>1</sup>, MMed; Shuoxin Ma<sup>2</sup>, BEng

<sup>1</sup>Department of Pathology Center, Shanghai General Hospital, Shanghai Jiao Tong University School of Medicine (originally named “Shanghai First People’s Hospital”), Shanghai, China

<sup>2</sup>TerryDr Info Technology Co., Ltd, Nanjing, Jiangsu, China

\* these authors contributed equally

**Corresponding Author:**

Shuoxin Ma, BEng

TerryDr Info Technology Co., Ltd

Room A3-701

#180 Ruanjiandadao, Yuhuatai District

Nanjing, Jiangsu, 210000

China

Phone: 86 13813998278

Email: [sxma@terrydr.com](mailto:sxma@terrydr.com)

## Abstract

**Background:** The aim was to develop scalable Whole Slide Imaging (sWSI), a WSI system based on mainstream smartphones coupled with regular optical microscopes. This ultra-low-cost solution should offer diagnostic-ready imaging quality on par with standalone scanners, supporting both oil and dry objective lenses of different magnifications, and reasonably high throughput. These performance metrics should be evaluated by expert pathologists and match those of high-end scanners.

**Objective:** The aim was to develop scalable Whole Slide Imaging (sWSI), a whole slide imaging system based on smartphones coupled with optical microscopes. This ultra-low-cost solution should offer diagnostic-ready imaging quality on par with standalone scanners, supporting both oil and dry object lens of different magnification. All performance metrics should be evaluated by expert pathologists and match those of high-end scanners.

**Methods:** In the sWSI design, the digitization process is split asynchronously between light-weight clients on smartphones and powerful cloud servers. The client apps automatically capture FoVs at up to 12-megapixel resolution and process them in real-time to track the operation of users, then give instant feedback of guidance. The servers first restitch each pair of FoVs, then automatically correct the unknown nonlinear distortion introduced by the lens of the smartphone on the fly, based on pair-wise stitching, before finally combining all FoVs into one gigapixel VS for each scan. These VSs can be viewed using Internet browsers anywhere. In the evaluation experiment, 100 frozen section slides from patients randomly selected among in-patients of the participating hospital were scanned by both a high-end Leica scanner and sWSI. All VSs were examined by senior pathologists whose diagnoses were compared against those made using optical microscopy as ground truth to evaluate the image quality.

**Results:** The sWSI system is developed for both Android and iPhone smartphones and is currently being offered to the public. The image quality is reliable and throughput is approximately 1 FoV per second, yielding a 15-by-15 mm slide under 20X object lens in approximately 30-35 minutes, with little training required for the operator. The expected cost for setup is approximately US \$100 and scanning each slide costs between US \$1 and \$10, making sWSI highly cost-effective for infrequent or low-throughput usage. In the clinical evaluation of sample-wise diagnostic reliability, average accuracy scores achieved by sWSI-scan-based diagnoses were as follows: 0.78 for breast, 0.88 for uterine corpus, 0.68 for thyroid, and 0.50 for lung samples. The respective low-sensitivity rates were 0.05, 0.05, 0.13, and 0.25 while the respective low-specificity rates were 0.18, 0.08, 0.20, and 0.25. The participating pathologists agreed that the overall quality of sWSI was generally on par with that produced by high-end scanners, and did not affect diagnosis in most cases. Pathologists confirmed that sWSI is reliable enough for standard diagnoses of most tissue categories, while it can be used for quick screening of difficult cases.

**Conclusions:** As an ultra-low-cost alternative to whole slide scanners, diagnosis-ready VS quality and robustness for commercial usage is achieved in the sWSI solution. Operated on main-stream smartphones installed on normal optical microscopes, sWSI readily offers affordable and reliable WSI to resource-limited or infrequent clinical users.

**KEYWORDS**

mobile health; image processing; cloud computing for health care; whole slide imaging

## Introduction

Virtual slides (VSs) generated from whole slide imaging (WSI) systems are an essential component of digitized diagnostic processes, as they provide extended fields-of-view (FoVs) under microscopes without handling specimens physically [1,2]. In addition to providing a basis for automated analysis [3-5], the reliability and efficiency of VSs is widely considered on par with traditional light microscopy (or even superior in certain cases [6-9]), making it one of the main trends in both digital pathology and bioinformatics [10-12]. The conversion from direct observation to digitization has been dramatic [13], even though the technology is still under heavy development and shows performance bottlenecks that may not be broken in the near future [14,15].

In practice, however, automated scanners that are commonly used to capture and process such data cost approximately US \$50,000 or more up-front, even for low-frequency usage. In many developing countries, this financial obstacle alone has significantly impeded modernizing related departments in hospitals [16], such as that of pathology in China. Lacking digitization undermines productivity and diagnostic accuracy, commonly leading to poorer administrative attention and tighter budgets, thus forming a vicious cycle.

In recent years, two alternative solutions have attracted much academic and commercial interest. One solution is aborting the automation feature, thus leaving the operator to control the microscope manually, reducing the product package to a dedicated digital camera and software [17-19], and costing as little as US \$10,000. Other attempts have made use of smartphones, which not only have integrated capturing and processing abilities, but are also widely distributed among clinical professionals, thus lowering the start-up cost to near zero. A small number of products in the latter category in both research and commercial stages have been evaluated by clinical professionals [20], but to the knowledge of the authors, all are made exclusively for relatively expensive iPhones and are not yet commercially available. Although rarely explained explicitly, robustness of full automation and guarantees of successful VS generation could be serious obstacles between publishable research and commercial products. Additionally, diversity in hardware and operating systems might be the reason that Android phones, although dominating handset markets in developing countries, are largely ignored.

In this paper, a WSI system on mainstream smartphones that just became publicly available with commercial-quality and

low cost (named scalable WSI; sWSI [21]), is introduced and evaluated. sWSI offers fast and reliable WSI on most handsets, average Androids or flagship iPhones alike, reducing the up-front cost to approximately US \$100 and the average service cost per scan is as low as US \$1. Pathologists recognize sWSI as an attractive alternative to stand-alone scanners, especially for quick scans (eg, frozen sections) as well as medium/low-frequency usages.

Beyond technical development of the sWSI system, this research also included evaluating it with cryosections, also known as the frozen section procedure. Cryosectioning is widely used in oncological surgeries, which require significantly faster preparation and diagnoses compared to traditional histology techniques. Frozen section samples have much lower technical quality, making them very difficult to analyze for whole slide scanners [22], but this process offers a good challenge to test the limits of sWSI.

## Methods

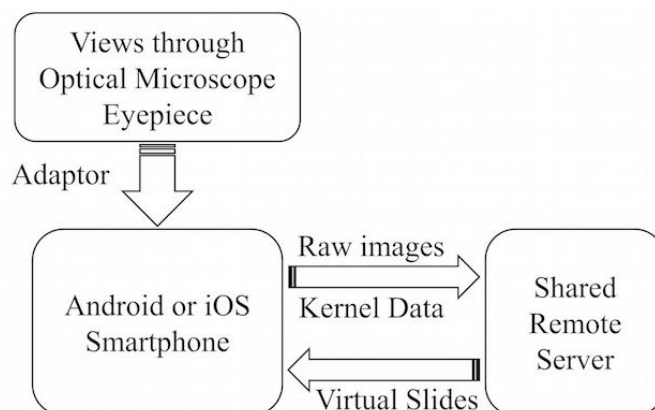
### System Overview

There are two essential and costly components in a typical WSI scanner: the capturing unit, typically a set of lenses with a distortion-calibrated digital eyepiece; and on-board or external high-performance computers. Like any dedicated devices, since both parts are specifically built for the system, they are mostly nonproductive when the system is idle and thus waste much of their value when underused. Unfortunately, this is commonly the case for smaller hospitals in which complicated pathological diagnoses occur only occasionally. This situation, coupled with consumer electronics' performance approaching that of medical-grade tools, led to the idea of creating sWSI with the structure illustrated in Figure 1.

### Hardware

To provide full WSI functionality at a dramatically lower cost, sWSI aims to reduce the cost of both hardware necessities. For the optical part, sWSI reversibly upgrades existing microscopes with built-in cameras of smartphones and compensates for the unknown optical distortions computationally, as discussed in detail in the *On-the-fly Image Distortion Correction* subsection below. For the computing part, sWSI utilizes smartphones for light-weight real-time processing and transfers the major bulk of the work to shared remote servers to allow temporal-multiplexing for improving utilization rate and cost-sharing.

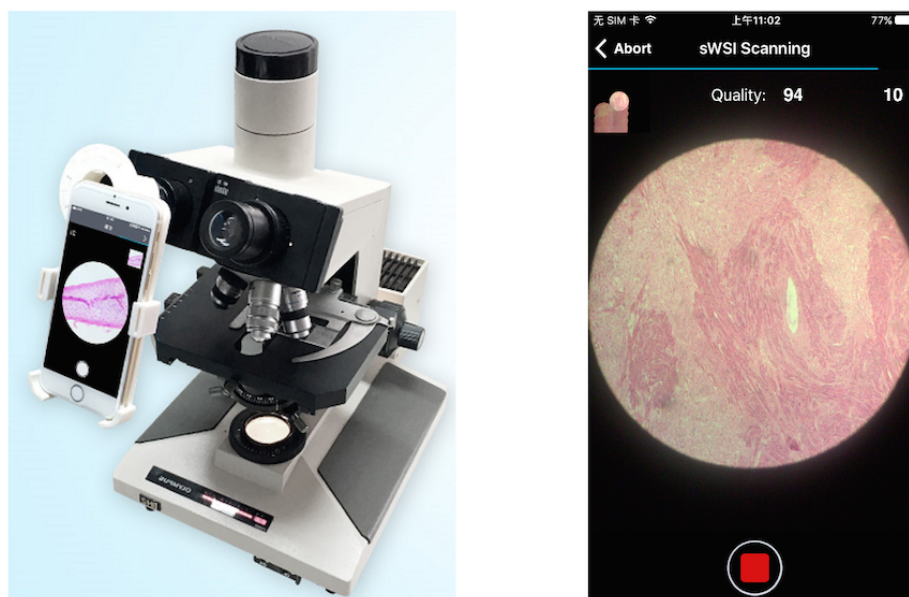


**Figure 1.** Simplified sWSI solution structure.**Table 1.** Minimal and optional hardware specifications.

Item	Minimal Specification	Recommended Specification
Operating System Version	Android 4.2 or iOS 9	N/A
Central Processing Unit	Dual-core @ 1.2 GHz	Quad-core @ 2.4 GHz
Camera	3 megapixels	12 megapixels

Although the prices of mainstream smartphones vary widely, much of the cost comes in the form of user-friendly features (eg, security or battery life) that are largely irrelevant to sWSI. Thanks to the fast expansion of smartphone markets, their cameras, which used to be the critical link in such clinical applications, are now on par with main-stream dedicated digital eyepieces [23,24]. Overall, newer smartphone models can easily meet the minimal requirements listed in Table 1 at prices as low as US \$100. It should also be noted that the higher-end smartphones that meet the optional specification for better performance may be bought at deep discounts as *used* or *refurbished*, which may suffer short battery life or a repaired screen, but do not affect the performance of sWSI.

Most professionals in research and health care services already own a handset that meets the criteria listed in Table 1, so sWSI requires installing only one adapter for each pair of existing smartphone and optical microscope. These microscope-smartphone adaptors are available with many commercial options as well as open-source designs for do-it-yourself 3-dimensional printing [25], although the ones specifically built for each phone model are preferred, to minimize the need for adjusting camera-eyepiece alignment and to block disrupting light sources. One setup is demonstrated in Figure 2, utilizing a used iPhone 6 costing US \$200 installed on an Olympus BH2-BHS microscope with a scalScope adaptor, which took about 15 seconds to set up, and was used for the clinical evaluation discussed later.

**Figure 2.** Typical hardware setup (left) and user interface (right).

## Software

In addition to image compressing, transferring and VS synthesizing is needed in any whole slide scanning systems, and the software in sWSI is also responsible for automatically measuring and compensating hardware diversity. Unfortunately, fully localizing many of these functions is beyond the reach of mass produced mobile devices. Synthesizing the VS from FoVs requires at least several gigabytes of random access memory and sequentially processing hundreds of FoVs at full resolution can take an hour or more on a mobile central processing unit (CPU). The VSs will be stored remotely anyway, so there is little extra cost in moving the bulk of processing onto remote servers, as implemented in sWSI. The downside of this distributed computing model is introducing significant risks of failure by splitting the processing work-flow into asynchronous halves, but in sWSI this is solved, as explained in *Fail-Proof Distributed Processing* subsection below.

Another practical issue worth noting is that due to architecture and driver support issues beyond the scope of this paper, most

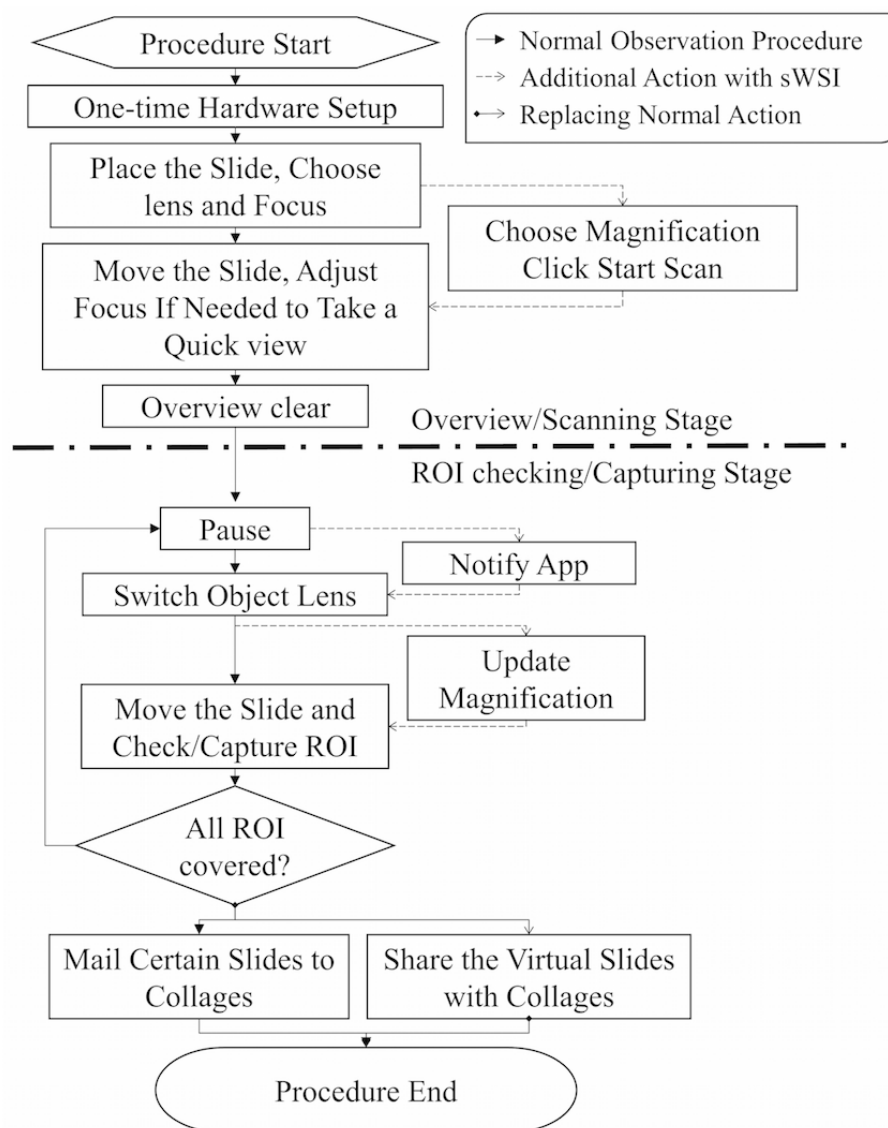
Android phones only support JPEG image capture at higher resolution, which cannot be directly processed pixel-by-pixel. This issue significantly constrains data flux since each FoV taken must go through an extra encoding-decoding routine costing several hundred milliseconds, depending on CPU power and resolution. As a result, the sWSI Android app limits the capturing resolution to approximately 3 megapixels and generally achieves throughput of approximately 1 to 3 FoVs per second, except for rare models with drivers offering high-resolution pixel data of images captured.

## Fail-Proof Distributed Processing

### Basic Scan Procedures and Interaction

In sWSI, a smartphone client app is responsible for gathering user inputs, capturing and processing the FoVs, and guiding users interactively, with a user interface during the scan (Figure 2). There is very little difference between the scanning procedure with sWSI shown in Figure 3 and that practiced by most microscope users, except for choosing a few parameters.

**Figure 3.** Microscope observation procedures adapted for sWSI. ROI: region of interest.



**Table 2.**

Error code	Description
Moving too fast	The translation is too far, so the KP matching in SURF may be unreliable
Lost	No reliable translation can be obtained: the causes cannot be further distinguished by the machine but should be noticeable to the users, such as moving so fast that there is little overlapping between the current pair of FOVs or the camera is out of focus
Touching a boundary	There are few KPs detected, so the FoV is likely near a boundary
No error	The translation is reliable

### Real-Time Feedback on Clients

The client's share of processing focuses on speed and robustness instead of accuracy, and therefore uses down-sampled copies of camera input. An algorithm roughly estimates pairwise translation of FoVs by stitching each captured FoV with the last one through key point (KP) detection and matching with the speeded-up robust features (SURF) algorithm [26,27]. This translation is then used in three ways: updating a mini-map illustrating current location on the slide, feeding a finite-state machine to manage the kernel asynchronously, and providing feedback to users as guidance for operating the microscope. The feedback and their trigger descriptions are listed in Table 2.

With users following the hints, sWSI essentially creates a closed feedback loop that allows scan-time interference against potential failure, such as inability to focus properly on thick samples or to track positioning on barren regions. This mechanism prevents most flops due to sample preparation and user operation before spending a long time in completing the scan, which is a common issue with automatic scanners.

### Full Resolution Processing With A-Priori Knowledge on Servers

The cloud servers in sWSI are the primary powerhouses of computation. With full resolution FoVs and scan results from clients, servers restitch the adjacent FoVs at maximal accuracy, correct distortion, and generate the virtual whole slide. The asynchronous two-staged stitching performed respectively on

the clients and servers, however, has inherent weak spots on both stability and efficiency.

The FoV pairwise stitching is based on KP detection and matching, whose outcome in turn is resolution-dependent. As a result, such outcomes in down-sampled and original resolution may potentially be significantly inconsistent. In many cases, as in almost every VS constructed from 100 FoVs or more with prototypes of sWSI, the full-resolution stitching produced unreliable matching at least once.

Conversely, by the law of large numbers, it is desirable to match as many KP pairs as possible for accurate estimation of the FoV-wise matching function, especially where this function has high degrees of freedom (as is the case of sWSI where raw images are nonlinearly distorted in unknown patterns). The computational cost of brute-and-force KP matching, however, grows quadratically with the number of KPs.

To resolve both issues at once, sWSI employs an a-priori knowledge-based SURF KP detection and a matching algorithm on the server. SURF detects KPs from a virtual image pyramid that has lower resolution on higher layers. In sWSI, instead of detecting with one threshold across all layers, multiple thresholds are chosen adaptively as described in Figure 4.

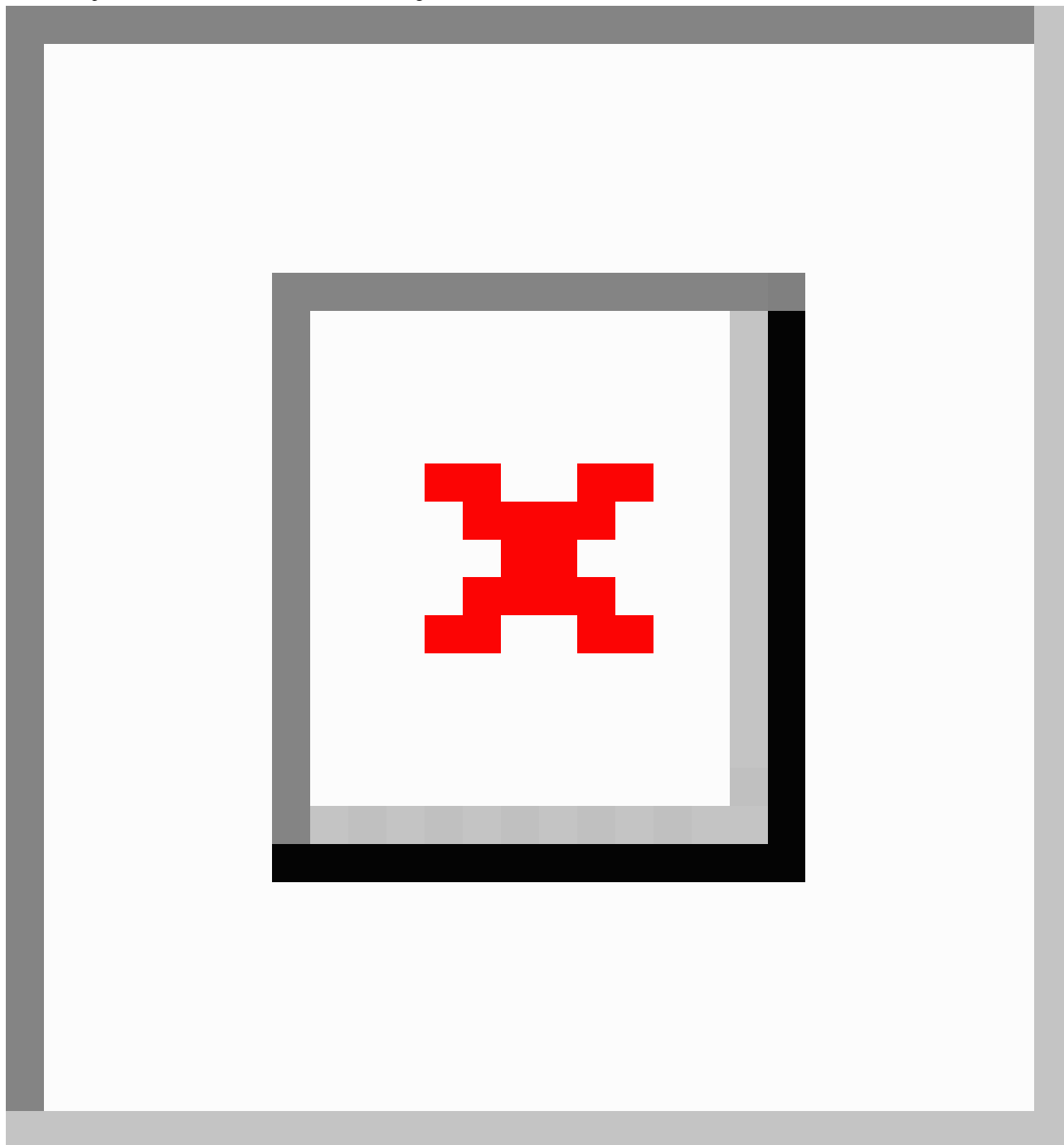
Afterwards, instead of brute-and-force matching by calculating difference of all pairs of KPs and picking the optimal set of matches, sWSI selectively calculates those within a constant distance from the coordinates indicated by the scan stage translation with up-sampling and assumes all others are infinitely large. Figure 5 further details this process.

**Figure 4.** Adaptive thresholding in SURF.

First, one threshold  $t_{r_{ds}}$  is picked to ensure at least  $p_h n_{kp}$  KPs are detected on layers  $[l_{ds}, l_{max}]$  in total, where  $p_h \geq 1$  is a constant multiplier,  $n_{kp}$  is the number of KPs detected in the down-sampled copy during scan,  $l_{max}$  is the index of the upper most layer, and  $l_{ds}$  is derived from the power-of-two down-sample rate during scan  $r_{ds}$  as

$$l_{ds} = \log_2 r_{ds}, l_{ds} \in \mathbb{N} \quad (1)$$

Next, threshold  $t_i, i \in [0, l_{ds})$  for detection on layer  $i$  is adaptively chosen so that  $p_l n_{kp}$  KPs are detected on each lower layer, where  $p_l > 0$  is another constant. With this thresholding approach, most KPs on the scan stage can be detected at full resolution with additional ones from lower layers that are more localized but with higher resolution, while the total number of KPs is controlled by  $p_h$  and  $p_l$  and would not overexpand.

**Figure 5.** Computational benefits of selective KP matching.

Since KP matching takes a large number of float point operations and consumes a large portion of time, this reduction boosts the overall efficiency of sWSI by over 50%.

### ***On-the-Fly Image Distortion Correction***

When stitching each FoV pair to match KPs, the projection function can be in any format as long as it minimizes error without overfitting. Combining all FoVs into a single continuous view, however, requires the projection to be linear so the nonlinear distortion must be corrected first. If the distortion is not corrected, the order of the stacked-up nonlinear transfer function of each FoV onto the whole slide will keep growing by each FoV and become very inefficient to solve.

Designed to fit any combination of microscope and smartphone model, sWSI assumes a generalized high-order polynomial (HOP) inverse-distortion model [28], which mathematically approximates any function with marginal error if the order is sufficiently high, as proven by Taylor's theorem [29]. Specifically, it is assumed that there exists a constant but unknown HOP projection function for each scan procedure that maps the raw FoVs into a corrected 2-dimensional space, where any matched pixel pairs in overlapping FoVs share the same phase difference for that FoV pair. After the raw FoVs are corrected by a HOP function, each adjacent FoV pair can be stitched with translation onto each other with small error. In sWSI, this HOP projection matrix is solved iteratively based

on FoV-pair-wise KP matching, formulated as described in [Figure 6](#).

### Clinical Evaluation Setup

To assess the diagnosis-readiness of sWSI VSs of challenging cases, a clinical evaluation experiment was carried out in Pathology Center, Shanghai General Hospital, School of Medicine, Shanghai Jiaotong University (SJTU-PC) between February 23rd, 2017 and April 15th, 2017. A total of 100 frozen section slides (from one of the five categories listed in [Table 3](#)) gathered among inpatients between February 23<sup>rd</sup>, 2017 and April 1st, 2017 were randomly selected as the test dataset. The slides varied significantly in size and shape, were prepared routinely by technicians in the department, and may have had common issues with frozen sections, such as unevenness and folding. Consideration of both possibilities of tumors as well as benign lesions were included in the diagnosis, following generally accepted practice standards [30-32].

All samples went through three diagnosis procedures using optical microscopy, sWSI VSs, and high-end standalone scanner

VSs under a 20X objective lens. In each procedure, each sample was examined by one or two pathologists independently. These diagnoses were then compared against each other to obtain agreement statistics. The full procedure is illustrated in [Figure 7](#). This would be the first of a planned sequence of experiments evaluating sWSI on a wide range of samples, so we focused on using iPhones (on which the international version of sWSI is supported). The smartphones used were: one iPhone 5s, one iPhone 6, and one iPhone 6 plus, purchased from the second-hand market between US \$120 and \$200. The microscopes for sWSI scanning were: an Olympus BH2-BHS, an Olympus CX21, and a Phoenix PH50-1B43L-PL, all being low-end bio-microscopes. The scanner used in this experiment was Aperio AT2 from Leica, which is a high-end model. The microscope for optical microscopy was an Olympus BX51, which is another high-end model. Pathologists A, B, C and D are all senior faculty members from SJTU-PC, while the work of pathologist E was carried out by 10 senior and respected pathologists invited from top hospitals across China, as listed in *Acknowledgements*.

**Table 3.** Sample categories, counts, and notes.

Category	Counts	Notes
Breast	20	Requiring median image quality
Uterine corpus	20	Requiring median image quality
Thyroid	28	Requiring high image quality; one sample of scanner VS missing
Lung	31	Requiring very high quality; one sample of sWSI VS missing (physically damaged)
Ovary	1	Accidentally included, as it was assumed to be a thyroid sample; diagnosis only considered in calculating overall statistics



**Figure 6.** On-the-fly distortion correction based on matched KPs.

First, assume the HOP model has  $O_p$  orders and name the two FoVs in the  $i$ th FoV pair as  $\text{srcFoV}_i$  and  $\text{dstFoV}_i$ , whereas  $\text{srcFoV}_i$  is stitched onto  $\text{dstFoV}_i$ . For point  $j$  on  $\text{srcFoV}_i$  with a 2-dimensional coordinate  $\tilde{x}_{i,j} = [x_{i,j,1}, x_{i,j,2}]$ , its polynomial expansion kernel is derived as [28]  $\tilde{x}_{i,j} = \tilde{\phi}(x_{i,j,1}, x_{i,j,2})$ , where

$$\tilde{\phi}_{O_p}(u, v) = (1, u, v, u^2, uv, v^2, \dots, v^{O_p}). \quad (2)$$

Thus, the number of dimensions of  $\tilde{x}_{i,j}$  is  $N_e = \frac{(O_p+1)(O_p+2)}{2}$ . Similarly, the point's exact match in  $\text{dstFoV}_i$  has a coordinate and kernel  $\tilde{y}_{i,j}$  and  $\tilde{y}_{i,j}$ , respectively. For simpler notations, also define  $N_{e-1} = N_e - 1$ .

Next, note the correction projection matrix as  $\beta \in \mathbb{R}^{N_e \times 2}$ . The linear projection used to stitch the pair after correction is an affine one in the form of  $A_i = [\tilde{T}_i^T, R_i^T]$ , where  $\tilde{T}_i^T \in \mathbb{R}^{1 \times 2}$  and  $R_i^T \in \mathbb{R}^{2 \times 2}$  are the translation and rotation components, respectively. The whole model would ideally satisfy

$$\tilde{x}_{i,j} \beta R_i + \tilde{T}_i - \tilde{y}_{i,j} \beta = 0 \quad (3)$$

for all point pairs across all FoV pairs, where  $\beta$  is constant and  $\{\tilde{T}_i^T, R_i^T\}$  are FoV-pair-dependent but point-pair-independent.

In reality, correction error exists and the process turns into solving a constrained optimization problem

$$\beta, \{\tilde{T}_i^T, R_i^T\} = \arg \min_{\beta, \{\tilde{T}_i^T, R_i^T\}} \sum_i \mathcal{F}_i(\beta, \tilde{T}_i^T, R_i^T) \quad (4)$$

where

$$\mathcal{F}_i(\beta, \tilde{T}_i^T, R_i^T) = \sum_j [\|\tilde{x}_{i,j} \beta R_i + \tilde{T}_i - \tilde{y}_{i,j} \beta\|^2 + \lambda \|\tilde{x}_{i,j} \beta - \tilde{x}_{i,j}\|^2]. \quad (5)$$

The regulation term  $\lambda \|\tilde{x}_{i,j} \beta - \tilde{x}_{i,j}\|^2$  here prevents the projections from collapsing into all zeros and  $\lambda = 0.001$  is used. It should be noted that based on the assumption that  $\beta$  should correct and only correct nonlinear distortions, the elements in

$$\beta = \begin{bmatrix} \beta_{00} & \beta_{10} & \beta_{20} & \dots & \beta_{N_{e-1}0} \\ \beta_{01} & \beta_{11} & \beta_{21} & \dots & \beta_{N_{e-1}1} \end{bmatrix} = \begin{bmatrix} \beta_{00} & \beta_{10} & \beta_{20} & \dots & \beta_0 \\ \beta_{01} & \beta_{11} & \beta_{21} & \dots & \beta_0 \end{bmatrix} \quad (6)$$

satisfy

$$\beta_{00} = \beta_{01} = \beta_{20} = \beta_{21} = 0, \beta_{10} = \beta_{21} = 1. \quad (7)$$

Then, the multi-variable nonlinear equation of equation (4) can be solved by iteratively fixing either  $\beta$  or  $\{\tilde{T}_i^T, R_i^T\}$  and finding the least-mean-square solution of the other until convergence. Specifically, by freezing  $\beta$  equation (4) can be derived into

$$[\tilde{T}_i^T, R_i^T]^T = \left( \sum_i \tilde{y}_{i,j} \beta \tilde{x}_{i,j}^T \right) \left( \sum_i \tilde{x}_{i,j} \tilde{x}_{i,j}^T \right)^{-1} \quad (8)$$

where  $\tilde{x}_{i,j} = [1, \tilde{x}_{i,j} \beta]$ . By keeping  $\{\tilde{T}_i^T, R_i^T\}$  constant and splitting the elements as

$$\tilde{T}_i = [T_{i,0} \quad T_{i,1}], R_i = \begin{bmatrix} R_{i,00} & R_{i,01} \\ R_{i,10} & R_{i,11} \end{bmatrix}, \quad (9)$$

elements in  $\beta$  can be solved as

$$\begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \left( \sum_{i,j} (M_{i,j}^T M_{i,j})^{-1} \right) \left( \sum_{i,j} M_{i,j}^T L_{i,j} \right) \quad (10)$$

where

$$\begin{aligned} M_{i,j} &= \begin{bmatrix} M_{i,j,00}(\tilde{x}, \tilde{y}) & M_{i,j,01}(\tilde{x}, \tilde{y}) \\ M_{i,j,10}(\tilde{x}, \tilde{y}) & M_{i,j,11}(\tilde{x}, \tilde{y}) \end{bmatrix} \\ \tilde{L}_{i,j} &= \begin{bmatrix} \tilde{L}_{i,j,0}(\tilde{x}_{i,j}, \tilde{y}_{i,j}) \\ \tilde{L}_{i,j,1}(\tilde{x}_{i,j}, \tilde{y}_{i,j}) \end{bmatrix} \end{aligned} \quad (11)$$

Omitting subscription  $(\cdot)_{i,j}$  and  $(\cdot)(\tilde{x}, \tilde{y})$  for simplicity, elements in equation (11) are calculated as

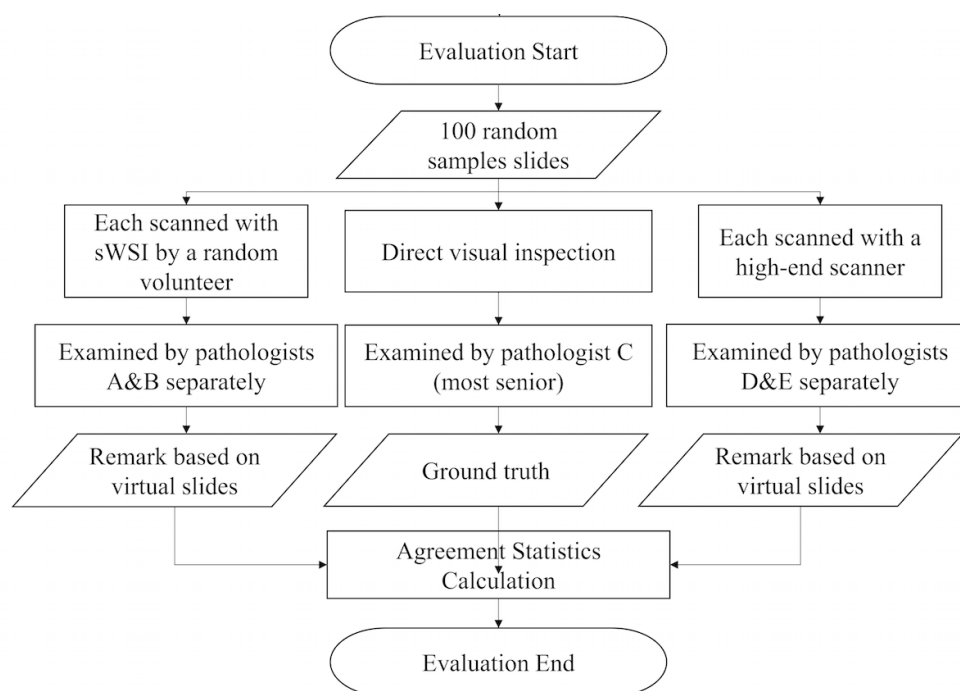
$$\begin{aligned} M_{00} &= (R_{00}\tilde{x} - \tilde{y})^T(R_{00}\tilde{x} - \tilde{y}) + R_{00}^T\tilde{x}^T\tilde{x} + \lambda\tilde{x}^T\tilde{x} \\ M_{01} &= R_{01}(R_{01}\tilde{x} - \tilde{y})^T\tilde{x} + R_{01}\tilde{x}^T(R_{11}\tilde{x} - \tilde{y}) \\ M_{10} &= R_{01}\tilde{x}^T(R_{01}\tilde{x} - \tilde{y}) + R_{01}(R_{11}\tilde{x} - \tilde{y})^T\tilde{x} \\ M_{11} &= (R_{11}\tilde{x} - \tilde{y})^T(R_{11}\tilde{x} - \tilde{y}) + R_{11}^T\tilde{x}^T\tilde{x} + \lambda\tilde{x}^T\tilde{x} \end{aligned} \quad (12)$$

$$\begin{aligned} \tilde{L}_0 &= -\tilde{T}_0(R_{00}\tilde{x} - \tilde{y})^T - R_{01}\tilde{T}_1\tilde{x}^T \\ \tilde{L}_1 &= -R_{10}\tilde{T}_0\tilde{x}^T - \tilde{T}_1(R_{11}\tilde{x} - \tilde{y})^T \end{aligned} \quad (13)$$

where  $\tilde{x} = \tilde{\phi}(x_1, x_2)$  and  $\tilde{y} = \tilde{\phi}(y_1, y_2)$  are respective subvectors of  $\tilde{x}$  and  $\tilde{y}$  calculated as

$$\tilde{\phi}_{O_p}(u, v) = (u^2, uv, v^2, \dots, v^{O_p}). \quad (14)$$

From experiments, it is shown that  $O_p = 7$  is generally sufficient and the model takes approximately 10 iterations to converge.

**Figure 7.** sWSI clinical evaluation procedure.

As part of the evaluation, quantifying the learning curve of manual scanning with sWSI was covered in this experiment. The VS quality depends on the sample being in focus and even the scan speed in the scanning process, so it is positively linked to the level of skill of operator or inversely to the time consumed. Presumably, as the operator gains experience, they would scan faster while maintaining the same quality. An intuitive way to measure this assumption is by comparing the time consumption of carefully scanning a unit area of samples under a fixed objective lens magnification against the number of slides scanned by operator. To this end, a total of 15 interns, medical school students, and assistant technicians who were well trained to use microscopes but had no prior experience with sWSI volunteered to scan the slides with sWSI. With duplication, each participant scanned 10 of the 100 slides after watching a demonstration, with the approximate sample size and time consumption recorded. Only one of the duplicated scans of the same sample was randomly selected for accuracy evaluation.

## Results

### Sensitivity and Specificity

The accuracy of diagnosis in this evaluation experiment was measured on a slide-wise level, since the classifications of regions in the sample can be subjective and pathologists usually consolidate multiple patterns across the whole sample into a conclusion. The sensitivity is defined as whether a pathologist

can use a VS to correctly identify all critical regions of interest across the sample. The specificity is defined as whether the pathologist can correctly analyze each region of interest and identify patterns based on a VS.

To represent the performance based on the sensitivity and specificity defined above, each diagnosis was classified into one of three categories: accurate, low sensitivity (LSen), or low specificity (LSpe). For each VS, if the pathologist missed any critical region of interest, the VS was deemed as LSen. With the VS past the sensitivity check, if the pathologist then correctly identified patterns of all regions of interest, it was classified as being accurate; otherwise it was LSpe. When averaged, these metrics were weighted by their sample counts.

Using optical microscopy as the ground truth, the sWSI system achieved good performance similar to that of the standalone scanner in most sample categories except lung, as summarized in Table 4, Table 5, Table 6, and Table 7. For easier comparison, the per-category metrics are plotted in Figure 8, Figure 9, Figure 10, and Figure 11, with the average in Figure 12. Screen shots of VS regions where sWSI provided adequate or poor image quality are illustrated in Figure 13 and Figure 14. Although the performance was not ideal in absolute terms, all pathologists involved in this experiment firmly agreed that sWSI performed very well for these frozen section samples after viewing the results in retrospect. The pathologists also suggested that sWSI is clinically reliable for daily use on easier and more common samples, such as margins or paraffin sections, which are not covered in this experiment.



**Table 4.** Accuracy of diagnosis based on sWSI VS by pathologist A. Results are presented as ratios between 0.00 and 1.00.

Category	Accurate	LSen	LSpe
Breast	0.70 (14/20)	0.05 (1/20)	0.25 (5/20)
Uterine corpus	0.75 (15/20)	0.10 (2/20)	0.15 (3/20)
Thyroid	0.68 (19/28)	0.11 (3/28)	0.21 (6/28)
Lung	0.43 (13/31)	0.30 (9/31)	0.27 (8/31)
Ovary	0.00 (0/1)	1.00 (1/1)	0.00 (0/1)
Average	0.61 (61/100)	0.16 (16/100)	0.22 (22/100)

**Table 5.** Accuracy of diagnosis based on sWSI VS by pathologist B. Results are presented as ratios between 0.00 and 1.00.

Category	Accurate	LSen	LSpe
Breast	0.85 (17/20)	0.05 (1/20)	0.10 (2/20)
Uterine corpus	1.00 (20/20)	0.00 (0/20)	0.00 (0/20)
Thyroid	0.68 (19/28)	0.14 (4/28)	0.18 (5/28)
Lung	0.57 (17/31)	0.20 (6/31)	0.23 (7/31)
Ovary	0.00 (0/1)	0.00 (0/1)	1.00 (1/1)
Average	0.74 (73/100)	0.11 (11/100)	0.15 (15/100)

**Table 6.** Accuracy of diagnosis based on scanner VS by pathologist D. Results are presented as ratios between 0.00 and 1.00.

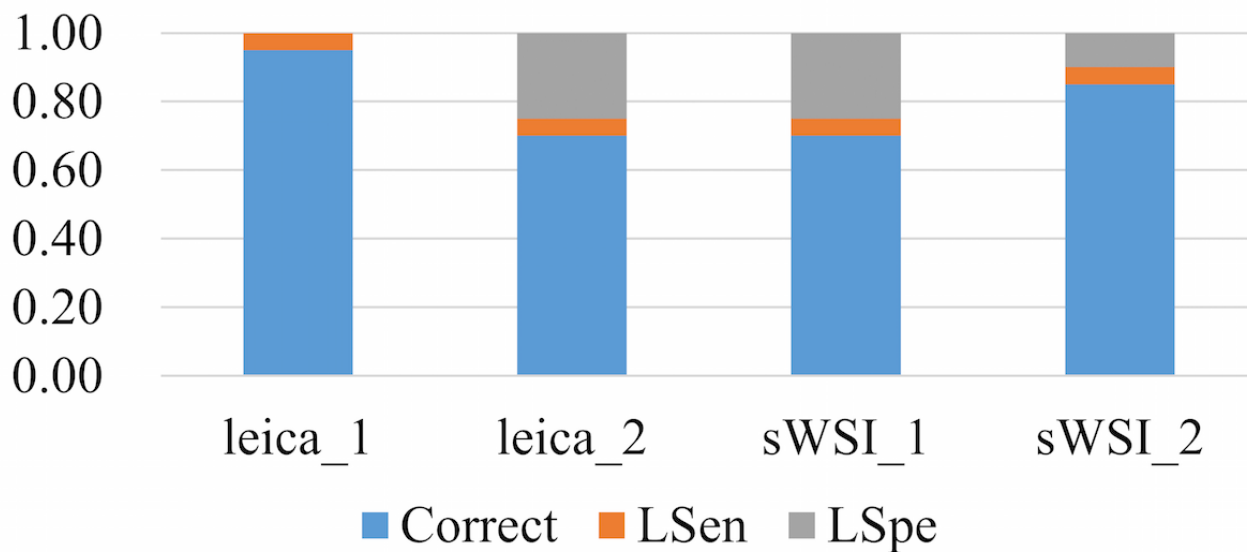
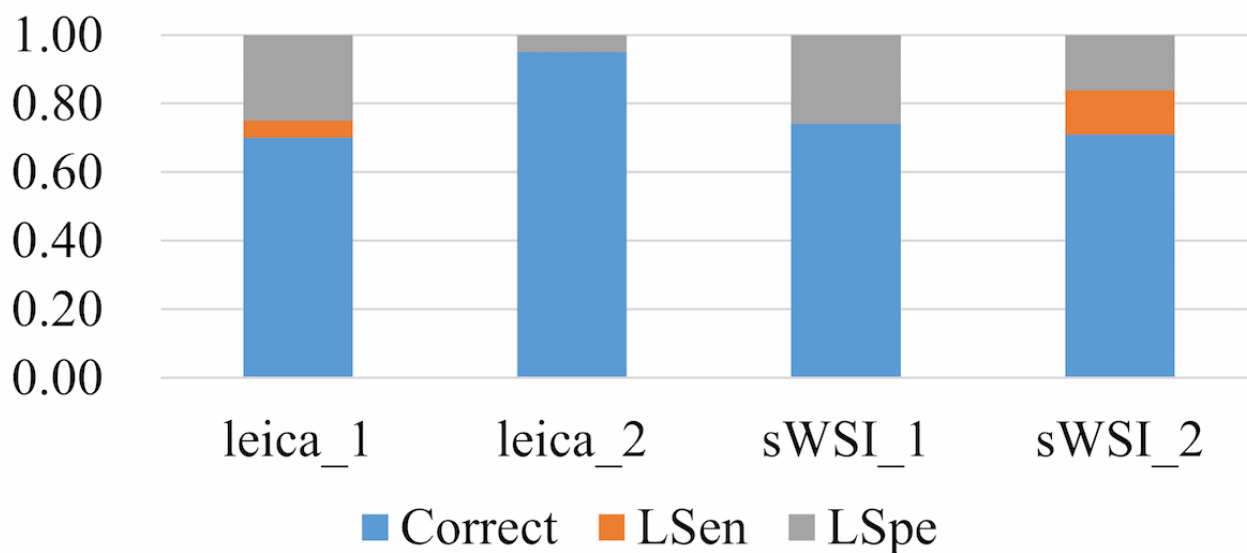
Category	Accurate	LSen	LSpe
Breast	0.95 (19/20)	0.05 (1/20)	0.00 (0/20)
Uterine corpus	0.95 (19/20)	0.00 (0/20)	0.05 (1/20)
Thyroid	0.85 (23/28)	0.00 (0/28)	0.15 (4/28)
Lung	0.77 (24/31)	0.16 (5/31)	0.06 (2/31)
Ovary	1.00 (1/1)	0.00 (0/1)	0.00 (0/1)
Average	0.87 (86/100)	0.06 (6/100)	0.07 (7/100)

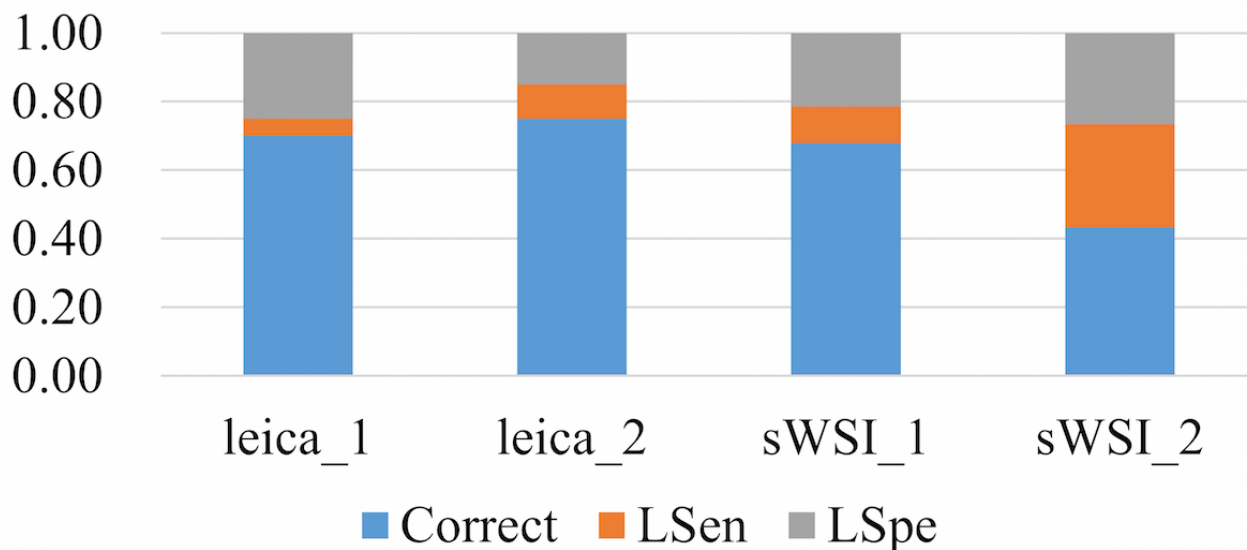
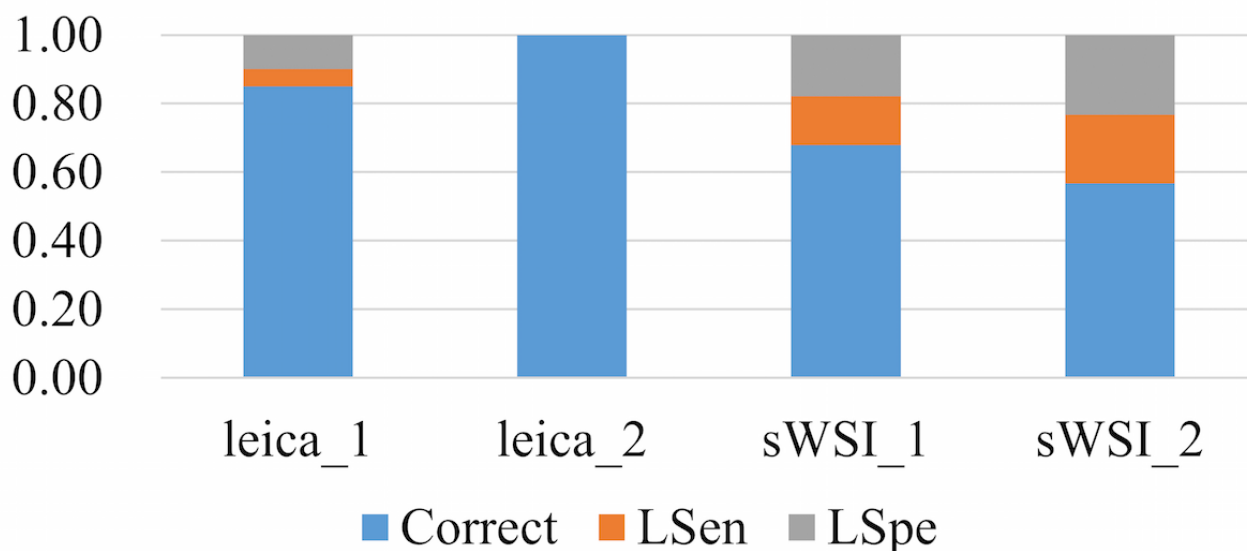
**Table 7.** Accuracy of diagnosis based on scanner VS by pathologist E. Results are presented as ratios between 0.00 and 1.00.

Category	Accurate	LSen	LSpe
Breast	0.70 (14/20)	0.05 (1/20)	0.25 (5/20)
Uterine corpus	0.95 (19/20)	0.00 (0/20)	0.05 (1/20)
Thyroid	0.74 (20/28)	0.00 (0/28)	0.26 (7/28)
Lung	0.71 (22/31)	0.13 (4/31)	0.16 (5/31)
Ovary	0.00 (0/1)	0.00 (0/1)	1.00 (1/1)
Average	0.75 (75/100)	0.05 (5/100)	0.19 (19/100)

For validation purposes, all VSs in this experiment captured by sWSI and the scanner can be viewed online through the links provided in [Multimedia Appendix 1](#). Details of each sample and the diagnosis, including the ground truth and those made

based on standalone scanners or sWSI, are provided in [Multimedia Appendix 2](#). Limited by time, this record is mostly in its original language (Chinese) but may be translated and requested from the authors.

**Figure 8.** Performance based on scanner and sWSI virtual slides. Sample type: breast. LSen: low sensitivity; LSpe: low specificity.**Figure 9.** Performance based on scanner and sWSI virtual slides. Sample type: uterine corpus. LSen: low sensitivity; LSpe: low specificity.

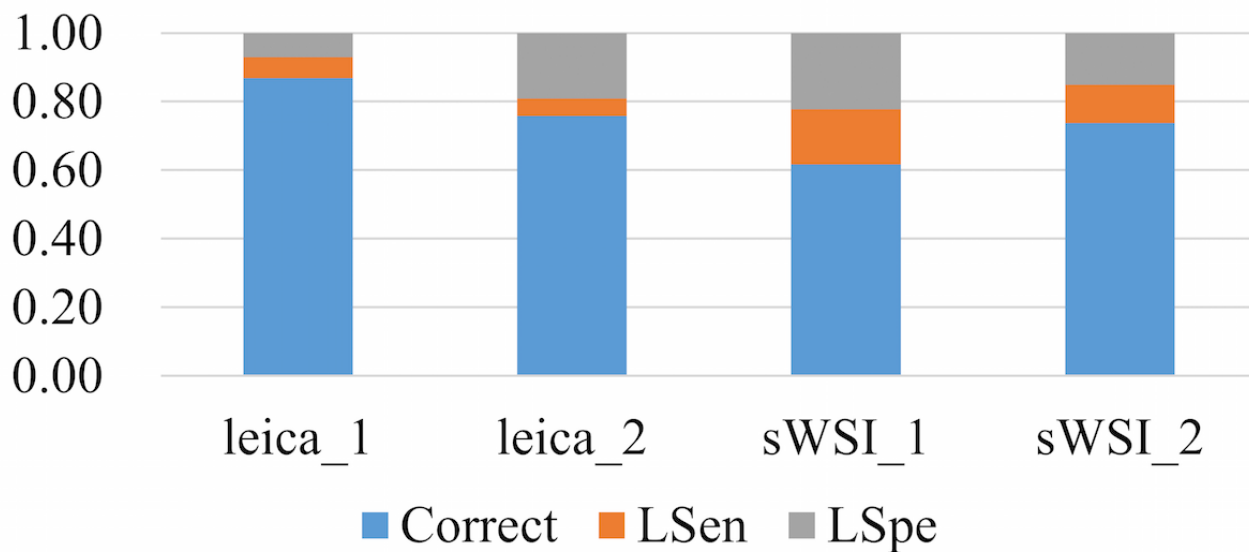
**Figure 10.** Performance based on scanner and sWSI virtual slides. Sample type: thyroid. LSen: low sensitivity; LSpe: low specificity.**Figure 11.** Performance based on scanner and sWSI virtual slides. Sample type: lung. LSen: low sensitivity; LSpe: low specificity.

### Learning Curve

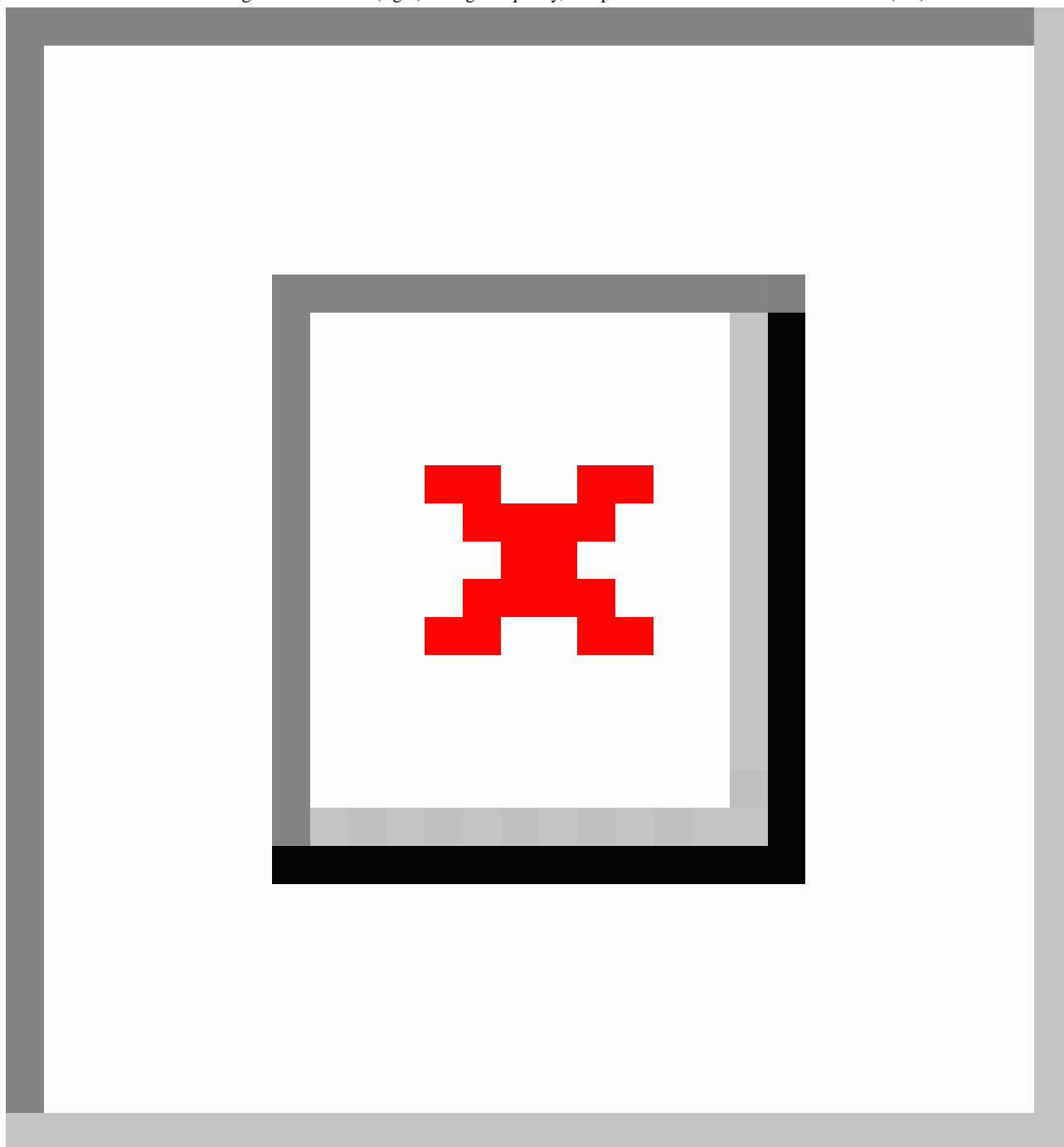
The learning curve, which is inversely approximated by the scan time normalized for a 15-by-15 mm sample being plotted against the number of samples scanned, is shown in Figure 15. In the experiment, productivity varied significantly among operators, as the slower ones consumed up to 50% more time at the beginning of the experiment. However, the slower operators caught up swiftly through practice and eventually

reached just over 35 minutes for every 15-by-15 mm sample, which is comparable to automatic scanners. Most operators reached a stable level of proficiency after scanning just 5 slides. It should be noted that the 20X objective lens for frozen section samples is a relatively challenging application, as the samples are often uneven and require frequent focus adjustments. As a result, the scan speed in Figure 15 may be considered as a worst-case scenario.

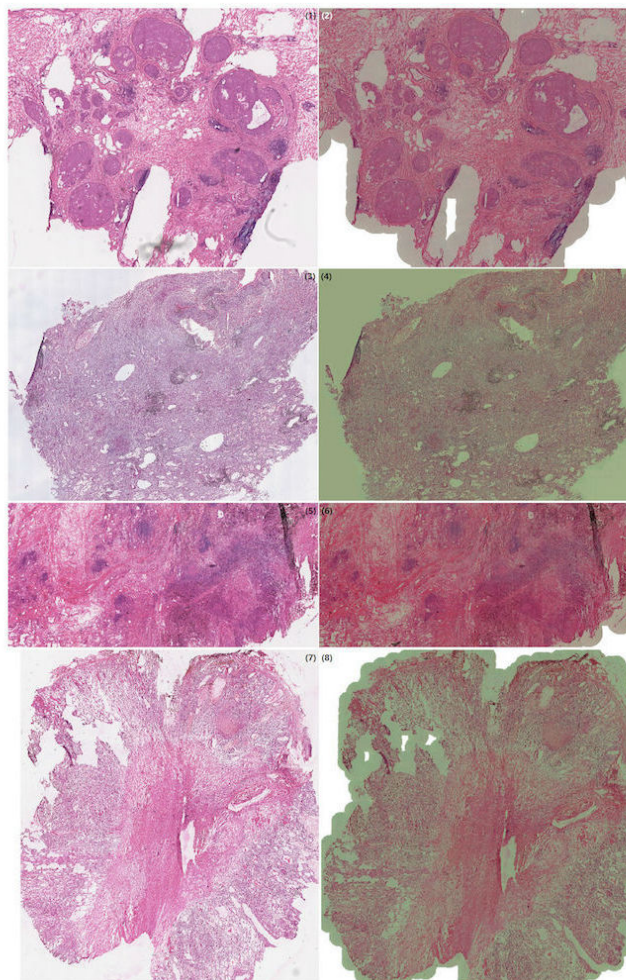
**Figure 12.** Performance based on scanner and sWSI virtual slides, averaged. LSen: low sensitivity; LSpe: low specificity.



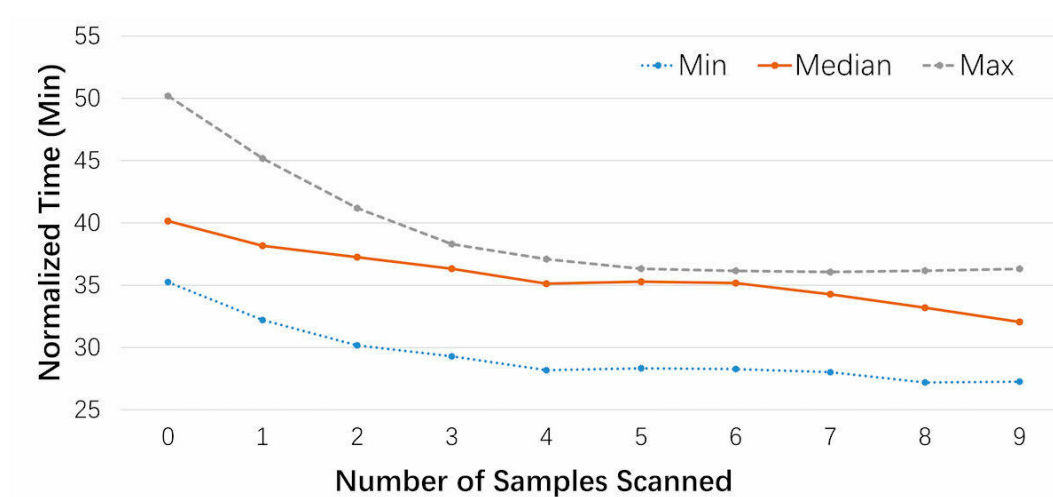
**Figure 13.** Selected virtual slide regions from sWSI (right) with good quality, compared to those from the Leica scanner (left).



**Figure 14.** Selected virtual slide regions from sWSI (right) with poor quality, compared to those from the Leica scanner (left).



**Figure 15.** Normalized scan time versus number of samples scanned.



## Discussion

### Current Limitations

First, a significant performance gap exists among different sample categories that are challenging and require very high image quality. Based on the experimental results, sWSI diagnoses of leiomyoma of the uterus, adenomyosis, papillary

carcinoma of the thyroid, invasive breast carcinoma, and fibroadenoma of the breast are relatively more reliable than standard procedures, but those for follicular carcinoma of the thyroid, intraductal papillary neoplasms of the breast, and lung cancer are much worse. On the positive side, since the quality can be assessed after scanning and before inspection by the pathologists, these mistakes may be largely avoided in practice,

especially those of LSpe (which should be analyzed by optical microscopy). However, this finding does reveal the fact that the current version of sWSI may not be suitable to all types of samples, or at least that testing may be required before formally adopting sWSI for applications requiring the highest quality images.

Second, similar to standalone scanners, sWSI is strongly affected by the physical preparation quality of the samples. Common problems include samples being broken into multiple pieces, varying thickness, and poor staining quality.

Finally, sWSI suffers from an inferior setup of hardware and environments, which comes with its focus on cost and scalability, and may make its competition against standalone scanners challenging. One such example is data storage: standalone scanners are mostly designed to work with high-speed wired and local networks. Since sWSI provides wireless connection-based data transmission and Web browser-based viewing, limits on bandwidth may severely affect image quality. Another example is that VSs produced by scanners are displayed on specialized work station monitors with high coverage of red/green/blue color space and 4K resolution, while those of sWSI are reviewed on normal monitors, or even smartphones with small screens.

The design of the experiment can be improved in future work. Based on discussions with participating pathologists, it can be safely assumed that there would be little error in diagnoses made by senior pathologists. However, from experimental results and comments, there are significant differences in the level of confidence that pathologists have regarding dubious cases where the provided VSs alone are not sufficient for a firm decision. In practical procedures, pathologists would undertake further investigation with other techniques, such as paraffin embedding and sectioning, which is more reliable (but far more time consuming) than frozen sectioning. In this experiment, participants were required to make binary decisions when judging thresholds that were different from each other. As a result, using multiple probabilities instead of a firm answer might be a better alternative.

## Future Work

There are a number of technical issues to be solved in future research and development. First, some parameters on Android phones cannot be controlled through publicly available programming interfaces for older Android operating systems. Weakened control may lead to improper configurations, such as a long exposure time causing blur. Second, the OpenGL driver which offers general-purpose graphic processing unit (GPGPU) computing potential is very difficult to work with and produces unexpected results on many smartphone models for unknown reasons. Preliminary research using GPGPU on iPhones brought a dramatic boost in processing speed over 60%, but certain models behave improperly. Finally, the real-time movement tracking is inaccurate and leads to location mismatch in subsequent mini-map construction. Although this issue does

not affect the VS generation, the mis-drawing in the mini-map can be confusing.

Limited by time and development cost, the international version of sWSI is only offered on iOS. The Chinese version (Tai Rui Jing Xia), which has a complicated still-FoV capturing tool and clinical report system, supports both Android and iOS but may not be accessible from abroad. In follow-up studies, both versions will be evaluated and compared. More quantitative metrics may also be covered (eg, the measure of absolute image quality) even though most pathologist users suggest such measurements have very limited implications in clinical settings.

On the clinical side, only the basic scanning function of sWSI was evaluated in this experiment, leaving sWSI's many other useful functions to be tested. These functions include swift scanning at a lower magnification then adding static FoVs at higher magnification to cover both speed and detail, recording z-stacks of thick smears with video clips, and potential application of sWSI in fluorescent or dark-field microscopy. The validation of these functions, as well as experiments with expanded sample sizes, are planned in future work.

## Conclusions

In this paper, an ultra-low-cost WSI system named sWSI (with clients hosted on mainstream Android and iOS smartphones) was introduced and analyzed. Compared to automatic scanners and high-end computer-based solutions, this alternative dramatically reduces the setup cost to as low as US \$100 per unit with service costs of US \$1-\$10 per scan. Although sWSI may not replace existing dedicated devices, it could become a reliable alternative that weighs more on cost-effectiveness.

By employing distributed image processing, both robustness and efficiency are covered. Through high performance computing and real-time feedback, user friendliness is optimized with minimal manual input, leaving most interface-kernel coordination (and even image distortion correction) fully automated. Based on clinical evaluation using 100 frozen section samples, sWSI is considered adequate by expert pathologists for making a diagnosis for most sample types. The overall accuracy based on sWSI VSs is slightly poorer than that based on high-end scanners, which is the current solution for digitizing whole slides. This gap is largely attributed to low specificity for samples requiring higher levels of detail (eg, lungs), which means that most of the inadequacy of sWSI comes from relatively low image quality, which can be identified before diagnoses are made.

In the experiment, 15 operators with no prior experience with sWSI learned to use the manual scanning very quickly. User throughput stably reached between 27 and 36 minutes per 15-by-15 mm sample area under a 20X objective lens after scanning 9 slides each, similar to that of mid-to-low-end scanners. In low-frequency usage situations (eg, remote or low-tier hospitals), this level of labor may be considered cost-effective, given the vast financial savings from deploying sWSI instead of scanners.



## Acknowledgments

sWSI is a commercial product of TerryDr Info Technology Co. Ltd (TerryDr). TerryDr provided sWSI and related data storage services for this project and have authorized publication of the technical contents in this paper. SJTU-PC provided anonymized clinical test data and recruited volunteers for this project. The authors sincerely appreciate the help and service of volunteers and supporting staff of SJTU-PC that are not listed in this paper.

The authors would also like to thank the following pathologists for their contribution to the diagnoses in the reported clinical evaluation: Fen Li from Chengdu Second People's Hospital, China; Zhijuan Guo from Cancer Hospital of Nei Menggu Autonomous Region, China; Xiulan Liu from the First People's Hospital of Neijiang; Huijuan Fang from People's Hospital of Zhengzhou; Yongli Gan from Ningbo Clinical Pathology Diagnosis Center; Shuangmei Zou from Cancer Hospital Chinese Academy of Medical Sciences; Chunyan Gu from the Third People's Hospital of Nantong; Shuping Zhou from the Central Hospital of Yongzhou; Haiyan Shi from the First People's Hospital of Foshan; and Xiaomei Li from Taian City Central Hospital.

## Conflicts of Interest

Author SM is a cofounder of TerryDr Info Technology Co. Ltd, the company that developed and is marketing sWSI as a commercial service.

## Multimedia Appendix 1

Links to virtual slides produced by sWSI and the Leica scanner.

[[XLSX File \(Microsoft Excel File\), 17KB - mhealth\\_v5i9e132\\_app1.xlsx](#)]

## Multimedia Appendix 2

Original diagnosis (Chinese).

[[XLSX File \(Microsoft Excel File\), 29KB - mhealth\\_v5i9e132\\_app2.xlsx](#)]

## References

1. Parwani A, Farahani N, Pantanowitz L. Whole slide imaging in pathology: advantages, limitations, and emerging perspectives. *PLMI* 2015 Jun;23. [doi: [10.2147/PLMI.S59826](#)]
2. Ghaznavi F, Evans A, Madabhushi A, Feldman M. Digital imaging in pathology: whole-slide imaging and beyond. *Annu Rev Pathol* 2013 Jan 24;8:331-359. [doi: [10.1146/annurev-pathol-011811-120902](#)] [Medline: [23157334](#)]
3. Bhargava R, Madabhushi A. Emerging themes in image informatics and molecular analysis for digital pathology. *Annu Rev Biomed Eng* 2016 Jul 11;18:387-412 [[FREE Full text](#)] [doi: [10.1146/annurev-bioeng-112415-114722](#)] [Medline: [27420575](#)]
4. Kwak JT, Hewitt SM. Multiview boosting digital pathology analysis of prostate cancer. *Comput Methods Programs Biomed* 2017 Apr;142:91-99. [doi: [10.1016/j.cmpb.2017.02.023](#)] [Medline: [28325451](#)]
5. Lloyd MC, Monaco JP, Bui MM. Image analysis in surgical pathology. *Surg Pathol Clin* 2016 Jun;9(2):329-337. [doi: [10.1016/j.path.2016.02.001](#)] [Medline: [27241112](#)]
6. Houghton JP, Wilson CP, Dolaghan MJ. Digital pathology imaging offers more benefits than glass slides and microscopes. *BMJ* 2016 Jul 12;i3813. [doi: [10.1136/bmj.i3813](#)]
7. Vodovnik A. Diagnostic time in digital pathology: a comparative study on 400 cases. *J Pathol Inform* 2016;7:4 [[FREE Full text](#)] [doi: [10.4103/2153-3539.175377](#)] [Medline: [26955502](#)]
8. Boyce B. Whole slide imaging: uses and limitations for surgical pathology and teaching. *Biotech Histochem* 2015 Jul;90(5):321-330. [doi: [10.3109/10520295.2015.1033463](#)] [Medline: [25901738](#)]
9. Neil DA, Roberts IS, Bellamy CO, Wigmore SJ, Neuberger JM. Improved access to histopathology using a digital system could increase the organ donor pool and improve allocation. *Transpl Int* 2014 Aug;27(8):759-764 [[FREE Full text](#)] [doi: [10.1111/tri.12320](#)] [Medline: [24629110](#)]
10. Bueno G, Fernández-Carrobles MM, Deniz O, García-Rojo M. New trends of emerging technologies in digital pathology. *Pathobiology* 2016;83(2-3):61-69. [doi: [10.1159/000443482](#)] [Medline: [27100343](#)]
11. García-Rojo M. International clinical guidelines for the adoption of digital pathology: a review of technical aspects. *Pathobiology* 2016;83(2-3):99-109. [doi: [10.1159/000441192](#)] [Medline: [27100834](#)]
12. Griffin J, Treanor D. Digital pathology in clinical use: where are we now and what is holding us back? *Histopathology* 2017 Jan;70(1):134-145. [doi: [10.1111/his.12993](#)] [Medline: [27960232](#)]
13. Hartman D, Pantanowitz L, McHugh J, Piccoli A, OLeary MJ, Lauro GR. Enterprise implementation of digital pathology: feasibility, challenges, and opportunities. *J Digit Imaging* 2017 Jan 23. [doi: [10.1007/s10278-017-9946-9](#)] [Medline: [28116576](#)]

14. Williams BJ, DaCosta P, Goacher E, Treanor D. A systematic analysis of discordant diagnoses in digital pathology compared with light microscopy. *Arch Pathol Lab Med* 2017 May 03. [doi: [10.5858/arpa.2016-0494-OA](https://doi.org/10.5858/arpa.2016-0494-OA)] [Medline: [28467215](https://pubmed.ncbi.nlm.nih.gov/28467215/)]
15. Randell R, Ruddell RA, Treanor D. Barriers and facilitators to the introduction of digital pathology for diagnostic work. *Stud Health Technol Inform* 2015;216:443-447. [Medline: [26262089](https://pubmed.ncbi.nlm.nih.gov/26262089/)]
16. Saliba V, Legido-Quigley H, Hallik R, Aaviksoo A, Car J, McKee M. Telemedicine across borders: a systematic review of factors that hinder or support implementation. *Int J Med Inform* 2012 Dec;81(12):793-809. [doi: [10.1016/j.ijmedinf.2012.08.003](https://doi.org/10.1016/j.ijmedinf.2012.08.003)] [Medline: [22975018](https://pubmed.ncbi.nlm.nih.gov/22975018/)]
17. Gherardi A, Bevilacqua A. Real-time whole slide mosaicing for non-automated microscopes in histopathology analysis. *J Pathol Inform* 2013;4(Suppl):S9 [FREE Full text] [doi: [10.4103/2153-3539.109867](https://doi.org/10.4103/2153-3539.109867)] [Medline: [23766945](https://pubmed.ncbi.nlm.nih.gov/23766945/)]
18. Piccinini F, Bevilacqua A, Lucarelli E. Automated image mosaics by non-automated light microscopes: the MicroMos software tool. *J Microsc* 2013 Dec;252(3):226-250 [FREE Full text] [doi: [10.1111/jmi.12084](https://doi.org/10.1111/jmi.12084)] [Medline: [24111790](https://pubmed.ncbi.nlm.nih.gov/24111790/)]
19. Pradhan D, Monaco S, Parwani A, Ahmed I, Duboy J, Pantanowitz L. Evaluation of panoramic digital images using Panoptiq for frozen section diagnosis. *J Pathol Inform* 2016;7:26 [FREE Full text] [doi: [10.4103/2153-3539.181770](https://doi.org/10.4103/2153-3539.181770)] [Medline: [27217976](https://pubmed.ncbi.nlm.nih.gov/27217976/)]
20. Auguste L, Palsana D. Mobile Whole Slide Imaging (mWSI): a low resource acquisition and transport technique for microscopic pathological specimens. *BMJ Innovations* 2015;1:137-143. [doi: [10.1136/bmjinnov-2015-000040](https://doi.org/10.1136/bmjinnov-2015-000040)]
21. TerryDr Info Technology. Scalable Whole Slide Imaging, sWSI. 2017. URL: <http://en.mydigipath.com/> [accessed 2017-05-07] [WebCite Cache ID 6qGv4LfHd]
22. Thrall MJ, Wimmer JL, Schwartz MR. Validation of multiple whole slide imaging scanners based on the guideline from the College of American Pathologists Pathology and Laboratory Quality Center. *Arch Pathol Lab Med* 2015 May;139(5):656-664. [doi: [10.5858/arpa.2014-0073-OA](https://doi.org/10.5858/arpa.2014-0073-OA)] [Medline: [25927149](https://pubmed.ncbi.nlm.nih.gov/25927149/)]
23. Ozdalga E, Ozdalga A, Ahuja N. The smartphone in medicine: a review of current and potential use among physicians and students. *J Med Internet Res* 2012 Sep 27;14(5):e128 [FREE Full text] [doi: [10.2196/jmir.1994](https://doi.org/10.2196/jmir.1994)] [Medline: [23017375](https://pubmed.ncbi.nlm.nih.gov/23017375/)]
24. Boulos MN, Brewer AC, Karimkhani C, Buller DB, Dellavalle RP. Mobile medical and health apps: state of the art, concerns, regulatory control and certification. *Online J Public Health Inform* 2014;5(3):229 [FREE Full text] [doi: [10.5210/ojphi.v5i3.4814](https://doi.org/10.5210/ojphi.v5i3.4814)] [Medline: [24683442](https://pubmed.ncbi.nlm.nih.gov/24683442/)]
25. Roy S, Pantanowitz L, Amin M, Seethala RR, Ishtiaque A, Yousem SA, et al. Smartphone adapters for digital photomicrography. *J Pathol Inform* 2014;5(1):24 [FREE Full text] [doi: [10.4103/2153-3539.137728](https://doi.org/10.4103/2153-3539.137728)] [Medline: [25191623](https://pubmed.ncbi.nlm.nih.gov/25191623/)]
26. Bay H, Tuytelaars T, Van Gool GL. SURF: speeded up robust features. *LNCS* 2006:404-417. [doi: [10.1007/11744023\\_32](https://doi.org/10.1007/11744023_32)]
27. Juan L, Oubong G. SURF applied in panorama image stitching. In: Proceedings of the international conference on image processing. 2010 Sep 26 Presented at: International conference on image processing; Sept 26-29, 2010; Hong Kong, China p. 495-499. [doi: [10.1109/IPTA.2010.5586723](https://doi.org/10.1109/IPTA.2010.5586723)]
28. Kaynig V, Fischer B, Müller E, Buhmann JM. Fully automatic stitching and distortion correction of transmission electron microscope images. *J Struct Biol* 2010 Aug;171(2):163-173. [doi: [10.1016/j.jsb.2010.04.012](https://doi.org/10.1016/j.jsb.2010.04.012)] [Medline: [20450977](https://pubmed.ncbi.nlm.nih.gov/20450977/)]
29. Jeffreys H, Jeffreys BS. Methods of mathematical physics, 3rd edition. Cambridge, UK: Cambridge University Press; 1999.
30. Tavassoli F, Devilee P. Pathology and genetics of tumours of the breast and female genital organs. In: World Health Organization: Tumours of the Breast and Female Genital Organs. UK: Oxford University Press; 2010.
31. Lakhani SR, Ellis IO, Schnitt SJ, Tan PH, van de Vijver MJ. WHO Classification of Tumours of the Breast (IARC WHO Classification of Tumours). In: WHO Classification of Tumours of the Breast, 4th edition. Geneva: World Health Organization; 2014.
32. Fritz A, Jack A, Parkin DM, Percy C, Shanmugarathan S, Sobin L, et al. International classification of diseases for oncology, 3rd edition, 1st revision. In: International classification of diseases for oncology: ICD-O. Geneva: World Health Organization; 2000.

## Abbreviations

**FoV:** field-of-view

**GPGPU:** general-purpose graphic processing unit

**HOP:** high-order polynomial

**KP:** key point

**LSen:** low sensitivity

**LSpe:** low specificity

**SJTU-PC:** Pathology Center, Shanghai General Hospital, School of Medicine, Shanghai Jiaotong

**SURF:** speeded-up robust features

**sWSI:** scalable Whole Slide Imaging

**VS:** virtual slide

**WSI:** whole slide imaging

*Edited by G Eysenbach; submitted 19.06.17; peer-reviewed by A Manteghinejad, G Tognola; comments to author 15.07.17; revised version received 28.07.17; accepted 14.08.17; published 15.09.17.*

*Please cite as:*

*Yu H, Gao F, Jiang L, Ma S*

*Development of a Whole Slide Imaging System on Smartphones and Evaluation With Frozen Section Samples*

*JMIR Mhealth Uhealth 2017;5(9):e132*

URL: <http://mhealth.jmir.org/2017/9/e132/>

doi: [10.2196/mhealth.8242](https://doi.org/10.2196/mhealth.8242)

PMID: [28916508](https://pubmed.ncbi.nlm.nih.gov/28916508/)

©Hong Yu, Feng Gao, Liren Jiang, Shuoxin Ma. Originally published in JMIR Mhealth and Uhealth (<http://mhealth.jmir.org>), 15.09.2017. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mhealth and uhealth, is properly cited. The complete bibliographic information, a link to the original publication on <http://mhealth.jmir.org/>, as well as this copyright and license information must be included.

Original Paper

# Improvements in Patient Acceptance by Hospitals Following the Introduction of a Smartphone App for the Emergency Medical Service System: A Population-Based Before-and-After Observational Study in Osaka City, Japan

Yusuke Katayama<sup>1</sup>, MD; Tetsuhisa Kitamura<sup>2</sup>, MD, DrPH; Kosuke Kiyohara<sup>3</sup>, DrPH; Taku Iwami<sup>4</sup>, MD, PhD; Takashi Kawamura<sup>4</sup>, MD, PhD; Junichi Izawa<sup>5</sup>, MD, DrPH; Koichiro Gibo<sup>6</sup>, MD, MSc (Statistics); Sho Komukai<sup>7</sup>, PhD; Sumito Hayashida<sup>8</sup>; Takeyuki Kiguchi<sup>4</sup>, MD, PhD; Mitsuo Ohnishi<sup>1</sup>, MD, PhD; Hiroshi Ogura<sup>1</sup>, MD, PhD; Takeshi Shimazu<sup>1</sup>, MD, PhD

<sup>1</sup>Department of Traumatology and Acute Critical Medicine, Osaka University Graduate School of Medicine, Suita, Japan

<sup>2</sup>Division of Environmental Medicine and Population Sciences, Department of Social and Environmental Medicine, Osaka University Graduate School of Medicine, Suita, Japan

<sup>3</sup>Department of Public Health, Tokyo Women's Medical University, Tokyo, Japan

<sup>4</sup>Kyoto University Health Services, Kyoto, Japan

<sup>5</sup>Intensive Care Unit, Department of Anesthesiology, The Jikei University School of Medicine, Tokyo, Japan

<sup>6</sup>Department of Emergency Medicine, Okinawa Prefectural Chubu Hospital, Uruma, Japan

<sup>7</sup>Clinical Research Center, Saga University Hospital, Saga, Japan

<sup>8</sup>Osaka Municipal Fire Department, Osaka, Japan

## Corresponding Author:

Takeshi Shimazu, MD, PhD

Department of Traumatology and Acute Critical Medicine

Osaka University Graduate School of Medicine

2-15 Yamada-oka

Suita,

Japan

Phone:

Email: [shimazu@hp-emerg.med.osaka-u.ac.jp](mailto:shimazu@hp-emerg.med.osaka-u.ac.jp)

## Abstract

**Background:** Recently, the number of ambulance dispatches has been increasing in Japan, and it is therefore difficult for hospitals to accept emergency patients smoothly and appropriately because of the limited hospital capacity. To facilitate the process of requesting patient transport and hospital acceptance, an emergency information system using information technology (IT) has been built and introduced in various communities. However, its effectiveness has not been thoroughly revealed. We introduced a smartphone app system in 2013 that enables emergency medical service (EMS) personnel to share information among themselves regarding on-scene ambulances and the hospital situation.

**Objective:** The aim of this study was to assess the effects of introducing this smartphone app on the EMS system in Osaka City, Japan.

**Methods:** This retrospective study analyzed the population-based ambulance records of Osaka Municipal Fire Department. The study period was 6 years, from January 1, 2010 to December 31, 2015. We enrolled emergency patients for whom on-scene EMS personnel conducted hospital selection. The main endpoint was the difficulty experienced in gaining hospital acceptance at the scene. The definition of difficulty was making  $\geq 5$  phone calls by EMS personnel at the scene to hospitals until a decision to transport was determined. The smartphone app was introduced in January 2013, and we compared the patients treated from 2010 to 2012 (control group) with those treated from 2013 to 2015 (smartphone app group) using an interrupted time-series analysis to assess the effects of introducing this smartphone app.

**Results:** A total of 600,526 emergency patients for whom EMS personnel selected hospitals were eligible for our analysis. There were 300,131 emergency patients in the control group (50.00%, 300,131/600,526) from 2010 to 2012 and 300,395 emergency

patients in the smartphone app group (50.00%, 300,395/600,526) from 2013 to 2015. The rate of difficulty in hospital acceptance was 14.19% (42,585/300,131) in the control group and 10.93% (32,819/300,395) in the smartphone app group. No change over time in the number of difficulties in hospital acceptance was found before the introduction of the smartphone app (regression coefficient:  $-2.43$ , 95% CI  $-5.49$  to  $0.64$ ), but after its introduction, the number of difficulties in hospital acceptance gradually decreased by month (regression coefficient:  $-11.61$ , 95% CI  $-14.57$  to  $-8.65$ ).

**Conclusions:** Sharing information between an ambulance and a hospital by using the smartphone app at the scene was associated with decreased difficulty in obtaining hospital acceptance. Our app and findings may be worth considering in other areas of the world where emergency medical information systems with IT are needed.

(*JMIR Mhealth Uhealth* 2017;5(9):e134) doi:[10.2196/mhealth.8296](https://doi.org/10.2196/mhealth.8296)

## KEYWORDS

emergency medicine; emergency medical services; mobile health; telemedicine; public health

## Introduction

In Japan, when emergency patients call for emergency medical service (EMS) at the scene, on-scene EMS personnel assess the patient's condition and then transport the patient to a hospital that can accept and treat him or her [1]. In this process, ambulances can transport the patient to the hospital only after obtaining permission from the selected hospital via a phone call [1]; this permission is defined as hospital acceptance in Japan [2]. Recently, the number of emergency patients transported to a hospital by EMS has been increasing and exceeding the hospital capacity in Japan. Therefore, it is becoming more difficult to obtain permission and transport and accept emergency patients smoothly and appropriately, especially for severely ill patients and pregnant women [3]. Indeed, our previous study revealed that prehospital factors such as patient's age and time of day were associated with difficulty in hospital acceptance at the scene by analyzing the ambulance records in Osaka City, Japan [4].

Digital information devices such as smartphones and tablet computers have been developing dramatically, and various medical information systems for EMS and medical institutions have also been introduced with the use of these devices in Japan [5]. If EMS personnel at the scene could see the real-time situation of patient transport and hospital acceptance by using a mobile app for smartphones and iPad, they would be able to transport emergency patients to the hospital more smoothly. However, it has not been sufficiently assessed whether the introduction of such information systems would improve the emergency patient transport process by an EMS.

Osaka City is the largest city in western Japan, and there are about 200,000 emergency dispatches every year. We developed a medical information system with smartphone app for an EMS system to facilitate hospital selection and the transport of emergency patients. We call this medical information system as ORION (Osaka emergency information Research Intelligent Operation Network system). It has been in operation in Osaka since January 2013. By analyzing the population-based ambulance records of the Osaka Municipal Fire Department (OMFD) before and after the introduction of ORION, this study aimed to assess the effects of the introduction of this medical information system for an EMS on the difficulty in obtaining hospital acceptance.

## Methods

### Study Design, Population, and Setting

This was a retrospective, population-based, observational study using ambulance records of the OMFD in Osaka City, Japan. The study period was 6 years, from January 1, 2010 to December 31, 2015. Among all emergency dispatches, this study enrolled emergency patients for whom EMS personnel at the scene selected the hospital, and it excluded those who were not transported or were transported to hospitals requested by the patients or their family and those who were transported between hospitals. This study was approved by the ethics committees of Osaka University Graduate School of Medicine and Kyoto University Graduate School of Medicine. The ambulance records of the OMFD were considered administrative records, and the requirement of obtaining patients' informed consent was waived. The researchers dealt only with anonymous data that were not linkable to the patients.

### EMS System and Hospitals in Osaka City

Osaka City, the largest metropolitan community in western Japan, had a population of about 2.7 million in 2017 and covers an area of 222 km<sup>2</sup>. The annual number of emergency patients transported by an EMS in Osaka City is about 200,000. The municipal EMS system is basically the same as that in the other areas of Osaka Prefecture, as previously described [6]. Briefly, an EMS system is operated exclusively by the OMFD and is activated by calling 119. The OMFD had 25 fire stations (60 ambulances) and one dispatch center in 2016. Usually, each ambulance has a crew of 3 emergency providers, including at least one emergency lifesaving technician who is a highly-trained prehospital emergency care provider authorized to use an automated external defibrillator to insert an intravenous line and administer adrenaline and to place advanced airway management [7]. Osaka City had 184 hospitals (32,645 beds) in 2015 [8]. Among those, 99 hospitals, including 6 critical care centers, were designated to accept life-threatening emergency patients from ambulances. During the study period, emergency dispatchers in Osaka City did not make phone calls to hospitals for acceptance; only ambulances crews at the scene selected appropriate hospitals, including critical care medical centers for the emergency patients.



## Smartphone App in the ORION System

EMS personnel at the scene operate a smartphone app connected to the ORION system for each emergency patient. When EMS personnel launch this app and register an emergency patient, the app screen for recording the prehospital time course of the patient's transport is active ([Multimedia Appendix 1](#)). When EMS personnel touch the button labeled "arrival at the scene," the time of arrival at the scene is recorded, and then the location is also recorded by activating the global positioning system on the smartphone. Next, when they leave the scene and touch the button labeled "departure from the scene," the time is recorded and the status of the ambulance changes to "transporting to hospital." On touching the button labeled "arrival at the hospital," the arrival time is recorded, and the status changes to "during treatment of patient." Ambulance statuses from the scene to the hospital are registered in the ORION cloud server and are also reflected on the screen of the medical institution list displayed on the app of other ambulances. For hospital selection, when an EMS personnel at the scene touches the button labeled "to patient check list" in the ambulance status screen, the app screen for recording the patient's status such as vital signs and background becomes active ([Multimedia Appendix 2](#)). EMS personnel can choose symptoms displayed on the app screen that match the patient's complaints, and then the appropriate treating hospitals are listed based on the patient's condition. For example, [Multimedia Appendix 3](#) shows the app screenshot for "chest pain," on which EMS personnel can check items such as "ST-T change" and "dyspnea" for patients with chest pain. In Osaka, medical institutions are categorized based on the feasibility of treating these conditions [9]. When EMS personnel select the check items on this screen, this app shows the list of hospitals that can conduct treatment for potential etiology (or disease) such as emergency percutaneous catheter intervention. In the screen listing the medical institutions ([Multimedia Appendix 4](#)), EMS personnel can select the appropriate hospital for the emergency patient. Colored circles in the screen illustrate the status of emergency patients in each hospital as follows: A red circle means that another ambulance is transporting a patient to that hospital, a yellow circle means that medical staff are now treating a patient transported by another ambulance, and a blue circle means that the hospital is currently neither receiving nor treating any patient. When EMS personnel at the scene select an appropriate hospital for the patient, considering the status of the medical institutions displayed on the app screen, and touch the name of a medical institution, the hospital is automatically called. After receiving consent of hospital acceptance from the hospital, the on-scene EMS personnel start to transport the patient to the hospital. If the hospital rejects the request from EMS personnel, the number of telephone calls required until a receiving hospital is determined is also automatically totaled. EMS personnel at the scene can call critical care centers without filling in the patient's

checklist in the app if they judge that the patient is in critical condition.

The smartphone app data are accumulated in the ORION cloud server, and data managers in cooperation with dispatched EMS personnel directly input or upload the ambulance record of each emergency patient so that it can be merged with the app data. Furthermore, each hospital also directly inputs or uploads the patient's data such as diagnosis and prognosis after hospital acceptance. All of these data, which comprise the smartphone app data, ambulance data, and hospital data, are merged in the ORION cloud server and managed as one large database in Osaka. To collect data from OMFD as well as emergency hospitals, we used a highly confidential line such as a virtual private network (VPN) rather than the Internet, and the server that could safely store massive data from these institutions was separated from the normal Internet line. In addition, we built up two backup servers in addition to the main server to avoid the loss of the ORION database. Analysis of the ORION data is fed back to every fire department and emergency hospital. Public health departments in Osaka will also be able to examine the effect of health policy on emergency medical system using these data ([Figure 1](#)).

This smartphone app of the ORION system was introduced in all areas of Osaka City at the same time on 1<sup>st</sup> January, 2013 and has been working as of July 2017. In Osaka City, the other emergency medical system did not change during the study period except for the introduction of the ORION system.

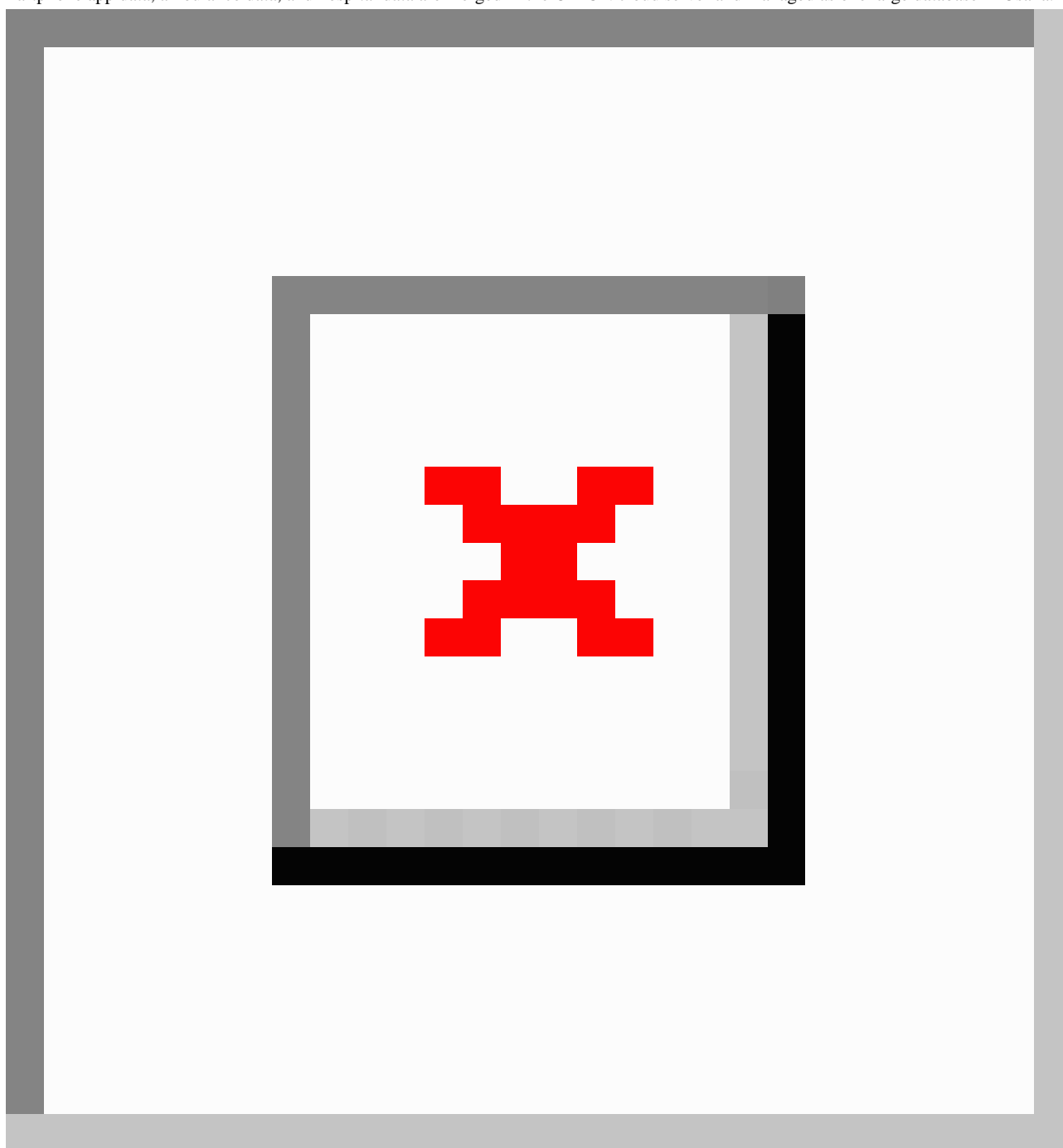
## Data Collection and Quality Control

Data were uniformly collected using specific forms that included age, sex, foreigner, Glasgow Coma Scale (GCS), chronological factors such as the time of day and day of the week, the time course of transportation such as time of the call and hospital arrival, reason for the EMS call, and the total number of phone calls made to hospitals by EMS personnel at the scene. The data were completed by EMS personnel in cooperation with the physicians caring for the patient, transferred to the EMS Information Center of OMFD, and then checked by the investigators. If the dataset was incomplete, the investigators returned it to the responsible EMS personnel for completion of the data.

## Endpoint

The main endpoint was the difficulty in hospital acceptance. In this study, we defined the difficulty in hospital acceptance as the case in which EMS personnel at the scene needed to make  $\geq 5$  phone calls to medical institution before the hospital accepted the patient according to the guidelines regarding the transport and hospital acceptance of emergency patients in Osaka City [9].

**Figure 1.** System configuration of Osaka emergency information Research Intelligent Operation Network system (ORION). All of the data consisting of smartphone app data, ambulance data, and hospital data are merged in the ORION cloud server and managed as one large database in Osaka.



### Statistical Analysis

As a primary analysis, we evaluated changes in the number of the difficulties in hospital acceptance for each month before and after the introduction of the smartphone app, with the use of interrupted time-series analysis to evaluate the introduction effect of a smartphone app on the difficulty in hospital acceptance [10]. In this study, we set the time point as January 2013, when this smartphone app was introduced and assessed the relationship between the number of difficulties in hospital acceptance by month and the elapsed time (months) since the study initiation, adjusting the month as a covariate to take seasonality into consideration. On the basis of previous studies [4,11], we also conducted subgroups such as age group (children

<15 years, adults aged 15-64 years, and the elderly aged  $\geq 65$  years), sex (male or female), time of the day (daytime or nighttime), and day of the week (weekday or weekend or holiday). Furthermore, we assessed the introduction effect especially in emergency cases such as out-of-hospital cardiac arrest, traffic accident, injury by assault, and self-induced drug abuse or gas poisoning or trauma.

Patient and EMS characteristics between the two groups (<5 and  $\geq 5$  phone calls) were assessed by chi-square test for categorical variables and the Wilcoxon test for continuous variables. In this study, we defined emergency patients enrolled in the period from 2013 to 2015 after the introduction of the ORION system as the smartphone app group, that is, the group



on which the smartphone app was used. As a sensitive analysis, we calculated the adjusted odds ratios (AORs) and 95% CIs with the use of a multivariable logistic regression model. We also considered potential confounding factors that existed before the EMS personnel made contact with the patient. These factors included age group (children <15 years, adults aged 15-64 years, and the elderly aged ≥65 years), sex (male or female), foreigner (yes or no), disturbance of consciousness (defined as GCS ≤8, or not), time of the day (daytime or nighttime), day of the week (weekday or weekend or holiday), seasonality (January-March, April-June, July-September, and October-December), use of the smartphone app (yes or no), and reason for the EMS call [1,12]. Reasons why a patient, the patient's family, or bystanders called an ambulance included internal disease; gynecological disease; traffic accident involving vehicle, ship, or aircraft; industrial accident; sports-related disease and injury; asphyxia; trauma by assault; self-induced drug abuse or gas poisoning or trauma; other trauma; and others.

All tests were two-tailed, and *P* values of <.05 were considered statistically significant. Statistical analyses were performed using Statistical Package for the Social Sciences (SPSS) statistical package V.22.0J (IBM Corp).

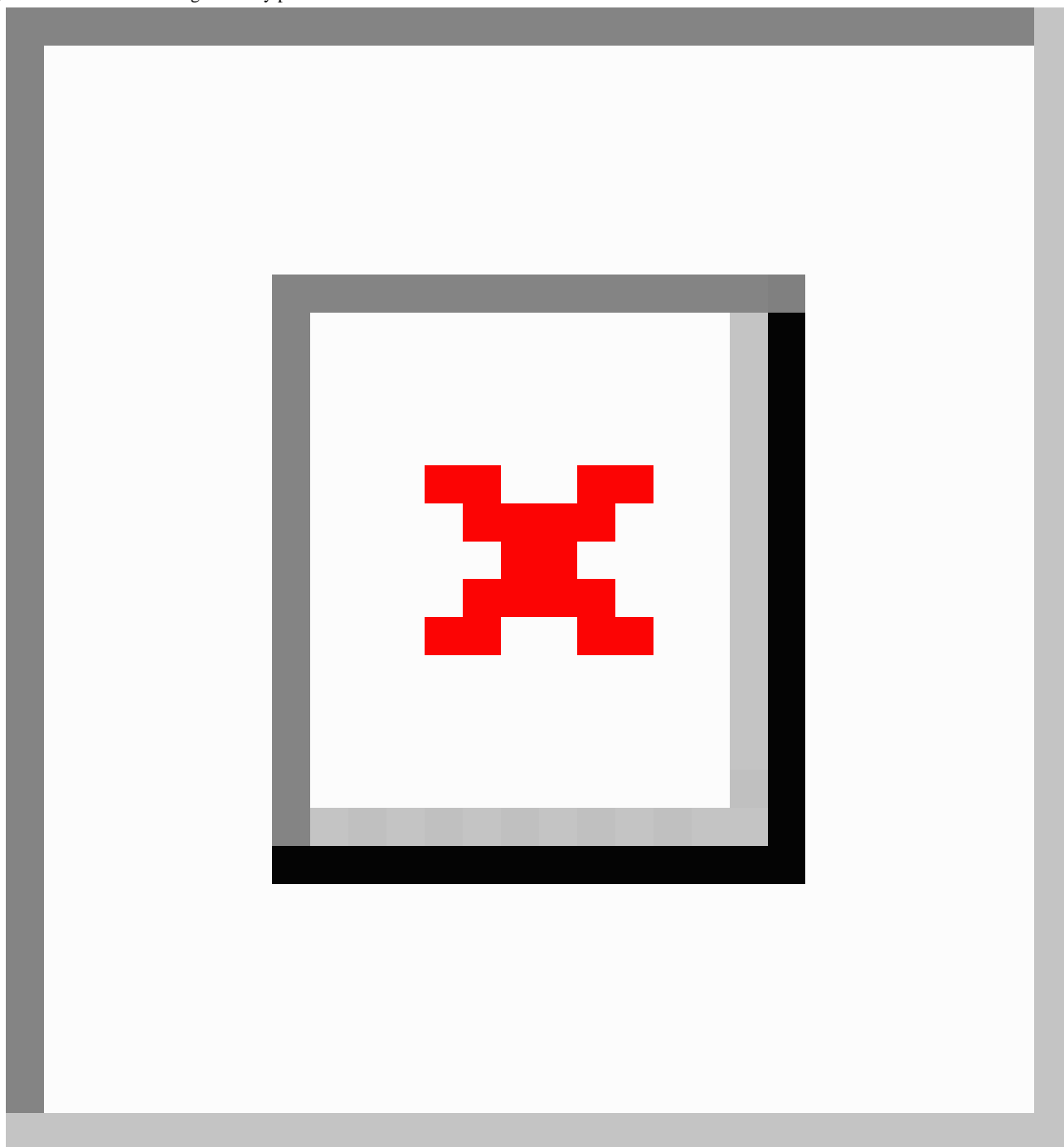
## Results

### Study Population

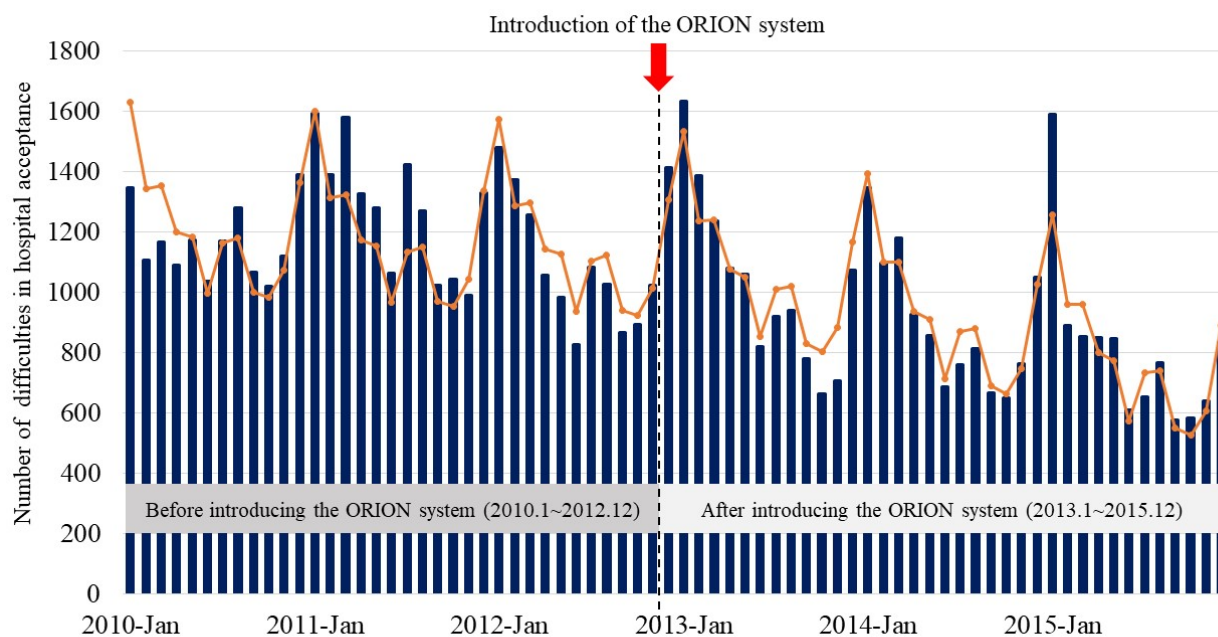
Figure 2 shows the flow of the enrolled patients during the study period from 2010 to 2015 in Osaka City. A total of 1,294,549 emergency dispatches were documented in Osaka City during the study period. A total of 600,526 emergency patients were included for our analysis (300,131 patients [50.00%] in the control group and 300,395 patients [50.00%] in the smartphone app group) after excluding 369,479 patients who were transported to the specific hospitals requested by patients or their family; 63,808 patients who underwent interhospital transport; 260,689 patients who were not transported by ambulance; and 47 patients with other reasons.

### Outcome

Figure 3 shows the number of patients who had difficulty in hospital acceptance (blue bars) by month and the predicted number of difficulties in hospital acceptance calculated from a regression formula with interrupted time-series design (orange line). The graph also shows that seasonality existed in the number of difficulties in hospital acceptance per month.

**Figure 2.** Patient flow during the study periods.

**Figure 3.** The number of difficulties experienced in hospital acceptance by month and the predicted number of difficulties in hospital acceptance by interrupted time-series analysis. The numbers of patients who had difficulty in hospital acceptance are shown by month with blue bars, and the predicted numbers of difficulties in hospital acceptance calculated from a regression formula with interrupted time-series design are shown by the orange line.



**Table 1.** Results of multiple linear regression analysis to detect association between the introduction of the smartphone app for the emergency medical service (EMS) system and the number of difficulties in hospital acceptance per month.

Object	Time trend before the introduction of the smartphone app (2010-2012) (change per month)			Time trend after the introduction of the smartphone app (2013-2015) (change per month)			Change in trends between pre- and postintervention period (2010-2015) (change per month)			$R^2$
	Regression coefficient <sup>a</sup>	95% CI	<i>P</i> value	Regression coefficient <sup>a</sup>	95% CI	<i>P</i> value	Regression coefficient <sup>a</sup>	95% CI	<i>P</i> value	
All	-2.43	-5.49 to 0.64	.118	-11.61	-14.57 to -8.65	<.001	-9.18	-14.56 to -3.81	.001	.810
<b>Subgroup</b>										
Children	-0.67	-0.89 to -0.45	<.001	-0.54	-0.75 to -0.33	<.001	0.13	-0.25 to 0.52	.484	.723
Adult	-1.94	-3.62 to -0.25	.025	-7.00	-8.62 to -5.37	<.001	-5.06	-8.01 to -2.11	.001	.776
Elderly	0.18	-1.31 to 1.67	.807	-4.26	-6.87 to -1.65	.002	-4.08	-5.52 to -2.64	<.001	.839
Daytime	-0.95	-2.04 to 0.14	.087	-3.63	-4.69 to -2.57	<.001	-2.68	-4.59 to -0.77	.007	.801
Nighttime	-1.48	-3.65 to 0.69	.178	-7.99	-10.09 to -5.89	<.001	-6.51	-10.32 to -2.70	.001	.788
Weekday	-1.81	-3.67 to 0.06	.058	-6.94	-8.74 to -5.14	<.001	-5.13	-8.41 to -1.86	.003	.795
Weekend/Holiday	-0.62	-2.26 to 1.01	.450	-4.67	-6.25 to -3.09	<.001	-4.05	-6.92 to -1.19	.006	.774
Out-of-hospital cardiac arrest	0.01	-0.09 to 0.11	.827	-0.20	-0.30 to -0.11	<.001	-0.22	-0.39 to -0.04	.018	.791
Traffic accident	-0.26	-0.68 to 0.15	.205	-1.46	-1.85 to -1.06	<.001	-1.19	-1.91 to -0.47	.002	.617
Trauma by assault	-0.05	-0.26 to 0.15	.598	-0.34	-0.53 to -0.14	.001	-0.28	-0.64 to 0.07	.115	.346
Drug abuse, gas poisoning and trauma by self-injury	-0.44	-0.62 to -0.27	<.001	-0.40	-0.57 to -0.23	<.001	0.04	-0.26 to 0.35	.778	.663

<sup>a</sup>Regression model was adjusted for seasonal effects.

No change over time in the number of difficulties in hospital acceptance was found before the introduction of the smartphone app (regression coefficient: -2.43, 95% CI -5.49 to 0.64), but after its introduction, the number of difficulties in hospital acceptance gradually decreased by month (regression coefficient: -11.61, 95% CI -14.57 to -8.65; Table 1). The regression coefficient of the changes in trends (slope change) between pre- and post-intervention periods was -9.18 (95% CI -14.56 to -3.81). In the subgroup analyses (Table 1), after the introduction of the smartphone app, the number of difficulties in hospital acceptance gradually decreased by month among adults (regression coefficient: -7.00, 95% CI -8.62 to -5.37) and the elderly (regression coefficient: -4.26, 95% CI -6.87 to -1.65). However, as for children, before its introduction, the number of difficulties in hospital acceptance gradually decreased (regression coefficient: -0.67, 95% CI -0.89 to -0.45), and the introduction effect before and after its introduction was not observed in this group (regression coefficient: 0.13, 95% CI -0.25 to -0.52). In all chorological groups, the number of difficulties in hospital acceptance gradually decreased after the introduction of the smartphone app. As for emergency cases,

after the introduction of the smartphone app, the number of difficulties in hospital acceptance gradually decreased by month in out-of-hospital cardiac arrest (regression coefficient: -0.20, 95% CI -0.30 to -0.11) and traffic accident (regression coefficient: -1.46, 95% CI -1.85 to -1.06). However, no change over time in the number of difficulties in hospital acceptance was found before and after the introduction of the smartphone app in trauma by assault (regression coefficient: -0.28, 95% CI -0.64 to 0.07) and drug abuse or gas poisoning or trauma by self-injury (regression coefficient: 0.04, 95% CI -0.26 to 0.35).

Patient and EMS characteristics before and after the introduction of the smartphone app are shown in Table 2. Patients in the smartphone app group were more likely to be older and female, have a lower proportion of disturbance of consciousness, and have a higher proportion of occurrence on weekend and holiday compared with the control group. The number of foreigners treated after the introduction of the smartphone app was higher than that before its introduction. Although the time interval from patient call to contact was similar between the two groups,

the time interval from patient call to hospital arrival in the smartphone app group was longer than that of the control group.

**Table 2.** Patient characteristics before and after the introduction of the smartphone app for emergency medical service (EMS).

Characteristics	Before the introduction of the smartphone app for EMS <sup>a</sup> (2010-2012) n=300,131	After the introduction of the smartphone app for EMS (2013-2015) n=300,395	P value
Age, median (IQR) <sup>b</sup>	49 (24-74)	50 (25-75)	<.001
<b>Age group in years, n (%)</b>			<.001
Children aged ≤14 years	27,892 (9.29)	26,656 (8.87)	
Adults aged 15-64 years	171,316 (57.08)	164,959 (54.91)	
Elderly aged ≥65 years	100,923 (33.63)	108,778 (36.21)	
Male, n (%)	168,559 (56.16)	164,826 (54.87)	<.001
Foreigner, n (%)	542 (0.18)	1227 (0.41)	<.001
Disturbance of consciousness (GCS <sup>c</sup> ≤8), n (%)	16,721 (5.57)	16,331 (5.44)	.026
<b>Time of day, n (%)</b>			.897
Daytime (9:00 am-5:00 pm)	125,885 (41.94)	126,071 (41.97)	
Nighttime (5:00 pm-9:00 am)	174,246 (58.06)	174,322 (58.03)	
<b>Day of week, n (%)</b>			<.001
Weekday	190,796 (63.57)	188,838 (62.86)	
Weekend or holiday	109,335 (36.43)	111,555 (37.14)	
<b>Seasonality, n (%)</b>			<.001
January-March	73,534 (24.50)	75,573 (25.16)	
April-June	72,148 (24.04)	72,339 (24.08)	
July-September	78,701 (26.22)	77,211 (25.70)	
October-December	75,748 (25.24)	75,271 (25.06)	
<b>Reason for EMS call, n (%)</b>			<.001
Internal disease	185,196 (61.71)	180,097 (59.95)	
Gynecological disease	3040 (1.01)	3190 (1.06)	
Traffic accident by car, ship, or aircraft	41,834 (13.94)	38,438 (12.80)	
Injury, toxication, and disease by industrial accident	3373 (1.12)	3756 (1.25)	
Sports-related disease and injury	2362 (0.79)	2533 (0.84)	
Asphyxia	1315 (0.44)	1421 (0.47)	
Trauma by assault	51,480 (17.15)	53,662 (17.86)	
Drug abuse, gas poisoning, and trauma by self-injury	6047 (2.01)	5560 (1.85)	
Other injury	4806 (1.60)	4149 (1.38)	
Others	678 (0.23)	587 (0.20)	
Time from patient's call to contact by EMS in minutes, median (IQR)	5 (3-6)	5 (3-6)	<.001
Time from patient's call to hospital arrival in minutes, median (IQR)	29 (23-39)	31 (24-41)	<.001

<sup>a</sup>EMS: emergency medical service.

<sup>b</sup>IQR: interquartile range.

<sup>c</sup>GCS: Glasgow Coma Scale.

**Table 3.** Number of phone calls and time interval for hospital selection before and after the introduction of the smartphone app for emergency medical service (EMS).

Outcome	Before the introduction of the smartphone app for EMS <sup>a</sup> (2010-2012) n=300,131	After the introduction of the smartphone app for EMS (2013-2015) n=300,395	P value
Number of phone calls until hospital acceptance, median (IQR <sup>b</sup> )	2 (1-3)	1 (1-3)	<.001
Time interval of hospital selection by EMS at the scene in minutes, median (IQR)	4 (2-10)	4 (3-9)	.012
Number of cases needing only one call by EMS until hospital acceptance, n (%)	143,050 (47.66)	154,987 (51.59)	<.001
Number of cases needing ≥5 calls by EMS until hospital acceptance, n (%)	42,585 (14.19)	32,819 (10.93)	<.001
Time interval from EMS scene arrival to hospital arrival in minutes, median (IQR)	24 (16-32)	26 (18-34)	<.001

<sup>a</sup>EMS: emergency medical service.<sup>b</sup>IQR: interquartile range.**Table 4.** Sensitivity analysis of ≥5 calls to hospitals by on-scene emergency medical service (EMS) personnel before and after the introduction of the smartphone app by using a multivariable logistic regression analysis.

Outcome	Percentage of difficulty in hospital acceptance % (n/N)	OR <sup>a</sup> adjusted	95% CI	P value
<b>Introduction of a smartphone app for EMS<sup>b</sup></b>				
Before the introduction of a smartphone app	14.19 (42,585/300,131)	Reference		
After the introduction of a smartphone app	10.93 (32,819/300,395)	0.73	0.72-0.74	<.001

<sup>a</sup>OR: odds ratio.<sup>b</sup>EMS: emergency medical service.

Table 3 shows the number of phone calls and the time taken for hospital selection by EMS personnel with or without the smartphone app. The hospital selection time by EMS personnel until hospital acceptance was similar between the two groups. However, the median number of phone calls was lower in the smartphone app group than in the control group (1 [IQR: 1-3] vs 2 [IQR: 1-3] calls,  $P<.001$ ). The proportion of emergency patients who were accepted by the first hospital called was higher in the smartphone app group than in the control group (51.59% vs 47.66%,  $P<.001$ ), and the proportion of those requiring ≥5 phone calls until hospital acceptance was lower in the smartphone app group than in the control group (10.93% [32,819/300,395] vs 14.19% [42,585/300,131],  $P<.001$ ). The time interval from EMS scene arrival to hospital arrival was significantly longer in the smartphone app group than that in the control group (26 [IQR: 18-34] vs 24 [IQR: 16-32] min,  $P<.001$ ).

The results from a multivariable logistic regression analysis assessing the effects of the introduction of the smartphone app are shown in Table 4. The AOR for the difficulty in hospital acceptance before and after the introduction of the smartphone app was 0.73 (95% CI 0.72-0.74).

## Discussion

### Principal Findings

From the population-based ambulance records in Osaka City, Japan, we evaluated the changes in the number of difficulties in hospital acceptance by month before and after the introduction of the smartphone app with the use of interrupted time-series analysis. Although there were no significant changes in the number of difficulties in hospital acceptance before the introduction of the smartphone app, the number of difficulties in hospital acceptance after the introduction of the smartphone app gradually decreased over time. Therefore, considering our results that the number of difficulties in hospital acceptance gradually decreased by month after the introduction of the smartphone app, we believe that a change in health policy, such as the introduction of a smartphone app, appeared to gradually affect the practice on the front line after the app's introduction. In other words, it appeared to take time for the on-scene EMS personnel to make full use of this app.

Furthermore, we revealed that the introduction of the smartphone app for the EMS system in prehospital settings reduced the difficulty in obtaining hospital acceptance. The ORION system was comprehensively introduced and is operated in both emergency medical institutions and the municipal fire department in Osaka City, one of the biggest cities in Japan. When EMS personnel select an appropriate hospital for

emergency patients, this app enables them to share both the real-time information on the transport situation of other ambulances and the treatment status of other patients after transport. Considering this information, EMS personnel can transport emergency patients to the selected hospital according to patient severity, and the introduction of this app has led to a decrease in the difficulty in obtaining hospital acceptance in this area. Our findings showing improvement of the EMS system by the introduction of an IT system also reinforce the importance of IT in prehospital settings.

On the other hand, the time interval from EMS scene arrival to hospital arrival in this study was significantly longer in the smartphone app group than that in the control group. Although this study defined the difficulty in hospital acceptance as emergency cases required  $\geq 5$  phone calls to a medical institution before the hospital accepted the patient according to the guidelines in Osaka City [9], this result suggests that there might be emergency cases with longer time from EMS scene arrival to hospital arrival with less phone calls in the smartphone app group. However, we consider that it is not appropriate to evaluate the improvements of difficulty in hospital acceptance only by the time interval. For example, it would be important that EMS personnel at the scene are able to transport emergency patients with less phone calls to the distant appropriate hospital that can treat them by advanced procedures, even if the hospital arrival time prolonged. Therefore, in the future, data collection and analysis about treatments after hospital arrival, as well as the time interval from an EMS arrival to in-hospital treatments is also needed to evaluate the improvement effect of difficulty in hospital acceptance by this smartphone app.

Several previous studies have demonstrated that sharing information on the transport situation between medical institutions and ambulances can lead to improvement of an EMS system. McLeod and colleagues [13] reported that a medical information system that shared information about hospital capacity according to the severity of emergency patients reduced ambulance avoidance and improved patient outcome in Calgary City, Canada. Raaber and colleagues [14] demonstrated that obtaining information about the ambulance situation in the emergency department with the use of a geographic information system was associated with a reduction in the waiting time of the trauma team and medical emergency team and improvement in the nurses' workflow when each hospital received emergency patients in Horsens City, Central Denmark. The population density of Osaka City is approximately 12,000 people/km<sup>2</sup> and is much higher than that of Calgary City (1360 people/km<sup>2</sup>) and Horsens City (159 people/km<sup>2</sup>) [15-17], and as such, the 99 emergency medical institutions, including 6 critical care centers in this area, must receive over 220,000 emergency patients transported by ambulances every year [12]. Therefore, the EMS system of Osaka City is always congested, and the request for patient transport by EMS is more likely to exceed the capacity of hospital acceptance. Therefore, sharing real-time information between medical institutions and ambulances with the smartphone app was of help in conducting hospital selection more appropriately by EMS personnel at the scene in Osaka City. From the viewpoint of this and other previous studies,

regardless of the size of the city or the difference in EMS systems, the introduction of IT in prehospital settings would appear to contribute to facilitating the EMS system, including hospital acceptance or treatment after hospital arrival.

In addition, some studies also showed improvements in ambulance diversion with the use of the Internet. Lagoe and colleagues [18] reported on an Internet system monitoring the number of ambulance diversions and interhospital transports in Syracuse, New York, but this system updates the situation only once a day. Sprivulis and colleagues [19] revealed that sharing information about the situation of patient acceptance in each emergency department via the Internet improved ambulance diversion in Perth, Western Australia. This system enabled emergency physicians and nurses to share information about each emergency department at 5-min intervals via a patient tracking system, but the method of tracking patients was not illustrated in their paper. In our smartphone app, when EMS personnel record the time course, such as the time of arrival at the scene or hospital arrival, the ambulance status on the app also changes, and this information is shared on the ORION cloud server, and the status of other ambulances is synchronously updated on the smartphone app. By utilizing this function, EMS personnel can be simply apprised of information on patient transport and hospital acceptance in real time without bothering the hospital staff.

In children, the difficulty in hospital acceptance improved both before and after the introduction of the smartphone app, but change in trend between pre- and post-intervention period was not recognized in a subgroup analysis. As shown in our previous study, emergency medical system for pediatric patients has been well worked before its introduction [4], and the use of a smartphone app for children was not associated with difficulty in hospital acceptance in Osaka City. In cases with trauma by assault and self-injury, no changes in trend were also observed by the introduction of a smartphone app. Although the cooperation between the emergency department and the psychiatry department is necessary to accept self-injured emergency patients, its cooperative relationship has not been sufficiently built up in Osaka City. Therefore, the smartphone app as a tool to search for an appropriate hospital with both departments might not be effective. In Japan, police officers are rarely stationed in medical institutions, and cooperation between the police and medical institutions is not sufficiently established. Therefore, there might be few medical institutions that accept patients with crime-related injuries in Japan, even if EMS personnel selected an appropriate hospital for such patients with the smartphone app. Thus, the improvement effect on the difficulty in hospital acceptance by the introduction of a smartphone app differed in some subgroups because of various factors. Both IT and efforts to improve the patient acceptance system are needed to comprehensively improve the EMS system and further reduce the difficulty in hospital acceptance in the future.

## Limitations

This study has some inherent limitations. First, the purpose of this study was to assess whether the introduction of a smartphone app reduced the difficulty in hospital acceptance,



and we did not assess the effect on the prognosis of the emergency patients. The ORION system has been collecting in-hospital data including patient prognosis since 2015, and we will assess this aspect in the future. Second, we assessed the effect of the smartphone app's introduction based on the unified definition of the difficulty in hospital acceptance regardless of pathological condition, but it may be necessary to define and assess disease-specific difficulty in hospital acceptance. For example, the time interval from onset to call to the start of percutaneous coronary intervention for acute coronary syndrome is one example of an important index [20,21]. However, we could not obtain such information before the introduction of the smartphone app during the study period. Third, we did not have information about potential factors that could affect the improvements of difficulty in hospital acceptance, such as the adherence of the ORION use (ie, the actual rate of using a smartphone app), the decision making of on-scene EMS

personnel to select a hospital from the hospital list by this app, and the number of emergency departments (EDs), ED beds, and ED providers before and after intervention. Finally, this study was an observational study, and there may be unknown confounding factors associated with the difficulty in hospital acceptance.

## Conclusions

We developed a smartphone app for the EMS system that enables EMS personnel at the scene to share various information regarding patient transport by other ambulances or treatment of patients in medical institutions in Osaka City, Japan. Sharing of such information between the ambulances and hospitals in the prehospital setting was associated with decreasing difficulty in hospital acceptance. Our findings may be considered useful for developing an emergency medical information system using IT in other areas of the world.

---

## Acknowledgments

The authors are greatly indebted to all of the EMS personnel working in the Osaka Municipal Fire Department.

---

## Conflicts of Interest

None declared.

---

## Multimedia Appendix 1

ORION smartphone app screenshot for time records. When EMS personnel launch this app and register an emergency patient, the app screen for recording the prehospital time course regarding patient transport is active.

[JPG File, 169KB - [mhealth\\_v5i9e134\\_app1.jpg](#) ]

---

## Multimedia Appendix 2

ORION smartphone app screenshot for patient status. When EMS personnel at the scene touch the button "To patient check list" in the ambulance status screen, the app screen for recording patient status, such as vital signs and background, is active.

[JPG File, 139KB - [mhealth\\_v5i9e134\\_app2.jpg](#) ]

---

## Multimedia Appendix 3

ORION smartphone app screenshot for assessing a patient with chest pain. EMS personnel can choose symptoms displayed on the app screen that match the patient's complaints, and then the appropriate hospitals are listed based on the patient's condition.

[JPG File, 172KB - [mhealth\\_v5i9e134\\_app3.jpg](#) ]

---

## Multimedia Appendix 4

ORION smartphone app screenshot for the hospital list. When EMS personnel select the check items on the screen, the app shows the list of hospitals that can conduct necessary treatments.

[JPG File, 144KB - [mhealth\\_v5i9e134\\_app4.jpg](#) ]

---

## References

1. Ambulance Service Planning Office of Fire and Disaster Management Agency of Japan. FDMA. 2016. Effect of first aid for emergency patients URL: [https://www.fdma.go.jp/neuter/topics/houdou/h28/12/281220\\_houdou\\_2.pdf](https://www.fdma.go.jp/neuter/topics/houdou/h28/12/281220_houdou_2.pdf) [accessed 2017-08-19] [WebCite Cache ID 6qLkYFjBm]
2. Katayama Y, Kitamura T, Kiyohara K, Iwami T, Kawamura T, Hayashida S, et al. Evaluation of factors associated with the difficulty in finding receiving hospitals for traffic accident patients at the scene treated by emergency medical services: a population-based study in Osaka City, Japan. *Acute Med Surg* 2017 Jun 16:- Epub ahead of print. [doi: [10.1002/ams2.291](https://doi.org/10.1002/ams2.291) ]

3. FDMA. 2014. A report on the acceptance of emergency patients by medical institutions URL: [http://www.fdma.go.jp/neuter/about/shingi\\_kento/h27/kyukyu\\_arikata/03/sankou-01.pdf](http://www.fdma.go.jp/neuter/about/shingi_kento/h27/kyukyu_arikata/03/sankou-01.pdf) [WebCite Cache ID 6qLkwqTWq]
4. Katayama Y, Kitamura T, Kiyohara K, Iwami T, Kawamura T, Hayashida S, et al. Factors associated with the difficulty in hospital acceptance at the scene by emergency medical service personnel: a population-based study in Osaka City, Japan. *BMJ Open* 2016 Oct 26;6(10):e013849 [FREE Full text] [doi: [10.1136/bmjopen-2016-013849](https://doi.org/10.1136/bmjopen-2016-013849)] [Medline: [27798040](https://pubmed.ncbi.nlm.nih.gov/27798040/)]
5. FDMA. 2014. Report about future of emergency medical system in Japan URL: [http://www.fdma.go.jp/neuter/about/shingi\\_kento/h26/kyukyu\\_arikata/02/houkokusyo.pdf](http://www.fdma.go.jp/neuter/about/shingi_kento/h26/kyukyu_arikata/02/houkokusyo.pdf) [WebCite Cache ID 6qLmQZ559]
6. Irisawa T, Iwami T, Kitamura T, Nishiyama C, Sakai T, Tanigawa-Sugihara K, et al. An association between systolic blood pressure and stroke among patients with impaired consciousness in out-of-hospital emergency settings. *BMC Emerg Med* 2013 Dec 17;13:24 [FREE Full text] [doi: [10.1186/1471-227X-13-24](https://doi.org/10.1186/1471-227X-13-24)] [Medline: [24341562](https://pubmed.ncbi.nlm.nih.gov/24341562/)]
7. Kajino K, Kitamura T, Iwami T, Daya M, Ong ME, Nishiyama C, et al. Impact of the number of on-scene emergency life-saving technicians and outcomes from out-of-hospital cardiac arrest in Osaka City. *Resuscitation* 2014 Jan;85(1):59-64. [doi: [10.1016/j.resuscitation.2013.09.002](https://doi.org/10.1016/j.resuscitation.2013.09.002)] [Medline: [24036196](https://pubmed.ncbi.nlm.nih.gov/24036196/)]
8. Ministry of Health, Labour and Welfare. MHLW. 2015. Medical faculty survey URL: <http://www.mhlw.go.jp/toukei/saikin/hw/iryosd/15/dl/gaikyo.pdf> [accessed 2017-08-19] [WebCite Cache ID 6qLmfN11A]
9. Osaka Prefectural Government. Osaka. 2012. The guidelines regarding the transport and hospital acceptance of emergency patients in Osaka City URL: [http://www.pref.osaka.lg.jp/attach/2502/00118050/02\\_25\\_02\\_fuikiban\\_03\\_syou\\_03\\_setsu\\_06\\_kyu.pdf](http://www.pref.osaka.lg.jp/attach/2502/00118050/02_25_02_fuikiban_03_syou_03_setsu_06_kyu.pdf) [accessed 2017-08-19] [WebCite Cache ID 6qLndEign]
10. Lopez BJ, Cummins S, Gasparrini A. Interrupted time series regression for the evaluation of public health interventions: a tutorial. *Int J Epidemiol* 2016 Jun 09:- Epub ahead of print [FREE Full text] [doi: [10.1093/ije/dyw098](https://doi.org/10.1093/ije/dyw098)] [Medline: [27283160](https://pubmed.ncbi.nlm.nih.gov/27283160/)]
11. Kitamura T, Iwami T, Kawamura T, Nishiyama C, Sakai T, Tanigawa-Sugihara K, et al. Ambulance calls and prehospital transportation time of emergency patients with cardiovascular events in Osaka City. *Acute Med Surg* 2014 Mar 05;1(3):135-144. [doi: [10.1002/ams2.25](https://doi.org/10.1002/ams2.25)]
12. Osaka Municipal Fire Department. Osaka. 2015. Annual report of Osaka municipal fire department URL: <http://www.city.osaka.lg.jp/shobo/cmsfiles/contents/0000361/361615/2.pdf> [accessed 2017-08-19] [WebCite Cache ID 6qLo2ZaTX]
13. McLeod B, Zaver F, Avery C, Martin DP, Wang D, Jessen K, et al. Matching capacity to demand: a regional dashboard reduces ambulance avoidance and improves accessibility of receiving hospitals. *Acad Emerg Med* 2010 Dec;17(12):1383-1389 [FREE Full text] [doi: [10.1111/j.1553-2712.2010.00928.x](https://doi.org/10.1111/j.1553-2712.2010.00928.x)] [Medline: [21122023](https://pubmed.ncbi.nlm.nih.gov/21122023/)]
14. Raaber N, Duvald I, Riddervold I, Christensen EF, Kirkegaard H. Geographic information system data from ambulances applied in the emergency department: effects on patient reception. *Scand J Trauma Resusc Emerg Med* 2016 Mar 31;24:39 [FREE Full text] [doi: [10.1186/s13049-016-0232-5](https://doi.org/10.1186/s13049-016-0232-5)] [Medline: [27029399](https://pubmed.ncbi.nlm.nih.gov/27029399/)]
15. Osaka. Osaka City statistics information URL: <http://www.city.osaka.lg.jp/toshikeikaku/page/0000379310.html> [accessed 2017-08-19] [WebCite Cache ID 6sp0yN1QC]
16. Calgaryeconomicdevelopment. Calgary economic development URL: <http://www.calgaryeconomicdevelopment.com/research-and-reports/demographics-lp/demographics/> [accessed 2017-05-10] [WebCite Cache ID 6qLopG4zf]
17. DST. Population in Denmark URL: <http://www.dst.dk/en/Statistik/emner/befolkning-og-valg/befolkning-og-befolkningsfremskrivning/folketal> [WebCite Cache ID 6qLpYOINO]
18. Lagoe RJ, Kohlbrenner JC, Hall LD, Roizen M, Nadle PA, Hunt RC. Reducing ambulance diversion: a multihospital approach. *Prehosp Emerg Care* 2003;7(1):99-108. [Medline: [12540152](https://pubmed.ncbi.nlm.nih.gov/12540152/)]
19. Sprivilis P, Gerrard B. Internet-accessible emergency department workload information reduces ambulance diversion. *Prehosp Emerg Care* 2005;9(3):285-291. [doi: [10.1080/10903120590962094](https://doi.org/10.1080/10903120590962094)] [Medline: [16147477](https://pubmed.ncbi.nlm.nih.gov/16147477/)]
20. Shiomi H, Nakagawa Y, Morimoto T, Furukawa Y, Nakano A, Shirai S, CREDO-Kyoto AMI Investigators. Association of onset to balloon and door to balloon time with long term clinical outcome in patients with ST elevation acute myocardial infarction having primary percutaneous coronary intervention: observational study. *BMJ* 2012 May 23;344:e3257. [Medline: [22623632](https://pubmed.ncbi.nlm.nih.gov/22623632/)]
21. Sun CJ, Ribo M, Goyal M, Yoo AJ, Jovin T, Cronin CA, et al. Door-to-puncture: a practical metric for capturing and enhancing system processes associated with endovascular stroke care, preliminary results from the rapid reperfusion registry. *J Am Heart Assoc* 2014;3(2):e000859 [FREE Full text] [Medline: [24772523](https://pubmed.ncbi.nlm.nih.gov/24772523/)]

## Abbreviations

**AOR:** adjusted odds ratio  
**EMS:** emergency medical service  
**GCS:** Glasgow Coma Scale  
**IQR:** interquartile range  
**IT:** information technology  
**OMFD:** Osaka Municipal Fire Department  
**OR:** odds ratio  
**ORION:** Osaka emergency information Research Intelligent Operation Network system

**SPSS:** Statistical Package for the Social Sciences

**VPN:** virtual private network

**XML:** Extensible Markup Language

*Edited by G Eysenbach; submitted 26.06.17; peer-reviewed by T Goto, M Okubo, T Matsuyama; comments to author 12.07.17; revised version received 22.07.17; accepted 27.07.17; published 11.09.17.*

*Please cite as:*

*Katayama Y, Kitamura T, Kiyohara K, Iwami T, Kawamura T, Izawa J, Gibo K, Komukai S, Hayashida S, Kiguchi T, Ohnishi M, Ogura H, Shimazu T*

*Improvements in Patient Acceptance by Hospitals Following the Introduction of a Smartphone App for the Emergency Medical Service System: A Population-Based Before-and-After Observational Study in Osaka City, Japan*

*JMIR Mhealth Uhealth 2017;5(9):e134*

URL: <http://mhealth.jmir.org/2017/9/e134/>

doi: [10.2196/mhealth.8296](https://doi.org/10.2196/mhealth.8296)

PMID: [28893725](https://pubmed.ncbi.nlm.nih.gov/28893725/)

©Yusuke Katayama, Tetsuhisa Kitamura, Kosuke Kiyohara, Taku Iwami, Takashi Kawamura, Junichi Izawa, Koichiro Gibo, Sho Komukai, Sumito Hayashida, Takeyuki Kiguchi, Mitsuo Ohnishi, Hiroshi Ogura, Takeshi Shimazu. Originally published in JMIR Mhealth and Uhealth (<http://mhealth.jmir.org>), 11.09.2017. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mhealth and uhealth, is properly cited. The complete bibliographic information, a link to the original publication on <http://mhealth.jmir.org/>, as well as this copyright and license information must be included.

Original Paper

# Atrial Fibrillation Screening in Nonmetropolitan Areas Using a Telehealth Surveillance System With an Embedded Cloud-Computing Algorithm: Prospective Pilot Study

Ying-Hsien Chen<sup>1</sup>, MD; Chi-Sheng Hung<sup>1</sup>, MD, PhD; Ching-Chang Huang<sup>1</sup>, MD; Yu-Chien Hung<sup>2</sup>, MD; Juey-Jen Hwang<sup>1,3\*</sup>, MD, PhD; Yi-Lwun Ho<sup>1\*</sup>, MD, PhD

<sup>1</sup>Department of Internal Medicine, National Taiwan University Hospital, Taipei, Taiwan

<sup>2</sup>Department of Medicine, National Taiwan University Hospital JinShan Branch, New Taipei, Taiwan

<sup>3</sup>Department of Internal Medicine, National Taiwan University Hospital Yun-Lin Branch, Yun-Lin, Taiwan

\*these authors contributed equally

**Corresponding Author:**

Yi-Lwun Ho, MD, PhD

Department of Internal Medicine

National Taiwan University Hospital

No.7, Chung Shan S. Rd.

Taipei, 10002

Taiwan

Phone: 886 2 2312 3456 ext 63651

Fax: 886 2 2353 9346

Email: [ylho@ntu.edu.tw](mailto:ylho@ntu.edu.tw)

## Abstract

**Background:** Atrial fibrillation (AF) is a common form of arrhythmia that is associated with increased risk of stroke and mortality. Detecting AF before the first complication occurs is a recognized priority. No previous studies have examined the feasibility of undertaking AF screening using a telehealth surveillance system with an embedded cloud-computing algorithm; we address this issue in this study.

**Objective:** The objective of this study was to evaluate the feasibility of AF screening in nonmetropolitan areas using a telehealth surveillance system with an embedded cloud-computing algorithm.

**Methods:** We conducted a prospective AF screening study in a nonmetropolitan area using a single-lead electrocardiogram (ECG) recorder. All ECG measurements were reviewed on the telehealth surveillance system and interpreted by the cloud-computing algorithm and a cardiologist. The process of AF screening was evaluated with a satisfaction questionnaire.

**Results:** Between March 11, 2016 and August 31, 2016, 967 ECGs were recorded from 922 residents in nonmetropolitan areas. A total of 22 (2.4%, 22/922) residents with AF were identified by the physician's ECG interpretation, and only 0.2% (2/967) of ECGs contained significant artifacts. The novel cloud-computing algorithm for AF detection had a sensitivity of 95.5% (95% CI 77.2%-99.9%) and specificity of 97.7% (95% CI 96.5%-98.5%). The overall satisfaction score for the process of AF screening was 92.1%.

**Conclusions:** AF screening in nonmetropolitan areas using a telehealth surveillance system with an embedded cloud-computing algorithm is feasible.

(*JMIR Mhealth Uhealth* 2017;5(9):e135) doi:[10.2196/mhealth.8290](https://doi.org/10.2196/mhealth.8290)

**KEYWORDS**

atrial fibrillation; screen; cloud-computing algorithm; electrocardiography

## Introduction

### Health Threats From Atrial Fibrillation

Atrial fibrillation (AF), a common form of sustained arrhythmia that has a significant impact on population health, is now a growing public health problem [1]. According to the Rotterdam Study, a large European population-based study, the overall prevalence of AF is 5.5% in a population of 55 years and older, rising from 0.7% in the age group of 55 to 59 years to 17.8% in those aged 85 years and older [2]. Meanwhile, in the ATRIA study from the United States, a cross-sectional study of adults aged 20 years or older, the overall prevalence of diagnosed AF was 0.95%, ranging from 0.1% among adults younger than 55 years to 9.0% in persons aged 80 years or older [3]. Both studies consistently demonstrated that the incidence of AF increased with age and was higher in men than in women. The number of patients with AF is likely to increase 2.5-fold during the next 50 years, reflecting the growing proportion of elderly individuals [3].

AF is considered a degenerative disease triggered by interactions with various substrate patterns, and it shares strong epidemiological associations with other cardiovascular diseases such as heart failure and coronary artery disease, rheumatic heart disease, hypertension, and diabetes. The incidence of AF varies depending on the population studied. The overall rate of incidence is 9.9 per 1000 person-years in a population older than 55 years according to the Rotterdam Study [2], whereas the Framingham Heart Study reports that the annual incidence is 0.5 per 1000 person-years [4]. AF is considered a risk factor for stroke [5,6] and congestive heart failure [7], and patients newly diagnosed with AF have a higher mortality risk, especially within the first 4 months of diagnosis [8]. There is a near 5-fold increase in the incidence of stroke when AF is present [6], and the annual risk of stroke ranges from 2% to 18% depending on other risk factors [9].

### Atrial Fibrillation Screening

Antithrombotic therapies, including vitamin K antagonists (VKA) [10,11] and nonvitamin K antagonist oral anticoagulants (NOAC) [12-15], reduce the risk of stroke in patients with AF. Currently, there is no effective way to prevent or cure AF and undiagnosed AF is common, especially in older populations and for patients with heart failure [16]. Previously, undiagnosed AF was found in 1.4% of those aged >65 years, which suggests that opportunistic screening for silent AF may be cost-effective in elderly populations [17]. The European Society of Cardiology (ESC) 2016 guidelines recommended conducting such screening by pulse taking or electrocardiogram (ECG) rhythm strips [18]. Currently, screening of older populations can be achieved through short-term ECG, pulse palpation [19], single-lead ECG [20-22], and blood pressure (BP) measurement with patented

AF algorithm [23]. However, the sensitivity, accuracy, and accessibility of these modalities may affect the dissemination of AF screening, and the traditional 12-lead ECG has inherent limitations for its application to AF screening, especially in nonmetropolitan areas where the accessibility of health care is limited. The Telehealth Center of the National Taiwan University Hospital (NTUH) has conducted the fourth-generation telehealth service since 2009 for patients with cardiovascular diseases [24-26]. By using ECG recorders (DigiO2 Cardio Care ECG recorder, DigiO2 International Co., Ltd), ECG measurement has become convenient and feasible at a distance from health care organizations. We conducted a prospective AF screening study in a nonmetropolitan area using a DigiO2 Cardio Care ECG recorder with a telesurveillance system embedded with a cloud-computing algorithm. The main purpose of the study was to evaluate the feasibility and accuracy of AF screening in nonmetropolitan areas.

## Methods

The Taiwan ELEctroHEALTH (TELEHEALTH) study group conducted a prospective clinical study of AF screening in nonmetropolitan areas of Jinshan, Wanli, Shimen, and Sanzhi districts, New Taipei City, Taiwan. These areas were the northern coast of Taiwan with Yangmingshan National Park mountain barrier separating these areas from the metropolitan city (Taipei City) (Figure 1). The AF screening was conducted in the community during the advocacy activities held by the local health bureaus, various government agencies, and the National Taiwan University Hospital, Jinshan branch. A booth for AF screening was established. Local residents who attended the advocacy activity without active life-threatening medical conditions and aged older than 20 years were enrolled after obtaining their informed consent. Trained personnel would assist participants when performing ECG measurement according to the step-by-step instruction. The electrodes placement followed the manufacturer's recommendation and in accordance with the American Heart Association recommendation [27]. Misplacement of electrodes can cause significant alteration to wave amplitudes or morphology, which may invalidate the use of recording [28]. All participants obtained an ECG measurement using the ECG recorder and completed a questionnaire regarding their individual health status, medical condition, and satisfaction toward the process of AF screening and handheld ECG measurement.

The measured ECGs were transmitted to a Web-based telesurveillance system at the Telehealth Center. An independent cardiologist performed the physician-based ECG interpretations. The computer-based ECG auto-interpretation was executed automatically according to the cloud-computing algorithm developed by the TELEHEALTH study group [29]. The institutional review board at NTUH approved the study protocol.



**Figure 1.** Map of nonmetropolitan area of JinShan, Wanli, Shimen and Sanzhi district.

### Single-Lead ECG Recorder

AF screening was performed in a stepwise manner under the assistance of trained personnel. After launching the Android-based AF screening application on an Android tablet, step-by-step instructions for the screening were displayed. Participants were instructed to connect the card reader to the Android tablet. If the card reader was connected to the tablet without launching the application first, the application would launch once the card reader was connected and proceed to the next step automatically. In the next step, participants were instructed to insert their identity-specific National Health Insurance Cards into the card reader to assess their personal information. After confirming their identity, a confirmation message was displayed, followed by an illustration showing how to position the ECG electrodes (Figure 2). The right arm limb lead (yellow) had to be placed anywhere between the right shoulder and right arm, the left arm limb lead (black) anywhere between the left shoulder and left arm, and the left leg lead (red) anywhere below the left torso and above the left ankle.

With the electrodes attached to their appropriate positions, the participants were instructed to press the measurement button, and the ECG recorder (DiGiO2 Cardio Care ECG Recorder)

recorded a 15-second single-lead ECG. The measurements were transferred automatically from the ECG recorder to the tablet through a Bluetooth connection and could be explored instantaneously on the tablet through the application (Figure 2). The ECG was relayed from the tablet to the server at the Telehealth Center through a wireless local area network (WLAN) once the user had confirmed the upload. The ECG was then ready to be retrieved from the server at the Telehealth Center by the telesurveillance system, proceed to ECG auto-interpretation by the cloud-computing algorithm, and receive physician interpretation.

### Telesurveillance System

The telesurveillance system (Figure 3) is a Web-based platform developed by the Graduate Institute of Biomedical Electronics and Bioinformatics, National Taiwan University, Taiwan [29]. The telesurveillance system was operated under a service-oriented architecture framework with the Health Level Seven standard. Part of the function of the telesurveillance system includes exploring and reviewing biometric data, such as single-lead ECGs, BP, heart rate, and oximetry, and transferring the patient data to our Telehealth Center daily and on demand [26]. The physician's ECG interpretation was also recorded on this Web-based platform.

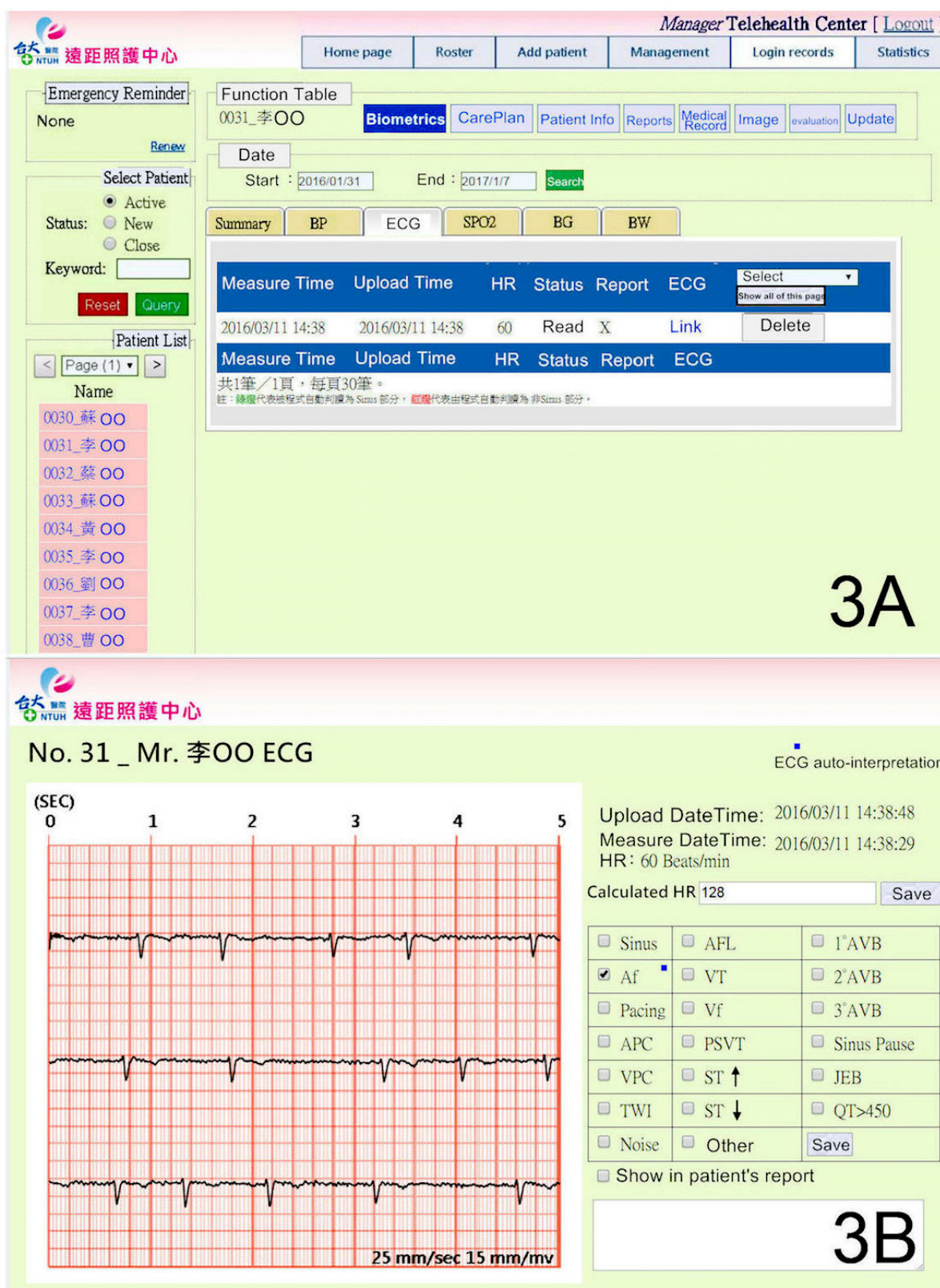


**Figure 2.** Steps in atrial fibrillation screening.

### Cloud-Computing Algorithm

The design and algorithm of the Web-based ECG auto-interpretation were described in a previous study [29]. After removing baseline noise by finite impulse response filter, the key features of the ECG waveforms extracted were processed by support vector machine or rule-based processing to construct

a classification model that can suggest diagnosis. A modified cloud-computing algorithm for determining AF was adopted during the study, where the detection of atrial premature complex (APC) or ventricular premature complex (VPC) was not included. Figure 3 demonstrates the result of ECG auto-interpretation by the cloud-computing algorithm on the telesurveillance system.

**Figure 3.** Tele-surveillance system and auto-interpretation by cloud-computing algorithm.

### ECG Quality Analysis and Artifacts Grading

An independent cardiologist evaluated the ECG quality while performing physician ECG interpretation. The quality of ECG was categorized from grade 0 to grade 3 artifacts. A grade 0 artifact represents excellent image quality without artifacts. A

grade 1 artifact represents an artifact percentage of <33% with recognizable P waves. A grade 2a artifact represents an artifact percentage of 33% to 66% or with a mild wandering baseline artifact with recognizable P waves. A grade 2b artifact represents an artifact percentage of >66% or with significant wandering baseline artifacts interfering with P wave recognition. A grade



3 artifact represents significant artifacts without recognizable P waves or QRS complexes. Figure 4 shows the illustrated examples of ECG artifacts in each categorization.

### Statistical Analysis

All continuous variables were expressed as mean (standard deviation [SD]) and categorical variables in numbers and percentages. Stata/SE 11.0 for Windows (StataCorp LP) was used for statistical analyses. The results of sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) were further stratified according to age.

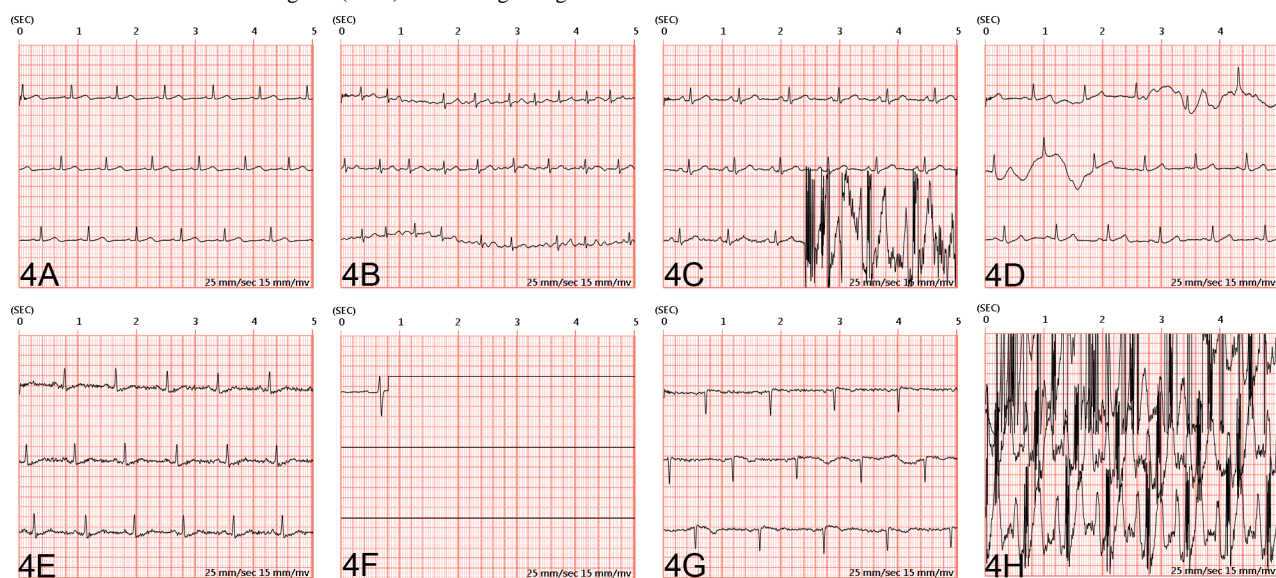
## Results

Between March 11, 2016 and September 8, 2016, 967 ECGs were recorded from 922 residents (age: 58.1 [SD 15.0] years; aged >65 years: 426/922, 46.2%; male participants: 337/922, 36.6%) in Jinshan, Wanli, Shimen, and Sanzhi districts, New Taipei City, Taiwan, through community-based AF screening. Among those who received ECG measurements, 885 participants received a single ECG test, whereas 34 participants received two ECG tests, and 3 participants received three or more ECG tests. Among the 967 ECG records, 807 (807/967, 83.5%) were categorized as grade 0 artifacts, 124 (124/967, 12.8%) ECGs were categorized as grade 1 artifacts, 26 (26/967, 2.7%) ECGs had grade 2a artifacts, 8 (8/967, 0.8%) ECGs had grade 2b artifacts, and 2 (2/967, 0.2%) ECGs were classified as grade 3 artifacts.

The results of physician's ECG interpretations demonstrated a sinus rhythm in 939 (939/967, 97.1%), including a sinus rhythm without ectopic beats in 907 patients (907/967, 93.8%), a sinus rhythm with APC in 15 patients (15/967, 1.6%), a sinus rhythm with VPC in 19 patients (19/967, 2.0%), and a sinus rhythm with both APC and VPC in 1 patient (1/967, 0.1%). There were 22 (22/967, 2.3%) AF rhythms, including AF rhythm with VPC in 3 (3/967, 0.3%). A paced rhythm in 1 (1/967, 0.1%) and a junctional rhythm in 3 (3/967, 0.3%) were identified by physician's-determined ECG interpretations. Two ECGs (2/967, 0.2%) had no physician-determined ECG interpretations because of the presence of grade 3 ECG artifacts. The estimated prevalence of AF from our study population is 2.4% (22/922).

The results of the ECG auto-interpretation demonstrated no AF in 922 ECG measurements (92/967, 95.3%) and AF in 45 ECG measurements (45/967, 4.7%). After excluding ECG measurements with grade 3 artifacts, the overall performance for the AF screening cloud-computing algorithm had a sensitivity of 95.5% (95% CI 77.2%-99.9%) and a specificity of 97.7% (95% CI 96.5%-98.5%), with a PPV of 48.8% (95% CI 38.5%-59.3%), and an NPV value of 99.9% (95% CI 99.3%-100.0%) for detecting a disease prevalence (by ECG numbers) of 2.3% (22/965). When stratified to participants aged older than 65 years, the cloud-computing algorithm for AF screening had a sensitivity of 94.4% (95% CI 72.7%-99.9%) and a specificity of 96.4% (95% CI 94.2%-98.0%), with a PPV of 53.1% (95% CI 40.5%-65.4%) and an NPV of 99.8% (95% CI 98.4%-100.0%) for detecting a disease prevalence (by ECG numbers) of 4.1% (Table 1).

**Figure 4.** Illustrative electrocardiogram (ECG) and noise grading.



**Table 1.** Performance of the cloud-computing algorithm stratified according to age.

Result of cloud-computing algorithm	Age ≤65 years (n=526)	Age >65 years (n=439)	Overall (N=965)
Sensitivity, % (95% CI)	100.0 (95% CI 39.8%-100.0%)	94.4 (95% CI 72.7%-99.9%)	95.5 (95% CI 77.2%-99.9%)
Specificity, % (95% CI)	98.7 (95% CI 97.3%-99.5%)	96.4 (95% CI 94.2%-98.0%)	97.7 (95% CI 96.5%-98.5%)
Positive predictive value, % (95% CI)	36.4 (95% CI 21.5%-54.4%)	53.1 (95% CI 40.5%-65.4%)	48.8 (95% CI 38.5%-59.3%)
Negative predictive value, % (95% CI)	100.0	99.8 (95% CI 98.4%-100.0%)	99.9 (95% CI 99.3%-100.0%)
Disease prevalence (by ECG numbers), % (n/N)	0.8 (4/526)	4.1 (18/439)	2.3 (22/965)
Disease prevalence (by screening resident numbers), % (n/N)	0.8 (4/496)	4.2 (18/426)	2.4 (22/922)

## False Positive Analysis

A total of 22 false positive ECGs from the cloud-computing algorithm were identified. We analyzed the ECG quality and ECG characteristics. In addition, 16 (72.7%) ECG measurements had good image quality (grade 0 artifact), 3 (13.3%) had grade 1 artifacts, and another 3 (13.3%) had grade 2a artifacts. Other ECG features in these 22 false positives include a sinus rhythm with APC in 8 patients (36.4%), a sinus rhythm with VPC in 8 (36.4%), and a QRS complex voltage <0.5 mV in 9 (40.9%). There was only 1 false negative ECG from the cloud-computing algorithm.

## Satisfaction Questionnaire

In total, 825 satisfaction questionnaires were obtained from 922 participants, with an overall satisfaction score of 92.1%. Regarding AF screening convenience, 91.2% (752/825) of participants rated 4 or 5 on a scale of 1 to 5, and 89.5% (738/825) of the participants would recommend others to receive AF screening in the future.

## Discussion

### Principal Findings

We conducted a prospective AF screening study on nonmetropolitan areas through a single-lead ECG recorder and a telehealth surveillance system with an embedded cloud-computing algorithm. We compared the results obtained from ECG auto-interpretation by the cloud-computing algorithm and a cardiologist and found that the ECG recorder can obtain high-quality ECG images and that the ECG auto-interpretation by the cloud-computing algorithm for AF detection has a sensitivity of 95.5% and a specificity of 97.7%, with a relative low PPV of 48.8%. The overall satisfaction score for the process of AF screening was 92.1%.

With the prevalence and incidence of AF increasing with age, AF is a growing public health problem [1]. AF management has recently evolved to include high-risk patient identification via the CHA<sub>2</sub>DS<sub>2</sub>-VASc score, bleeding risk evaluation via the HAS-BLED score, and the SAME-TT2R<sub>2</sub> score for the initial selection of VKA or NOAC therapy [30]. The etiology for AF comprises complex pathophysiological changes in the atrium, including stretch-induced atrial fibrosis, hypocontractility, fatty infiltration, inflammation, remodeling, ischemia, and ion channel dysfunction.

There is no effective method to prevent AF, although some retrospective analyses from large randomized trials showed a lower incidence of new-onset AF in patients receiving angiotensin-converting enzyme inhibitors or angiotensin receptor blockers [31,32]. The early detection of AF and timely treatment before the first complications occur remain the best practice according to contemporary practices. According to the 2016 ESC guidelines, opportunistic screening for AF by pulse taking or the application of ECG rhythm strips in patients >65 years of age is now a class I indication [18]. However, traditional pulse palpation can be unreliable, and 12-lead ECG recordings can be cumbersome and might not readily be available or accessible to put into practice for AF screening practices. Other modalities such as AF detection during automated BP measurement, handheld ECG machines, mobile phone ECGs, and finger-probe instruments are thus under investigation for AF screening.

### Pulse Palpation

The classical sign of AF by pulse palpation is an irregular pulse. Sanmartin et al conducted a campaign for information and diagnosis of AF through pulse palpation. Among 1532 participants with a mean age of 73 (SD 7) years, 4.11% (63/1532) were identified with AF, including 1.11% (17/1532) with newly diagnosed AF [33]. Cooke G et al investigated three studies (2385 patients) that compared pulse palpation with ECG. The estimated sensitivity of pulse palpation ranged from 91% to 100% and specificity ranged from 70% to 77%. Pooled sensitivity was 94% (95% CI 84%-97%), and pooled specificity was 72% (95% CI 69%-75%). Given that pulse palpation has a high sensitivity but relatively low specificity for AF detection, it was considered a suitable tool for ruling out AF [34]. The diagnosis of AF still requires rhythm documentation using ECG [18,35].

### Twelve-Lead ECG

The typical pattern of AF on an ECG would be irregular RR intervals and no discernible, distinct P waves. Although economic analyses have concluded the cost-effectiveness of either annual screening [36] or opportunistic screening [17] by using a 12-lead ECG in those aged ≥65 years, the accessibility and higher cost for a 12-lead ECG may limit the dissemination of systemic screening. Other screening modalities are now under investigation for feasibility, accuracy, cost-effectiveness, and the potential to replace the 12-lead ECG.

## Screening for AF With Automated Blood Pressure Measurement

A specific algorithm for AF detection during automated BP measurement was developed and implemented in a novel oscillometric device (Microlife WatchBP Home-A). According to a meta-analysis composed of 6 studies ( $n=2332$ ) performed by Verberk et al, the highest diagnostic accuracy for AF detection would be provided by using the Microlife BP monitor to take three sequential readings with at least two detecting AF, giving an estimated pooled sensitivity of 0.98 (95% CI 0.95-1.00) and specificity of 0.92 (95% CI 0.88-0.96) [23]. In 2013, the UK National Institute for Health and Care Excellence recommended this device for AF screening during routine office BP measurement in primary care for patients aged  $\geq 65$  years [37]. Although AF detection with routine-automated BP measurement could be a potential screening tool in the elderly people, it still requires confirmation by ECG [18,35].

## Handheld Single-Lead ECG Device

Desteghe et al evaluated the usability, accuracy, and cost-effectiveness of 2 handheld single-lead ECG devices (MyDiagnostick and AliveCor) for AF screening in a hospital population. The performance of the automated algorithm of each device was evaluated against a full 12-lead or 6-lead ECG recording. In the study, handheld recordings were not possible in 7% to 21.4% of hospital patients because they were unable to hold the devices properly. Both automated algorithms for each device had suboptimal sensitivity and specificity results. The sensitivity for MyDiagnostick was 81.8% to 89.5%, with a specificity of 94.2% to 95.7%. For AliveCor, the sensitivity was 54.5% to 78.9%, with a specificity of 97.5% to 97.9% [38].

For handheld DigiO2 Cardio Care ECG recorder, it is optional to use either dry contact electrodes or adhesive electrodes for ECG measurement; we preferred to use adhesive electrodes for ECG measurements to guarantee a better ECG quality and to test the accuracy of the cloud-computing algorithm. Therefore, use of adhesive electrodes for ECG measurement ensures that no participants are excluded from AF screening for being unable to hold the device. Moreover, even though an automated algorithm was embedded in the handheld single-lead ECG device in another study [38], our novel cloud-computing algorithm was embedded in a telesurveillance system, allowing a more comprehensive algorithm and a greater storage capacity for AF detection measurements.

## Finger-Probe Instruments

Lewis et al analyzed the application of a plethysmographic analysis of fingertip pulses in the detection of AF. A 12-lead ECG was recorded immediately for comparison when the finger probe was disconnected. The device detected all cases of AF (100% sensitivity), and a specificity of 91.9% (8.1% false positives) was obtained [39]. The finger probe may provide a potential tabletop instrument that allows for AF screening; however, it still requires a confirmatory ECG.

## Novel AF Detection Modalities

Photoplethysmography (PPG) is an optical method to measure changes in tissue blood volume caused by the pressure pulse. By placing a finger in contact with a mobile phone camera, the

PPG waveform can be acquired through the light intensity reflected from a finger illuminated by the light-emitting diode mobile flash [40]. Chan et al [41] investigated the ability of PPG to diagnose AF in real-world situations. By using Cardiio Rhythm mobile app, the diagnostic sensitivity of the Cardiio Rhythm for AF detection was 92.9% (95% CI 77%-99%), which was higher than that of the AliveCor automated algorithm (71.4%; 95% CI 51%-87%). The specificities of Cardiio Rhythm and the AliveCor automated algorithm were comparable [41].

Nemati et al [42] proposed a noise-resistant machine learning approach to detect AF from noisy ambulatory PPG recorded from the wrist wearable technology. The preliminary result showed a sensitivity of 97% and specificity of 94% in 46 study subjects. Couderc et al proposed another novel technology for contactless detection of AF by using facial video recordings. The video plethysmographic signal acquired using a standard Web camera was extracted. A novel quantifier of pulse variability called the pulse harmonic strength was introduced to detect the presence of AF, which showed a 20% detection error rate [43]. Meanwhile, these new modalities still require confirmatory ECG for AF diagnosis.

## Strengths and Limitations

Although multiple modalities demonstrated a potential to be used in AF screening, some strengths and key features differentiated our study from others. First, we used adhesive electrodes to receive ECG signals to obtain the best ECG quality possible because we understand that the ECG is essential for AF diagnosis [18,35]. The ECG measurement first appears on site and can be explored through the application. Artifact ECGs can be identified before the final submission, and patients can repeat the ECG measurement when artifacts are presented. Although we use adhesive electrodes for ECG measurement, participants do not need to take off their clothes as with the traditional 12-lead ECG. By avoiding embarrassment and inconvenience brought by the removal of clothing, women should be more willing to receive AF screening. In fact, more women than men participated in our study (73.4%).

Second, our novel cloud-computing algorithm was embedded in a telesurveillance system, but not in the single-lead ECG devices per se, allowing a more comprehensive algorithm for AF detection. Moreover, the single-lead ECG recorder was used for ECG measurement only, and all measurements were transferred and stored in the cloud. There is only a requirement for temporary storage, which is a great advantage when AF screening is conducted in a large population where large amounts of ECG data are expected.

Third, the participant's identification and ECG were matched and recorded electronically, which minimized the occurrence of error during data filing.

Fourth, the performance of AF screening through the ECG recorder and cloud-computing algorithm is satisfactory, with a high sensitivity (95.5%), specificity (97.7%), and NPV (99.9%). The result supports its use for AF screening in the future.

Fifth, the satisfaction questionnaire administered in our study received a high satisfaction score when graded by participants.



This study has several limitations. First, this is a single-arm study without comparison groups or randomization design. Second, our ECG measurements were not compared with the current gold standard 12-lead ECG. In this study, the AF screening was performed on a community basis, which made comparisons with a 12-lead ECG impossible. To compensate for this shortage, all single-lead ECGs were measured through adhesive electrodes to receive the highest ECG quality possible and make the use of a 12-lead ECG less necessary. Third, we measured a 15-second single-lead ECG during each AF screening by DigiO2 Cardio Care ECG recorder. Comparing with the other handheld ECG that measures 30 seconds ECG tracing, shorter ECG tracing in this study may raise the concern of diagnostic power for AF. The best method to compensate the shortage would be obtaining high-quality ECG for reference. In addition, we are able to preview the ECG instantaneously on the tablet before uploading to the server at the Telehealth Center. Repeat measurement was allowed if the ECG measurement was composed of artifacts. Fourth, although this study demonstrates

that the ECG auto-interpretation by cloud-computing algorithm for AF detection has satisfactory sensitivity and specificity, the PPV of 48.8% is relatively low. A confirmatory examination is needed for the screened positive results from the ECG auto-interpretation by the cloud-computing algorithm. Fifth, an independent cardiologist performed all the physician-based ECG interpretations in the study. Potential interpretation error and bias may exist, as the interpretation of (single lead) ECG tracings can still vary between cardiologists. Most of the studies assigned 2 independent electrophysiologists for reviewing the ECGs [38,44]. Sixth, we did not evaluate the cost-effectiveness in this study. Potential expenditure could come from personnel expenses, adhesive electrodes, tablets, WLAN, and so on. Future studies for the cost-effectiveness should be performed before broadly applying our methodology for AF screening.

## Conclusions

It is feasible to conduct AF screening in nonmetropolitan areas using an ECG recorder with a telehealth surveillance system containing an embedded cloud-computing algorithm.

## Acknowledgments

We would like to acknowledge the contributions from Jen-Kuang Lee, MD; Dr Kuan-Chih Huang, MD; Hung-Ju Lin, MD; Po-Chih Lin, MD; Ms Shu-Ting Lee; and Ms Shiou-Chi Jeng for their assistance and support in this study. We would also like to thank NTUH for funding the study (106-A136).

## Authors' Contributions

The contributions of the authors are described as following: conception and design by YL-H and JJ-H; data collection and analysis by YH-C, CC-H, and YC-H; statistical analysis and interpretation by CS-H; drafting and revision of the manuscript by YH-C; and the final approval of the manuscript submitted by YL-H. YL-H and JJ-H contributed equally to the manuscript.

## Conflicts of Interest

None declared.

## References

1. Chugh SS, Blackshear JL, Shen WK, Hammill SC, Gersh BJ. Epidemiology and natural history of atrial fibrillation: clinical implications. *J Am Coll Cardiol* 2001 Feb;37(2):371-378 [FREE Full text] [Medline: [11216949](#)]
2. Heeringa J, van DK, Hofman A, Kors JA, van HG, Stricker BH, et al. Prevalence, incidence and lifetime risk of atrial fibrillation: the Rotterdam study. *Eur Heart J* 2006 Apr;27(8):949-953. [doi: [10.1093/eurheartj/ehi825](#)] [Medline: [16527828](#)]
3. Go AS, Hylek EM, Phillips KA, Chang Y, Henault LE, Selby JV, et al. Prevalence of diagnosed atrial fibrillation in adults: national implications for rhythm management and stroke prevention: the AnTicoagulation and Risk Factors in Atrial Fibrillation (ATRIA) Study. *JAMA* 2001 May 09;285(18):2370-2375. [Medline: [11343485](#)]
4. Kim E, Yin X, Fontes JD, Magnani JW, Lubitz SA, McManus DD, et al. Atrial fibrillation without comorbidities: prevalence, incidence and prognosis (from the Framingham Heart Study). *Am Heart J* 2016 Jul;177:138-144 [FREE Full text] [doi: [10.1016/j.ahj.2016.03.023](#)] [Medline: [27297859](#)]
5. Wang TJ, Massaro JM, Levy D, Vasan RS, Wolf PA, D'Agostino RB, et al. A risk score for predicting stroke or death in individuals with new-onset atrial fibrillation in the community: the Framingham Heart Study. *JAMA* 2003 Aug 27;290(8):1049-1056. [doi: [10.1001/jama.290.8.1049](#)] [Medline: [12941677](#)]
6. Wolf PA, Abbott RD, Kannel WB. Atrial fibrillation as an independent risk factor for stroke: the Framingham Study. *Stroke* 1991 Aug;22(8):983-988 [FREE Full text] [Medline: [1866765](#)]
7. Wang TJ, Larson MG, Levy D, Vasan RS, Leip EP, Wolf PA, et al. Temporal relations of atrial fibrillation and congestive heart failure and their joint influence on mortality: the Framingham Heart Study. *Circulation* 2003 Jun 17;107(23):2920-2925 [FREE Full text] [doi: [10.1161/01.CIR.0000072767.89944.6E](#)] [Medline: [12771006](#)]
8. Miyasaka Y, Barnes ME, Bailey KR, Cha SS, Gersh BJ, Seward JB, et al. Mortality trends in patients diagnosed with first atrial fibrillation: a 21-year community-based study. *J Am Coll Cardiol* 2007 Mar 06;49(9):986-992 [FREE Full text] [doi: [10.1016/j.jacc.2006.10.062](#)] [Medline: [17336723](#)]



9. Feinberg WM, Blackshear JL, Laupacis A, Kronmal R, Hart RG. Prevalence, age distribution, and gender of patients with atrial fibrillation. Analysis and implications. *Arch Intern Med* 1995 Mar 13;155(5):469-473. [Medline: [7864703](#)]
10. Lip GY, Hart RG, Conway DS. Antithrombotic therapy for atrial fibrillation. *BMJ* 2002 Nov 02;325(7371):1022-1025 [[FREE Full text](#)] [Medline: [12411366](#)]
11. Hart RG, Pearce LA, Aguilar MI. Meta-analysis: antithrombotic therapy to prevent stroke in patients who have nonvalvular atrial fibrillation. *Ann Intern Med* 2007 Jun 19;146(12):857-867. [Medline: [17577005](#)]
12. Connolly SJ, Ezekowitz MD, Yusuf S, Eikelboom J, Oldgren J, Parekh A, et al. Dabigatran versus warfarin in patients with atrial fibrillation. *N Engl J Med* 2009 Sep 17;361(12):1139-1151. [doi: [10.1056/NEJMoa0905561](#)] [Medline: [19717844](#)]
13. Patel MR, Mahaffey KW, Garg J, Pan G, Singer DE, Hacke W, et al. Rivaroxaban versus warfarin in nonvalvular atrial fibrillation. *N Engl J Med* 2011 Sep 08;365(10):883-891. [doi: [10.1056/NEJMoa1009638](#)] [Medline: [21830957](#)]
14. Granger CB, Alexander JH, McMurray JJV, Lopes RD, Hylek EM, Hanna M, et al. Apixaban versus warfarin in patients with atrial fibrillation. *N Engl J Med* 2011 Sep 15;365(11):981-992. [doi: [10.1056/NEJMoa1107039](#)] [Medline: [21870978](#)]
15. Ruff CT, Giugliano RP, Antman EM, Crugnale SE, Bocanegra T, Mercuri M, et al. Evaluation of the novel factor Xa inhibitor edoxaban compared with warfarin in patients with atrial fibrillation: design and rationale for the Effective aNticoagulation with factor xA next GEneration in Atrial Fibrillation-Thrombolysis In Myocardial Infarction study 48 (ENGAGE AF-TIMI 48). *Am Heart J* 2010 Oct;160(4):635-641. [doi: [10.1016/j.ahj.2010.06.042](#)] [Medline: [20934556](#)]
16. Davis RC, Hobbs FD, Kenkre JE, Roalfe AK, Iles R, Lip GY, et al. Prevalence of atrial fibrillation in the general population and in high-risk groups: the ECHOES study. *Europace* 2012 Nov;14(11):1553-1559. [doi: [10.1093/europace/eus087](#)] [Medline: [22490371](#)]
17. Hobbs FD, Fitzmaurice DA, Mant J, Murray E, Jowett S, Bryan S, et al. A randomised controlled trial and cost-effectiveness study of systematic screening (targeted and total population screening) versus routine practice for the detection of atrial fibrillation in people aged 65 and over. The SAFE study. *Health Technol Assess* 2005 Oct;9(40):iii-iv, ix [[FREE Full text](#)] [Medline: [16202350](#)]
18. Kirchhof P, Benussi S, Kotecha D, Ahlsson A, Atar D, Casadei B, et al. 2016 ESC Guidelines for the management of atrial fibrillation developed in collaboration with EACTS. *Eur Heart J* 2016 Oct 07;37(38):2893-2962. [doi: [10.1093/eurheartj/ehw210](#)] [Medline: [27567408](#)]
19. Lowres N, Neubeck L, Redfern J, Freedman SB. Screening to identify unknown atrial fibrillation. A systematic review. *Thromb Haemost* 2013 Aug;110(2):213-222. [doi: [10.1160/TH13-02-0165](#)] [Medline: [23595785](#)]
20. Aronsson M, Svennberg E, Rosenqvist M, Engdahl J, Al-Khalili F, Friberg L, et al. Cost-effectiveness of mass screening for untreated atrial fibrillation using intermittent ECG recording. *Europace* 2015 Jul;17(7):1023-1029. [doi: [10.1093/europace/euv083](#)] [Medline: [25868469](#)]
21. Levin L, Husberg M, Sobocinski PD, Kull VF, Friberg L, Rosenqvist M, et al. A cost-effectiveness analysis of screening for silent atrial fibrillation after ischaemic stroke. *Europace* 2015 Feb;17(2):207-214. [doi: [10.1093/europace/euu213](#)] [Medline: [25349228](#)]
22. Jacobs MS, Kaasenbrood F, Postma MJ, van HM, Tieleman RG. Cost-effectiveness of screening for atrial fibrillation in primary care with a handheld, single-lead electrocardiogram device in the Netherlands. *Europace* 2016 Oct 12. [doi: [10.1093/europace/euw285](#)] [Medline: [27733465](#)]
23. Verberk WJ, Omboni S, Kollias A, Stergiou GS. Screening for atrial fibrillation with automated blood pressure measurement: research evidence and practice recommendations. *Int J Cardiol* 2016 Jan 15;203:465-473 [[FREE Full text](#)] [doi: [10.1016/j.ijcard.2015.10.182](#)] [Medline: [26547741](#)]
24. Ho Y, Yu J, Lin Y, Chen Y, Huang C, Hsu T, et al. Assessment of the cost-effectiveness and clinical outcomes of a fourth-generation synchronous telehealth program for the management of chronic cardiovascular disease. *J Med Internet Res* 2014;16(6):e145 [[FREE Full text](#)] [doi: [10.2196/jmir.3346](#)] [Medline: [24915187](#)]
25. Chen Y, Lin Y, Hung C, Huang C, Yeih D, Chuang P, et al. Clinical outcome and cost-effectiveness of a synchronous telehealth service for seniors and nonseniors with cardiovascular diseases: quasi-experimental study. *J Med Internet Res* 2013;15(4):e87 [[FREE Full text](#)] [doi: [10.2196/jmir.2091](#)] [Medline: [23615318](#)]
26. Hung C, Yu J, Lin Y, Chen Y, Huang C, Lee J, et al. Mortality benefit of a fourth-generation synchronous telehealth program for the management of chronic cardiovascular disease: a longitudinal study. *J Med Internet Res* 2016 May 13;18(5):e102 [[FREE Full text](#)] [doi: [10.2196/jmir.5718](#)] [Medline: [27177497](#)]
27. Kligfield P, Gettes LS, Bailey JJ, Childers R, Deal BJ, Hancock EW, American Heart Association Electrocardiography and Arrhythmias Committee, Council on Clinical Cardiology, American College of Cardiology Foundation, Heart Rhythm Society. Recommendations for the standardization and interpretation of the electrocardiogram: part I: the electrocardiogram and its technology a scientific statement from the American Heart Association Electrocardiography and Arrhythmias Committee, Council on Clinical Cardiology; the American College of Cardiology Foundation; and the Heart Rhythm Society endorsed by the International Society for Computerized Electrocardiology. *J Am Coll Cardiol* 2007 Mar 13;49(10):1109-1127 [[FREE Full text](#)] [doi: [10.1016/j.jacc.2007.01.024](#)] [Medline: [17349896](#)]
28. Pahlm O, Haisty WK, Edenbrandt L, Wagner NB, Sevilla DC, Selvester RH, et al. Evaluation of changes in standard electrocardiographic QRS waveforms recorded from activity-compatible proximal limb lead positions. *Am J Cardiol* 1992 Jan 15;69(3):253-257. [Medline: [1731468](#)]

29. Ho T, Huang C, Lin C, Lai F, Ding J, Ho Y, et al. A telesurveillance system with automatic electrocardiogram interpretation based on support vector machine and rule-based processing. *JMIR Med Inform* 2015 May 07;3(2):e21 [FREE Full text] [doi: [10.2196/medinform.4397](https://doi.org/10.2196/medinform.4397)] [Medline: [25953306](https://pubmed.ncbi.nlm.nih.gov/25953306/)]
30. Lip GY, Lane DA. Stroke prevention in atrial fibrillation: a systematic review. *JAMA* 2015 May 19;313(19):1950-1962. [doi: [10.1001/jama.2015.4369](https://doi.org/10.1001/jama.2015.4369)] [Medline: [25988464](https://pubmed.ncbi.nlm.nih.gov/25988464/)]
31. Healey JS, Baranchuk A, Crystal E, Morillo CA, Garfinkle M, Yusuf S, et al. Prevention of atrial fibrillation with angiotensin-converting enzyme inhibitors and angiotensin receptor blockers: a meta-analysis. *J Am Coll Cardiol* 2005 Jun 07;45(11):1832-1839 [FREE Full text] [doi: [10.1016/j.jacc.2004.11.070](https://doi.org/10.1016/j.jacc.2004.11.070)] [Medline: [15936615](https://pubmed.ncbi.nlm.nih.gov/15936615/)]
32. Schneider MP, Hua TA, Böhm M, Wachtell K, Kjeldsen SE, Schmieder RE. Prevention of atrial fibrillation by Renin-Angiotensin system inhibition a meta-analysis. *J Am Coll Cardiol* 2010 May 25;55(21):2299-2307 [FREE Full text] [doi: [10.1016/j.jacc.2010.01.043](https://doi.org/10.1016/j.jacc.2010.01.043)] [Medline: [20488299](https://pubmed.ncbi.nlm.nih.gov/20488299/)]
33. Sanmartín M, Fraguera FF, Martín-Santos A, Moix BP, García-Ruiz A, Vázquez-Caamaño M, Grupo RIA (Registros en Insuficiencia Cardíaca y Aterosclerosis). A campaign for information and diagnosis of atrial fibrillation: "pulse week". *Rev Esp Cardiol (Engl Ed)* 2013 Jan;66(1):34-38. [doi: [10.1016/j.recresp.2012.05.012](https://doi.org/10.1016/j.recresp.2012.05.012)] [Medline: [22964060](https://pubmed.ncbi.nlm.nih.gov/22964060/)]
34. Cooke G, Doust J, Sanders S. Is pulse palpation helpful in detecting atrial fibrillation? a systematic review. *J Fam Pract* 2006 Feb;55(2):130-134. [Medline: [16451780](https://pubmed.ncbi.nlm.nih.gov/16451780/)]
35. January CT, Wann LS, Alpert JS, Calkins H, Cigarroa JE, Cleveland JC, et al. 2014 AHA/ACC/HRS guideline for the management of patients with atrial fibrillation: a report of the American College of Cardiology/American Heart Association Task Force on practice guidelines and the Heart Rhythm Society. *Circulation* 2014 Dec 02;130(23):e199-e267 [FREE Full text] [doi: [10.1161/CIR.0000000000000041](https://doi.org/10.1161/CIR.0000000000000041)] [Medline: [24682347](https://pubmed.ncbi.nlm.nih.gov/24682347/)]
36. Maeda K, Shimbo T, Fukui T. Cost-effectiveness of a community-based screening programme for chronic atrial fibrillation in Japan. *J Med Screen* 2004;11(2):97-102. [doi: [10.1258/096914104774061092](https://doi.org/10.1258/096914104774061092)] [Medline: [15153326](https://pubmed.ncbi.nlm.nih.gov/15153326/)]
37. Willits I, Keltie K, Craig J, Sims A. WatchBP Home A for opportunistically detecting atrial fibrillation during diagnosis and monitoring of hypertension: a NICE Medical Technology Guidance. *Appl Health Econ Health Policy* 2014 Jun;12(3):255-265 [FREE Full text] [doi: [10.1007/s40258-014-0096-7](https://doi.org/10.1007/s40258-014-0096-7)] [Medline: [24664995](https://pubmed.ncbi.nlm.nih.gov/24664995/)]
38. Desteghe L, Raymaekers Z, Lutin M, Vijgen J, Dilling-Boer D, Koopman P, et al. Performance of handheld electrocardiogram devices to detect atrial fibrillation in a cardiology and geriatric ward setting. *Europace* 2017 Jan;19(1):29-39. [doi: [10.1093/europace/euw025](https://doi.org/10.1093/europace/euw025)] [Medline: [26893496](https://pubmed.ncbi.nlm.nih.gov/26893496/)]
39. Lewis M, Parker D, Weston C, Bowes M. Screening for atrial fibrillation: sensitivity and specificity of a new methodology. *Br J Gen Pract* 2011 Jan;61(582):38-39 [FREE Full text] [doi: [10.3399/bjgp11X548956](https://doi.org/10.3399/bjgp11X548956)] [Medline: [21401988](https://pubmed.ncbi.nlm.nih.gov/21401988/)]
40. Scully CG, Lee J, Meyer J, Gorbach AM, Granquist-Fraser D, Mendelson Y, et al. Physiological parameter monitoring from optical recordings with a mobile phone. *IEEE Trans Biomed Eng* 2012 Feb;59(2):303-306 [FREE Full text] [doi: [10.1109/TBME.2011.2163157](https://doi.org/10.1109/TBME.2011.2163157)] [Medline: [21803676](https://pubmed.ncbi.nlm.nih.gov/21803676/)]
41. Chan P, Wong C, Poh YC, Pun L, Leung WW, Wong Y, et al. Diagnostic performance of a smartphone-based photoplethysmographic application for atrial fibrillation screening in a primary care setting. *J Am Heart Assoc* 2016 Jul 21;5(7):- [FREE Full text] [doi: [10.1161/JAHA.116.003428](https://doi.org/10.1161/JAHA.116.003428)] [Medline: [27444506](https://pubmed.ncbi.nlm.nih.gov/27444506/)]
42. Nemati S, Ghassemi MM, Ambai V, Isakadze N, Levantsevych O, Shah A, et al. Monitoring and detecting atrial fibrillation using wearable technology. *Conf Proc IEEE Eng Med Biol Soc* 2016 Aug;2016:3394-3397. [doi: [10.1109/EMBC.2016.7591456](https://doi.org/10.1109/EMBC.2016.7591456)] [Medline: [28269032](https://pubmed.ncbi.nlm.nih.gov/28269032/)]
43. Couderc J, Kyal S, Mestha LK, Xu B, Peterson DR, Xia X, et al. Detection of atrial fibrillation using contactless facial video monitoring. *Heart Rhythm* 2015 Jan;12(1):195-201. [doi: [10.1016/j.hrthm.2014.08.035](https://doi.org/10.1016/j.hrthm.2014.08.035)] [Medline: [25179488](https://pubmed.ncbi.nlm.nih.gov/25179488/)]
44. Mant J, Fitzmaurice DA, Hobbs FD, Jowett S, Murray ET, Holder R, et al. Accuracy of diagnosing atrial fibrillation on electrocardiogram by primary care practitioners and interpretative diagnostic software: analysis of data from screening for atrial fibrillation in the elderly (SAFE) trial. *BMJ* 2007 Aug 25;335(7616):380 [FREE Full text] [doi: [10.1136/bmj.39227.551713.AE](https://doi.org/10.1136/bmj.39227.551713.AE)] [Medline: [17604299](https://pubmed.ncbi.nlm.nih.gov/17604299/)]

## Abbreviations

**AF:** atrial fibrillation  
**APC:** atrial premature complex  
**BP:** blood pressure  
**ECG:** electrocardiogram  
**ESC:** European Society of Cardiology  
**NOAC:** nonvitamin K antagonist oral anticoagulants  
**NPV:** negative predictive value  
**NTUH:** National Taiwan University Hospital  
**PPG:** photoplethysmography  
**PPV:** positive predictive value  
**SD:** standard deviation

**TELEHEALTH:** Taiwan ELEctroHEALTH

**VKA:** vitamin K antagonists

**VPC:** ventricular premature complex

**WLAN:** wireless local area network

*Edited by G Eysenbach; submitted 23.06.17; peer-reviewed by L Desteghe, S Omboni, M Lang; comments to author 15.07.17; revised version received 15.08.17; accepted 15.08.17; published 26.09.17.*

*Please cite as:*

*Chen YH, Hung CS, Huang CC, Hung YC, Hwang JJ, Ho YL*

*Atrial Fibrillation Screening in Nonmetropolitan Areas Using a Telehealth Surveillance System With an Embedded Cloud-Computing Algorithm: Prospective Pilot Study*

*JMIR Mhealth Uhealth 2017;5(9):e135*

URL: <https://mhealth.jmir.org/2017/9/e135/>

doi: [10.2196/mhealth.8290](https://doi.org/10.2196/mhealth.8290)

PMID: [28951384](https://pubmed.ncbi.nlm.nih.gov/28951384/)

©Ying-Hsien Chen, Chi-Sheng Hung, Ching-Chang Huang, Yu-Chien Hung, Juey-Jen Hwang, Yi-Lwun Ho. Originally published in JMIR Mhealth and Uhealth (<http://mhealth.jmir.org>), 26.09.2017. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mhealth and uhealth, is properly cited. The complete bibliographic information, a link to the original publication on <http://mhealth.jmir.org/>, as well as this copyright and license information must be included.

Original Paper

# Tablet-Based Patient-Centered Decision Support for Minor Head Injury in the Emergency Department: Pilot Study

Navdeep Singh<sup>1,2</sup>, BSAB, BSA; Erik Hess<sup>3</sup>, MSc, MD; George Guo<sup>2</sup>, BS; Adam Sharp<sup>4,5</sup>, MS, MD; Brian Huang<sup>4</sup>, MPH; Maggie Breslin<sup>6</sup>, MDes; Edward Melnick<sup>2</sup>, MHS, MD

<sup>1</sup>Medical College of Georgia, AU/UGA Medical Partnership, Athens, GA, United States

<sup>2</sup>Department of Emergency Medicine, Yale School of Medicine, New Haven, CT, United States

<sup>3</sup>Department of Emergency Medicine, Mayo Clinic, Rochester, MN, United States

<sup>4</sup>Department of Research and Evaluation, Kaiser Permanente Southern California, Pasadena, CA, United States

<sup>5</sup>Department of Emergency Medicine, Los Angeles Medical Center, Kaiser Permanente Southern California, Los Angeles, CA, United States

<sup>6</sup>School of Visual Arts, New York, NY, United States

**Corresponding Author:**

Edward Melnick, MHS, MD

Department of Emergency Medicine

Yale School of Medicine

464 Congress Ave, Suite 260

New Haven, CT, 06519

United States

Phone: 1 203 737 6454

Fax: 1 203 785 4580

Email: [edward.melnick@yale.edu](mailto:edward.melnick@yale.edu)

## Abstract

**Background:** The Concussion or Brain Bleed app is a clinician- and patient-facing electronic tool to guide decisions about head computed tomography (CT) use in patients presenting to the emergency department (ED) with minor head injury. This app integrates a patient decision aid and clinical decision support (using the Canadian CT Head Rule, CCHR) at the bedside on a tablet computer to promote conversations around individualized risk and patients' specific concerns within the ED context.

**Objective:** The objective of this study was to describe the use of the Concussion or Brain Bleed app in a high-volume ED and to establish preliminary efficacy estimates on patient experience, clinician experience, health care utilization, and patient safety. These data will guide the planning of a larger multicenter trial testing the effectiveness of the Concussion or Brain Bleed app.

**Methods:** We conducted a prospective pilot study of adult (age 18-65 years) patients presenting to the ED after minor head injury who were identified by participating clinicians as low risk by the CCHR. The primary outcome was patient knowledge regarding the injury, risks, and CT use. Secondary outcomes included patient satisfaction, decisional conflict, trust in physician, clinician acceptability, system usability, Net Promoter scores, head CT rate, and patient safety at 7 days.

**Results:** We enrolled 41 patients cared for by 29 different clinicians. Patient knowledge increased after the use of the app (questions correct out of 9: pre-encounter, 3.3 vs postencounter, 4.7; mean difference 1.4, 95% CI 0.8-2.0). Patients reported a mean of 11.7 (SD 13.5) on the Decisional Conflict Scale and 92.5 (SD 12.0) in the Trust in Physician Scale (both scales range from 0 to 100). Most patients were satisfied with the app's clarity of information (35, 85%), helpfulness of information (36, 88%), and amount of information (36, 88%). In the 41 encounters, most clinicians thought the information was somewhat or extremely helpful to the patient (35, 85%), would want to use something similar for other decisions (27, 66%), and would recommend the app to other providers (28, 68%). Clinicians reported a mean system usability score of 85.1 (SD 15; scale from 0 to 100 with 85 in the "excellent" acceptability range). The total Net Promoter Score was 36.6 (on a scale from -100 to 100). A total of 7 (17%) patients received a head CT in the ED. No patients had a missed clinically important brain injury at 7 days.

**Conclusions:** An app to help patients assess the utility of CT imaging after head injury in the ED increased patient knowledge. Nearly all clinicians reported the app to be helpful to patients. The high degree of patient satisfaction, clinician acceptability, and system usability support rigorous testing of the app in a larger multicenter trial.

(JMIR Mhealth Uhealth 2017;5(9):e144) doi:[10.2196/mhealth.8732](https://doi.org/10.2196/mhealth.8732)

## KEYWORDS

clinical decision support; decision aids; head injury, minor; medical informatics; spiral computed tomography; health services overuse; patient-centered outcomes research

## Introduction

One-third of patients with minor head injury receive head computed tomography (CT) that may not be clinically indicated [1-6]. These potentially avoidable CTs do not change management. However, they do increase health care costs, exposure to ionizing radiation, and length of stay in the emergency department (ED) [7]. The American Board of Internal Medicine and the American College of Emergency Physicians' Choosing Wisely initiative have recognized this and recommend avoiding unnecessary head CTs in patients with minor head injuries as the top national priority for addressing CT overuse in emergency care [1]. The Canadian CT Head Rule (CCHR) is a clinical decision rule that was developed using a rigorous, evidence-based derivation and validation process to identify minor head injury in patients at risk for clinically important structural brain injuries and the need for neurosurgical intervention. A total of 7 history and physical criteria are used as indications for CT based on their association with these risks. This rule was designed to safely reduce head CT use in patients with minor head injury. It has been validated to be 100% sensitive in detecting patients needing neurosurgical intervention. Additionally, the CCHR outperforms other decision rules with the highest specificity in its class [8-11].

Implementing the CCHR with traditional computerized clinical decision support (CDS) has had a modest effect (5%-8%) on decreasing CT use in these patients [12,13]. Since one-third of CTs in minor head injury patients are potentially avoidable and traditional CDS has had limited effect on reducing these scans, it has been hypothesized that nonclinical factors (such as fear of litigation, physician personality, fear of missed diagnoses,

financial incentives, paucity of information, and patient expectations) also contribute to CT overuse in these patients [14,15]. Qualitative research on this topic revealed that physician-based empathic factors such as establishing trust and engaging patients by identifying and addressing their concerns are essential to reduce CT overuse [15,16].

We previously developed a clinician- and patient-facing electronic tool to guide decisions about CT use in ED patients with minor head injury, called Concussion or Brain Bleed [17,18]. This app integrates a patient decision aid and CDS (using the CCHR) at the bedside on a tablet computer to promote conversations around individualized risk and patients' specific concerns within the ED's clinical constraints [19,20]. Although intended primarily for use in low-risk patients, the app includes pathways for moderate- and high-risk patients as well. Figure 1 presents the conceptual workflow of the app: (1) welcome screen, (2) injury evaluator (CDS portion), (3) risk visualization, (4) risk discussion with conversation prompts such as "You can't see concussion on CT?," (5) considerations, and (6) integration back to traditional workflow (the paper handout given to patients after using the intervention is publicly available [21]).

Figure 2 presents the risk visualization screen for the low-risk pathway. The long-term implementation goal for this patient-centered decision support tool is to safely and effectively reduce CT use for patients with minor head injury while simultaneously improving the patient experience. In the trial presented here, our objective was to describe the use of the Concussion or Brain Bleed app in a high-volume ED and to establish preliminary efficacy estimates on patient experience, clinician experience, health care utilization, and patient safety.

**Figure 1.** Conceptualization of the workflow and potential pathways for the Concussion or Brain Bleed app. CT: computed tomography; EHR: electronic health record.





**Figure 2.** Risk visualization screen shot for low-risk patients from the Concussion or Brain Bleed app. CT: computed tomography.

## Methods

### Study Design, Setting, and Population

We performed a prospective pilot study with a convenience sample of 41 ED patients with minor head injury. Patients were enrolled over a 6-week period (May 23 to July 3, 2017). Patients and clinicians who were eligible and willing to participate used the Concussion or Brain Bleed app and completed a survey to determine the app's baseline efficacy on patient experience, clinician experience, health care utilization, and patient safety. Participants were patients and clinicians recruited from an urban, academic level I trauma center ED with 103,000 patient-visits per year. Eligible patients were adults (age 18–65 years) presenting to the ED who had experienced blunt head injury within the last 24 hours who were determined to be at low risk by the CCHR (see Figure 3) and were being considered for head CT imaging by the treating clinician. Patients who were pregnant, non-English speaking, in police custody, undergoing psychiatric evaluation, or found to have drug or alcohol intoxication were excluded. Eligible clinicians were attending physicians, fellows, residents, and midlevel providers caring for eligible patients. We recruited clinicians from the 48

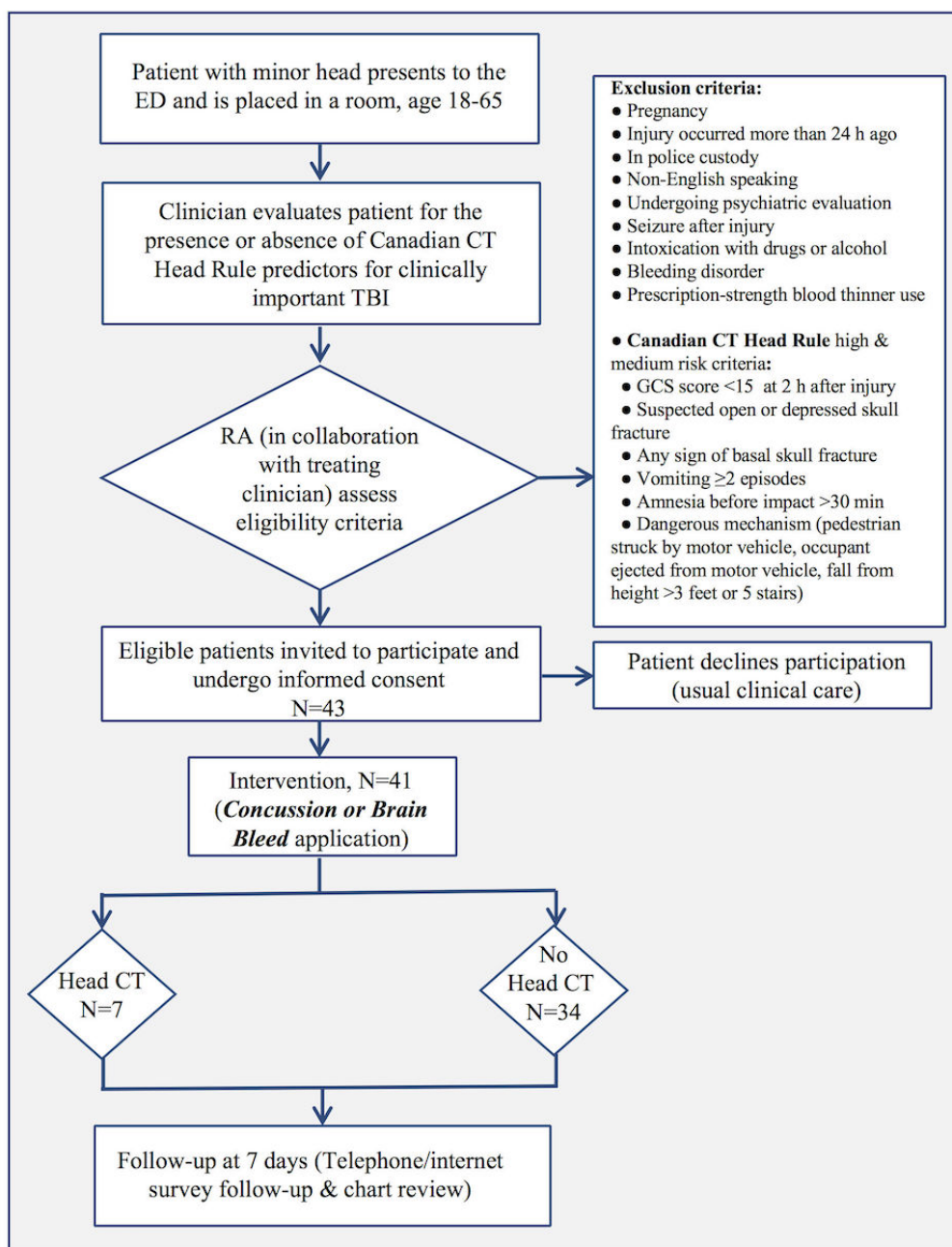
attending physician faculty, 58 resident physicians, and 47 midlevel providers. The study protocol was approved by our institutional review board (IRB), the Yale Human Investigation Committee.

### Participant Identification, Recruitment, and Enrollment

A research assistant (RA; NS) reviewed an electronic patient tracking board at regular intervals to identify potentially eligible patients based on a chief complaint potentially consistent with head trauma. Next, the RA worked with the clinician assigned to the patient's care team to determine whether the patient met inclusion criteria (either before or after the initial clinical evaluation). Next, the clinician and patient were informed about the study and asked if they would be willing to participate. The participating clinician and patient provided verbal consent as specified by the IRB-approved protocol. We collected all data using the Web-based survey tool Qualtrics Survey Tool (Qualtrics, LLC), on Yale's electronic patient health information-approved and certified, licensed online platform. Clinicians were compensated for participation in the study with a US \$10 gift card to a coffee shop.



**Figure 3.** Flow diagram for patient identification and enrollment in the flow of patient care with number of patients identified, enrolled, and receiving computed tomography (CT) in the emergency department (ED). GCS: Glasgow Coma Scale; RA: research assistant; TBI: traumatic brain injury.



## Training

The RA gave participating clinicians a brief (<2 minute) tutorial of the Concussion or Brain Bleed app prior to using it the first time. This individualized, just-in-time training provided an opportunity to highlight each section of the app and to demonstrate its navigation. The clinician was given an opportunity to ask any additional questions or to repeat sections of the training as needed until they felt comfortable with its use. Although RAs were available at the point of care to assist with any technical issues or difficulty navigating the app on an as-needed basis, they refrained from interfering with the actual use of the app to observe an accurate representation of its use in routine care.

## Patient and Clinician Characteristics

We collected patient demographics by self-report at the time of enrollment, including age, sex, race, ethnicity, highest level of education, insurance status, and household income. Patient literacy and numeracy were assessed immediately before use of the app using the validated Subjective Literacy Scale and Subjective Numeracy Scale [22-24]. The Subjective Literacy Scale comprises 3 items, each rated on a 5-point Likert scale and summed into a total score of 3-15. The Subjective Numeracy Scale consists of 8 items that assesses comfort in working with numbers, each rated on a 6-point Likert scale with an overall score ranging from 6 to 48.

We collected clinician characteristics by self-report following the clinician's first use of the tool. Clinician characteristics that were collected included demographics, years practicing

emergency medicine, medical degree or role, and details on personal technology use.

## Outcome Measures

Outcome selection was informed by a similar study performed by Hess et al [25] using a paper-based, shared decision-making aid in a pediatric population to compare the decision aid's effectiveness with usual care on (1) parent knowledge regarding their child's risk, diagnostic options, and risks associated with CT, (2) parent engagement in the decision-making process, (3) degree of conflict parents experience related to feeling uninformed, (4) patient and clinician satisfaction, (5) rate of clinically important traumatic brain injury at 7 days, (6) proportion of patients in whom a CT scan was obtained, and (7) 7-day health care utilization [25]. That study selected outcomes based on input from key stakeholders, including patient representatives, practicing clinicians, researchers (including shared decision-making experts), and health policy decision makers. Patient knowledge was selected as the primary outcome for that study based on input from patient representatives. For our study reported here, we selected patient knowledge as the primary outcome and other secondary outcomes based on this precedent from the pediatric shared decision-making study, including the Decisional Conflict Scale, the Trust in Physician Scale, similar satisfaction, health care utilization, and patient safety outcomes [25].

## Patient Outcomes

### Patient Knowledge

We assessed patient knowledge using a pre- and postvisit survey administered immediately before and after the clinical encounter (Multimedia Appendix 1) [25]. In the survey, 9 questions assessed patients' knowledge regarding concussion, their individual risk of structural brain injury, the available diagnostic options, the risks related to radiation exposure associated with a head CT scan, the potential for a CT scan to identify incidental abnormalities that may require further investigation, and reasons to return to the ED for reevaluation should their symptoms worsen after ED discharge. We calculated the percentage of knowledge questions answered correctly to determine the mean difference between knowledge scores before and after use of the intervention.

### Decisional Conflict

We measured the patient's degree of conflict with the decision of whether to get a CT scan using the validated Decisional Conflict Scale [25-28]. The 16 items on this scale are scored on a scale 0 to 4; the items are summed, divided by 16, and then multiplied by 25. The scale ranges from 0 to 100, where higher scores reflect patient uncertainty about the choice.

### Trust in the Physician

We measured patients' trust in their clinician using the validated Trust in Physician Scale [25,28-30]. This scale has 10 items, which are scored on a scale of 1 to 5; the items are summed, divided by 10, and then multiplied by 100. The scale ranges from 0 to 100, where higher values reflect higher levels of trust in their clinician.

## Patient Satisfaction

We measured patients' satisfaction with the way information was shared during the encounter by asking 5 questions using a 7-point Likert scale. For the analysis, we classified satisfaction into satisfied/very satisfied versus other responses.

## Clinician Outcomes

### Clinician Satisfaction

We assessed clinician satisfaction immediately after the patient encounter via a questionnaire regarding the helpfulness of the app and the clinician's satisfaction with the way information was shared on a 7-point Likert scale. For the analysis, we classified satisfaction into satisfied/very satisfied versus other responses.

### System Usability Scale

The System Usability Scale consists of a 10-item questionnaire on a 5-point Likert scale that gives a reliable assessment of usability [31]. The 10 items of the System Usability Scale are scored on a scale of 0 to 4, with each even-numbered question reverse coded. The items are summed and then multiplied by 2.5. Scores range from 0 to 100, where higher scores indicate higher usability.

### Net Promoter Score

The Net Promoter Score has been employed across industries to measure how willing a user is to recommend a product or service to others [32]. A higher score on this scale ranging from -100 to 100 can indicate a greater growth rate of the corresponding product or service. We determined the score by first asking the clinician user on a scale from 0 to 10 (0=not likely at all, 10=extremely likely) "How likely are you to recommend the Concussion or Brain Bleed application to a colleague?" If a clinician answered 9 or 10, we categorized them as a "promoter"—someone who would enthusiastically recommend the app to others. If a clinician answered 6 or lower, we considered them to be a "detractor"—someone who would potentially give a negative review to others. The Net Promoter Score is calculated by subtracting the percentage of detractors from the percentage of promoters. We calculated a total Net Promoter Score factoring in all encounters in which the app was used, as well as a first-time user Net Promoter Score and a second-time user Net Promoter Score.

### Fidelity Score

We assessed the fidelity with which the intervention was delivered and used as intended using a fidelity checklist of 8 intended actions (see Multimedia Appendix 2). The fidelity checklist has been used in the absence of the intervention to check for contamination in the usual-care arm of a trial [25].

## Health Care Utilization and Patient Safety

CT scans were obtained at the ED clinicians' discretion and interpreted by site faculty radiologists. The main health care utilization outcome was the proportion of patients for whom head CT was obtained in the ED. We also collected data at the time of the ED visit (and confirmed by chart review) on (1) whether the patient was admitted to the hospital, (2) acute findings on CT if obtained, and (3) whether the clinician

reported that they would have made the same decision regarding CT imaging without using the app. The RA contacted enrolled patients by telephone or email starting at 7 days after the index ED visit to ensure no outcomes were missed. The 7-day follow-up was based on timing of delayed clinical deterioration and our previous work [8,25].

### Analysis Plan

Results are reported using descriptive statistics and stratified by patient and clinician outcomes. The unit of analysis was the ED encounter. We defined change in patient knowledge as the mean difference of questions answered correctly pre- and postencounter. We performed analysis in Microsoft Excel (version 2016; Microsoft Corporation) on data exported from the Qualtrics Survey Tool. We made every effort to minimize the occurrence of missing data. We attempted to verify (or ascertain, if missing) items self-reported by patients at the 7-day follow-up by medical record review. We report rates of missing data as well as known reasons for missing data. We conducted secondary exploratory analyses of variables predictive of the odds of CT imaging, patient knowledge, and trust in physician using univariate logistic and linear regression with SAS (version 9.3; SAS Institute).

## Results

### Patient and Clinician Characteristics

We enrolled 41 of 43 identified patients (see [Figure 1](#); recruitment rate 95%) in the 6-week study period with a mean age of 34.9 years (range 18-59; see [Table 1](#)). The majority of patients were female (26, 63%), were not of Hispanic or Latino origin (31, 76%), and identified high school or general educational diploma or less as their highest level of education (24, 59%). The mean patient subjective literacy score was 12.4 (SD 2.8), and mean subjective numeracy score was 30.4 (SD 8.5).

Of 33 eligible clinicians, 29 (recruitment rate 88%) caring for eligible patients agreed to participate. The mean clinician age was 34 years (range 24-51; see [Table 2](#)). The majority of clinicians were female (15, 52%), not of Hispanic or Latino origin (36, 90%), white (20, 69%) and physicians (MDs) (16, 55%). There were 11 (38%) clinicians with a Physician Assistant degree and 2 (7%) with an Advanced Practice Registered Nurse (nurse practitioner) degree. The mean (range) years of experience practicing emergency medicine (including residency) was 5.8 (0-24). All clinicians owned a personal smartphone (29, 100%) and most owned a personal tablet computer (21, 72%). The majority of clinicians (24, 83%) also indicated they spent over 30 hours a week on a computer, tablet, or smartphone.

### Patient Experience

Mean (SD) knowledge assessment scores increased from 3.3 (1.9) out of 9 pre-encounter to 4.7 (2.1) postencounter ([Multimedia Appendix 1](#)), with mean difference of 1.4 (95% CI 0.8-2.0, see [Table 3](#)). The mean (SD) patient decisional conflict score was 11.7 (13.5), and the mean (SD) trust in physician score was 92.5 (12). Both scales are from 0 to 100.

Patient satisfaction scores showed that a majority of patients were satisfied with the clarity of information (35, 85%), helpfulness of the information (36, 88%), and amount of information (36, 88%). The majority of patients also said that they would recommend the app to others (36, 88%) and would want to use something similar for other clinical decisions (26, 63%).

The mean (SD) fidelity score was 6.7 (1.8; see [Table 5](#)) out of the 8 intended actions that the app aimed to elicit. Clinicians most consistently described the different risk levels portrayed on the risk visualization pictograph (95%). Clinicians least frequently elicited the patient or caregiver's concerns (61%).

### Health Care Utilization and Patient Safety

In the 41 encounters in which the app was used, 7 patients (17%; see [Table 6](#)) received a head CT in the ED. Since these patients were at low risk, all 7 CTs scans were not recommended based on the CCHR criteria. Of the 7 CTs, the 3 most frequently cited reasons for obtaining CT were referring physician request (5/7, 71%), mechanism of injury (3/7, 43%), and headache (3/7, 43%). In 100% of cases in which the app was used, clinicians reported they would make the same decision without the app. No patients were admitted to the hospital (0, 0%). Follow-up data were collected via phone call from 34 patients (83%), email from 4 patients (10%), and chart review for the remaining 3 patients (7%). At 7-day follow-up, 4 patients (10%) had returned to an ED, 14 patients (34%) had visited a physician's office or clinic, 1 patient (2%) did both, and 22 patients (54%) did neither. Further testing or procedures were obtained for 5 patients (12%) within 7 days following the encounter, and 2 patients (5%) underwent neuroimaging within 7 days. No patient had acute findings on CT in the ED or on follow-up imaging (0%).

### Secondary Analyses

On secondary analyses of variables predictive of the odds of CT imaging, fidelity with the concerns portion of the intervention (odds ratio 0.19, 95% CI 0.03-1.15,  $P=.07$ ), not having low literacy (odds ratio 0.23, 95% CI 0.04-1.26,  $P=.09$ ), and system usability score above average (odds ratio 0.24, 95% CI 0.03-1.83,  $P=.17$ ) trended toward significance but these results were not statistically significant. Patient knowledge and trust in the physician yielded no statistically significant results. Variables that trended toward significance for change in patient knowledge from pre- to postencounter in univariate analysis were white patient race (variable 1.14, 95% CI -0.14 to 2.41,  $P=.08$ ), fidelity with the concerns portion of the intervention (variable -0.83, 95% CI -2.13 to 0.47,  $P=.21$ ), not having low literacy and not having low numeracy (both with variable 0.64, 95% CI -0.66 to 1.94,  $P=.33$ ) but these results were not statistically significant. Variables that trended toward significance for trust in physician on univariate analysis were white patient race (variable 5.89, 95% CI -1.29 to 13.08,  $P=.11$ ), fidelity with the concerns portion of the intervention (variable 3.59, 95% CI -3.74 to 10.91,  $P=.34$ ), and not having low literacy (variable 3.83, 95% CI -4.02 to 11.68,  $P=.34$ ) but these results were not statistically significant.

**Table 1.** Patient characteristics.

Characteristics	Data
Participants recruited, n	43
Participants enrolled, n (%)	41 (95)
Age (years), mean (range)	34.9 (18-59)
Female, n (%)	26 (63)
<b>Race, n (%)</b>	
Black or African American	15 (37)
White	17 (42)
Asian	1 (2)
American Indian or Alaska Native	1 (2.4)
Other	9 (22)
<b>Ethnicity, n (%)</b>	
Hispanic or Latino origin	10 (24)
Not of Hispanic or Latino origin	31 (76)
<b>Education, n (%)</b>	
Some high school or less	4 (10)
High school graduate	20 (49)
Some college	12 (29)
College graduate or more	5 (12)
<b>Insurance, n (%)</b>	
Private/HMO <sup>a</sup>	21 (51)
Medicaid only	17 (42)
Medicare only	0 (0)
Medicare + Medicaid	1 (2)
Uninsured	2 (5)
<b>Annual household income (US \$), n (%)</b>	
<20,000	8 (20)
20,000-29,999	6 (15)
30,000-39,999	6 (15)
40,000-59,999	4 (10)
60,000-79,999	7 (17)
80,000-99,999	5 (12)
≥100,000 or more	5 (12)
Subjective Literacy Scale score, mean (SD)	12.4 (2.8)
Subjective Numeracy Scale score, mean (SD)	30.4 (8.5)

<sup>a</sup>HMO: health maintenance organization.

**Table 2.** Clinician characteristics.

Characteristics	Data
Participants recruited, n	33
Participants enrolled, n (%)	29 (88)
Age (years), mean (range)	34 (24-51)
Female, n (%)	15 (52)
<b>Race, n (%)</b>	
White	20 (69)
Asian	8 (28)
Other	2 (7)
<b>Ethnicity, n (%)</b>	
Hispanic or Latino origin	3 (10)
Not of Hispanic or Latino origin	36 (90)
<b>Medical degree, n (%)</b>	
Advanced Practice Registered Nurse	2 (7)
Physician Assistant	11 (38)
Physician (MD)	16 (55)
Experience (years), mean (range)	5.8 (0-24)
<b>Technology use, n (%)</b>	
<b>Time (hours) spent on a computer, tablet, or smartphone per week</b>	
<10	0 (0)
10-20	2 (7)
20-30	3 (10)
30-40	8 (28)
>40	16 (55)
<b>Preferred method of contact</b>	
Call on landline	0 (0)
Call on mobile phone	9 (31)
Email	0 (0)
Text	20 (69)
Other	0 (0)
<b>Mobile technology use</b>	
Personal tablet computer	21 (72)
Personal smartphone	29 (100)

**Table 3.** Patient experience outcomes and results.

Outcome	Data
<b>Patient knowledge</b>	
<b>Knowledge (no. questions correct out of 9), mean (SD)</b>	
Pre-encounter	3.3 (1.9)
Postencounter	4.7 (2.1)
Mean difference (95% CI)	1.4 (0.8-2.0)
<b>Decisional conflict and trust</b>	
Decisional Conflict Scale (scale of 0-100), mean (SD)	11.7 (13.5)
Trust in Physician Scale (scale of 0-100), mean (SD)	92.5 (12.0)
<b>Patient satisfaction, n (%)</b>	
<b>Amount of information</b>	
Too little (1-2)	0 (0)
Just right (3-5)	36 (88)
Too much (6-7)	5 (12)
<b>Clarity of information</b>	
Satisfied (5-7)	35 (85)
Unsatisfied (1-4)	6 (15)
<b>Helpfulness of information</b>	
Satisfied (5-7)	36 (88)
Unsatisfied (1-4)	5 (12)
<b>Would recommend to others</b>	
Yes (1-3)	36 (88)
Not sure/no (4-7)	5 (12)
<b>Would want to use for other decisions</b>	
Yes (1-3)	26 (63)
Not sure/no (4-7)	15 (37)



**Table 4.** Clinician experience outcomes and results.

Outcome	Data
<b>System usability and net promoter scores</b>	
System Usability Scale score (scale of 0-100), mean (SD)	85.1 (15.0)
Total Net Promoter Score (scale of –100 to 100)	36.6
First-time user Net Promoter Score	31.0
Second-time user Net Promoter Score	50.0
<b>Clinician acceptability, n (%)</b>	
<b>Helpfulness of the information</b>	
Not helpful at all (1-2)	1 (2)
Somewhat helpful (3-5)	16 (39)
Extremely helpful (6-7)	24 (59)
<b>Would want to use for other decisions</b>	
Yes (1-2)	27 (66)
Not sure (3-5)	13 (32)
No (6-7)	1 (2)
<b>Would recommend to others</b>	
Yes (1-2)	28 (68)
Not sure (3-5)	13 (32)
No (6-7)	0 (0)

**Table 5.** Fidelity score and compliance with delivery of the intervention as intended.

Fidelity of Use Assessment Question	n (%)
Did the clinician describe how the severity of the injury was evaluated using the Canadian CT <sup>a</sup> Head Rule?	37 (90)
Did the clinician describe the risk as a natural frequency (eg, “of 100 people like you, 6 will...”)?	37 (90)
Did the clinician describe the different risk levels portrayed on the risk visualization pictograph?	39 (95)
Did the clinician explain the difference between concussion and brain bleed?	31 (76)
Did the clinician explain what kinds of injuries can and cannot be seen on a CT scan?	33 (81)
Did the clinician elicit the patient and/or caregiver’s concerns?	25 (61)
Did the clinician discuss the patient and/or caregiver’s specific concerns?	35 (85)
<b>(Follow-up Discussion)</b>	
(If no CT performed) Did the clinician discuss what to expect after leaving the ED?	36 (88)
(If CT performed) Did the clinician discuss issues to consider before getting a CT scan?	0 (0)
Total score out of 8 possible, mean (SD)	6.7 (1.6)

<sup>a</sup>CT: computed tomography.

**Table 6.** Health care utilization and patient safety results.

Outcome	n (%)
Head CT <sup>a</sup> obtained in the ED <sup>b</sup>	7 (17)
Clinician would make same decision without the app	41 (100)
Admitted to the hospital	0 (0)
Acute findings on CT in ED	0 (0)
ED return visit within 7 days	4 (10)
Physician office or clinic visit within 7 days	14 (34)
Both ED return visit and physician office or clinic visit within 7 days	1 (2)
Neither ED return visit nor physician office or clinic visit within 7 days	22 (54)
Neuroimaging within 7 days	2 (5)
Acute findings on neuroimaging within 7 days	0 (0)

<sup>a</sup>CT: computed tomography.

<sup>b</sup>ED: emergency department.

## Discussion

In patients with low-risk minor head injury who were being considered for CT head imaging in the ED, use of the Concussion or Brain Bleed app in this prospective interventional pilot study resulted in increased patient knowledge and was associated with a low rate of CT use, high trust in the physician, low patient decisional conflict, high clinician Net Promoter Score, and high system usability score without any adverse events in patients. We found the app to be acceptable to both patients and clinicians.

### Comparison With Other Studies

Our trial's setup was similar to those of other ED shared decision-making trials for adult patients with chest pain and pediatric patients with head injury [25,28]. The high trust in physician and low decisional conflict scores reported here establish baseline efficacy of the Concussion or Brain Bleed app. These scores are consistent with those of previous ED trials of paper-based decision aids for adult ED patients with chest pain (trust in physician: mean 89.5, SD 13.4 versus this study, 92.5, SD 12.0; decisional conflict: mean 43.5, SD 11.3 versus this study, 11.7, SD 13.5) and parents of pediatric ED patients with head injury (results to be reported soon) [25,28]. Although the results have not yet been formally reported, our population had similar but slightly lower literacy and numeracy than the trial studying parents of pediatric ED patients with head injury described in the Outcome Measures subsection above [25].

Traditional implementation strategies lead to increased CT use in minor head injury [33]. On the other hand, traditional CDS has had only a modest effect (5%-8%) on decreasing the rate of CT overuse (35%) in these patients [2,5]. The overuse rate in our study of 17% cuts this rate in half. Based on our previous qualitative work, we hypothesize that this additional decrease was due to the intervention's ability to engage patients and address nonclinical factors (eg, identifying and addressing patient concerns and increasing physician trust). However, the number of patients enrolled in this study was limited and was a convenience sample [15].

The intervention's System Usability Scale and Net Promoter scores were also high. To put them in context, a system usability score of 85.1 has been correlated with the adjective rating of "excellent" or a grade of A+ [34,35]. Amazon.com is a frequently used website that has been found to have a similar system usability score [36]. Furthermore, the Net Promoter Score of 36.6 indicates a greater rate of users who were promoters than detractors of the product and, therefore, suggests the product's growth potential [32].

### Meaning of the Study

Overuse of CT in minor head injury is complex and multifactorial, including both clinical and nonclinical contributing factors [14,15]. Traditional implementation strategies such as CDS can address clinical factors such as a lack of awareness of the evidence [37]. However, these strategies have had limited success for this decision, likely due to nonclinical factors such as patients' concerns with their condition and care [12-15]. Findings of this study suggest that patients can be educated and engaged in the ED setting in decisions about CT imaging for low-risk minor head injury using a health information technology interface that supports the clinician-patient relationship (rather than getting in its way) [17,38,39]. Specifically, if these findings are confirmed in a larger effectiveness trial, it would imply that successful adoption of the Concussion or Brain Bleed app could help address nonclinical factors that contribute to overuse of CT in minor head injury that are not addressed with traditional implementation strategies and traditional CDS.

### Strengths and Weaknesses of the Study

Unlike traditional implementation efforts, this intervention systematically aims to use technology at the bedside to engage, educate, and reassure patients. This pilot study took place at a single site, so the results may not be generalizable to other EDs. Similarly, unique infrastructure already in place in our ED (but not part of the intervention) could have contributed to the app's success. This study was conducted by 1 RA who was responsible for enrollment, clinician training, and data collection. An RA provides internal consistency but could be prone to bias based

on the RA's level of performance. Enrollment primarily occurred in the evenings, which is similar to our previous findings on enrollment for head injury patients in the ED [40]. The patients enrolled were representative of the patient population seen in our ED, which serves an urban, underserved population with low literacy and numeracy.

This pilot study has shown that it is feasible to use an integrated decision aid with CDS on a tablet computer at the bedside in the ED to engage, educate, and reassure low-risk minor head injury patients about CT and concussion. This finding is promising but, without a control arm, a conclusion cannot be drawn regarding the intervention's efficacy in reducing potentially avoidable CT scans. Although only 2 patients and 4 clinicians declined to participate, enrollment of a convenience sample may also have introduced a self-selection bias of clinicians and patients who were more amenable to this type of approach. For example, the clinicians were relatively young and tech savvy (average age 34 years and 69% using text messaging as their preferred method of personal communication). Since diffusion of innovations benefits from early adoption by a population that is likely to be receptive to change and technology, we believe this is a necessary first step to adoption [41].

One of the top priorities of the Concussion or Brain Bleed app is to have the clinician identify and address the patient's specific concerns. Therefore, we were troubled to note that fidelity with eliciting concerns was the lowest fidelity score of the 8 intended actions that the intervention aimed to elicit (Table 5). To address this, we revised the Risk Discussion screen as discussed in Multimedia Appendix 3.

Based on the secondary analyses, fidelity with identifying and addressing patients' concerns trended to significance for being predictive of CT imaging rate (odds ratio 0.19,  $P=.07$ ), change in patient knowledge (variable  $-0.83$ ,  $P=.21$ ), and trust in physician (variable 5.89,  $P=.11$ ) but these results were not statistically significant. These findings are consistent with our qualitative research that identifying and addressing patients' concerns influences overuse of CT in low-risk minor head injury patients [15]. The findings of the secondary analysis of fidelity with the intervention were consistent with verbal feedback we received from users that it was difficult to *both* educate and address patient concerns due to time constraints. This study was not powered to detect which variables were predictive of outcomes. However, these estimates give us a sense of the direction of association that may exist. The results reported here

will help to determine the sample size of future effectiveness research comparing this intervention versus usual care.

## Unanswered Questions and Future Research

In this pilot study, research staff were available to coordinate use of the Concussion and Brain Bleed app in appropriate patients. Given the competing demands in the ED context, in the absence of research staff there would be multiple barriers to its use, adoption, and integration into routine ED care. Although clinicians reported in every use of the intervention that the app did not affect their clinical decision whether to obtain CT imaging, we maintain that the Concussion or Brain Bleed app has the potential to safely reduce CT imaging in low-risk minor head injury patients. Future research should focus on assessing and optimizing the context for implementation of the Concussion or Brain Bleed app into routine ED care. Identifying barriers and facilitators for how best to embed this complex innovation as part of routine care could optimize its reach, effectiveness, adoption, implementation, and maintenance in routine care [42,43]. For example, a qualitative analysis could explore the reasons that some physicians approved of the tool but would not recommend it to others. Once these factors are identified and optimized, our plan to compare the effectiveness of the app versus usual care could more fully determine its effects on patient experience, clinician experience, health care utilization, and patient safety. If the app is effective, our next goal would be to scale the intervention for dissemination and implementation to outside sites. At the time of this publication, the Concussion or Brain Bleed app is also being adapted for use in Canada with plans to study it there in a comparative effectiveness trial as well. Another category of unanswered questions to explore further would be the concept of patient-centered decision support—for example, which clinical decisions are appropriate for patient decision aids versus traditional CDS versus patient-centered integrated solutions like the one presented here.

## Conclusion

An app to help patients assess the utility of CT imaging after head injury in the ED increased patient knowledge, was associated with a low rate of CT overuse, and was reported to be “extremely helpful” to patients. The high degree of patient satisfaction and clinician acceptability, and high system usability scores are evidence to support the need for rigorous testing of the app in future research that could optimize its implementation into routine ED care and measure its effectiveness compared with usual care.

## Acknowledgments

This project was supported by grant number K08HS021271 from the Agency for Healthcare Research and Quality. The content is solely the responsibility of the authors and does not necessarily represent the official views of the Agency for Healthcare Research and Quality. The authors would like to express their gratitude to Kevin Lopez for computer programming of the intervention and Caitlin Johnson for facilitating the IRB process.

## Conflicts of Interest

None declared.

## Multimedia Appendix 1

Patient Knowledge Assessment Questionnaire.

[JPG File, 53KB - [mhealth\\_v5i9e144\\_app1.jpg](#)]

## Multimedia Appendix 2

Fidelity checklist.

[PDF File (Adobe PDF File), 20KB - [mhealth\\_v5i9e144\\_app2.pdf](#)]

## Multimedia Appendix 3

Revisions to Risk Discussion screen based on pilot user feedback.

[PDF File (Adobe PDF File), 247KB - [mhealth\\_v5i9e144\\_app3.pdf](#)]

## References

1. American College of Emergency Physicians. Choosing Wisely. Philadelphia, PA: ABIM Foundation; 2014. Ten things physicians and patients should question URL: <http://www.choosingwisely.org/societies/american-college-of-emergency-physicians/> [accessed 2017-09-19] [WebCite Cache ID 6tb1stKNR]
2. Melnick ER, Szlezak CM, Bentley SK, Dziura JD, Kotlyar S, Post LA. CT overuse for mild traumatic brain injury. *Jt Comm J Qual Patient Saf* 2012 Nov;38(11):483-489. [Medline: [23173394](#)]
3. Korley FK, Morton MJ, Hill PM, Mundangepfufu T, Zhou T, Mohareb AM, et al. Agreement between routine emergency department care and clinical decision support recommended care in patients evaluated for mild traumatic brain injury. *Acad Emerg Med* 2013 May;20(5):463-469 [FREE Full text] [doi: [10.1111/acem.12136](#)] [Medline: [23672360](#)]
4. Parma C, Carney D, Grim R, Bell T, Shoff K, Ahuja V. Unnecessary head computed tomography scans: a level 1 trauma teaching experience. *Am Surg* 2014 Jul;80(7):664-668. [Medline: [24987897](#)]
5. Sharp AL, Nagaraj G, Ripberger EJ, Shen E, Swap CJ, Silver MA, et al. Computed tomography use for adults with head injury: describing likely avoidable emergency department imaging based on the Canadian CT Head Rule. *Acad Emerg Med* 2017 Dec;24(1):22-30. [doi: [10.1111/acem.13061](#)] [Medline: [27473552](#)]
6. Melnick ER. Big versus small data and the generalizability of the rate of computed tomography overuse in minor head injury. *Acad Emerg Med* 2017 Mar;24(3):391-392. [doi: [10.1111/acem.13084](#)] [Medline: [27617902](#)]
7. Korley FK, Pham JC, Kirsch TD. Use of advanced radiology during visits to US emergency departments for injury-related conditions, 1998-2007. *JAMA* 2010 Oct 06;304(13):1465-1471. [doi: [10.1001/jama.2010.1408](#)] [Medline: [20924012](#)]
8. Stiell IG, Wells GA, Vandemheen K, Clement C, Lesiuk H, Laupacis A, et al. The Canadian CT Head Rule for patients with minor head injury. *Lancet* 2001 May 05;357(9266):1391-1396. [Medline: [11356436](#)]
9. Stiell IG, Wells GA. Methodologic standards for the development of clinical decision rules in emergency medicine. *Ann Emerg Med* 1999 Apr;33(4):437-447. [Medline: [10092723](#)]
10. Papa L, Stiell IG, Clement CM, Pawlowicz A, Wolfram A, Braga C, et al. Performance of the Canadian CT Head Rule and the New Orleans Criteria for predicting any traumatic intracranial injury on computed tomography in a United States level I trauma center. *Acad Emerg Med* 2012 Jan;19(1):2-10 [FREE Full text] [doi: [10.1111/j.1553-2712.2011.01247.x](#)] [Medline: [22251188](#)]
11. Smits M, Dippel DWJ, Nederkoorn PJ, Dekker HM, Vos PE, Kool DR, et al. Minor head injury: CT-based strategies for management--a cost-effectiveness analysis. *Radiology* 2010 Feb;254(2):532-540. [doi: [10.1148/radiol.2541081672](#)] [Medline: [20093524](#)]
12. Sharp AL, Huang BZ, Tang T, Shen E, Melnick ER, Venkatesh AK, et al. Implementation of the Canadian CT Head Rule and its association with use of computed tomography among patients with head injury. *Ann Emerg Med* 2017 Jul 21;33:1505-1514. [doi: [10.1016/j.annemergmed.2017.06.022](#)] [Medline: [28739290](#)]
13. Ip IK, Raja AS, Gupta A, Andruchow J, Sodickson A, Khorasani R. Impact of clinical decision support on head computed tomography use in patients with mild traumatic brain injury in the ED. *Am J Emerg Med* 2015 Mar;33(3):320-325. [doi: [10.1016/j.ajem.2014.11.005](#)] [Medline: [25572644](#)]
14. Probst MA, Kanzaria HK, Schriger DL. A conceptual model of emergency physician decision making for head computed tomography in mild head injury. *Am J Emerg Med* 2014 Jun;32(6):645-650 [FREE Full text] [doi: [10.1016/j.ajem.2014.01.003](#)] [Medline: [24560384](#)]
15. Melnick ER, Shafer K, Rodulfo N, Shi J, Hess EP, Wears RL, et al. Understanding overuse of computed tomography for minor head injury in the emergency department: a triangulated qualitative study. *Acad Emerg Med* 2015 Dec;22(12):1474-1483 [FREE Full text] [doi: [10.1111/acem.12824](#)] [Medline: [26568523](#)]
16. Melnick ER. How to make less more: empathy can fill the gap left by reducing unnecessary care. *BMJ* 2015 Nov 04;351:h5831. [Medline: [26537887](#)]



17. Melnick ER, Lopez K, Hess EP, Abujarad F, Brandt CA, Shiffman RN, et al. Back to the bedside: developing a bedside aid for concussion and brain injury decisions in the emergency department. *EGEMS (Wash DC)* 2015;3(2):1136. [doi: [10.13063/2327-9214.1136](https://doi.org/10.13063/2327-9214.1136)] [Medline: [26290885](#)]
18. Melnick ER, Hess EP, Guo G, Breslin M, Lopez K, Pavlo AJ, et al. Patient-centered decision support: formative usability evaluation of integrated clinical decision support with a patient decision aid for minor head injury in the emergency department. *J Med Internet Res* 2017 May 19;19(5):e174 [FREE Full text] [doi: [10.2196/jmir.7846](https://doi.org/10.2196/jmir.7846)] [Medline: [28526667](#)]
19. Montori VM, Breslin M, Maleska M, Weymiller AJ. Creating a conversation: insights from the development of a decision aid. *PLoS Med* 2007 Aug;4(8):e233 [FREE Full text] [doi: [10.1371/journal.pmed.0040233](https://doi.org/10.1371/journal.pmed.0040233)] [Medline: [17683195](#)]
20. Melnick ER, Probst MA, Schoenfeld E, Collins SP, Breslin M, Walsh C, et al. Development and testing of shared decision making interventions for use in emergency care: a research agenda. *Acad Emerg Med* 2016 Dec;23(12):1346-1353. [doi: [10.1111/acem.13045](https://doi.org/10.1111/acem.13045)] [Medline: [27457137](#)]
21. Consumer Reports, Choosing Wisely, American College of Emergency Physicians. ConsumerHealthChoices. Yonkers, NY: Consumer Reports, Inc Do you need a CT scan for a head injury? URL: <http://consumerhealthchoices.org/wp-content/uploads/2016/10/ChoosingWiselyHeadInjuryACEP-ER.pdf> [accessed 2017-09-20] [WebCite Cache ID 6tccmMewD]
22. Fagerlin A, Zikmund-Fisher BJ, Ubel PA, Jankovic A, Derry HA, Smith DM. Measuring numeracy without a math test: development of the Subjective Numeracy Scale. *Med Decis Making* 2007;27(5):672-680. [doi: [10.1177/0272989X07304449](https://doi.org/10.1177/0272989X07304449)] [Medline: [17641137](#)]
23. Zikmund-Fisher BJ, Smith DM, Ubel PA, Fagerlin A. Validation of the Subjective Numeracy Scale: effects of low numeracy on comprehension of risk communications and utility elicitation. *Med Decis Making* 2007;27(5):663-671. [doi: [10.1177/0272989X07303824](https://doi.org/10.1177/0272989X07303824)] [Medline: [17652180](#)]
24. McNaughton C, Wallston KA, Rothman RL, Marcovitz DE, Storrow AB. Short, subjective measures of numeracy and general health literacy in an adult emergency department. *Acad Emerg Med* 2011 Nov;18(11):1148-1155 [FREE Full text] [doi: [10.1111/j.1553-2712.2011.01210.x](https://doi.org/10.1111/j.1553-2712.2011.01210.x)] [Medline: [22092896](#)]
25. Hess EP, Wyatt KD, Kharbanda AB, Louie JP, Dayan PS, Tzimenatos L, et al. Effectiveness of the head CT choice decision aid in parents of children with minor head trauma: study protocol for a multicenter randomized trial. *Trials* 2014 Jun 25;15:253 [FREE Full text] [doi: [10.1186/1745-6215-15-253](https://doi.org/10.1186/1745-6215-15-253)] [Medline: [24965659](#)]
26. O'Connor AM. Validation of a decisional conflict scale. *Med Decis Making* 1995;15(1):25-30. [Medline: [7898294](#)]
27. Koedoot N, Molenaar S, Oosterveld P, Bakker P, de Graeff A, Nooy M, et al. The decisional conflict scale: further validation in two samples of Dutch oncology patients. *Patient Educ Couns* 2001 Dec 01;45(3):187-193. [Medline: [11722854](#)]
28. Hess EP, Hollander JE, Schaffer JT, Kline JA, Torres CA, Diercks DB, et al. Shared decision making in patients with low risk chest pain: prospective randomized pragmatic trial. *BMJ* 2016 Dec 05;355:i6165 [FREE Full text] [Medline: [27919865](#)]
29. Anderson LA, Dedrick RF. Development of the Trust in Physician scale: a measure to assess interpersonal trust in patient-physician relationships. *Psychol Rep* 1990 Dec;67(3 Pt 2):1091-1100. [doi: [10.2466/pr0.1990.67.3f.1091](https://doi.org/10.2466/pr0.1990.67.3f.1091)] [Medline: [2084735](#)]
30. Thom DH, Ribisl KM, Stewart AL, Luke DA. Further validation and reliability testing of the Trust in Physician Scale. The Stanford Trust Study Physicians. *Med Care* 1999 May;37(5):510-517. [Medline: [10335753](#)]
31. Brooke J. SUS-a quick and dirty usability scale. *Usability Eval Ind* 1996;189(194):4-7.
32. Reichheld F. The one number you need to grow. *Harvard Bus Rev* 2003;81(12):46-55.
33. Stiell IG, Clement CM, Grimshaw JM, Brison RJ, Rowe BH, Lee JS, et al. A prospective cluster-randomized trial to implement the Canadian CT Head Rule in emergency departments. *CMAJ* 2010 Oct 05;182(14):1527-1532 [FREE Full text] [doi: [10.1503/cmaj.091974](https://doi.org/10.1503/cmaj.091974)] [Medline: [20732978](#)]
34. Bangor A, Kortum P, Miller J. Determining what individual SUS scores mean: adding an adjective rating scale. *J Usability Stud* 2009;4(3):114-123.
35. Sauro J, Lewis JR. Quantifying the User Experience: Practical Statistics for User Research. Waltham, MA: Morgan Kaufmann; 2012.
36. Bangor A, Kortum P. Usability ratings for everyday products measured with the System Usability Scale. *Int J Hum Comput Interact* 2013;29(2):76.
37. Glasziou P, Haynes B. The paths from research to improved health outcomes. *Evid Based Nurs* 2005 Apr;8(2):36-38 [FREE Full text] [Medline: [15830412](#)]
38. Gellert G, Webster S, Gilleen J, Melnick E, Kanzaria H. Should US doctors embrace electronic health records? *BMJ* 2017 Dec 24;356:j242. [Medline: [28119282](#)]
39. Ratanawongsa N, Barton JL, Lyles CR, Wu M, Yelin EH, Martinez D, et al. Association between clinician computer use and communication with patients in safety-net clinics. *JAMA Intern Med* 2016 Jan;176(1):125-128 [FREE Full text] [doi: [10.1001/jamainternmed.2015.6186](https://doi.org/10.1001/jamainternmed.2015.6186)] [Medline: [26619393](#)]
40. Mbachu S, Pieribone V, Bechtel KA, McCarthy ML, Melnick E. Optimizing recruitment and retention of adolescents in emergency department research: findings from concussion biomarker pilot. *Am J Emerg Med* 2017 Sep 14. [doi: [10.1016/j.ajem.2017.09.014](https://doi.org/10.1016/j.ajem.2017.09.014)] [Medline: [28918968](#)]
41. Rogers E. Diffusion of Innovations. New York, NY: Free Press Glencoe; 1962.

42. Glasgow RE, Vogt TM, Boles SM. Evaluating the public health impact of health promotion interventions: the RE-AIM framework. *Am J Public Health* 1999 Sep;89(9):1322-1327. [Medline: [10474547](#)]
43. Glasgow RE, Lichtenstein E, Marcus AC. Why don't we see more translation of health promotion research to practice? Rethinking the efficacy-to-effectiveness transition. *Am J Public Health* 2003 Aug;93(8):1261-1267. [Medline: [12893608](#)]

## Abbreviations

**CCHR:** Canadian CT Head Rule

**CDS:** clinical decision support

**CT:** computed tomography

**ED:** emergency department

**IRB:** institutional review board

**RA:** research assistant

*Edited by G Eysenbach; submitted 14.08.17; peer-reviewed by T Agresta, F De Backere; comments to author 06.09.17; revised version received 07.09.17; accepted 08.09.17; published 28.09.17.*

*Please cite as:*

*Singh N, Hess E, Guo G, Sharp A, Huang B, Breslin M, Melnick E*

*Tablet-Based Patient-Centered Decision Support for Minor Head Injury in the Emergency Department: Pilot Study*

*JMIR Mhealth Uhealth* 2017;5(9):e144

URL: <http://mhealth.jmir.org/2017/9/e144/>

doi: [10.2196/mhealth.8732](https://doi.org/10.2196/mhealth.8732)

PMID: [28958987](https://pubmed.ncbi.nlm.nih.gov/28958987/)

©Navdeep Singh, Erik Hess, George Guo, Adam Sharp, Brian Huang, Maggie Breslin, Edward Melnick. Originally published in JMIR Mhealth and Uhealth (<http://mhealth.jmir.org>), 28.09.2017. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mhealth and uhealth, is properly cited. The complete bibliographic information, a link to the original publication on <http://mhealth.jmir.org/>, as well as this copyright and license information must be included.



Review

# Mobile Phone Interventions for Sleep Disorders and Sleep Quality: Systematic Review

Jong Cheol Shin<sup>1</sup>, MS; Julia Kim<sup>2</sup>, RD, MPH; Diana Grigsby-Toussaint<sup>1,2</sup>, MPH, PhD

<sup>1</sup>Department of Kinesiology and Community Health, University of Illinois-Urbana Champaign, Champaign, IL, United States

<sup>2</sup>Division of Nutritional Sciences, University of Illinois-Urbana Champaign, Urbana, IL, United States

**Corresponding Author:**

Diana Grigsby-Toussaint, MPH, PhD

Department of Kinesiology and Community Health

University of Illinois-Urbana Champaign

1206 S Fourth Street

Champaign, IL, 61820

United States

Phone: 1 2173339207

Fax: 1 2173332765

Email: [dgrigs1@illinois.edu](mailto:dgrigs1@illinois.edu)

## Abstract

**Background:** Although mobile health technologies have been developed for interventions to improve sleep disorders and sleep quality, evidence of their effectiveness remains limited.

**Objective:** A systematic literature review was performed to determine the effectiveness of mobile technology interventions for improving sleep disorders and sleep quality.

**Methods:** Four electronic databases (EBSCOhost, PubMed/Medline, Scopus, and Web of Science) were searched for articles on mobile technology and sleep interventions published between January 1983 and December 2016. Studies were eligible for inclusion if they met the following criteria: (1) written in English, (2) adequate details on study design, (3) focus on sleep intervention research, (4) sleep index measurement outcome provided, and (5) publication in peer-reviewed journals.

**Results:** An initial sample of 2679 English-language papers were retrieved from five electronic databases. After screening and review, 16 eligible studies were evaluated to examine the impact of mobile phone interventions on sleep disorders and sleep quality. These included one case study, three pre-post studies, and 12 randomized controlled trials. The studies were categorized as (1) conventional mobile phone support and (2) utilizing mobile phone apps. Based on the results of sleep outcome measurements, 88% (14/16) studies showed that mobile phone interventions have the capability to attenuate sleep disorders and to enhance sleep quality, regardless of intervention type. In addition, mobile phone intervention methods (either alternatively or as an auxiliary) provide better sleep solutions in comparison with other recognized treatments (eg, cognitive behavioral therapy for insomnia).

**Conclusions:** We found evidence to support the use of mobile phone interventions to address sleep disorders and to improve sleep quality. Our findings suggest that mobile phone technologies can be effective for future sleep intervention research.

(*JMIR Mhealth Uhealth* 2017;5(9):e131) doi:[10.2196/mhealth.7244](https://doi.org/10.2196/mhealth.7244)

**KEYWORDS**

mHealth; apps; mobile health; sleep

## Introduction

Sleep disorders are an important public health problem that affects approximately 50 to 70 million people in the United States [1]. Sleep disorders are defined as having abnormal sleep behaviors, including insomnia, sleep apnea, restless leg syndrome, and narcolepsy [2]. Those who have chronic sleep

disorders are at a greater risk for obesity, diabetes, hypertension, cardiovascular disease, stroke, and depression [3-6].

In addition to sleep disorders, insufficient sleep and irregular sleep patterns are also risk factors for obesity [7], impaired cardiovascular function [8,9], and diabetes [10]. According to the 2011-2014 report from the National Center for Health Statistics [11], 31.7% of US adults do not meet the National Sleep Foundation's recommendation for at least 7 hours of sleep

per night. Moreover, the 2011 Sleep in America poll reported that 63% of Americans did not meet the recommendations for sleep time during weekdays [12]. Given that sleep disorders and insufficient sleep increase the risk for chronic diseases, sleep interventions that improve sleep quality are increasingly becoming more important.

In 2011, approximately 39% of Americans, 72% of whom were adolescents, used mobile phones immediately before sleeping [12]. With the increasing use of mobile technology, mobile health (mHealth) is increasingly being used as a practical intervention tool in medicine and public health [13-19]. The term mHealth refers to the provision of health care services and delivery of personal health information using mobile technology such as mobile phones [20]. The portability of mobile devices used in mHealth addresses issues related to accessibility. This allows health care practitioners to provide services to patients regardless of geographical location [15,21]. Specifically, mHealth technology has additional unique functions that are not typically found in traditional care, such as text messaging-tailored medical advice or individualized phone alarms to encourage specific health behaviors (eg, exercising or taking medications) [22]. Regarding sleep, accelerometers in mobile phones can be used to measure and to evaluate sleep patterns [23], and voice recordings can detect abnormal sleep behavior such as snoring and sleep talking [24].

The potential and importance of mHealth technology in health care and health interventions are evident through more than 100,000 health apps in the app store, and a US \$26 billion estimated mHealth market size in 2017 [25]. The 2016 mHealth App Developer Economics report [26] indicated that 53% of new apps were designed to improve or address various health conditions. In sum, mHealth is expected to play a significant role in health care due to its ability to be easily integrated into health care services and intervention studies.

## Research Objective

Although new mHealth technologies are assumed to be able to improve the quality and quantity of sleep, limited examinations of behavioral sleep interventions using mobile phones exist [13,22]. For example, previous studies [27-31] have focused on examining the technological aspects of mobile phones only (ie, calibration and validation studies, device feasibility). Moreover, previous studies examining the use of the mobile phone's effectiveness for addressing sleep quality or quantity have only focused on one specific sleep disorder, such as sleep disturbance [32] or obstructive sleep apnea [33,34]. Therefore, the primary purpose of this review was to determine the effectiveness for sleep disorder improvement. We hypothesized that interventions with mobile technology have a positive impact on improving sleep quality and various sleep disorders (eg, sleep apnea, snoring, or insomnia).

## Methods

### Study Selection Criteria

This study included articles if they met the following criteria:

1. Study design: randomized controlled trials (RCTs), pre-post studies, and case-control studies;

2. Sleep intervention study: using mHealth technology (eg, a mobile device or app);
3. Intervention outcome: sleep outcome measurement (eg, Insomnia Severity Index [ISI], Pittsburgh Sleep Quality Index [PSQI], or Epworth Sleepiness Scale [ESS]);
4. Language: written in English; and
5. Article type: peer-reviewed publications.

### Search Strategy

Articles published between January 1983 (the year of the first handheld and commercial cellular phone from Motorola [35]) and December 2016 were searched from four electronic databases, including EBSCOhost: Academic Search Complete (ASC) & Cumulative Index to Nursing and Allied Health Literature (CINAHL), PubMed/Medline, Scopus, and Web of Science. The journal search process occurred from February 18 to 19, 2017. To identify components of three research topics, namely, "sleep," "mHealth," and "study design," the specified search keywords were used as a search algorithm (see [Multimedia Appendix 1](#)).

After searching the electronic databases, one author (JCS) selected articles that were published in peer-reviewed journals and excluded books, case reports, conference proceedings, product reviews, newspapers, patents, serials, and theses. Secondly, the same author removed duplicate articles from the combined search results and screened articles based on the title and abstract. In addition, JCS hand-searched for relevant articles in the JMIR search engine on February 28, 2017. Once all full-text articles were found, two researchers (JCS and JK) reviewed each article using the eligibility criteria and eliminated unrelated articles. The entire procedure followed the guidelines for Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) [36].

### Data Extraction

The following information was extracted from each article included in the systematic review: first author's name, publication year, the country where the study was conducted, sample size, sample age, sample characteristics, study period, retention rate, study design, the technology used in the intervention, and measurement of sleep outcome. The retention rate was calculated using the number of participants who completed all assessments for the intervention study divided by the number of individuals who were originally recruited. Intervention technology was identified using two approaches: (1) intervention using mobile phone app and (2) supplementary mobile phone usage with traditional intervention. Additional information was extracted by using the empirical approach of statistical analysis such as effect size, standard error, and standard deviation for the difference before and after interventions.

### Statistical Analysis

The effect size and standard deviation for each study's sleep outcome measurement were obtained from data extraction. With each effect size, there was a difference of calculation for the total effect size based on the type of study. For example, the effect size of pre-post test was only considered the difference between pretest scores and posttest scores. On the other hand,

the effect size of RCTs was calculated using the difference between the intervention and the control groups' effect sizes. Based on the total effect size, sample size, and standard error, two-sample *t* tests were performed to determine the mobile phone intervention effectiveness.

### Study Quality Assessment

The study quality assessment tool was derived from Zhu and An [37], and then revised after discussion and peer review by the researchers for the purpose of this study. The quality of research was assessed using the following criteria: (1) the research question and objectives were stated clearly, (2) mobile health and/or mHealth was defined, (3) a control group was included, (4) participants were randomly recruited from a well-defined population, (5) sample size was more than 30, (6) attrition was analyzed and determined not to significantly differ by respondents' baseline characteristics between control and experiment groups (<20%), (7) baseline characteristics between control and intervention groups were similar, (8) the intervention period was at least 4 weeks, (9) the sleep disorder measurement tools were shown to be reliable and valid in previous studies, and (10) demographic information was available to control for potential confounders in future analysis. A sample size of 30 was chosen due to a guarantee of a normal sample distribution based on the central limit theorem. The total study quality score was determined by summing items 1 to 10, and categories were created to define a general quality of research as poor (score 0-3); moderate (score 4-6), and high (score 7-10).

## Results

### Study Selection

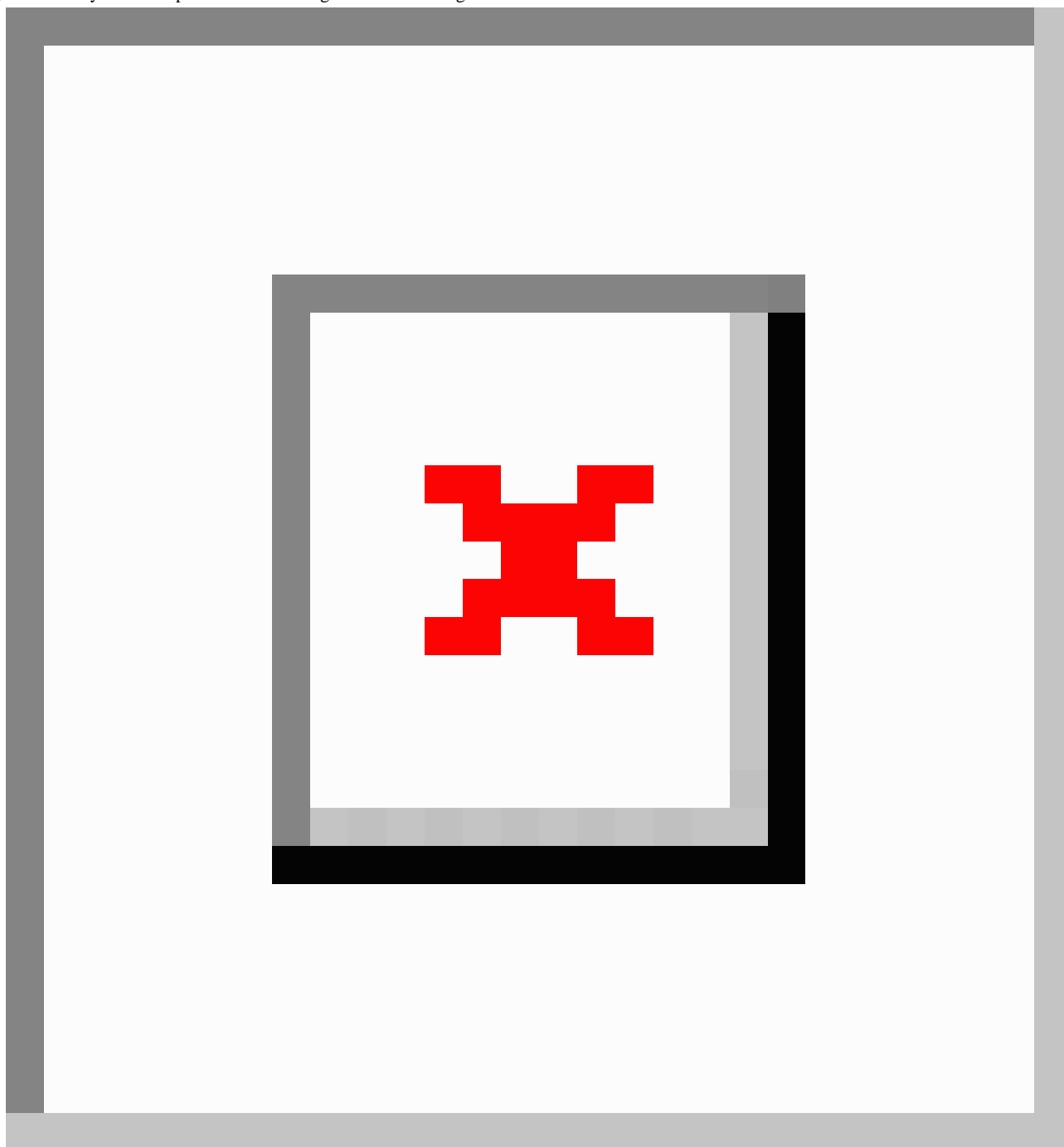
The results of the literature search are summarized in Figure 1. A total of 2674 articles were initially selected from four electronic databases (EBSCOhost: CINAHL&ASC, PubMed/Medline, Scopus, and Web of Science). After excluding 440 duplicate articles, 896 articles were excluded from the first screening of peer-reviewed journal articles and 1339 articles were left. After the first screening, five articles were included from hand searching, and 1304 were excluded based on the title

(n=1180) and abstract (n=124). All 39 remaining articles were then downloaded as PDF files for full-text peer review and discussion [13-16,21,22,38-70]. We excluded 23 articles after the full-text screening for the following reasons: (1) a device or program feasibility study (n=2) [66,69], (2) duplicated article (n=1) [21], (3) not an intervention study (n=5) [39,43,47,49,53], (4) not a mobile phone intervention (n=7) [14,42,48,56,58,63,68], (5) no sleep outcome (n=3) [16,44,65], and (6) a protocol or trial (n=5) [40,45,50-52]. The entire selection procedure was independently performed by two researchers (JCS and JK), and then the research team resolved differences by discussion. From individual selections, the interrater reliability analysis using the kappa statistic was performed with peer reviewers to determine consistency among raters. The interrater reliability was found to be Cohen kappa=.80, representing substantial agreement [20]. Therefore, a total of 16 articles that satisfied the inclusion criteria were used for the final analysis.

### Study Characteristics

Table 1 provides summary information of individual studies. A total of 16 studies were included in the analysis: one case study [46], three pre-post studies [13,15,41], and 11 RCTs [21,22,38,54,55,57,59,61,62,64,67,70]. Of the 12 RCTs, three studies [62,64,67] used mobile phone apps as an intervention tool; one study [21] used text messages as part of the intervention. RCTs ranged in total sample size from 30 to 502 and had a larger sample size than pre-post studies, which ranged from three [13] to 12 [41]. The study period for RCTs was from 4 weeks to 6 months, whereas for pre-post studies it was between 2 and 5 weeks. For the study region, eight studies were performed in the United States [13,15,21,41,55,61,62,70], two studies in Canada [54,67], two studies in Asia (Hong Kong [57] and Taiwan [46]), and four studies in Europe (Finland [38], France [64], Netherlands [21], and Sweden [59]). Most of the studies were published after 2012, except one study that was published in 2006 [67]. Three studies [15,46,64] focused on an elderly population, whereas three studies [21,67,70] focused on a young adult population.

**Figure 1.** Study selection procedure according to the PRISMA guidelines.



**Table 1.** Basic characteristics of the studies of mobile phone interventions on sleep disorders (N=16).

Author(s), Year	Study region	Sample characteristics <sup>a</sup>	Sample size	Mean age	Study period	Intervention methods <sup>b</sup>	Study design <sup>c</sup>	Sleep outcome measurement <sup>d</sup>
Anttalainen et al, 2014 [38]	Finland	OSA patients	111	55.27	3 m	Phone + CPAP	RCT	AHI, ESS
Babson et al, 2015 [13]	United States	Veterans with CUD	3	47	2 w	App: “CBT-I Coach” for iOS	Pre-post	PSQI
Bauer et al, 2012 [41]	United States	Metropolitan area adults	12	32	4 w	App: “ShutEye” for Android	Pre-post	ESS
Chen et al, 2015 [46]	Taiwan	Elderly female	1	64	5 w	App: “Win-Win A Sleep”	Case study	SSR
Filion et al, 2015 [21]	United States	Young adult smokers	116	18-25	6 w	Text message	RCT	SQ_PSQI, TST
Fox et al, 2012 [54]	Canada	OSA patients	75	53.54	3 m	Phone + CPAP	RCT	AHI, ESS
Freeman et al, 2015 [55]	United States	Breast cancer survivors	102	55.44	3 m	Phone-supported teleconference	RCT	PSQI
Ho et al, 2014 [57]	Hong Kong	Insomnia patients	149	38.5	12 w	Phone + CBT-I	RCT	PSQI, ISI, DBAS, SE, SQ, SOL, WASO, TST
Jernelov et al, 2012 [59]	Sweden	Adults with insomnia	133	47.9	6 w	Phone + bibliotherapy	RCT	ISI, DBAS, SE, SQ, SOL, WASO, TST, BTS, SRBQ
Koffel et al, 2016 [62]	United States	Veterans	18	48.5	5 w	Apps: “CBT-I Coach”	RCT	ISI
Lichstein et al, 2013 [15]	United States	Rural area adults with insomnia	5	65.8	5 w	Apps: “Skype” + CBT	Pre-post	ISI, HRSD, NWAK, SOL, SQ, WASO
Kauffman 2016 [61]	United States	Menopausal status with insomnia	106	54.85	24 w	Phone + CBT-I	RCT	ISI, PSQI, SE, SOL, TST, WASO
Mendelson et al, 2014 [64]	France	OSA patients	107	63	4 w	Researcher-built app + CPAP	RCT	ESS
Stremmler et al, 2006 [67]	Canada	First-time mothers	30	31.85	6 w	Phone + sleep education	RCT	GSDS
van Dongen et al, 2014 [22]	Netherlands	Airline pilots	502	40.9	6 m	App: “More energy”	RCT	PSQI
Vuletic et al., 2016 [70]	United States	Soldiers With MTBI	356	29.35	6 m	Telephone-based problem-solving treatment	RCT	PSQI, SE, SOL, SQ

<sup>a</sup>CUD: cannabis use disorders; MTBI: mild traumatic brain injury; OSA: obstructive sleep apnea.

<sup>b</sup>CBT-I: Cognitive Behavioral Therapy for Insomnia; CPAP: continuous positive airway pressure.

<sup>c</sup>RCT: randomized controlled trial.

<sup>d</sup>AHI: Apnea-Hypopnea Index; BTS: bed time stress; DBAS: Dysfunctional Beliefs and Attitudes about Sleep scale; ESS: Epworth Sleepiness Scale; GSDS: General Sleep Disturbance Scale; HRSD: Hamilton Rating Scale for Depression with sleep; ISI: Insomnia Severity Index; PSQI: Pittsburgh Sleep Quality Index; NWAK: number of awakenings; SE: sleep efficiency; SOL: sleep-onset latency; SQ: sleep quality; SQ\_PSQI: extracted sleep quality score based on PSQI; SRBQ: Sleep-Related Behavior Questionnaire; SSR: Sleep Satisfaction Rate; TST: total sleep time; WASO: wakefulness after sleep onset.



**Table 2.** Summary of sleep measurement tools and study design for mobile phone sleep intervention included in the systematic review (N=16).

Measurement	Scales	Case study & pre-post test (n=4)	RCT (n=12)  Standard treatment: CBT-I & CPAP (n=5) <sup>a</sup>	Other recognized treatment (n=7)	Waitlist (n=4) <sup>b</sup>
Apnea-Hypopnea Index (AHI)	Score (total apneas event/TST)	0	2 [38,54]	0	0
Bed time stress (BTS)	Scored on scale 0-5	0	0	1 [59]	1 [59]
Dysfunctional Beliefs and Attitudes about Sleep Scale (DBAS)	30 items with scale 1-10 (total: 300)	0	1 [57]	1 [59]	2 [57,59]
Epworth Sleepiness Scale (ESS)	8 items with scale 0-3 (total: 24)	1 [41]	2 [38,54]	1 [64]	0
General Sleep Disturbance Scale (GSDS)	21 items with scale 0-7 (total: 147)	0	0	1 [67]	0
Hamilton Rating Scale for Depression with sleep (HRSD)	Total 21 items: <sup>c</sup> 10 items with scale 0-4, 2 items with scale 0-3, and 10 items with scale 0-2 (total: 66)	1 [15]	0	0	0
Insomnia Severity Index (ISI)	7 items with scale 0-4 (total: 28)	1 [15]	3 [15,57,61]	2 [59,62]	2 [57,59]
Number of awakenings (NWAK)	Frequency of awakening	1 [15]	0	0	0
Pittsburgh Sleep Quality Index (PSQI)	7 components calculated from 9 questions with scale 0-3 (total: 21)	1 [13]	2 [57,61]	2 [55,70]	3 [22,55,57]
Sleep efficiency (SE) <sup>c</sup>	Percentage (TST/total time in bed)	0	2 [57,61]	2 [59,70]	2 [57,59]
Sleep-onset latency (SOL)	Minutes	1 [15]	2 [57,61]	2 [59,70]	2 [57,59]
Sleep quality (SQ) <sup>c</sup>	Scored on scale 1-5	1 [15]	1 [57]	2 [59,70]	2 [57,59]
Extracted sleep quality score based on PSQI (SQ_PSQI)	8 items with scale 1-4 (total: 32)	0	0	1 [21]	1 [21]
Sleep-Related Behavior Questionnaire (SRBQ)	32 items with scale 1-5 (total: 160)	0	0	1 [59]	1 [59]
Sleep Satisfaction Rate (SSR) <sup>c</sup>	Scored on scale 0-3	1 [46]	0	0	0
Total sleep time (TST)	Hours	0	2 [57,61]	2 [21,59]	3 [21,57,59]
Wakefulness after initial sleep onset (WASO)	Minutes	1 [15]	2 [57,61]	1 [59]	2 [57,59]

<sup>a</sup>CBT-I: Cognitive Behavioral Therapy for Insomnia; CPAP: continuous positive airway pressure.

<sup>b</sup>Three articles have two comparison groups (waitlist vs other).

<sup>c</sup>Higher scores indicate lesser severity.

Table 2 shows the number of studies based on study design and sleep measurement tools. Several questionnaires were used to evaluate various aspects of sleep quality and quantity.

Specifically, the Pittsburgh Sleep Quality Index (PSQI) [13,21,22,55,57,61,70] had a list of items to evaluate sleep quality and it was the most frequently used measure in our review. In addition, three studies used a sleep quality [15,57,59] measure that consisted of a simple question about the participant's perception about sleep quality using a five-point scale. Moreover, five studies [15,21,57,59,61] used measures of sleep quantity: (1) sleep-onset latency (n=4; the length of time to transition from full wakefulness to sleep completely) [15,57,59,61], (2) number of awakenings (n=1, the frequency of awakening during sleep) [15], (3) wake after sleep onset (n=4; the amount of time [eg, minutes] one wakes up during sleep) [15,57,59,61], and (4) total sleep time (n=4; total length of sleep time) [21,57,59,61]. Total sleep time is also used to

calculate sleep efficiency [57,59,61]; that is, the percentage of total sleep time per total bed time. For the measurement of behavioral aspects of sleep, some researchers used the Epworth Sleepiness Scale (ESS) (n=4) [38,41,54,64] to capture participant sleepiness. Also two additional studies measured other specific sleep-related behaviors, such as the Sleep-Related Behavior Questionnaire (SRBQ) [59], the Dysfunctional Beliefs and Attitudes about Sleep Scale (DBAS) [57,59], and Sleep Satisfaction Rate (SSR) [46]. Finally, some sleep measurement tools focused on specific sleep disorders such as the severity of insomnia. The Apnea-Hypopnea Index (AHI) [38,54] is used to evaluate sleep apnea, and the General Sleep Disturbance Scale (GSDS) [67] is meant to examine factors that interrupt sleep. Two studies [15,59] used tools to examine the emotional aspects of sleep, such as bed time stress (BTS) levels [59], and the Hamilton Rating Scale for Depression with sleep (HRSD) [15]. All sleep outcome measurements were conducted by subjective measurement (ie, self-report questionnaire).

**Table 3.** Summary statistics: quality assessment of each journal and the effects of intervention through *t* test by intervention type and sleep outcome measurement.

Author(s), year	Quality score	Study design: intervention type <sup>a</sup>	Sleep measurement <sup>b</sup>	Effect size, mean (SD)	95% CI	<i>P</i>
Anttalainen et al, 2014 [38]	7	RCT: mobile + CPAP vs standard CPAP	AHI	-1.9 (2.9)	-2.4, -1.4	<.001
			ESS	0.0 (1.7)	-0.3, 0.3	>.99
Babson et al, 2015 [13]	3	Pre-post	PSQI	-1.5 (2.4)	-6.2, 3.2	NA
Bauer et al, 2012 [41]	5	Pre-post	ESS	-1.7 (1.3)	-2.4, -0.9	.001
Chen et al, 2015 [46]	2	Case study	SSR			NA
Filion et al, 2015 [21]	8	RCT: smoking prevention text vs sleep or activity	SQ_PSQI	-0.2 (1.7)	-0.5, 0.1	.14
			TST (weekday)	0.6 (0.5)	0.5, 0.6	<.001
		Promotion text	TST (weekend)	0.6 (0.7)	0.5, 0.7	<.001
Fox et al, 2012 [54]	8	RCT: mobile + CPAP vs standard CPAP	AHI	-1.9 (4.3)	-2.9, -0.9	<.001
			ESS	-0.9 (5.1)	-2.1, 0.3	.13
Freeman et al, 2015 [55]	9	RCT: mobile vs waitlist	PSQI	-2.6 (1.5)	-2.9, -2.2	<.001
		RCT: mobile vs standard treatment	PSQI	-0.7 (1.4)	-1.0, -0.4	<.001
Ho et al, 2014 [57]	8	RCT: mobile + CBT-I vs waitlist	DBAS	-23.5 (15.1)	-25.5, -21.5	<.001
			ISI	-3.1 (1.8)	-3.3, -2.9	<.001
			PSQI	-2.3 (1.5)	-2.5, -2.1	<.001
			SE	4.5 (4.8)	3.8, 5.2	<.001
			SOL	-7.4 (12.9)	-9.2, -5.6	<.001
			SQ	-0.2 (0.2)	-0.2, -0.2	<.001
			TST	0.0 (8.8)	-1.2, 1.2	.99
			WASO	-10.8 (14.8)	-12.8, -8.8	<.001
		RCT: mobile + CBT-I vs standard CBT-I	DBAS	-0.3 (16.0)	-2.5, 1.9	<.001
			ISI	-1.1 (1.8)	-1.3, -0.9	<.001
			PSQI	-0.8 (1.5)	-1.0, -0.6	<.001
			SE	1.2 (4.8)	0.5, 1.9	<.001
			SOL	-7.1 (12.9)	-8.9, -5.3	<.001
			SQ	-0.1 (0.2)	-0.1, -0.1	<.001
			TST	-0.1 (8.8)	-1.3, 1.1	.74
			WASO	-7.7 (14.7)	-9.7, -5.7	<.001
Jernelov et al, 2012 [59]	9	RCT: mobile + bibliotherapy vs waitlist	BTS	-0.9 (0.4)	-1.0, -0.8	<.001
			DBAS	-56.4 (9.8)	-58.4, -54.4	<.001
			ISI	-8.8 (1.6)	-9.1, -8.5	<.001
			SE	15.3 (7.1)	13.8, 16.8	<.001
			SOL	-30.7 (20.4)	-35.0, -26.4	<.001
			SQ	0.9 (0.2)	0.9, 0.9	<.001
			SRBQ	-21.9 (4.8)	-22.9, -20.9	<.001
			TST	0.4 (0.5)	0.3, 0.5	<.001
			WASO	-30.1 (18.9)	-34.1, -26.1	<.001
		RCT: mobile+ bibliotherapy vs bibliotherapy only	BTS	-0.9 (0.4)	-1.0, -0.8	<.001
			DBAS	-37.0 (11.0)	-39.3, -34.7	<.001

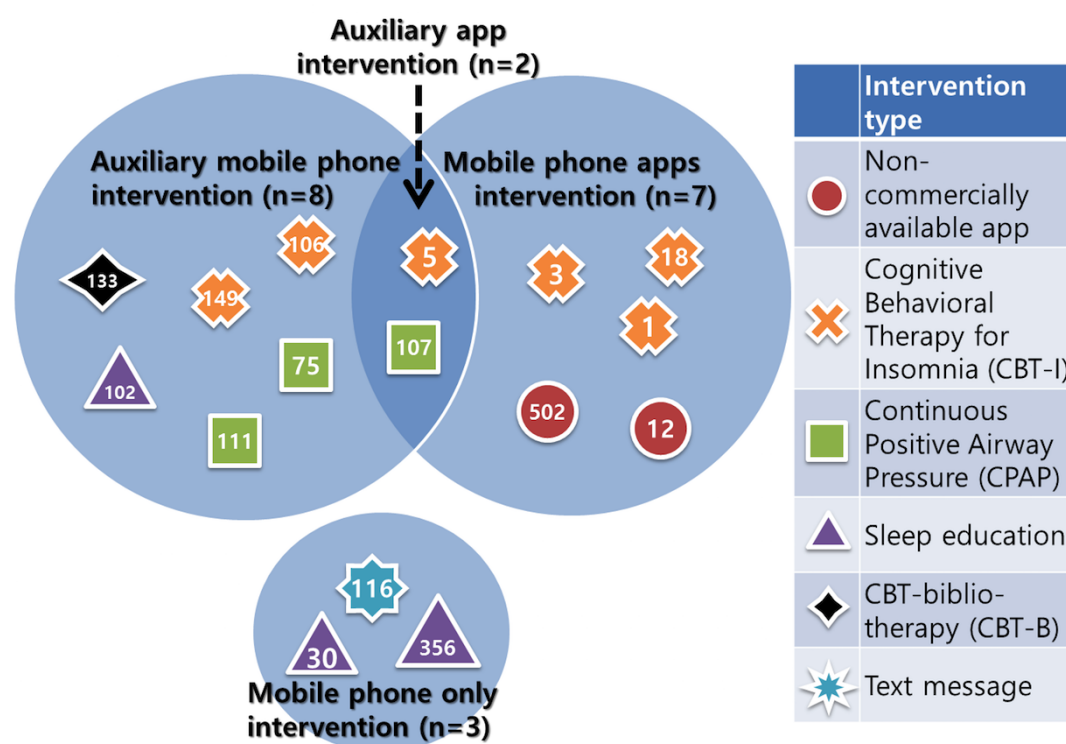
Author(s), year	Quality score	Study design: intervention type <sup>a</sup>	Sleep measurement <sup>b</sup>	Effect size, mean (SD)	95% CI	P
			ISI	-4.5 (1.5)	-4.8, -4.2	<.001
			SE	10.2 (7.0)	8.7, 11.7	<.001
			SOL	-14.8 (21.8)	-19.3, -10.3	<.001
			SQ	0.5 (0.2)	0.5, 0.5	<.001
			SRBQ	-15.3 (5.4)	-16.4, -14.2	<.001
			TST	0.0 (0.5)	-0.1, 0.1	.69
			WASO	-21.1 (19.4)	-25.1, -17.1	<.001
Koffel et al, 2016 [62]	8	RCT: app use vs non app	ISI	3.4 (4.6)	1.2, 5.5	.008
Lichstein et al, 2013 [15]	6	Pre-post	HRSD	-7.0 (3.6)	-10.2, -3.8	.01
			ISI	-11.3 (3.3)	-14.2, -8.3	.002
			NWAK	-1.7 (2.1)	-3.5, 0.1	.14
			SOL	-18.4 (20.9)	-36.7, -0.1	.12
			SQ	0.5 (0.1)	0.4, 0.6	.001
			WASO	-23.8 (9.2)	-31.9, -15.7	.004
Kauffman 2016 [61]	8	RCT: mobile + CBT-I vs menopause education	ISI	-4.0 (5.6)	-5.1, -2.9	<.001
			PSQI	-1.6 (3.0)	-2.2, -1.0	<.001
			SE	3.2 (12.3)	0.8, 5.6	.009
			SOL	-11.9 (36.3)	-18.8, -5.0	.001
			TST	0.2 (1.2)	0.0, 0.4	.08
			WASO	-6.8 (47.3)	-15.8, 2.2	.14
Mendelson et al, 2014 [64]	8	RCT: app + CPAP vs standard CPAP	ESS	-0.2 (1.5)	-0.5, 0.1	.17
Stremmler et al, 2006 [67]	8	RCT: mobile + education vs education	GSDS	-13.3 (8.2)	-16.2, -10.4	<.001
van Drongelen et al, 2014 [22]	9	RCT: app use vs non app	PSQI	-0.6 (1.3)	-0.7, -0.4	<.001
Vuletic et al, 2016 [70]	8	RCT: mobile + education vs education	PSQI	-1.5 (1.5)	-1.7, -1.4	<.001

<sup>a</sup>CBT-I: cognitive behavioral therapy for insomnia; CPAP: continuous positive airway pressure.

<sup>b</sup>AHI: Apnea-Hypopnea Index; BTS: bed time stress; DBAS: Dysfunctional Beliefs and Attitudes about Sleep Scale; ESS: Epworth Sleepiness Scale; GSDS: General Sleep Disturbance Scale; HRSD: Hamilton Rating Scale for Depression with sleep; ISI: Insomnia Severity Index; PSQI: Pittsburgh Sleep Quality Index; NWAK: number of awakenings; SE: sleep efficiency; SOL: sleep-onset latency; SQ: sleep quality; SQ\_PSQI: extracted sleep quality score based on PSQI; SRBQ: Sleep-Related Behavior Questionnaire; SSR: Sleep Satisfaction Rate; TST: total sleep time; WASO: wakefulness after initial sleep onset.

Table 3 provides a summary of the effectiveness of each sleep intervention. It is categorized by sleep outcome measurement, and the statistics can be comparable if there are more than two sleep measures. Specifically, half of studies [13,22,41,46,62,64,67,70] used a single sleep outcome measurement, while others had at least two or more measurements. Of nine studies [13,15,38,46,54,57,61,62,64] of standard treatment, five studies compared a mobile phone

intervention and standard treatment intervention: (1) continuous positive airway pressure (CPAP) [38,54,64] and (2) cognitive behavioral therapy for insomnia (CBT-I) [57,61]. Five studies [13,22,41,46,61] focused on a mobile phone app. Two of these studies [13,61] used a “CBT-I coach” app that was developed based on CBT-I, which is considered one of the traditional standard treatments for insomnia (Figure 2).

**Figure 2.** Sample size based on the intervention methods (N=16).**Table 4.** Quality assessment of studies to determine the impact of sleep intervention with mobile technology on sleep disorders.

Criteria item	Score, mean (SD)
1. Research question and objective were stated clearly	0.94 (0.24)
2. Definition of telehealth and/or mHealth was stated	0.18 (0.39)
3. A control group was included	0.82 (0.39)
4. Participants were randomly recruited from well-defined population	0.82 (0.39)
5. Sample size was >30	0.76 (0.44)
6. Attrition was analyzed and determined not to significantly differ by respondents' baseline characteristics between control and experiment groups (<20%)	0.47 (0.51)
7. Baseline characteristics between control and intervention groups were similar	0.76 (0.44)
8. The intervention period was at least 4 weeks	0.94 (0.24)
9. The sleep disorder measurement tools were shown to be reliable and valid in previously published studies	0.88 (0.33)
10. Demographic information is available to control potential confounders for future analysis	0.82 (0.39)
Total study quality score <sup>a</sup>	7.41 (0.44)

<sup>a</sup>By summing up items 1 to 10 (range 3-10).

## Study Quality

Table 4 summarizes the results of the quality assessment of the studies included in this review. The quality score of the final articles ranged from 3 to 10 out of a possible score of 10, with an overall mean score of 7.41 (SD 0.44). The distribution of quality scores differed substantially across each criterion. Of 16 articles, 12 articles were considered high quality [21,22,38,54,55,57,59,61,62,64,67,70], two articles moderate quality [15,41], and two articles poor quality [13,46]. Most of the studies had clear research questions and objectives [13,15,21,22,38,41,46,54,55,57,59,62,64,67,70], and used valid

sleep measurement tools [13,15,21,22,38,41,54,55,57,59,61,62,64,70] for outcome measures after 4 weeks [15,21,22,38,41,46,54,55,57,59,61,62,64,67,70]. All RCTs [21,22,38,54,55,57,59,61,62,64,67,70] satisfied items 3 (ie, existence of control group) and 5 (ie, sample size greater than 30). The mean intraclass correlation (ICC) between peer reviewers for quality assessment was ICC 0.987 (95% CI 0.964-0.996,  $P<.001$ ).

## Effectiveness of Mobile Phones Usage as a Sleep Intervention Tool

The mobile phone can be used as a tool to effectively deliver and enhance traditional behavioral interventions due to its

portability. We found eight articles that described the advantages of using mobile phones to improve traditional intervention methods for sleep disorders [38,54,55,57,59,61,67,70].

### **Mobile Phone Usage as Alternative Intervention Tools**

Mobile phones are a tool for intervention studies due to the portability for the participant. Within the results of our review, two studies focused on the effectiveness of mobile phones as intervention tools.

Stremmler et al [67] was the first published research article that used a mobile phone as a behavioral-educational strategy tool for a sleep intervention. The focus of the study was to evaluate the feasibility, acceptability, and effects of a mobile phone intervention on sleep in the early postpartum period. During the 6-week intervention period, the intervention group was provided a well-established booklet, in-depth counseling, and a nurse to call for sleep advice, whereas the control group was only provided a one-page pamphlet and brief meeting session. Because the primary target subjects were first-time mothers and their children, researchers used GSDS to evaluate the sleep disturbances in employed women. As a result, the sleep intervention group had lower GSDS scores as much as 13.3 points compared to the control group ( $t_{28}=-8.84$ ,  $P<.001$ ).

Similarly, Vuletic et al [70] had two intervention groups: (1) a group with 12 biweekly “telecounseling” and (2) an “education-only” group with an educational brochure. The research team developed the telephone-based problem-solving treatment, which was a sleep intervention that related to sleep quality to improve post-deployment soldier’s traumatic brain injury. Since the researchers used various sleep outcome measurements such as SOL, total sleep time, sleep disturbance, and daytime dysfunction that were part of PSQI, only PSQI was included for data analysis. After 6 months of intervention, the intervention group with telecounseling had significantly lower scores than the education-only intervention group (effect size=-1.54, SE 0.19;  $t_{274}=-16.62$ ,  $P<.001$ ).

### **Mobile Phone Used as Auxiliary Equipment**

Both CPAP and CBT-I are considered standard treatments for obstructive sleep apnea (OSA) [38,54] and sleep disturbances [13,57,62]. Three studies [38,54,61] applied mobile phones as supplementary tools in addition to these standard treatments.

Fox and colleagues [54] conducted a RCT to determine the effectiveness and adherence of telemonitoring and CPAP combined. To demonstrate the effectiveness of the intervention, AHI and ESS were used to measure sleep outcome. Researchers found that participants in the intervention group had an improvement in adherence to CPAP, but no significant differences were found for AHI and ESS between the control and intervention groups. Based on  $t$  test results, in addition, the effect size of AHI showed significantly decreased ( $t_{73}=-3.82$ ,  $P<.001$ ).

Anttalainen et al [38] had a similar research method as Fox and colleagues’ [54], such as intervention duration, study design, intervention methods, presence of a control group, and sleep outcome measurement. Anttalainen et al did not find a statistically significant difference in ESS between groups,

whereas participants in the intervention group had a significantly higher score on the AHI than standard in-person treatment ( $t_{109}=-6.91$ ,  $P<.001$ ). Because ESS between standard CPAP and mobile-supported CPAP were not significantly different, the effects of mobile-supported CPAP did not differ from standard CPAP.

Kauffman [61] conducted a RCT to determine the effectiveness of telephone-based CBT-I to determine the effectiveness of telephone-based CBT-I compared to menopause education control. The target population was menopausal middle-aged women with moderate insomnia ( $ISI>15$ ). Based on the  $t$  test results, there was a significantly different effect size for the index type sleep measurements between the telephone-based CBT-I and menopause education control ( $ISI$ :  $t_{104}=-7.33$ ,  $P<.001$ ;  $PSQI$ :  $t_{104}=-5.54$ ,  $P=.007$ ), whereas sleep time measures were not significant such as sleep onset latency, wake after sleep onset, total sleep time, and sleep efficiency.

### **Multidimensional Approach to Mobile Phone Usage**

Three studies [55,57,59] examined the effectiveness of mobile phone usage as an alternative intervention tool and as an auxiliary equipment.

Jernelov et al [59] evaluated the effectiveness of mobile phone support with CBT bibliotherapy (CBT). To examine the effectiveness of mobile phone support, a total of 133 participants were randomly allocated to three comparable groups (44 for CBT bibliotherapy with mobile support; 45 for only CBT bibliotherapy; 44 for the control group). After the 6-week intervention, a total of 116 participants completed the follow-up assessment (40 for CBT bibliotherapy with mobile support; 37 for CBT bibliotherapy; 39 for the control group), with an attrition rate of 12.78%. Participants in the mobile-supported CBT bibliotherapy group had significant improvements in all sleep measurements compared to the control group (BTS:  $t_{86}=-20.80$ ,  $P<.001$ ; DBAS:  $t_{86}=-53.96$ ,  $P<.001$ ; ISI:  $t_{86}=-50.64$ ,  $P<.001$ ; SE:  $t_{85}=20.18$ ,  $P<.001$ ; SOL:  $t_{85}=-14.06$ ,  $P<.001$ ; SQ:  $t_{85}=39.59$ ,  $P<.001$ ; SRBQ:  $t_{86}=-42.81$ ,  $P<.001$ ; TST:  $t_{85}=6.77$ ,  $P<.001$ ; WASO:  $t_{85}=-14.83$ ,  $P<.001$ ), and even CBT bibliotherapy only (BTS:  $t_{87}=-21.95$ ,  $P<.001$ ; DBAS:  $t_{87}=-31.74$ ,  $P<.001$ ; ISI:  $t_{87}=-28.88$ ,  $P<.001$ ; SE:  $t_{87}=13.78$ ,  $P<.001$ ; SOL:  $t_{87}=-6.39$ ,  $P<.001$ ; SQ:  $t_{87}=24.41$ ,  $P<.001$ ; SRBQ:  $t_{87}=-26.77$ ,  $P<.001$ ; WASO:  $t_{87}=-10.26$ ,  $P<.001$ ), except TST ( $t_{87}=0.685$ ,  $P=.84$ ).

Although Ho et al [57] had a similar study design and approach to Jernelov and colleagues [59], Ho et al had a different intervention tool: the comparison between the self-help CBT-I with the mobile phone support and self-help CBT-I only. Of importance was that this study had a relatively higher attrition rate (49%) because this study was community based. To be specific, although the baseline participants were high ( $N=302$ ; 103 for CBT-I with mobile support; 104 for CBT-I; 105 for control group), the total number of participants that completed the follow-up assessment was similar with other RCTs ( $n=149$ ; 44 for CBT-I with mobile support; 39 for CBT-I; 66 for control group). The intervention had two follow-up sessions (at 4 weeks and at 12 weeks), but the waitlist group did not have 12-week



evaluation. Therefore, the statistics in Table 3 were based on the follow-up session at 4 weeks to compare whether both interventions significantly improved the sleep measurement score compared to the initial assessment. From the  $t$  test, several sleep outcome measures indicated that mobile phone-supported intervention had a statistically significant improvement compared to the waitlist control group (PSQI:  $t_{206}=-21.93$ ,  $P<.001$ ; ISI:  $t_{206}=-24.37$ ,  $P<.001$ ; DBAS:  $t_{206}=-22.48$ ,  $P<.001$ ; SOL:  $t_{206}=-8.27$ ,  $P<.001$ ; WASO:  $t_{206}=-10.56$ ,  $P<.001$ ; SE:  $t_{206}=13.57$ ,  $P<.001$ ; and SQ:  $t_{206}=-14.14$ ,  $P<.001$ ) and also the standard treatment intervention (PSQI:  $t_{205}=-7.63$ ,  $P<.001$ ; ISI:  $t_{205}=-8.64$ ,  $P<.001$ ; SOL:  $t_{205}=-7.94$ ,  $P<.001$ ; WASO:  $t_{205}=-7.53$ ,  $P<.001$ ; SE:  $t_{205}=3.62$ ,  $P<.001$ ; and SQ:  $t_{205}=-7.85$ ,  $P<.001$ ).

Freedman et al [55] established the “Envision the Rhythms of Life” program, which is a group intervention to increase the quality of life for breast cancer survivors, and also to evaluate its advantages on aspects of quality of life. A total of 118 participants were randomly assigned to three groups (48 participants for the in-person delivery; 23 participants for the mobile phone-supported teleconferencing; 47 participants for waitlists) at the beginning of the study, and 102 participants remained (40 participants for the in-person delivery; 19 participants for the mobile phone-supported teleconferencing; 43 participants for the waitlists) at the end of the study. According to the  $t$  test, participants in the mobile phone-supported group had a significant increase in PSQI scores among breast cancer survivors compared to both control groups (standard treatment: mean  $-0.68$ , SD  $1.39$ ;  $t_{73}=-4.11$ ,  $P<.001$ ; waitlist: mean  $-2.55$ , SD  $1.53$ ;  $t_{73}=-13.97$ ,  $P<.001$ ). In addition, authors performed the linear multilevel modeling analysis, and found that PSQI was only considerable for group effect ( $P<.001$ ), not for time ( $P=.35$ ) or group $\times$ time ( $P=.30$ ).

### Text Message for Sleep Intervention

Filion et al [21] used a text message-based intervention for young adult smokers to promote better sleep and physical activity habits. A total of 164 baseline participants were assigned to two groups: (1) sleep/physical activity group ( $n=63$ ) and (2) smoking cessation group ( $n=101$ ). A total of 129 participants completed the study (smoking cessation group:  $n=77$ ; sleep group:  $n=44$ ). Participants in the sleep/physical activity group received sleep- and activity-related messages for 6 weeks, whereas the smoking cessation group received quitting smoking-related messages. Although the difference of the PSQI score between the smoking group (mean  $-1.47$ , SD  $1.77$ ) and sleep group (mean  $-1.7$ , SD  $1.48$ ) was not significant ( $P=.60$ ), the amount of sleep was significantly different ( $P=.03$ ). Specifically, the amount of sleep (hours) for the sleep/physical activity group increased (mean  $0.52$ , SD  $0.44$ ), whereas sleep among the smoking cessation group decreased (mean  $-0.03$ , SD  $0.46$ ).

### Effectiveness of Mobile Phone Apps as a Sleep Intervention Tool

As a measurement tool for sleep disorder intervention, mobile phone apps are able to perform diverse functions (ie, tracking [13,41,46,62], sleep advice for behavioral change [13,15,41,46],

and optimized alarms [13,22,62]). In our review, seven studies evaluated the effectiveness of these app-based interventions [13,15,22,41,46,62,64].

### Mobile Phone Apps as Auxiliary Equipment

Lichstein et al [15] used the Skype app for teleconferencing in addition to CBT as a sleep intervention. The purpose of the study was to examine the feasibility and effectiveness of interventions to address challenges faced by rural older adults with comorbid conditions, namely insomnia and depression. The researchers demonstrated the limitation of mobile phone apps for older population (age: mean  $65.8$ , SD  $10.4$  years). From the  $t$  test, participants in the Skype group had significant improved sleep quality (sleep quality:  $t_4=9.7$ ,  $P<.001$ ), less insomnia (ISI:  $t_4=-7.54$ ,  $P=.002$ ), less wakefulness after sleep onset ( $t_4=-5.77$ ,  $P=.004$ ), and lower HRSD scores ( $t_4=-4.33$ ,  $P=.01$ ). Despite the improvement in sleep quality, this study had a higher attrition rate; only five of 18 baseline participants completed the entire study (Multimedia Appendix 2).

Mendelson et al [64] evaluated whether a combination of the mobile phone app using telemedicine and CPAP had better performance compared to standard care (CPAP only). After 4 weeks of intervention, a total 82 of 107 baseline participants (76.6%) completed the entire test procedure, and 54 participants adhered to CPAP for more than 4 hours. Although the ESS was statistically significantly different between pretest and posttest (combined intervention: mean  $-2.3$ , SD  $4.0$ ,  $P<.05$ ; standard care: mean  $-2.1$ , SD  $4.1$ ,  $P<.001$ ), the effectiveness of treatment did not differ in each intervention.

### Mobile Phone Apps Developed by Researchers to Improve Sleep Behavior

Bauer et al [41] were the first to assess the use of mobile phone apps for sleep intervention programs. The research team developed the “Shuteye” app in 2011, which offered real-time recommendations to promote awareness related to healthy sleep behaviors. They also highlighted the following as being important for effective sleep apps: design of the app, adherence of mobile phone, appearance, awareness of app usage, and learnability (ie, user-friendly interface). For the intervention study, ESS was measured for 12 participants before and after the intervention. We calculated the changes in participant scores pre- and postintervention based on the raw data provided by the research team. Based on the data in the articles, we performed the paired  $t$  test, and ESS decreased after intervention (mean  $-1.67$ , SD  $1.32$ ;  $t_{11}=-4.36$ ,  $P=.001$ ), although this was not found to be statistically significant, most likely due to the small sample size.

Van Drongelen et al [22] conducted a RCT addressing sleep using a mobile phone app with international aircraft pilots. The “More Energy” app was invented to reduce pilots’ fatigue and also to improve their healthy behavior and sleep. During the 6 months of intervention, a total of 390 participants (77.7% of 502 baseline participants: 191 in the intervention group, 199 in waitlists) completed the study. The intervention program showed statistically significant fatigue improvement, and also showed a positive effect on sleep. We calculated the difference between the intervention group (score decreased by mean  $0.2$ , SD  $1.2$ )

and waitlists (score increased by mean 0.37, SD 1.32); the PSQI score significantly improved ( $\beta = -0.59$ ,  $P = .001$ ).

### Mobile Phone Apps With CBT-I

Three studies [13,46,62] used mobile phone apps to reform CBT-I, which is one of the traditional standard treatments for improving sleep behavior.

Babson et al [13] used a mobile phone app version of CBT-I to determine the feasibility and efficacy for veterans using cannabis to treat sleep disorders. A total of four participants were included in the study (two in the intervention group; two in the control group). According to the  $t$  tests result, the mobile phone intervention did not significantly affect the PSQI score ( $t_1 = -0.87$ ,  $P = .48$ ). However, this study was considered a pre-post study because all participants in the control group dropped out during the test procedure. In addition to the loss of a follow-up evaluation in the control group, several limitations with the study, such as small sample size and the rarity of the target population (ie, veterans using cannabis treatment for sleep disorders), caused the authors' skeptical conclusion of the ineffectiveness of mHealth intervention.

Chen et al [46] used the "Win-Win a Sleep" app, which was developed from CBT-I, and revealed effectiveness and feasibility of sleep interventions. In this case study, a participant who was female aged 63 years suffered from insomnia and used the app during a 5-week test period. Because the participant took hypnotic medications, researchers also considered the impact of decreasing the amount of drugs used to sleep treat disorders. The participant had better sleep satisfaction, from no preference (mean 1.5, SD 0.7) to good (mean 2.0, SD 0.0), based on the result of a Likert scale change (score range between 0="very bad" and 3="very good").

Koffel et al [62] recruited from the general population to conduct a RCT to determine the feasibility and acceptability of CBT-I coach apps. A total of 18 participants evaluated their ISI scores at each therapy session (five times), and also reported the adherence rate. We did not perform a  $t$  test because there was no distinguishable statistic of ISI included in the article. However, ISI scores tended to decrease based on figures that were analyzed by hierarchical linear models. There was a significant difference ( $t_{15} = 3.04$ ,  $P = .008$ ) in ISI scores between baseline and after the third session, even though the standard CBT-I group had lower ISI scores than the app users group.

### Lack of Data for Performing Meta-Analysis

Although meta-analysis is a great way to examine the effectiveness of interventions with a specific parameter, we did not have enough articles to conduct a meta-analysis. Sleep measurements were not included in all articles selected for the systematic literature review. Because more than four articles were required to perform a meta-analysis, according to Table 2, there were not enough articles when stratified by sleep measurement and study design. Also, a sample size of  $N=2$  or  $N=3$  is insufficient to perform the publication bias test (Begg rank correlation test or Egger regression test) and to determine the level of heterogeneity.

## Discussion

### Summary of Evidence

The purpose of our systematic review was to investigate whether mobile phones are a feasible and usable tool to improve sleep disorders and sleep quality in intervention studies. We also assumed that auxiliary mobile phone use helps to enhance the performance of existing intervention methods.

This review presented the articles based on two intervention methods. The first method was utilizing mobile phone apps and the second was conventional mobile phone support through telephone calls or text messages. Although many mobile phone apps were used independently, the conventional mobile phone supporting methods were (1) to use the mobile phone itself as only an intervention tool (eg teleconferencing or telecounseling), or (2) to combine with another treatment (eg, CBT-I, CBT-B, and CPAP) to enhance the effectiveness of sleep interventions. In our review, six studies were related to CBT-I [13,15,46,57,61,62] and three studies to CPAP [38,54,64]. This indicated that sleep intervention studies using mobile phones have been developed based on the standard treatment, which has already shown reliability [33,53,71-73].

Among the sleep outcome measurements, PSQI was the most frequently reported, and PSQI scores of all seven studies [13,21,22,55,57,61,70] were significantly decreased in the intervention group using mobile phones compared to standard treatment and a waitlist group. Participants in the intervention groups with mobile phones had a mean decrease in PSQI scores of 1.73 points compared to the waitlist group, and a mean decrease of 0.97 points compared to other treatment groups. Although the difference of PSQI was statistically significant, clinicians must consider the clinical significance of a one-point change in PSQI scores. According to the Buysse et al [74], a score of 5 of 21 is the cut-off point for poor sleep. Thus, the clinical significance of a one-point difference depends on the baseline score.

The ISI was the second most frequently used sleep outcome measurement tool, and all five articles using the measure [15,57,59,61,62] found a statistically significant decrease in ISI score in the mobile phone use intervention group compared to non-mobile phone use group. Participants in the intervention group with mobile phones had a mean decrease in ISI scores of 5.09 points compared to the waitlist groups [57,59,62], and a mean decrease of 3.2 points in ISI scores compared to other treatments [57,59,61]. Because a change of 7 points in the ISI is interpreted as a different level of insomnia, a 3- or 5-point decrease in ISI scores might be considered a critical improvement. Although ESS, WASO, SOL, and TST were included in at least four studies, there was at least more than one study that reported a lack of statistical significance of intervention effect [15,38,41,54,57,59,61]. Therefore, ISI and PSQI seem to be the most applicable measurement tools to investigate the effect of mobile phone intervention on sleep disorders.

In general, 87.5% (14/16) of the studies [15,21,22,38,41,54,55,57,59,61,62,64,67,70] reviewed supported

the evidence of capability and efficacy of mobile phone usage interventions. When mobile phones were used as auxiliary equipment, the intervention that applied telephone calls or mobile phone apps clearly demonstrated equal or enhanced efficacy for sleep quality and index score compared to the traditional sleep intervention [15,38,54,55,57,59,61,64,67,70]. This difference was seen not only for the waitlist control group [54,55,57,59], but also for the telephone-supported intervention and traditional treatment [15,38,61,64,67,70]. The advantage of mobile phone intervention was also illustrated in apps using intervention [13,22,41,46,62] or text message interventions [21].

### Limitations of Review

There are several limitations of the studies included in this review. First, there is no standardized study design, especially for the test period, procedure, and sleep intervention tools. For example, there is a limitation to use all mobile phone apps to compare each study because researchers used their own personally developed or nonpublicly available apps. As such, we were unable to make comparisons between interventions due to differences in app functions and interfaces. Due to the heterogeneity of the study design and sleep measurement tools, it was difficult to compare the effectiveness of the sleep app interventions. This will be an issue for reproducibility in further research. Second, although it is a common limitation for RCT study designs, the uniqueness of each study's target population limits further analysis and replication. For instance, results from studies with a cannabis disorder use group [13] and a post-deployment soldier with mild traumatic brain injury [70] were difficult to extrapolate to the general population. Third, there are few interventions using mobile technology that have been published in peer-reviewed journals, so we were limited with respect to the number of articles we could include in the study. Although many *t* test for each sleep measurement showed significant effectiveness of sleep interventions, the small number of studies for subgrouping limits performing a meta-analysis. Fourth, there is the possibility of missed articles. Because the mHealth market is growing rapidly, our search possibly missed some articles that used new mobile phone intervention tools.

On the other hand, it also might be possible to miss some articles during the search stage due to the usage of different language or jargon unique to the evolving mHealth industry.

### Conclusion

This study has several strengths. To the best of our knowledge, this study was the first to review the effectiveness of mobile phone interventions on sleep quality, quantity, and sleep disorders. By focusing exclusively on the mobile phone itself, it allows us to tailor future mHealth interventions for sleep. Also, our study examined various aspects of sleep measurement tools that account for sleep quality, sleep quantity, and many sleep disorders such as insomnia and sleep latency. Additionally, our study provided evidence of the potential of mobile-based interventions for improving sleep disorders. Along with current research that support the benefits of cost-efficiency of mHealth interventions, these findings provide an impetus for further research examining the empirical evidence of sleep interventions using mobile phones.

In conclusion, our systematic review supports the evidence that mobile technology-based interventions are an effective tool to improve symptoms of sleep disorders and quality of sleep than traditional intervention without mobile phone. Also, we suggest the following research design for future sleep intervention studies: (1) PSQI and ISI as sleep outcome measurements, (2) RCTs, (3) compare with standard treatment (ie, CPAP, CBT-I), and (4) compare to a waitlist control group. In addition to intervention methods, because mobile phone apps vary and many of these apps are not being studied, it is important to perform a content analysis on commercially available apps to determine common functionalities prior to undertaking interventions [75]. Our finding was not only applicable to those with sleep disorders who need clinical care, but also to medical professionals who are interested in ways to determine effective sleep interventions. Moving forward, app developers and sleep experts need to develop evidence-based guidelines with behavioral change components for sleep apps to maximize their efficacy and to take advantage of mobile phones to apply to existing standard treatments for sleep interventions.

### Acknowledgments

The authors would like to thank Dr Ruopeng An for his advice regarding systematic reviews.

### Authors' Contributions

JCS contributed to the concept and design for the current study, performed data analyses, and drafted the manuscript. JK participated in conception and design, data analysis of review procedure, and provided critical feedback on the draft of the manuscript. DGT provided guidance and critical feedback on the concept of the study and drafts of the manuscript. All authors contributed critical revision of the manuscript for important intellectual content and read and approved the final manuscript.

### Conflicts of Interest

None declared.

### Multimedia Appendix 1

Searching Keywords.

[PDF File (Adobe PDF File), 15KB - [mhealth\\_v5i9e131\\_app1.pdf](#)]



## Multimedia Appendix 2

Quality Assessment Score.

[PDF File (Adobe PDF File), 26KB - [mhealth\\_v5i9e131\\_app2.pdf](#)]

## References

1. Altevogt B, Colten H. Sleep Disorders and Sleep Deprivation: An Unmet Public Health Problem. Washington, DC: National Academies Press; 2006.
2. Cleveland Clinic. 2013. Common sleep disorders URL: <https://my.clevelandclinic.org/health/articles/common-sleep-disorders> [accessed 2017-02-18]
3. Gottlieb DJ, Yenokyan G, Newman AB, O'Connor GT, Punjabi NM, Quan SF, et al. Prospective study of obstructive sleep apnea and incident coronary heart disease and heart failure: the sleep heart health study. *Circulation* 2010 Jul 27;122(4):352-360 [FREE Full text] [doi: [10.1161/CIRCULATIONAHA.109.901801](https://doi.org/10.1161/CIRCULATIONAHA.109.901801)] [Medline: [20625114](#)]
4. Cappuccio FP, Taggart FM, Kandala N, Currie A, Peile E, Stranges S, et al. Meta-analysis of short sleep duration and obesity in children and adults. *Sleep* 2008 May;31(5):619-626 [FREE Full text] [Medline: [18517032](#)]
5. Gangwisch JE, Heymsfield SB, Boden-Albala B, Buijs RM, Kreier F, Pickering TG, et al. Sleep duration as a risk factor for diabetes incidence in a large U.S. sample. *Sleep* 2007 Dec;30(12):1667-1673 [FREE Full text] [Medline: [18246976](#)]
6. Macey PM, Woo MA, Kumar R, Cross RL, Harper RM. Relationship between obstructive sleep apnea severity and sleep, depression and anxiety symptoms in newly-diagnosed patients. *PLoS One* 2010 Apr 16;5(4):e10211 [FREE Full text] [doi: [10.1371/journal.pone.0010211](https://doi.org/10.1371/journal.pone.0010211)] [Medline: [20419135](#)]
7. Markwald RR, Melanson EL, Smith MR, Higgins J, Perreault L, Eckel RH, et al. Impact of insufficient sleep on total daily energy expenditure, food intake, and weight gain. *Proc Natl Acad Sci U S A* 2013 Apr 02;110(14):5695-5700 [FREE Full text] [doi: [10.1073/pnas.1216951110](https://doi.org/10.1073/pnas.1216951110)] [Medline: [23479616](#)]
8. Möller-Levet CS, Archer S, Bucca G, Laing EE, Slak A, Kabiljo R, et al. Effects of insufficient sleep on circadian rhythmicity and expression amplitude of the human blood transcriptome. *Proc Natl Acad Sci U S A* 2013 Mar 19;110(12):E1132-E1141 [FREE Full text] [doi: [10.1073/pnas.1217154110](https://doi.org/10.1073/pnas.1217154110)] [Medline: [23440187](#)]
9. Gottlieb DJ, Redline S, Nieto FJ, Baldwin CM, Newman AB, Resnick HE, et al. Association of usual sleep duration with hypertension: The sleep heart health study. *Sleep* 2006 Aug;29(8):1009-1014. [Medline: [16944668](#)]
10. Chaput J, Després JP, Bouchard C, Astrup A, Tremblay A. Sleep duration as a risk factor for the development of type 2 diabetes or impaired glucose tolerance: analyses of the Quebec Family Study. *Sleep Med* 2009 Sep;10(8):919-924. [doi: [10.1016/j.sleep.2008.09.016](https://doi.org/10.1016/j.sleep.2008.09.016)] [Medline: [19332380](#)]
11. National Center for Health Statistics. Table SLP-2. Adults aged 18 and over who met the Healthy People 2020 objective for sufficient sleep, by sex and selected characteristics: United States, annualized, 2011-2014. 2016. URL: [https://ftp.cdc.gov/pub/Health\\_Statistics/NCHS/NHIS/SHS/2011-2014\\_AHB\\_Table\\_SLP-2.pdf](https://ftp.cdc.gov/pub/Health_Statistics/NCHS/NHIS/SHS/2011-2014_AHB_Table_SLP-2.pdf) [accessed 2017-07-26] [WebCite Cache ID 6sGCJt8Y6]
12. National Sleep Foundation. 2011 Mar 7. Annual Sleep in America poll exploring connections with communications technology use and sleep URL: <https://sleepfoundation.org/media-center/press-release/annual-sleep-america-poll-exploring-connections-communications-technology-use-> [accessed 2017-07-27] [WebCite Cache ID 6sGD0sPz3]
13. Babson K, Ramo D, Baldini L, Vandrey R, Bonn-Miller M. Mobile app-delivered cognitive behavioral therapy for insomnia: feasibility and initial efficacy among veterans with cannabis use disorders. *JMIR Res Protoc* 2015 Jul 17;4(3):e87 [FREE Full text] [doi: [10.2196/resprot.3852](https://doi.org/10.2196/resprot.3852)] [Medline: [26187404](#)]
14. Holmqvist M, Vincent N, Walsh K. Web- vs. telehealth-based delivery of cognitive behavioral therapy for insomnia: a randomized controlled trial. *Sleep Med* 2014 Feb;15(2):187-195. [doi: [10.1016/j.sleep.2013.10.013](https://doi.org/10.1016/j.sleep.2013.10.013)] [Medline: [24461370](#)]
15. Lichstein K, Scogin F, Thomas S, DiNapoli E, Dillon H, McFadden A. Telehealth cognitive behavior therapy for co-occurring insomnia and depression symptoms in older adults. *J Clin Psychol* 2013 Oct;69(10):1056-1065 [FREE Full text] [doi: [10.1002/jclp.22030](https://doi.org/10.1002/jclp.22030)] [Medline: [24014056](#)]
16. Munafo D, Hevener W, Crocker M, Willes L, Sridasome S, Muhsin M. A telehealth program for CPAP adherence reduces labor and yields similar adherence and efficacy when compared to standard of care. *Sleep Breath* 2016 May;20(2):777-785 [FREE Full text] [doi: [10.1007/s11325-015-1298-4](https://doi.org/10.1007/s11325-015-1298-4)] [Medline: [26754933](#)]
17. Sparrow D, Aloia M, Demolles D, Gottlieb D. A telemedicine intervention to improve adherence to continuous positive airway pressure: a randomised controlled trial. *Thorax* 2010 Dec;65(12):1061-1066. [doi: [10.1136/thx.2009.133215](https://doi.org/10.1136/thx.2009.133215)] [Medline: [20880872](#)]
18. Nicholas J, Larsen M, Proudfoot J, Christensen H. Mobile apps for bipolar disorder: a systematic review of features and content quality. *J Med Internet Res* 2015 Aug 17;17(8):e198 [FREE Full text] [doi: [10.2196/jmir.4581](https://doi.org/10.2196/jmir.4581)] [Medline: [26283290](#)]
19. Bakker D, Kazantzis N, Rickwood D, Rickard N. Mental health smartphone apps: review and evidence-based recommendations for future developments. *JMIR Ment Health* 2016 Mar 01;3(1):e7 [FREE Full text] [doi: [10.2196/mental.4984](https://doi.org/10.2196/mental.4984)] [Medline: [26932350](#)]

20. American Telemedicine Association Pediatric Special Interest Group. American Academy of Pediatrics. What is telehealth? URL: <https://www.aap.org/en-us/professional-resources/practice-transformation/telehealth/Pages/What-is-Telehealth.aspx> [accessed 2017-07-27] [WebCite Cache ID 6sGDYVrTX]
21. Filion A, Darlington G, Chaput J, Ybarra M, Haines J. Examining the influence of a text message-based sleep and physical activity intervention among young adult smokers in the United States. *BMC Public Health* 2015 Jul 16;15:671 [FREE Full text] [doi: [10.1186/s12889-015-2045-2](https://doi.org/10.1186/s12889-015-2045-2)] [Medline: [26178640](https://pubmed.ncbi.nlm.nih.gov/26178640/)]
22. van Drongelen A, Boot CR, Hlobil H, Twisk JW, Smid T, van der Beek AJ. Evaluation of an mHealth intervention aiming to improve health-related behavior and sleep and reduce fatigue among airline pilots. *Scand J Work Environ Health* 2014 Nov;40(6):557-568 [FREE Full text] [doi: [10.5271/sjweh.3447](https://doi.org/10.5271/sjweh.3447)] [Medline: [25121620](https://pubmed.ncbi.nlm.nih.gov/25121620/)]
23. Bhat S, Ferraris A, Gupta D, Mozafarian M, DeBari VA, Gushway-Henry N, et al. Is there a clinical role for smartphone sleep apps? comparison of sleep cycle detection by a smartphone application to polysomnography. *J Clin Sleep Med* 2015 Jul 15;11(7):709-715 [FREE Full text] [doi: [10.5664/jcsm.4840](https://doi.org/10.5664/jcsm.4840)] [Medline: [25766719](https://pubmed.ncbi.nlm.nih.gov/25766719/)]
24. Stippig A, Hübers U, Emerich M. Apps in sleep medicine. *Sleep Breath* 2015 Mar;19(1):411-417. [doi: [10.1007/s11325-014-1009-6](https://doi.org/10.1007/s11325-014-1009-6)] [Medline: [24888483](https://pubmed.ncbi.nlm.nih.gov/24888483/)]
25. Krisch JA. The New York Times. 2015 Mar 16. Questioning the value of health apps URL: [https://well.blogs.nytimes.com/2015/03/16/health-apps-provide-pictures-if-not-proof-of-health/?\\_r=1](https://well.blogs.nytimes.com/2015/03/16/health-apps-provide-pictures-if-not-proof-of-health/?_r=1) [accessed 2017-07-27] [WebCite Cache ID 6sGDx081D]
26. Research 2 Guidance. mHealth App Developer Economics 2016. 2016 Oct. URL: <https://research2guidance.com/r2g/r2g-mHealth-App-Developer-Economics-2016.pdf> [accessed 2017-07-27] [WebCite Cache ID 6sGEeCCym]
27. Behar J, Roebuck A, Domingos J, Geder E, Clifford G. A review of current sleep screening applications for smartphones. *Physiol Meas* 2013 Jul;34(7):R29-R46. [doi: [10.1088/0967-3334/34/7/R29](https://doi.org/10.1088/0967-3334/34/7/R29)] [Medline: [23771214](https://pubmed.ncbi.nlm.nih.gov/23771214/)]
28. Ko P, Kientz J, Choe E, Kay M, Landis C, Watson N. Consumer sleep technologies: a review of the landscape. *J Clin Sleep Med* 2015 Dec 15;11(12):1455-1461 [FREE Full text] [doi: [10.5664/jcsm.5288](https://doi.org/10.5664/jcsm.5288)] [Medline: [26156958](https://pubmed.ncbi.nlm.nih.gov/26156958/)]
29. Kolla B, Mansukhani S, Mansukhani M. Consumer sleep tracking devices: a review of mechanisms, validity and utility. *Expert Rev Med Devices* 2016 May;13(5):497-506. [doi: [10.1586/17434440.2016.1171708](https://doi.org/10.1586/17434440.2016.1171708)] [Medline: [27043070](https://pubmed.ncbi.nlm.nih.gov/27043070/)]
30. Lee J, Finkelstein J. Consumer sleep tracking devices: a critical review. *Stud Health Technol Inform* 2015;210:458-460. [Medline: [25991187](https://pubmed.ncbi.nlm.nih.gov/25991187/)]
31. Zia S, Fields B. Sleep telemedicine: an emerging field's latest frontier. *Chest* 2016 Jun;149(6):1556-1565. [doi: [10.1016/j.chest.2016.02.670](https://doi.org/10.1016/j.chest.2016.02.670)] [Medline: [26970035](https://pubmed.ncbi.nlm.nih.gov/26970035/)]
32. Donker T, Petrie K, Proudfoot J, Clarke J, Birch M, Christensen H. Smartphones for smarter delivery of mental health programs: a systematic review. *J Med Internet Res* 2013 Nov 15;15(11):e247 [FREE Full text] [doi: [10.2196/jmir.2791](https://doi.org/10.2196/jmir.2791)] [Medline: [24240579](https://pubmed.ncbi.nlm.nih.gov/24240579/)]
33. Sawyer AM, Gooneratne NS, Marcus CL, Ofer D, Richards KC, Weaver TE. A systematic review of CPAP adherence across age groups: clinical and empiric insights for developing CPAP adherence interventions. *Sleep Med Rev* 2011 Dec;15(6):343-356 [FREE Full text] [doi: [10.1016/j.smrv.2011.01.003](https://doi.org/10.1016/j.smrv.2011.01.003)] [Medline: [21652236](https://pubmed.ncbi.nlm.nih.gov/21652236/)]
34. Verbraecken J. Applications of evolving technologies in sleep medicine. *Breathe* 2013 Dec 01;9(6):442-455. [doi: [10.1183/20734735.012213](https://doi.org/10.1183/20734735.012213)]
35. Anjarwalla T. CNN. 2010 Jul 9. Inventor of cell phone: we knew someday everybody would have one URL: <http://www.cnn.com/2010/TECH/mobile/07/09/cooper.cell.phone.inventor/index.html> [accessed 2017-07-27] [WebCite Cache ID 6sGFJWtMx]
36. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med* 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: [10.1371/journal.pmed.1000097](https://doi.org/10.1371/journal.pmed.1000097)] [Medline: [19621072](https://pubmed.ncbi.nlm.nih.gov/19621072/)]
37. Zhu H, An R. Impact of home-delivered meal programs on diet and nutrition among older adults: a review. *Nutr Health* 2013 Apr;22(2):89-103. [doi: [10.1177/0260106014537146](https://doi.org/10.1177/0260106014537146)] [Medline: [24916974](https://pubmed.ncbi.nlm.nih.gov/24916974/)]
38. Anttalainen U, Melkko S, Hakko S, Laitinen T, Saaresranta T. Telemonitoring of CPAP therapy may save nursing time. *Sleep Breath* 2016 Dec;20(4):1209-1215. [doi: [10.1007/s11325-016-1337-9](https://doi.org/10.1007/s11325-016-1337-9)] [Medline: [27043327](https://pubmed.ncbi.nlm.nih.gov/27043327/)]
39. Arora T, Broglia E, Thomas GN, Taheri S. Associations between specific technologies and adolescent sleep quantity, sleep quality, and parasomnias. *Sleep Med* 2014 Feb;15(2):240-247. [doi: [10.1016/j.sleep.2013.08.799](https://doi.org/10.1016/j.sleep.2013.08.799)] [Medline: [24394730](https://pubmed.ncbi.nlm.nih.gov/24394730/)]
40. Baig M, Antonescu-Turcu A, Ratarasarn K. Impact of sleep telemedicine protocol in management of sleep apnea: a 5-year VA experience. *Telemed J E Health* 2016 May;22(5):458-462. [doi: [10.1089/tmj.2015.0047](https://doi.org/10.1089/tmj.2015.0047)] [Medline: [26974884](https://pubmed.ncbi.nlm.nih.gov/26974884/)]
41. Bauer JS, Consolvo S, Greenstein B, Schooler J, Wu E, Watson NF, et al. ShutEye: encouraging awareness of healthy sleep recommendations with a mobile, peripheral display. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2012 May 5 Presented at: CHI '12: SIGCHI Conference on Human Factors in Computing Systems; May 5-10, 2012; Austin, TX p. 1401-1410.
42. Blom K, Tarkian TH, Wiklund T, Danlycke E, Forssén M, Söderström A, et al. Internet-vs. group-delivered cognitive behavior therapy for insomnia: a randomized controlled non-inferiority trial. *Behav Res Ther* 2015 Jul;70:47-55 [FREE Full text] [doi: [10.1016/j.brat.2015.05.002](https://doi.org/10.1016/j.brat.2015.05.002)] [Medline: [25981329](https://pubmed.ncbi.nlm.nih.gov/25981329/)]



43. Buman MP, Epstein DR, Gutierrez M, Herb C, Hollingshead K, Huberty JL, et al. BeWell24: development and process evaluation of a smartphone “app” to improve sleep, sedentary, and active behaviors in US Veterans with increased metabolic risk. *Transl Behav Med* 2016 Sep;6(3):438-448 [[FREE Full text](#)] [doi: [10.1007/s13142-015-0359-3](https://doi.org/10.1007/s13142-015-0359-3)] [Medline: [27528532](#)]
44. Camacho M, Chang ET, Fernandez-Salvador C, Capasso R. Treatment of snoring with a nasopharyngeal airway tube. *Case Rep Med* 2016;2016:3628716 [[FREE Full text](#)] [doi: [10.1155/2016/3628716](https://doi.org/10.1155/2016/3628716)] [Medline: [27795710](#)]
45. Cavallera C, Pagnini F, Rovaris M, Mendozzi L, Pugnetti L, Garegnani M, et al. A telemedicine meditation intervention for people with multiple sclerosis and their caregivers: study protocol for a randomized controlled trial. *Trials* 2016 Jan 04;17:4 [[FREE Full text](#)] [doi: [10.1186/s13063-015-1136-9](https://doi.org/10.1186/s13063-015-1136-9)] [Medline: [26729330](#)]
46. Chen Y, Hung Y, Chen H. Mobile application-assisted cognitive behavioral therapy for insomnia in an older adult. *Telemed J E Health* 2016 Apr;22(4):332-334. [doi: [10.1089/tmj.2015.0064](https://doi.org/10.1089/tmj.2015.0064)] [Medline: [26393902](#)]
47. Cho YM, Lim HJ, Jang H, Kim K, Choi JW, Shin C, et al. A follow-up study of the association between mobile phone use and symptoms of ill health. *Environ Health Toxicol* 2016;32:e2017001 [[FREE Full text](#)] [doi: [10.5620/eh.t.2017001](https://doi.org/10.5620/eh.t.2017001)] [Medline: [28111420](#)]
48. Craven O, Hughes CA, Burton A, Saunders MP, Molassiotis A. Is a nurse-led telephone intervention a viable alternative to nurse-led home care and standard care for patients receiving oral capecitabine? Results from a large prospective audit in patients with colorectal cancer. *Eur J Cancer Care (Engl)* 2013 May;22(3):413-419. [doi: [10.1111/ecc.12047](https://doi.org/10.1111/ecc.12047)] [Medline: [23527965](#)]
49. Cvetengros JA. One size does not fit all: matching patients with insomnia treatment modality. *Sleep Med* 2014 Feb;15(2):161-162. [doi: [10.1016/j.sleep.2013.12.002](https://doi.org/10.1016/j.sleep.2013.12.002)] [Medline: [24434167](#)]
50. van Drongelen A, van der Beek AJ, Hlobil H, Smid T, Boot C. Development and evaluation of an intervention aiming to reduce fatigue in airline pilots: design of a randomised controlled trial. *BMC Public Health* 2013 Aug 26;13:776 [[FREE Full text](#)] [doi: [10.1186/1471-2458-13-776](https://doi.org/10.1186/1471-2458-13-776)] [Medline: [23971514](#)]
51. Duncan M, Vandelanotte C, Trost S, Rebar AL, Rogers N, Burton NW, et al. Balanced: a randomised trial examining the efficacy of two self-monitoring methods for an app-based multi-behaviour intervention to improve physical activity, sitting and sleep in adults. *BMC Public Health* 2016 Jul 30;16:670 [[FREE Full text](#)] [doi: [10.1186/s12889-016-3256-x](https://doi.org/10.1186/s12889-016-3256-x)] [Medline: [27473327](#)]
52. Eaton LH, Gordon DB, Wyant S, Theodore BR, Meins AR, Rue T, et al. Development and implementation of a telehealth-enhanced intervention for pain and symptom management. *Contemp Clin Trials* 2014 Jul;38(2):213-220 [[FREE Full text](#)] [doi: [10.1016/j.cct.2014.05.005](https://doi.org/10.1016/j.cct.2014.05.005)] [Medline: [24846620](#)]
53. Espie C, Hames P, McKinstry B. Use of the Internet and mobile media for delivery of cognitive behavioral insomnia therapy. *Sleep Medicine Clinics* 2013 Sep;8(3):407-419. [doi: [10.1016/j.jsmc.2013.06.001](https://doi.org/10.1016/j.jsmc.2013.06.001)]
54. Fox N, Hirsch-Allen A, Goodfellow E, Wenner J, Fleetham J, Ryan CF, et al. The impact of a telemedicine monitoring system on positive airway pressure adherence in patients with obstructive sleep apnea: a randomized controlled trial. *Sleep* 2012 Apr 01;35(4):477-481 [[FREE Full text](#)] [doi: [10.5665/sleep.1728](https://doi.org/10.5665/sleep.1728)] [Medline: [22467985](#)]
55. Freeman LW, White R, Ratcliff CG, Sutton S, Stewart M, Palmer JL, et al. A randomized trial comparing live and telemedicine deliveries of an imagery-based behavioral intervention for breast cancer survivors: reducing symptoms and barriers to care. *Psychooncology* 2015 Aug;24(8):910-918 [[FREE Full text](#)] [doi: [10.1002/pon.3656](https://doi.org/10.1002/pon.3656)] [Medline: [25146413](#)]
56. Gehrman P, Shah MT, Miles A, Kuna S, Godleski L. Feasibility of group cognitive-behavioral treatment of insomnia delivered by clinical video telehealth. *Telemed J E Health* 2016 Dec;22(12):1041-1046. [doi: [10.1089/tmj.2016.0032](https://doi.org/10.1089/tmj.2016.0032)] [Medline: [27286369](#)]
57. Ho FY, Chung K, Yeung W, Ng TH, Cheng SK. Weekly brief phone support in self-help cognitive behavioral therapy for insomnia disorder: Relevance to adherence and efficacy. *Behav Res Ther* 2014 Dec;63:147-156. [doi: [10.1016/j.brat.2014.10.002](https://doi.org/10.1016/j.brat.2014.10.002)] [Medline: [25461790](#)]
58. Isetta V, Negrín MA, Monasterio C, Masa JF, Feu N, Álvarez A, Spanish Sleep Network. A Bayesian cost-effectiveness analysis of a telemedicine-based strategy for the management of sleep apnoea: a multicentre randomised controlled trial. *Thorax* 2015 Nov;70(11):1054-1061. [doi: [10.1136/thoraxjnl-2015-207032](https://doi.org/10.1136/thoraxjnl-2015-207032)] [Medline: [26310452](#)]
59. Jernelöv S, Lekander M, Blom K, Rydh S, Ljótsson B, Axelsson J, et al. Efficacy of a behavioral self-help treatment with or without therapist guidance for co-morbid and primary insomnia--a randomized controlled trial. *BMC Psychiatry* 2012 Jan 22;12:5 [[FREE Full text](#)] [doi: [10.1186/1471-244X-12-5](https://doi.org/10.1186/1471-244X-12-5)] [Medline: [22264332](#)]
60. Filion AJ, Darlington G, Chaput JP, Ybarra M, Haines J. Examining the influence of a text message-based sleep and physical activity intervention among young adult smokers in the United States. *BMC Public Health* 2015 Jul 16;15:671 [[FREE Full text](#)] [doi: [10.1186/s12889-015-2045-2](https://doi.org/10.1186/s12889-015-2045-2)] [Medline: [26178640](#)]
61. Kauffman R. Telephone-based CBT reduced insomnia severity more than menopause education in menopausal women. *Ann Intern Med* 2016 Sep 20;165(6):JC30. [doi: [10.7326/ACPJC-2016-165-6-030](https://doi.org/10.7326/ACPJC-2016-165-6-030)] [Medline: [27653727](#)]
62. Koffel E, Kuhn E, Petsoulis N, Erbes CR, Anders S, Hoffman JE, et al. A randomized controlled pilot study of CBT-I Coach: feasibility, acceptability, and potential impact of a mobile phone application for patients in cognitive behavioral therapy for insomnia. *Health Informatics J* 2016 Jun 27;1460458216656472. [doi: [10.1177/1460458216656472](https://doi.org/10.1177/1460458216656472)] [Medline: [27354394](#)]

63. Mason EC, Harvey AG. Insomnia before and after treatment for anxiety and depression. *J Affect Disord* 2014 Oct;168:415-421. [doi: [10.1016/j.jad.2014.07.020](https://doi.org/10.1016/j.jad.2014.07.020)] [Medline: [25108278](#)]
64. Mendelson M, Vivodtzev I, Tamišier R, Laplaud D, Dias-Domingos S, Baguet JP, et al. CPAP treatment supported by telemedicine does not improve blood pressure in high cardiovascular risk OSA patients: a randomized, controlled trial. *Sleep* 2014 Nov 01;37(11):1863-1870 [FREE Full text] [doi: [10.5665/sleep.4186](https://doi.org/10.5665/sleep.4186)] [Medline: [25364081](#)]
65. Salaffi F, Ciapetti A, Gasparini S, Atzeni F, Sarzi-Puttini P, Baroni M. Web/Internet-based telemonitoring of a randomized controlled trial evaluating the time-integrated effects of a 24-week multicomponent intervention on key health outcomes in patients with fibromyalgia. *Clin Exp Rheumatol* 2015;33(1 Suppl 88):S93-S101. [Medline: [25786050](#)]
66. Sedkaoui K, Leseux L, Pontier S, Rossin N, Leophonte P, Fraysse JL, et al. Efficiency of a phone coaching program on adherence to continuous positive airway pressure in sleep apnea hypopnea syndrome: a randomized trial. *BMC Pulm Med* 2015 Sep 14;15:102 [FREE Full text] [doi: [10.1186/s12890-015-0099-7](https://doi.org/10.1186/s12890-015-0099-7)] [Medline: [26370444](#)]
67. Stremler R, Hodnett E, Lee K, MacMillan S, Mill C, Ongcango L, et al. A behavioral-educational intervention to promote maternal and infant sleep: a pilot randomized, controlled trial. *Sleep* 2006 Dec;29(12):1609-1615. [Medline: [17252892](#)]
68. Taylor Y, Eliasson A, Andrada T, Kristo D, Howard R. The role of telemedicine in CPAP compliance for patients with obstructive sleep apnea syndrome. *Sleep Breath* 2006 Sep;10(3):132-138. [doi: [10.1007/s11325-006-0059-9](https://doi.org/10.1007/s11325-006-0059-9)] [Medline: [16565867](#)]
69. Van Devender M, Shang Y. Wake IQ: using a smartphone to reduce sleep inertia. 2013 Jan Presented at: IEEE Consumer Communications and Networking Conference; Jan 11-14, 2013; Las Vegas, NV.
70. Vuletic S, Bell KR, Jain S, Bush N, Temkin N, Fann JR, CONTACT Investigators. Telephone problem-solving treatment improves sleep quality in service members with combat-related mild traumatic brain injury: results from a randomized clinical trial. *J Head Trauma Rehabil* 2016;31(2):147-157. [doi: [10.1097/HTR.0000000000000221](https://doi.org/10.1097/HTR.0000000000000221)] [Medline: [26959668](#)]
71. Coma-Del-Corral MJ, Alonso-Álvarez ML, Allende M, Cordero J, Ordax E, Masa F, et al. Reliability of telemedicine in the diagnosis and treatment of sleep apnea syndrome. *Telemed J E Health* 2013 Jan;19(1):7-12 [FREE Full text] [doi: [10.1089/tmj.2012.0007](https://doi.org/10.1089/tmj.2012.0007)] [Medline: [23186084](#)]
72. Bang M, Timpka T, Eriksson H, Holm E, Nordin C. Mobile phone computing for in-situ cognitive behavioral therapy. *Stud Health Technol Inform* 2007;129(Pt 2):1078-1082. [Medline: [17911881](#)]
73. Kuhn E, Weiss B, Taylor K, Hoffman JE, Ramsey KM, Manber R, et al. CBT-I Coach: a description and clinician perceptions of a mobile app for cognitive behavioral therapy for insomnia. *J Clin Sleep Med* 2016 Apr 15;12(4):597-606 [FREE Full text] [doi: [10.5664/jcsm.5700](https://doi.org/10.5664/jcsm.5700)] [Medline: [26888586](#)]
74. Buysse D, Reynolds C, Monk T, Berman S, Kupfer D. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. *Psychiatry Res* 1989 May;28(2):193-213. [Medline: [2748771](#)]
75. Grigsby-Toussaint DS, Shin JC, Reeves DM, Beattie A, Auguste E, Jean-Louis G. Sleep apps and behavioral constructs: a content analysis. *Prev Med Rep* 2017 Jun;6:126-129. [doi: [10.1016/j.pmedr.2017.02.018](https://doi.org/10.1016/j.pmedr.2017.02.018)] [Medline: [28316907](#)]

## Abbreviations

**AHI:** Apnea-Hypopnea Index  
**BTS:** bed time stress  
**CINAHL:** Cumulative Index to Nursing and Allied Health Literature  
**CPAP:** continuous positive airway pressure  
**DBAS:** Dysfunctional Beliefs and Attitudes about Sleep Scale  
**ESS:** Epworth Sleepiness Scale  
**GSDS:** General Sleep Disturbance Scale  
**HRSD:** Hamilton Rating Scale for Depression with sleep  
**ISI:** Insomnia Severity Index  
**OSA:** obstructive sleep apnea  
**PRISMA:** Preferred Reporting Items for Systematic Reviews and Meta-Analysis  
**PSQI:** Pittsburgh Sleep Quality Index  
**RCT:** randomized controlled trial  
**SE:** sleep efficiency  
**SOL:** sleep-onset latency  
**SQ:** sleep quality  
**SRBQ:** Sleep-Related Behavior Questionnaire  
**SSR:** Sleep Satisfaction Rate  
**TST:** total sleep time  
**WASO:** wakefulness after initial sleep onset

*Edited by G Eysenbach; submitted 31.12.16; peer-reviewed by YS Bin, M Zhang, J Apolinário-Hagen, J Dudeney, Y Liu; comments to author 11.02.17; revised version received 18.04.17; accepted 17.06.17; published 07.09.17.*

*Please cite as:*

*Shin JC, Kim J, Grigsby-Toussaint D*

*Mobile Phone Interventions for Sleep Disorders and Sleep Quality: Systematic Review*

*JMIR Mhealth Uhealth 2017;5(9):e131*

URL: <http://mhealth.jmir.org/2017/9/e131/>

doi: [10.2196/mhealth.7244](https://doi.org/10.2196/mhealth.7244)

PMID: [28882808](https://pubmed.ncbi.nlm.nih.gov/28882808/)

©Jong Cheol Shin, Julia Kim, Diana Grigsby-Toussaint. Originally published in JMIR Mhealth and Uhealth (<http://mhealth.jmir.org>), 07.09.2017. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mhealth and uhealth, is properly cited. The complete bibliographic information, a link to the original publication on <http://mhealth.jmir.org/>, as well as this copyright and license information must be included.

Original Paper

# Effectiveness of SmartMoms, a Novel eHealth Intervention for Management of Gestational Weight Gain: Randomized Controlled Pilot Trial

Leanne M Redman<sup>1</sup>, MSc, PhD; L. Anne Gilmore<sup>1</sup>, PhD; Jeffrey Breaux<sup>2</sup>, MD; Diana M Thomas<sup>3</sup>, PhD; Karen Elkind-Hirsch<sup>2</sup>, MSc, PhD; Tiffany Stewart<sup>1</sup>, PhD; Daniel S Hsia<sup>1</sup>, MD; Jeffrey Burton<sup>1</sup>, MSc, PhD; John W Apolzan<sup>1</sup>, PhD; Loren E Cain<sup>1</sup>, MS; Abby D Altazan<sup>1</sup>, MS; Shelly Ragusa<sup>1</sup>, MS, RD, CDE; Heather Brady<sup>1</sup>, MS; Allison Davis<sup>1</sup>, MS; J. Mick Tilford<sup>4</sup>; Elizabeth F Sutton<sup>1</sup>, PhD; Corby K Martin<sup>1</sup>, PhD

<sup>1</sup>Pennington Biomedical Research Center, Baton Rouge, LA, United States

<sup>2</sup>Woman's Hospital, Baton Rouge, LA, United States

<sup>3</sup>United States Military Academy, West Point, NY, United States

<sup>4</sup>University of Arkansas for Medical Sciences, Little Rock, AR, United States

**Corresponding Author:**

Leanne M Redman, MSc, PhD

Pennington Biomedical Research Center

6400 Perkins Road

Baton Rouge, LA, 70808

United States

Phone: 1 2257630947

Fax: 1 763 3030

Email: [leanne.redman@pbrc.edu](mailto:leanne.redman@pbrc.edu)

## Abstract

**Background:** Two-thirds of pregnant women exceed gestational weight gain (GWG) recommendations. Because excess GWG is associated with adverse outcomes for mother and child, development of scalable and cost-effective approaches to deliver intensive lifestyle programs during pregnancy is urgent.

**Objective:** The aim of this study was to decrease the proportion of women who exceed the Institute of Medicine (IOM) 2009 GWG guidelines.

**Methods:** In a parallel-arm randomized controlled trial, 54 pregnant women (age 18–40 years) who were overweight (n=25) or obese (n=29) were enrolled to test whether an intensive lifestyle intervention (called SmartMoms) decreased the proportion of women with excess GWG, defined as exceeding the 2009 IOM guidelines, compared to no intervention (usual care group). The SmartMoms intervention was delivered through mobile phone (remote group) or in a traditional in-person, clinic-based setting (in-person group), and included a personalized dietary intake prescription, self-monitoring weight against a personalized weight graph, activity tracking with a pedometer, receipt of health information, and continuous personalized feedback from counselors.

**Results:** A significantly smaller proportion of women exceeded the IOM 2009 GWG guidelines in the SmartMoms intervention groups (in-person: 56%, 10/18; remote: 58%, 11/19) compared to usual care (85%, 11/13;  $P=.02$ ). The remote intervention was a lower cost to participants (mean US \$97, SD \$6 vs mean US \$347, SD \$40 per participant;  $P<.001$ ) and clinics (US \$215 vs US \$419 per participant) and with increased intervention adherence (76.5% vs 60.8%;  $P=.049$ ).

**Conclusions:** An intensive lifestyle intervention for GWG can be effectively delivered via a mobile phone, which is both cost-effective and scalable.

**Trial Registration:** Clinicaltrials.gov NCT01610752; <https://clinicaltrials.gov/ct2/show/NCT01610752> (Archived by WebCite at <http://www.webcitation.org/6sarNB4iW>)

(JMIR Mhealth Uhealth 2017;5(9):e133) doi:[10.2196/mhealth.8228](https://doi.org/10.2196/mhealth.8228)

**KEYWORDS**

pregnancy; gestational weight gain; lifestyle modification; intervention

## Introduction

More than two-thirds of pregnant women exceed the Institute of Medicine (IOM) 2009 gestational weight gain (GWG) recommendations [1]. According to weight management guidelines for obesity treatment [2], prenatal care should provide an ideal clinical framework for treatment delivery with frequent visits, weight recording, an established definition of acceptable weight gain, and opportunities for in-person counseling. However, despite several efforts to prevent excessive GWG in clinical trials, it remains unclear if lifestyle interventions can be efficacious, particularly in women with overweight or obesity [3,4]. Poor effectiveness in these trials is explained by intervention designs that fail to take advantage of the entire prenatal care continuum because program initiation is often delayed until mid or late gestation and weight management counseling and intervention are limited to one or two in-person sessions [3]. As more patients have access to mobile phones and 67% of pregnant women subscribe to electronic health information delivery during pregnancy [5], eHealth interventions designed to target healthy weight gain provide an opportunity for high-intensity and cost-effective interventions to be delivered to all patients throughout prenatal care. The aim of this study was to test whether a personalized gestational weight management program (SmartMoms) delivered in-person or via an intensity-matched mobile phone app could decrease the proportion of women with overweight and obesity that exceed the IOM 2009 guidelines for GWG by 25%.

## Methods

This study targeted overweight and obese (body mass index [BMI] 25.0–39.9 kg/m<sup>2</sup>) women aged 18 to 40 years expecting a singleton pregnancy in their first trimester. Women with a known fetal anomaly, hypertension (systolic >160 mm Hg or diastolic >90 mm Hg), history of or current psychotic or eating disorder, human immunodeficiency virus, preexisting diabetes (self-report or determined by glycated hemoglobin A<sub>1c</sub> and/or 75 g oral glucose tolerance test in the first trimester), or with contraindications to exercise (by PARmed-X and American College of Obstetricians and Gynecologists committee opinion #67 [6]) were excluded. With support of local obstetricians, participants were recruited from brochures placed in various clinics and by study staff during the patients' first prenatal appointment [7]. Participants were randomized by unblinded intervention staff equally to one of three groups between 10.4 to 13.6 weeks of gestation: (1) no intervention (usual care group), (2) receipt of the SmartMoms intervention in-person (in-person group), or (3) receipt of the SmartMoms intervention via mobile phone (remote group), with randomization stratified by pregravid BMI. The block randomization schedule and sealed numbered randomization envelopes were prepared by the biostatistician. Usual care (control) participants were under the usual care of their obstetrician and did not receive weight management services from the intervention team. The SmartMoms intervention was designed to assist an expectant mother in gaining weight within the recommended 2009 IOM guidelines for her respective BMI class. SmartMoms is grounded in the ability to objectively quantify dietary adherence to an

energy intake prescription based on measured body weight and to provide patients with data-driven feedback about their energy intake [8–12]. SmartMoms participants received dietary intake advice, exercise advice, and a weight graph created from the dynamic GWG models to determine the trimester-specific increase in energy intake required by each participant to adhere to the IOM GWG recommendations [13]. To promote these lifestyle changes, participants received a structured intervention that consisted of 18 lessons and behavior modification strategies. SmartMoms participants received behavior modification counseling weekly between weeks 13 and 24 of gestation and biweekly from week 25 until delivery. Importantly, the content of the lesson materials were identical and only the mode of delivery differed between the two intervention groups. Participants in both the intervention groups were provided with a wireless Internet-connected bathroom scale and a pedometer (in-person group: Omron Healthcare, Lake Forest, IL, USA; remote group: Fitbit Zip, FitBit, San Francisco, CA, USA) to self-monitor body weight and step counts daily. The mobile phone app is similar to a virtual weight management system described for weight loss in which body weight and daily steps are automatically transmitted in real time to personalized charts [14]. The SmartMoms intervention includes an IOM 2009 GWG weight graph personalized for each patient and behavioral modification tools including daily self-monitoring of weight, dietary intake, and physical activity [14]. SmartMoms participants were provided with an individualized calorie intake above their estimated prepregnancy energy requirement [13] or energy gap represented by an ideal weight gain zone [15], and were coached how to adjust energy intake and/or physical activity to adhere to the IOM 2009 GWG guidelines. The in-person group tracked step counts with pen and paper, and the IOM weight graph was reviewed in hard copy during counseling sessions with interventionists.

Clinic assessments were performed by certified staff who were blinded to group assignment. Maternal weight was measured fasting and in a hospital gown. Total GWG and GWG per week were calculated between the initial (10–13 weeks) and final (35–36 weeks) study visits. GWG per week was used to calculate the proportion of women based on prepregnancy BMI with recommended or excessive GWG per the 2009 IOM GWG guidelines [16].

Adherence to the SmartMoms intervention was defined as the percentage of days participants weighed and recorded step counts in comparison to the expected number of days. Study economics, including costs incurred for travel to and from treatment sessions and time spent with the counselor while accounting for session attendance and intervention adherence, were calculated for each participant. The clinic economics included cost of interventionist time (training, session preparation, participant contacts, routine staff meetings) and equipment (scale, pedometer) cost.

Statistical analyses were completed using SAS/STAT version 9.4 software of the SAS System for Windows (Cary, NC, USA). Sample size was based on the hypothesis that the proportion of overweight and obese pregnant women in the usual care group exceeding IOM guidelines for GWG would be 58% and that lowering this proportion by at least 25% would be clinically



significant. Intention-to-treat comparison of continuous variables (eg, GWG, birth weight) between the three treatment groups used one-way analysis of variance with post hoc pairwise intervention group comparisons. Comparisons of categorical variables (eg, prevalence of excess GWG) between the three treatment groups used Fisher exact test. Equality of adherence to IOM GWG guidelines was tested through one-sided *z* tests for proportions. Finally, differences in study costs and intervention adherence were assessed through two-sample *t* tests. All tests were performed with significance level  $\alpha=.05$ , and findings were considered significant when  $P<\alpha$ . Data are reported as least square (LS) mean and standard error (SE) unless otherwise noted.

## Results

Recruitment of participants from community clinics from February 1, 2013 to April 14, 2014, yielded three groups of pregnant women who were similar (Table 1). The majority of participants were white and nulliparous or primiparous. No study-related serious adverse events were reported.

## Gestational Weight Gain and Guideline Adherence

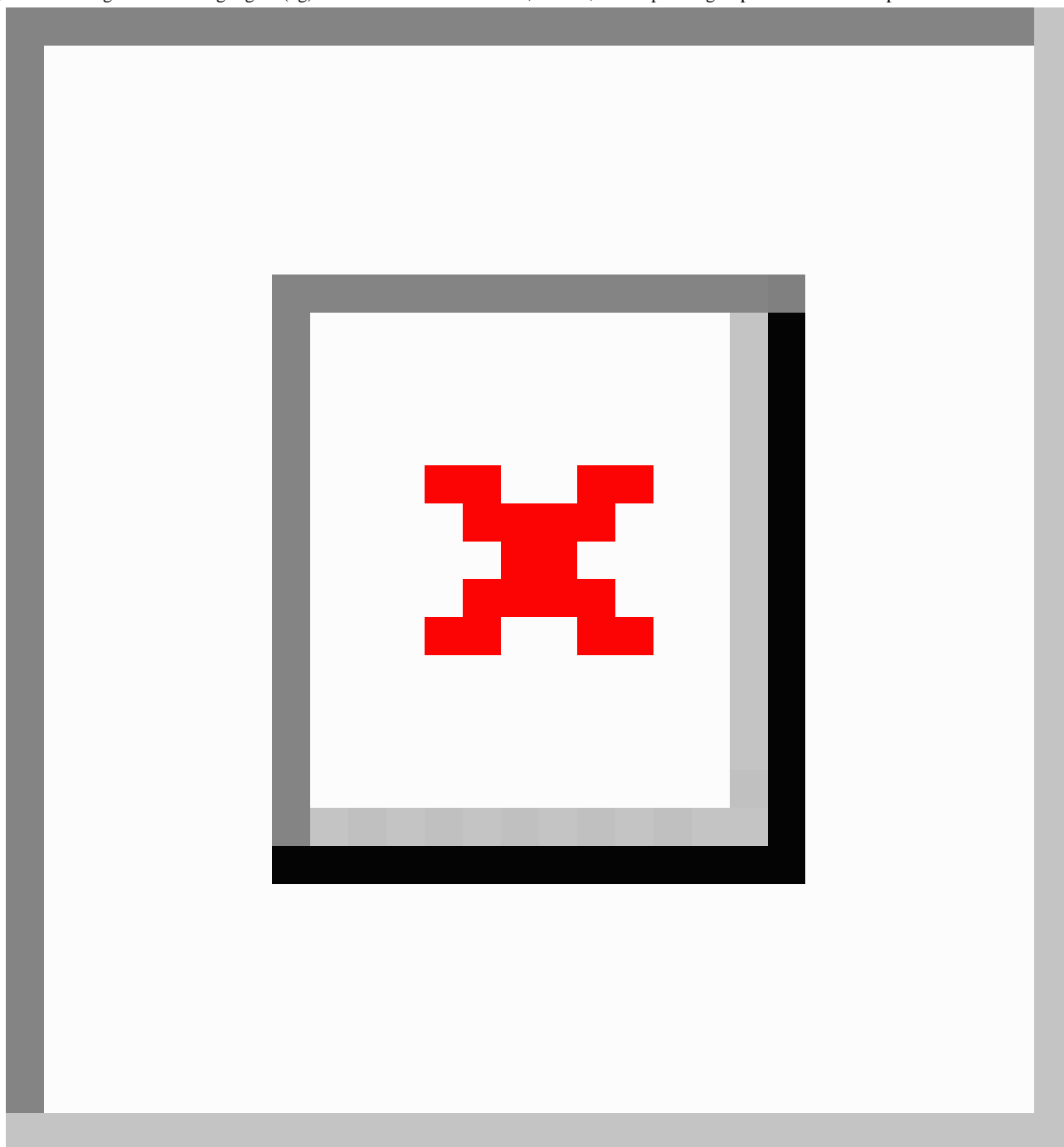
The SmartMoms intervention (in-person and remote groups combined) was effective at reducing GWG in overweight and obese pregnant women (usual care: LS mean 12.8, SE 1.5 kg; SmartMoms: LS mean 9.2, SE 0.9 kg;  $P=.04$ ). The in-person group gained significantly less total weight (Figure 1) during pregnancy than the usual care group (LS mean 8.0, SE 1.3 kg vs LS mean 12.8, SE 1.5 kg;  $P=.04$ ) and weight gain in the remote group was equivalent to the in-person group (LS mean 10.0, SE 1.3 kg;  $P=.04$  equivalence) and modestly lower than weight gain with usual care (LS mean 10.0, SE 1.2 kg vs LS mean 12.8, SE 1.5 kg;  $P=.07$ ). Compared to usual care, the rate of GWG was significantly lower in the in-person group (LS mean 0.49, SE 0.06 kg/week vs LS mean 0.31, SE 0.05 kg/week;  $P=.01$ ) and the rate of GWG in the in-person group was equivalent to the remote group (LS mean 0.39, SE 0.05 kg/week;  $P=.04$ ) within 200 grams of weight gained per week. The proportion of women with excess GWG (Figure 2) was significantly lower in the in-person (56%, 10/18;  $P=.03$ ) and remote groups (58%, 11/19;  $P=.04$ ) compared to usual care (84.6%, 11/13).

**Table 1.** Baseline characteristics by treatment group (N=54).

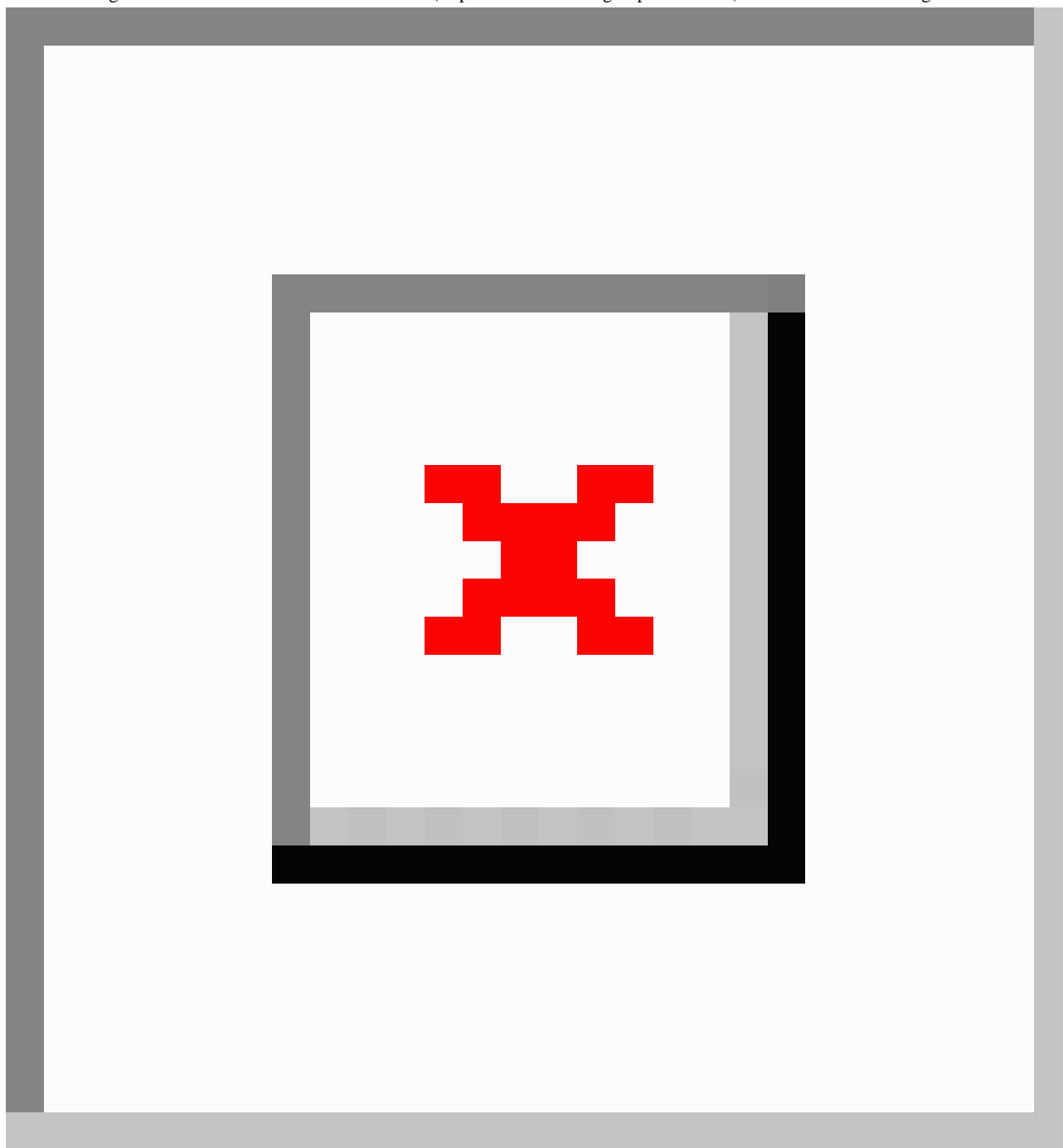
Characteristic	Usual care (n=17)	In-person (n=18)	Remote (n=19)	<i>P</i> <sup>a</sup>
Age (years), mean (SD)	29.5 (5.1)	29.2 (4.8)	29.0 (4.2)	.96
<b>Race, n (%)</b>				.24
Black	6 (35)	5 (28)	2 (11)	
White	11 (65)	11 (61)	16 (84)	
Other	0 (0)	2 (11)	1 (5)	
Weight (kg), mean (SD)	86.2 (12)	83.0 (12)	81.1 (13)	.45
Parity, n (%)	0.6 (1)	0.7 (1)	0.9 (1)	.54
<b>Pregravid BMI group, n (%)</b>				.84
Overweight (25.0-29.9 kg/m <sup>2</sup> )	9 (53)	8 (44)	8 (42)	
Obese (30.0-40.0 kg/m <sup>2</sup> )	8 (47)	10 (56)	11 (58)	
<b>Family income per year (US\$), n (%)</b>				.50
<\$5000-\$39,999	7 (41)	9 (50)	7 (37)	
\$40,000-\$99,999	5 (29)	7 (39)	5 (26)	
≥\$100,000	5 (29)	2 (11)	7 (37)	
<b>Education, n (%)</b>				.24
Some high school	0 (0)	1 (6)	0 (0)	
High school diploma/GED/1-3 years of college, business, or technical school	7 (41)	5 (28)	6 (32)	
College degree	4 (24)	7 (39)	11 (58)	
Postgraduate work	6 (35)	5 (28)	2 (11)	

<sup>a</sup>*P* values were derived from ANOVA for continuous variables and Fisher exact test for categorical variables.

**Figure 1.** Mean gestational weight gain (kg) for women in the usual care, remote, and in-person groups. The whiskers represent standard error.



**Figure 2.** Proportion of women in the usual care, remote, and in-person groups who had appropriate and excess gestational weight gain (GWG) based on the IOM 2009 guidelines. \*The SmartMoms intervention (in-person and remote groups combined) was effective at reducing GWG.

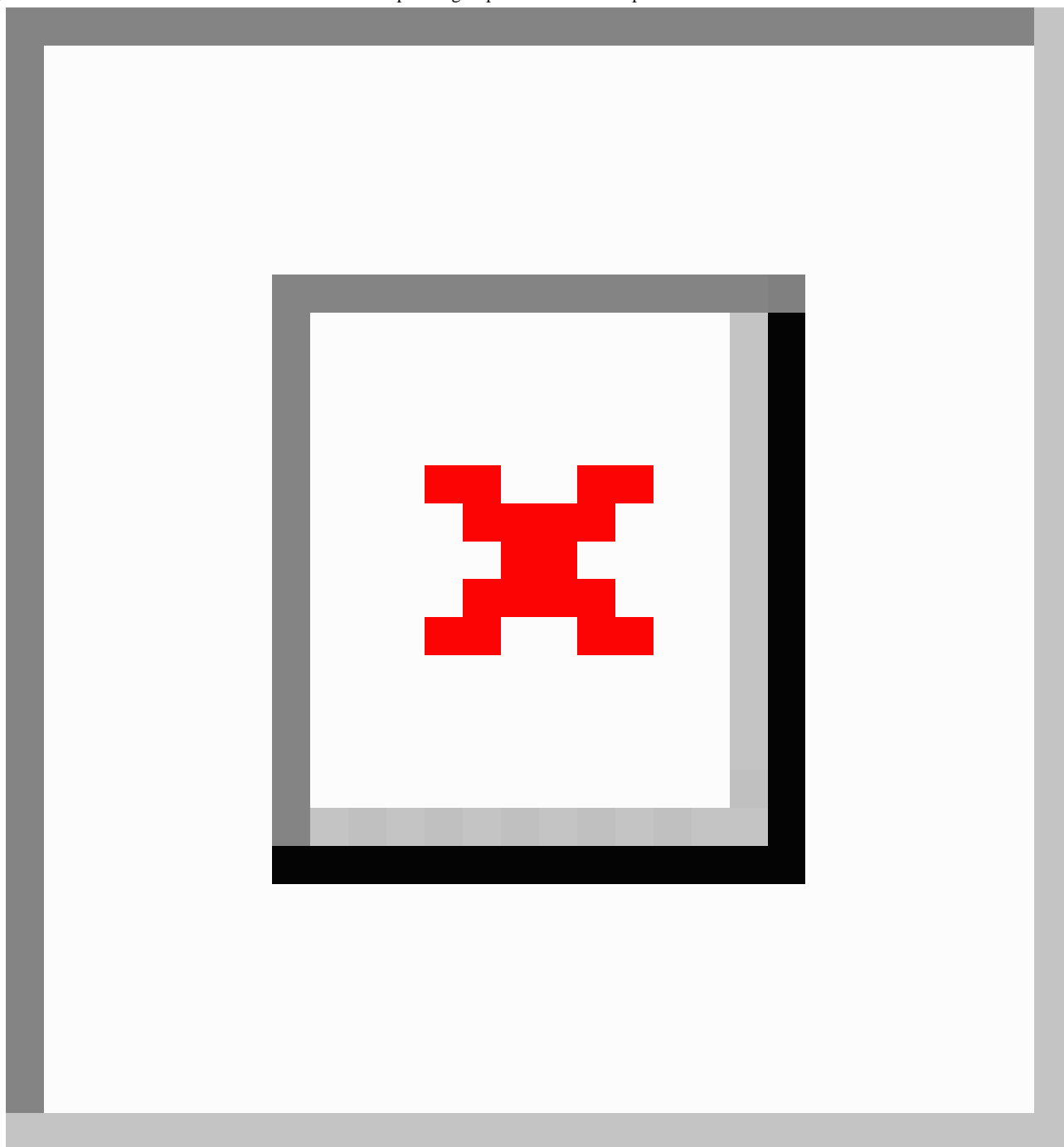


### Intervention Adherence and Study Economics

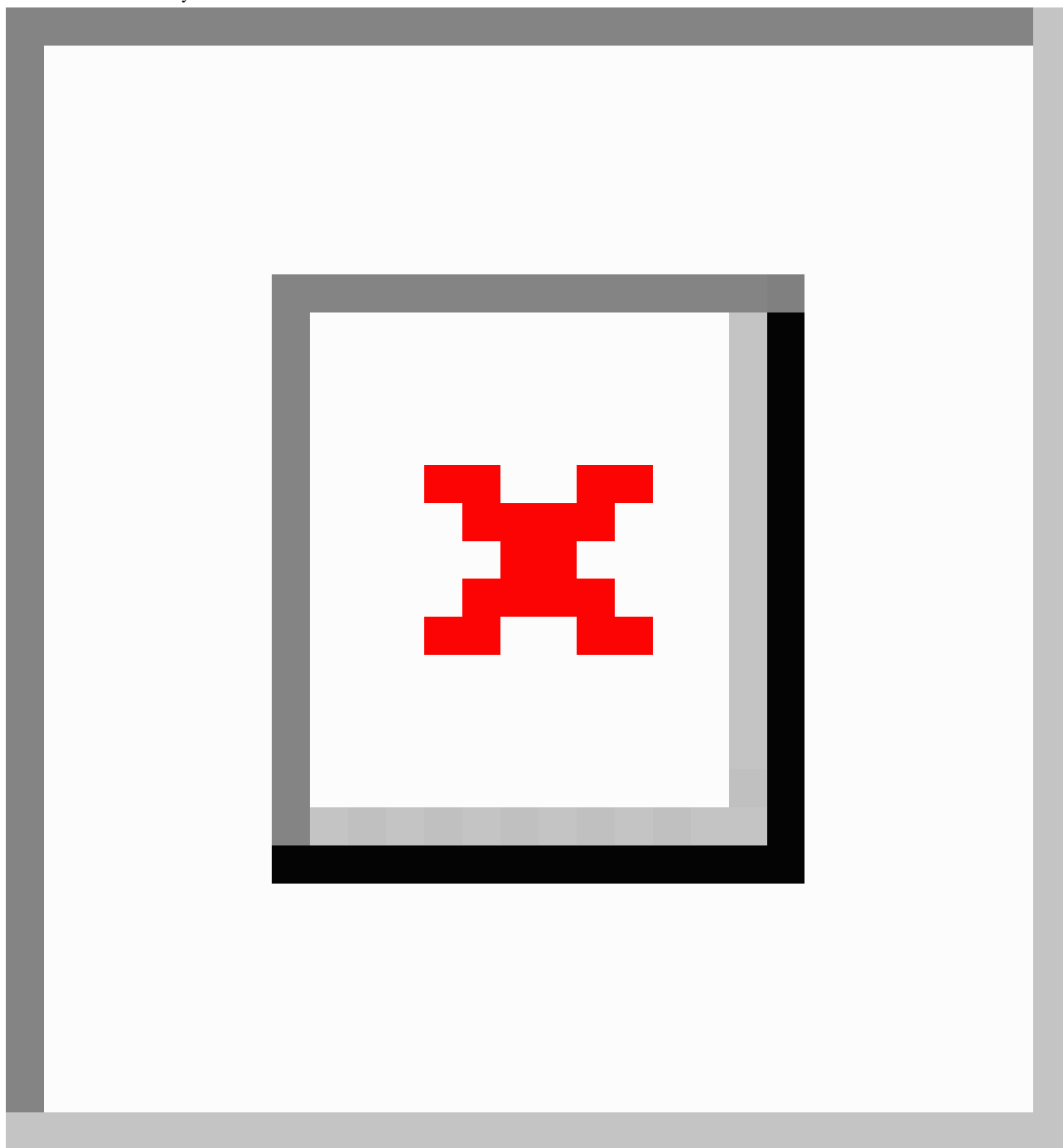
The in-person group recorded weight and step data (weight: mean 57.2%, SD 33.8%; step: mean 44.5%, SD 33.3%) less often than the remote group (weight: mean 71.2%, SD 24.1%; step: mean 72.5, SD 29.0%) and the in-person group attended mean 78% (SD 39%) of planned behavioral sessions. Therefore, overall intervention adherence (Figure 3) was greater in the

remote group than the in-person group (76.5% vs 60.8%;  $P=.049$ ). The intervention cost (Figure 4) to a participant in the remote group was 3.5 times less than the cost for a participant in the in-person group (mean US \$97, SD \$6 vs mean US \$347, SD \$40 per participant;  $P<.001$ ). Similarly, the clinic cost to deploy the remote intervention was 50% less than the cost to deploy the in-person intervention (US \$215 vs US \$419 per participant).

**Figure 3.** Intervention adherence for the remote and in-person groups. The whiskers represent standard error.



**Figure 4.** Mean intervention cost (US \$) for participants and clinics for the remote and in-person groups. The whiskers represent standard error. The intervention cost incurred by the clinic was fixed.



## Discussion

Lifestyle interventions to improve adherence to the IOM GWG recommendations have had modest success in reducing GWG [4,17] and little impact reducing the incidence of excess GWG [18]. The greatest success has been with a recommendation of caloric restriction [19]; however, due to popular beliefs such as the need to “eat for two” [20], caloric restriction is not widely accepted among patients, practitioners, or their support systems. Albeit in a small sample, we attribute the success of the SmartMoms intervention to its early initiation (13 weeks gestation) and intervention intensiveness being commensurate with weight management treatment for nonpregnant individuals

including self-monitoring with timely feedback, a dietary prescription to foster optimal weight change, and receipt of structured behavior change intervention through delivery of 18 lessons over a 24-week interval beginning at the second trimester.

When deployed remotely through a mobile phone, the SmartMoms intervention was just as effective at reducing the proportion of excess GWG when delivered in-person; however, it was found to be at least 50% more economical for patients and providers with a higher level of patient engagement or adherence. This eHealth intervention, including the provision of a personalized IOM GWG weight graph through Internet-connected devices, easily disseminates supportive health



information to patients, and remote patient communication provides an ideal framework for integration into an electronic health record system. Using estimates of interventionist time recorded throughout the study, it is estimated that approximately 30 to 50 new patients per clinician per month could be monitored simultaneously through the remote program by a single health care provider, such as a dietician or lifestyle coach, for universal delivery to all patients within a clinical practice. Similar telehealth services are covered by health care insurance

companies, including Medicaid, and are already used by health care facilities across the United States such as in the Veterans Affairs Hospital System for management of chronic health conditions [21]. The SmartMoms mobile phone intervention tested on community-based obstetrical patients could easily be integrated into standard clinical practice, thereby improving access to effective and efficient health care for millions of American women throughout the entire prenatal care continuum.

---

## Acknowledgments

We thank the LIFE-Moms consortium members for their contributions to the development and oversight of the common measures and procedures shared across the trials. We also thank numerous local obstetrician gynecologists who allowed us to recruit patients in their clinics and the study participants who dedicated their time to this research.

Funding was received through National Institutes of Health grants U01 DK094418 (LMR), 1 U54 GM104940 (LAG, JHB, JWA), P30 DK072476, and F31 HD084199 (EFS). This study (U01 DK094418) is part of the National Institutes of Health Lifestyle Interventions in Expectant Moms (LIFE-Moms) Consortium. LIFE-Moms is supported by the National Institutes of Health through the National Institute of Diabetes and Digestive and Kidney Diseases (U01 DK094418, U01 DK094463, U01 DK094416, 5U01 DK094466 [RCU]), the National Heart, Lung, and Blood Institute (U01 HL114344, U01 HL114377), the Eunice Kennedy Shriver National Institute of Child Health and Human Development (U01 HD072834), the National Center for Complementary and Integrative Health (NCCIH), the NIH Office of Research in Women's Health (ORWH), the Office of Behavioral and Social Science Research (OBSSR), the Indian Health Service, and the Intramural Research Program of the NIDDK. This study was also supported in part by U54 GM104940 from the National Institute of General Medical Sciences of the National Institutes of Health, which funds the Louisiana Clinical and Translational Science Center. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

---

## Authors' Contributions

LMR, CKM, and KEH designed the study. EFS, LEC, ADA, JB, and DH collected outcome data. LAG, JWA, SR, HB, AD, TS, and CKM conducted the interventions. JHB and JMT conducted analysis. LMR and LAG wrote the manuscript. All authors reviewed and approved the final draft.

---

## Conflicts of Interest

Drs Redman, Martin, and Thomas developed the trademarked approach of SmartMoms (a registered trademark of the Louisiana State University System). There are no direct benefits to these authors for publication of this manuscript and they have no financial affiliations with the companies who conducted the work to develop the SmartMoms Virtual Weight Management Suite, although they could financially benefit from any licensing of SmartMoms along with LSU-Pennington Biomedical Research Center and Montclair State University. All other authors have no conflicts of interest to declare.

---

## Multimedia Appendix 1

CONSORT-EHEALTH checklist (v1.6.1).

[[PDF File \(Adobe PDF File\), 596KB - mhealth\\_v5i9e133\\_app1.pdf](#)]

---

## References

1. Institute of Medicine and National Research Council Committee to Reexamine IOM Pregnancy Weight Guidelines. In: Rasmussen KM, Yaktine AL, editors. *Weight Gain During Pregnancy: Reexamining the Guidelines*. Washington, DC: National Academies Press; 2009.
2. Jensen M, Ryan DH, Apovian CM, Ard JD, Comuzzie AG, Donato KA, American College of Cardiology/American Heart Association Task Force on Practice Guidelines, Obesity Society. 2013 AHA/ACC/TOS guideline for the management of overweight and obesity in adults: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines and The Obesity Society. *J Am Coll Cardiol* 2014 Jul 01;63(25 Pt B):2985-3023. [doi: [10.1016/j.jacc.2013.11.004](#)] [Medline: [24239920](#)]
3. Flynn A, Dalrymple K, Barr S, Poston L, Goff LM, Rogozińska E, i-WIP (International Weight Management in Pregnancy) Collaborative Group. Dietary interventions in overweight and obese pregnant women: a systematic review of the content, delivery, and outcomes of randomized controlled trials. *Nutr Rev* 2016 May;74(5):312-328. [doi: [10.1093/nutrit/nuw005](#)] [Medline: [27083868](#)]

4. Oteng-Ntim E, Varma R, Croker H, Poston L, Doyle P. Lifestyle interventions for overweight and obese pregnant women to improve pregnancy outcome: systematic review and meta-analysis. *BMC Med* 2012 May 10;10:47 [FREE Full text] [doi: [10.1186/1741-7015-10-47](https://doi.org/10.1186/1741-7015-10-47)] [Medline: [22574949](https://pubmed.ncbi.nlm.nih.gov/22574949/)]
5. Childbirth Connection. How Do US Women Use the Internet and Other Sources of Pregnancy Information? A Listening to Mothers III Data Brief. 2016. URL: <http://transform.childbirthconnection.org/wp-content/uploads/2013/05/LTMI-DB-infosources.pdf> [accessed 2017-08-30] [WebCite Cache ID 6t6jNuEML]
6. ACOG Committee Obstetric Practice. ACOG Committee opinion. Number 267, January 2002: exercise during pregnancy and the postpartum period. *Obstet Gynecol* 2002 Jan;99(1):171-173. [Medline: [11777528](https://pubmed.ncbi.nlm.nih.gov/11777528/)]
7. Sutton E, Cain LE, Vallo PM, Redman LM. Strategies for successful recruitment of pregnant patients into clinical trials. *Obstet Gynecol* 2017 Mar;129(3):554-559. [doi: [10.1097/AOG.0000000000001900](https://doi.org/10.1097/AOG.0000000000001900)] [Medline: [28178062](https://pubmed.ncbi.nlm.nih.gov/28178062/)]
8. Pieper C, Redman L, Racette S, Roberts S, Bhapkar M, Rochon J, et al. Development of adherence metrics for caloric restriction interventions. *Clin Trials* 2011 Apr;8(2):155-164 [FREE Full text] [doi: [10.1177/1740774511398369](https://doi.org/10.1177/1740774511398369)] [Medline: [21385788](https://pubmed.ncbi.nlm.nih.gov/21385788/)]
9. Thomas DM, Ciesla A, Levine JA, Stevens JG, Martin CK. A mathematical model of weight change with adaptation. *Math Biosci Eng* 2009 Oct;6(4):873-887 [FREE Full text] [Medline: [19835433](https://pubmed.ncbi.nlm.nih.gov/19835433/)]
10. Thomas D, Das SK, Levine JA, Martin CK, Mayer L, McDougall A, et al. New fat free mass - fat mass model for use in physiological energy balance equations. *Nutr Metab (Lond)* 2010 May 09;7:39 [FREE Full text] [doi: [10.1186/1743-7075-7-39](https://doi.org/10.1186/1743-7075-7-39)] [Medline: [20459692](https://pubmed.ncbi.nlm.nih.gov/20459692/)]
11. Thomas DM, Martin CK, Heymsfield S, Redman LM, Schoeller DA, Levine JA. A simple model predicting individual weight change in humans. *J Biol Dyn* 2011 Nov;5(6):579-599 [FREE Full text] [doi: [10.1080/17513758.2010.508541](https://doi.org/10.1080/17513758.2010.508541)] [Medline: [24707319](https://pubmed.ncbi.nlm.nih.gov/24707319/)]
12. Thomas DM, Schoeller DA, Redman LA, Martin CK, Levine JA, Heymsfield SB. A computational model to determine energy intake during weight loss. *Am J Clin Nutr* 2010 Dec;92(6):1326-1331 [FREE Full text] [doi: [10.3945/ajcn.2010.29687](https://doi.org/10.3945/ajcn.2010.29687)] [Medline: [20962159](https://pubmed.ncbi.nlm.nih.gov/20962159/)]
13. Thomas DM, Navarro-Barrientos JE, Rivera DE, Heymsfield SB, Bredlau C, Redman LM, et al. Dynamic energy-balance model predicting gestational weight gain. *Am J Clin Nutr* 2012 Jan;95(1):115-122 [FREE Full text] [doi: [10.3945/ajcn.111.024307](https://doi.org/10.3945/ajcn.111.024307)] [Medline: [22170365](https://pubmed.ncbi.nlm.nih.gov/22170365/)]
14. Martin C, Gilmore LA, Apolzan JW, Myers CA, Thomas DM, Redman LM. Smartloss: a personalized mobile health intervention for weight management and health promotion. *JMIR Mhealth Uhealth* 2016 Mar 16;4(1):e18 [FREE Full text] [doi: [10.2196/mhealth.5027](https://doi.org/10.2196/mhealth.5027)] [Medline: [26983937](https://pubmed.ncbi.nlm.nih.gov/26983937/)]
15. Hill J, Peters JC, Wyatt HR. Using the energy gap to address obesity: a commentary. *J Am Diet Assoc* 2009 Nov;109(11):1848-1853 [FREE Full text] [doi: [10.1016/j.jada.2009.08.007](https://doi.org/10.1016/j.jada.2009.08.007)] [Medline: [19857625](https://pubmed.ncbi.nlm.nih.gov/19857625/)]
16. Gilmore LA, Redman LM. Weight gain in pregnancy and application of the 2009 IOM guidelines: toward a uniform approach. *Obesity (Silver Spring)* 2015 Mar;23(3):507-511 [FREE Full text] [doi: [10.1002/oby.20951](https://doi.org/10.1002/oby.20951)] [Medline: [25521748](https://pubmed.ncbi.nlm.nih.gov/25521748/)]
17. Dodd J, Grivell RM, Crowther CA, Robinson JS. Antenatal interventions for overweight or obese pregnant women: a systematic review of randomised trials. *BJOG* 2010 Oct;117(11):1316-1326 [FREE Full text] [doi: [10.1111/j.1471-0528.2010.02540.x](https://doi.org/10.1111/j.1471-0528.2010.02540.x)] [Medline: [20353459](https://pubmed.ncbi.nlm.nih.gov/20353459/)]
18. Kinnunen TI, Pasanen M, Aittasalo M, Fogelholm M, Hilakivi-Clarke L, Weiderpass E, et al. Preventing excessive weight gain during pregnancy - a controlled trial in primary health care. *Eur J Clin Nutr* 2007 Jul;61(7):884-891. [doi: [10.1038/sj.ejcn.1602602](https://doi.org/10.1038/sj.ejcn.1602602)] [Medline: [17228348](https://pubmed.ncbi.nlm.nih.gov/17228348/)]
19. Vesco K, Karanja N, King JC, Gillman MW, Leo MC, Perrin N, et al. Efficacy of a group-based dietary intervention for limiting gestational weight gain among obese women: a randomized trial. *Obesity (Silver Spring)* 2014 Sep;22(9):1989-1996 [FREE Full text] [doi: [10.1002/oby.20831](https://doi.org/10.1002/oby.20831)] [Medline: [25164259](https://pubmed.ncbi.nlm.nih.gov/25164259/)]
20. Kim LP, Koleilat M, Whaley SE. A qualitative study to examine perceptions and barriers to appropriate gestational weight gain among participants in the Special Supplemental Nutrition Program for Women Infants and Children Program. *J Pregnancy* 2016;2016:4569742 [FREE Full text] [doi: [10.1155/2016/4569742](https://doi.org/10.1155/2016/4569742)] [Medline: [27403341](https://pubmed.ncbi.nlm.nih.gov/27403341/)]
21. Benton S. Hospitals and Health Networks. 2016 Mar 3. Four reasons why telemedicine can improve patient engagement URL: <http://www.hhnmag.com/articles/6974-telemedicine-to-improve-patient-engagement> [accessed 2017-08-11] [WebCite Cache ID 6sjhXDWJe]

## Abbreviations

**BMI:** body mass index  
**GWG:** gestational weight gain  
**IOM:** Institute of Medicine

*Edited by G Eysenbach; submitted 16.06.17; peer-reviewed by D Downs, A Rickman; comments to author 13.07.17; revised version received 27.07.17; accepted 29.07.17; published 13.09.17.*

*Please cite as:*

Redman LM, Gilmore LA, Breaux J, Thomas DM, Elkind-Hirsch K, Stewart T, Hsia DS, Burton J, Apolzan JW, Cain LE, Altazan AD, Ragusa S, Brady H, Davis A, Tilford JM, Sutton EF, Martin CK

*Effectiveness of SmartMoms, a Novel eHealth Intervention for Management of Gestational Weight Gain: Randomized Controlled Pilot Trial*

*JMIR Mhealth Uhealth* 2017;5(9):e133

URL: <http://mhealth.jmir.org/2017/9/e133/>

doi: [10.2196/mhealth.8228](https://doi.org/10.2196/mhealth.8228)

PMID: [28903892](https://pubmed.ncbi.nlm.nih.gov/28903892/)

©Leanne M Redman, L. Anne Gilmore, Jeffrey Breaux, Diana M Thomas, Karen Elkind-Hirsch, Tiffany Stewart, Daniel S Hsia, Jeffrey Burton, John W Apolzan, Loren E Cain, Abby D Altazan, Shelly Ragusa, Heather Brady, Allison Davis, J. Mick Tilford, Elizabeth F Sutton, Corby K Martin. Originally published in JMIR Mhealth and Uhealth (<http://mhealth.jmir.org>), 13.09.2017. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mhealth and uhealth, is properly cited. The complete bibliographic information, a link to the original publication on <http://mhealth.jmir.org/>, as well as this copyright and license information must be included.

Original Paper

# A Smartphone App for Improvement of Colonoscopy Preparation (ColoprAPP): Development and Feasibility Study

Benjamin Walter<sup>1</sup>, MD; Roland Schmid<sup>2</sup>, Prof. Dr. med.; Stefan von Delius<sup>2</sup>, Prof. Dr. med.

<sup>1</sup>Universitätsklinik Ulm, Medizinische Klinik I, Universität Ulm, Ulm, Germany

<sup>2</sup>Klinikum rechts der Isar der Technischen Universität München, II. Medizinische Klinik und Poliklinik, Technische Universität München, München, Germany

**Corresponding Author:**

Benjamin Walter, MD

Universitätsklinik Ulm

Medizinische Klinik I

Universität Ulm

Albert-Einstein-Allee 23

Ulm,

Germany

Phone: 49 731 500 44818

Fax: 49 731 500 44812

Email: [Benjamin.Walter@uniklinik-ulm.de](mailto:Benjamin.Walter@uniklinik-ulm.de)

## Abstract

**Background:** Optimal bowel preparation is one of the major cornerstones for quality of colonoscopy. But poor bowel preparation still occurs in 10% to 25% of all patients. To optimize patient guidance, we developed a new smartphone app (ColoprAPP) for Android smartphones which guides and accompanies the patient starting 4 days before colonoscopy throughout the whole colonoscopy preparation procedure.

**Objective:** The objective of this study was to assess the function of a newly developed smartphone app for supporting colonoscopy preparation.

**Methods:** We carried out a prospective feasibility study including 25 patients undergoing outpatient colonoscopy at our hospital. As a control, we retrieved the data of 25 patients undergoing outpatient colonoscopy matching in age, sex, and indication for colonoscopy from our colonoscopy database. Patients were asked to download the smartphone app, ColoprAPP, in addition to being given the regular colonoscopy preparation leaflet. All colonoscopies were performed in the morning after using a split-dose preparation containing a polyethylene glycol-based purgative. The study was designed to test feasibility of the prototype, evaluate grade of bowel cleanliness (Boston bowel preparation scale [BBPS]), and assess patient satisfaction with the app.

**Results:** The smartphone app use was feasible in all patients. BBPS count as a marker for grade of bowel preparation was significantly higher in the smartphone app-supported group (mean 8.1 [SD 0.3] vs 7.1 [SD 0.4],  $P=.02$ ). Left (mean 2.8 [SD 0.1] vs 2.4 [SD 0.11],  $P=.02$ ) and transverse colon (mean 2.8 [SD 0.07] vs 2.4 [SD 0.11],  $P<.001$ ) revealed significantly higher BBPS counts in the smartphone app-supported group than in controls. Patient satisfaction with a smartphone app-supported colonoscopy preparation was high with an average numeric rating scale score for usefulness of 8.2 (visual analog scale 1-10).

**Conclusions:** A novel developed smartphone app for reinforced education of bowel cleansing was feasible and led to high BBPS scores and patient satisfaction.

**Trial registration:** ClinicalTrials.gov NCT02512328; <https://clinicaltrials.gov/ct2/show/NCT02512328> (Archived by WebCite at <http://www.webcitation.org/6sz3Kk26z>)

(JMIR Mhealth Uhealth 2017;5(9):e138) doi:[10.2196/mhealth.7703](https://doi.org/10.2196/mhealth.7703)

## KEYWORDS

bowel preparation; smartphone app; intestinal cleansing; patient education; colonoscopy; colonoscopy preparation

## Introduction

Optimal bowel preparation is a major cornerstone of colonoscopy quality [1]. Improvement of bowel preparation by split-dose purgative regime has already raised the quality [2]. Nevertheless, bowel preparation is still inadequate in a considerable amount of patients [3]. Many factors have already been described like socioeconomic and educational reasons [4]. Education and continuous guidance of patients has been shown to ensure quality of colonoscopy preparation and patient compliance. Numerous approaches using additional booklets, cartoons, educational videos, and text messaging have shown a direct benefit [5-8]. Nevertheless, all approaches still have their specific limitations toward the availability of the medium.

Worldwide, an enormous increase in the use of smartphones has taken place. Integration of this new medium could help optimize the preparation procedure. We therefore created a novel smartphone app to improve colonoscopy preparation. The primary purpose of the actual study was then to evaluate if this newly developed app for Android smartphones is feasible for

colonoscopy preparation. Secondary endpoints were patient satisfaction and bowel cleanliness.

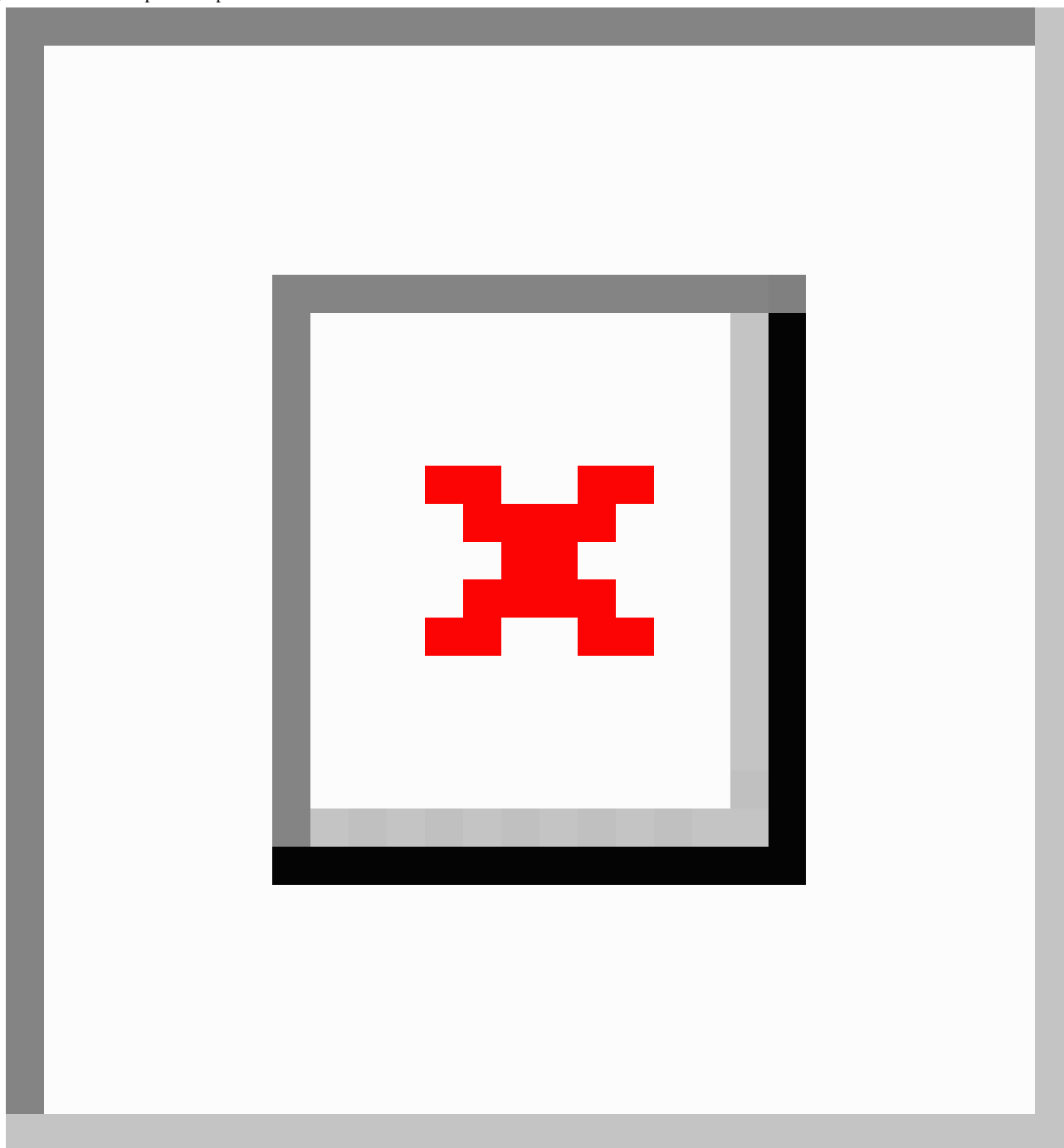
## Methods

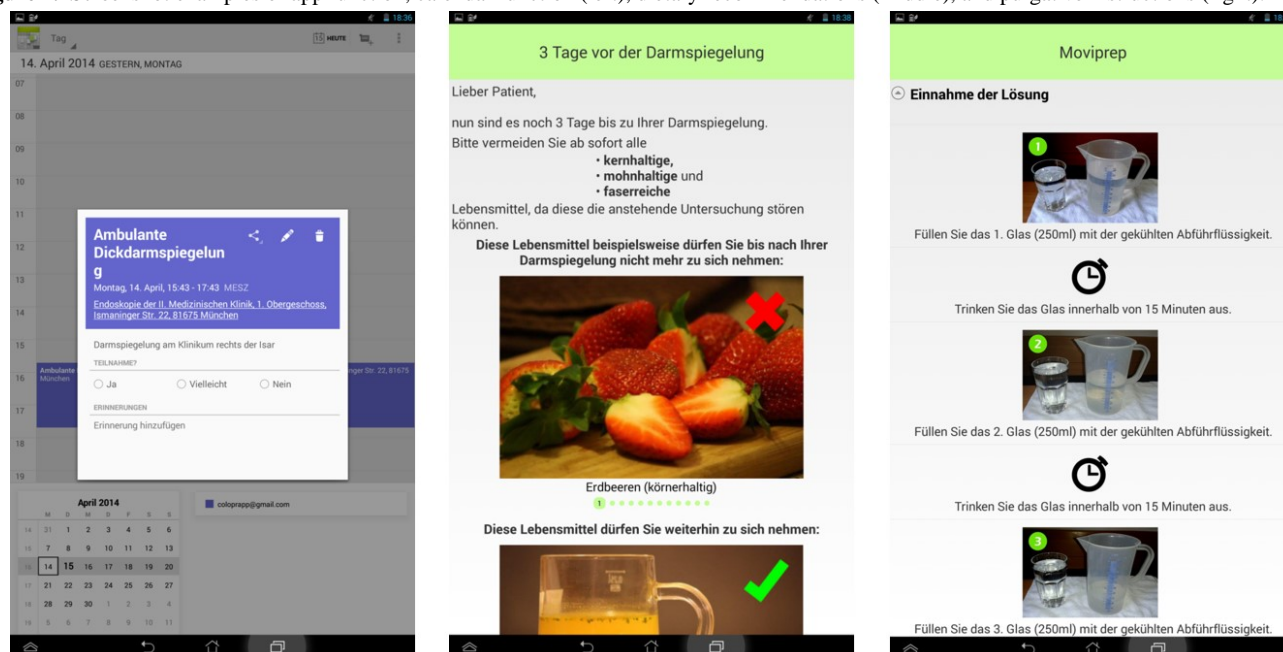
### Smartphone App Development

The Colonoscopy Preparation App (ColoprAPP) is an app (Figures 1 and 2) for mobile devices using the leading mobile operating system, Android. ColoprAPP is written in the official development language for Android called Java. The app does not need access to the Internet after download. Focus is on data security (more details about software development in Multimedia Appendix 1). The smartphone app was developed in cooperation with the University of Applied Sciences, Munich.

After the smartphone app was developed, this prospective study was conducted at the II. Medizinische Klinik, Klinikum rechts der Isar, Technische Universität München. The study protocol was approved by the local ethics committee. Informed consent was obtained from the study participants. The study is registered at ClinicalTrials.gov [NCT02512328].



**Figure 1.** Start desktop of ColoprAPP.

**Figure 2.** Screenshot examples of app function, calendar function (left), dietary recommendations (middle), and purgative instructions (right).

## Colonoscopy Preparation Scheme

At our hospital, the colonoscopy preparation standard is a regular polyethylene glycol–based split-dose regime (MoviPrep, Norgine SA). Explanations of the preparation procedure are given during the informed consent discussion several days prior to the endoscopy. Furthermore, a leaflet containing detailed diet and preparation recommendations is given to every patient undergoing colonoscopy. For quality assurance reasons, colonoscopy preparation is routinely measured by the Boston bowel preparation scale (BBPS) ( $\geq 5$  points as the threshold for sufficient bowel cleansing) [9].

## Feasibility Study Performance

A total of 25 patients undergoing outpatient colonoscopy were included in the smartphone app group. Inclusion criteria were outpatient colonoscopy, written informed consent, aged  $>18$  years, smartphone owner (Android), German (national) mobile phone provider. Exclusion criteria were, due to feasibility study design, very strict excluding patients with phenprocoumon therapy, diabetes mellitus with insulin therapy, pregnancy, recent neurologic illnesses, reported previous electrolyte disturbances, and smartphones with other operating systems than Android (eg, iOS, Microsoft).

The smartphone app was immediately downloaded, saved, and activated from a cloud software storage to the patient's smartphone after study inclusion. All colonoscopies took place in the morning, and 3 in-house endoscopists performed the colonoscopies. They did not get any information on participation status in app-supported colonoscopy preparation.

At the time of colonoscopy, a questionnaire was given to the smartphone app study group subjects asking for (1) history of prior colonoscopies (experience in bowel preparation), (2) when smartphone app was downloaded and used, (3) extent of

discomfort caused by preparation procedure, (4) whether the smartphone app was perceived as helpful, (5) whether smartphone app was perceived a hindrance toward preparation, (6) whether the patient would favor the use of smartphone app-supported colonoscopy preparation again, and (7) whether patients would recommend the smartphone app to friends or family members.

Satisfaction with the smartphone app was assessed using a numeric rating scale (NRS) from 1 to 10 (1=not helpful to 10=very helpful; counterpart: 1=not inhibitory to 10=very inhibitory).

For reasons of comparison, BBPS data of 25 patients from our outpatient colonoscopy database matching inclusion criteria and in age, sex, first colonoscopy or not status, and indication for colonoscopy who underwent outpatient colonoscopy at our institution during the study period were used as “matched pairs.” In case of more than one match, the one with a higher BBPS was chosen.

## Statistical Analysis

A total of 25 patients were planned to be included in the feasibility study. As a control, data from 25 patients matching in age, sex, and indication were taken from the local endoscopic database. Descriptive statistics were computed for all variables to provide means and standard deviations for continuous variables and frequencies for categorical variables. Total BBPS scores were calculated (smartphone app study group and controls). *P* values correspond to the Mann-Whitney U test, Student *t* test with Welch correction, and the paired *t* test. A *P* value  $< .05$  was considered statistically significant. The results for colon preparation were dichotomized to adequate preparation (BBPS total score 5–9) and inadequate colon preparation (BBPS total score  $< 5$ ). All statistical analyses were performed with the statistical software Graphpad Prism (Graphpad Software Inc).

**Table 1.** General patient characteristics.

Characteris- tic	n or mean
<b>Sex, n</b>	
Male	11
Female	14
Age, years, mean (SEM <sup>a</sup> )	44.1 (13.0)
<b>Age, years, mean</b>	
Male	49.4
Female	40.0
<b>First colonoscopy, n</b>	
Yes	14
No	11

<sup>a</sup>SEM: standard error of the mean.

**Table 2.** Indications for colonoscopy performed.

Indication for colonoscopy	n (%)
Colon cancer screening	11 (44)
Unclear abdominal pain	8 (32)
Inflammatory bowel disease (not first diagnosis)	4 (16)
Other	2 (8)

## Results

### Patient Characteristics

For the study, a total of 25/25 patients with outpatient colonoscopy were included. Male/female ratio 1:1.3 (11 male, 14 female). Mean age of participants 44.1 (SD 13.0) years (male: 49.4 years, female: 40.0 years) (Table 1). Indications for colonoscopy were colon cancer screening (11/25, 44%), unclear abdominal pain (8/25, 32%), inflammatory bowel disease (not first diagnosis) (4/25, 16%), other (2/25, 8%); 14/14 patients were having first colonoscopy, and 11/11 had already one or more colonoscopies in their medical history (Table 2). Data of matched pairs were taken from colonoscopy database at the hospital. Pairs matched in age, sex, indication for colonoscopy, and previous colonoscopies.

### Primary Endpoint

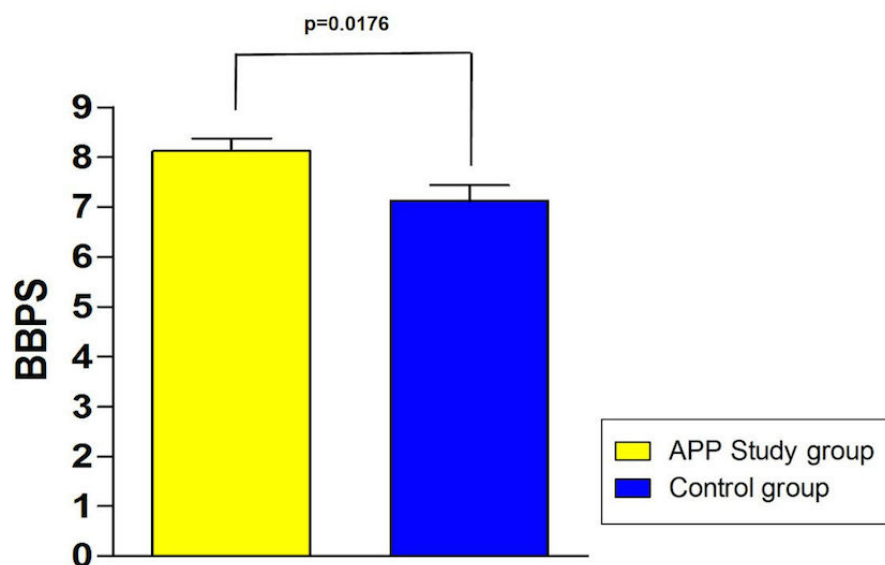
The smartphone app prototype was sufficiently working with stable function during the time period of colonoscopy preparation in all 25 participants. Programmed information for procedure and recommendations were provided to the participants in a correct time-adjusted manner.

### Secondary Endpoints

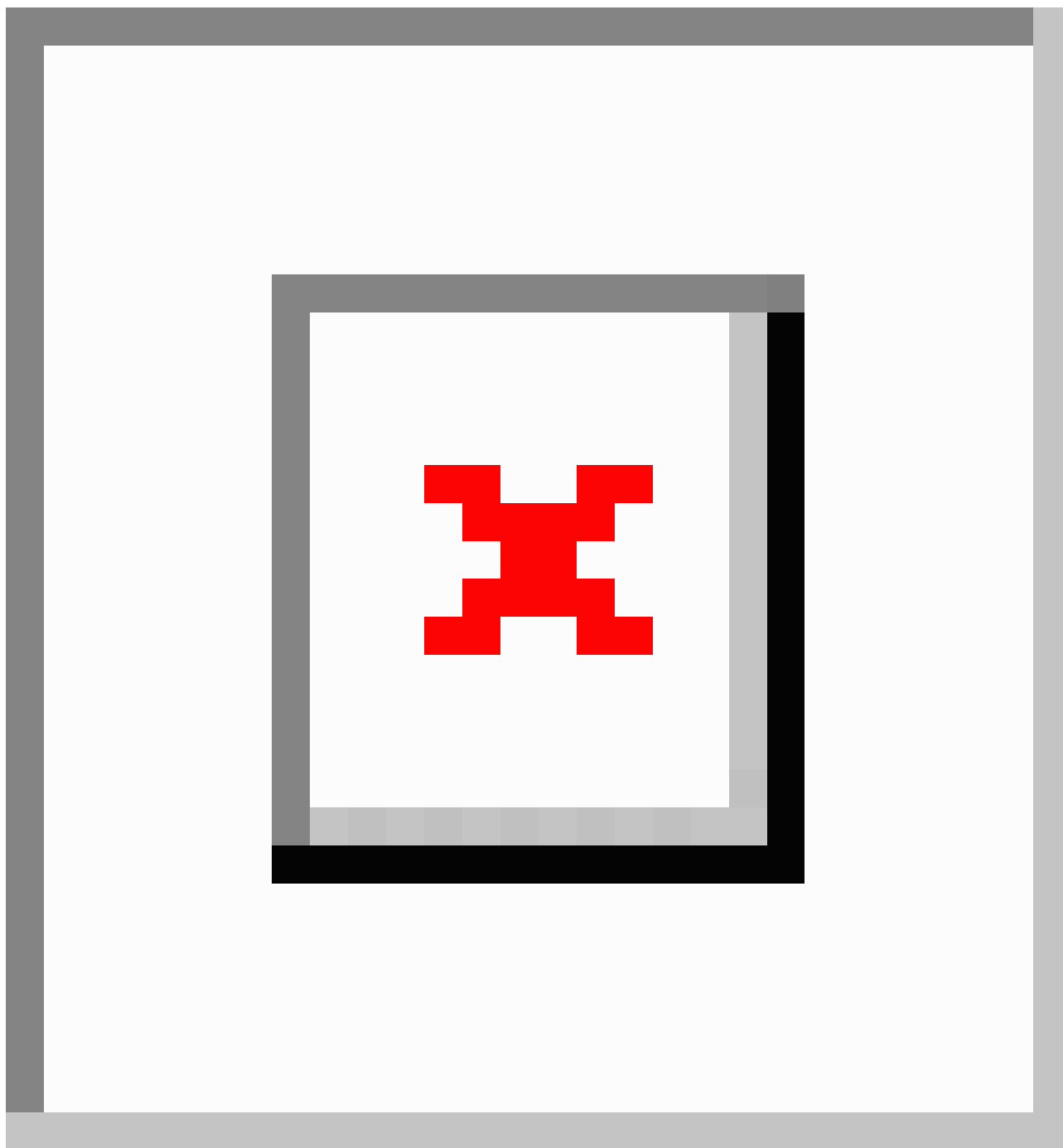
Mean BBPS score of the smartphone app study group was 8.1 (SD 0.25) versus 7.1 (SD 0.41) ( $P=.02$  for difference) (control group) (Figure 3). The mean BBPS score of the smartphone app study group of the left colon was 2.8 (SD 0.08) versus 2.4 (SD 0.11) ( $P=.02$ ). The mean BBPS score of the smartphone app group of the transverse colon was 2.8 (SD 0.07) versus 2.4 (SD 0.11) ( $P<.001$ ). The mean BBPS score of the smartphone app group of the right colon was 2.5 (SD 0.13) versus 2.3 (SD 0.11) ( $P=.36$ ) (Figure 4). With regard to an insufficient bowel preparation (BBPS <5), there was 1 patient in the control group, whereas all smartphone app study participants achieved a BBPS  $\geq 5$ .

All study participants of the smartphone app group would use the smartphone app again and recommend the system to their friends and relatives. When asked if the smartphone app was helpful to get the colonoscopy preparation done, patients reported an average NRS score for usefulness of 8.2. The smartphone app was not found to be a hindrance by an average NRS score for inhibitory effect of the smartphone app of 1.4 (25/25). Preparation for colonoscopy in general was reported to be unpleasant by an average NRS score of 6.1.

**Figure 3.** Total Boston bowel preparation scale score with significantly higher grade for smartphone app-supported group.



**Figure 4.** Average Boston bowel preparation scale score split into colonic regions (left, transverse, and right colon) with a significantly higher score for left and transverse colon.



## Discussion

### Principal Findings

Poor bowel preparation is associated with a bundle of problems and causes an even more reduced acceptability of colon cancer screening programs [10], so this remains a concern. To reach an optimal bowel preparation, close patient guidance and education are crucial [11]. Previous published studies evaluated different forms of interventions and media to improve patient education [1,2,5,6,8].

Smartphones may have an even higher impact for close and up-to-date guidance for our patients, as many people carry a

smartphone 24 hours a day. Smartphone apps are software platforms for smartphones which could be easily used for patient education. To date there is little research on the effect of smartphone apps for colonoscopy preparation. But data already published (only for iOS) seem to be very promising [12].

### Smartphone App Development

We designed a free-of-charge, offline working smartphone app (ColoprAPP) for Android smartphones with the objective to improve patient satisfaction and bowel preparation for colonoscopy. The app offers various educational tools and gives reminders on time-axis adjusted instructions for all steps of colonoscopy preparation (Figures 1 and 2).



## Feasibility Study

This study showed that using a smartphone app is feasible to support colonoscopy preparation and could be an add-on to improve patient education. Use of the app improved grade of bowel cleanliness when compared to matched controls without reinforced educational measures. The participants perceived the smartphone app service as very helpful and would recommend its use to friends and relatives.

## Study Limitations

This feasibility study is subject to some limitations. First, it had only a small number of participants, and second, the app was only available for Android-based smartphones. We note that matched pairs for comparison can include some bias. The use of smartphones varies by different factors including age, sex,

and socioeconomic factors, which by themselves influence the quality of bowel preparation. The real impact of a smartphone app-based reminder system on the quality of bowel cleansing has therefore to be shown in larger randomized studies. We also understand that statistics are very limited due to the small number of study participants, and therefore conclusions regarding improvement of bowel cleanliness are also very limited. Results could only be interpreted as a tendency.

## Conclusion

We conclude that a newly developed Android-based smartphone app for reinforced education to improve colonoscopy preparation is feasible. It works offline to maximize patient data safety. The provided 4 days of guidance containing dietary recommendations and recommendations for laxative intake supported patients to get through the colonoscopy preparation.

## Acknowledgments

The smartphone app was developed in cooperation with the University of Applied Sciences, Munich. Johann Döhring, Johannes Ulmer, Sebastian Klenk, and Alexander Jagemann developed the software as part of a university project under supervision of professors Jochen Hertle and Christian Greiner. Medical input and supervision during development was provided by Benjamin M Walter. This project was supported by the Gesellschaft für Gastroenterologie in Bayern; the funder did not influence development, trial performance, or results of the study.

## Conflicts of Interest

None declared.

## Multimedia Appendix 1

Technical details about the smartphone app.

[[PDF File \(Adobe PDF File\), 33KB - mhealth\\_v5i9e138\\_app1.pdf](#)]

## References

1. Corley DA, Jensen CD, Marks AR, Zhao WK, Lee JK, Doubeni CA, et al. Adenoma detection rate and risk of colorectal cancer and death. *N Engl J Med* 2014 Apr 03;370(14):1298-1306 [[FREE Full text](#)] [doi: [10.1056/NEJMoa1309086](#)] [Medline: [24693890](#)]
2. Beck DE. Bowel preparation for colonoscopy. *Clin Colon Rectal Surg* 2010 Feb;23(1):10-13 [[FREE Full text](#)] [doi: [10.1055/s-0030-1247851](#)] [Medline: [21286285](#)]
3. Lebwohl B, Kastrinos F, Glick M, Rosenbaum AJ, Wang T, Neugut AI. The impact of suboptimal bowel preparation on adenoma miss rates and the factors associated with early repeat colonoscopy. *Gastrointest Endosc* 2011 Jun;73(6):1207-1214 [[FREE Full text](#)] [doi: [10.1016/j.gie.2011.01.051](#)] [Medline: [21481857](#)]
4. Lebwohl B, Wang TC, Neugut AI. Socioeconomic and other predictors of colonoscopy preparation quality. *Dig Dis Sci* 2010 Jul;55(7):2014-2020. [doi: [10.1007/s10620-009-1079-7](#)] [Medline: [20082217](#)]
5. Park J, Kim T, Lee N, Kim H, Seo EH, Heo N, et al. The effectiveness of short message service to assure the preparation-to-colonoscopy interval before bowel preparation for colonoscopy. *Gastroenterol Res Pract* 2015;2015:628049 [[FREE Full text](#)] [doi: [10.1155/2015/628049](#)] [Medline: [25792978](#)]
6. Prakash SR, Verma S, McGowan J, Smith BE, Shroff A, Gibson GH, et al. Improving the quality of colonoscopy bowel preparation using an educational video. *Can J Gastroenterol* 2013 Dec;27(12):696-700 [[FREE Full text](#)] [Medline: [24340313](#)]
7. Liu X, Luo H, Zhang L, Leung FW, Liu Z, Wang X, et al. Telephone-based re-education on the day before colonoscopy improves the quality of bowel preparation and the polyp detection rate: a prospective, colonoscopist-blinded, randomised, controlled study. *Gut* 2014 Jan;63(1):125-130. [doi: [10.1136/gutjnl-2012-304292](#)] [Medline: [23503044](#)]
8. Walter BM, Klare P, Neu B, Schmid RM, von Delius S. Development and testing of an automated 4-day text messaging guidance as an aid for improving colonoscopy preparation. *JMIR Mhealth Uhealth* 2016 Jun 21;4(2):e75 [[FREE Full text](#)] [doi: [10.2196/mhealth.5289](#)] [Medline: [27329204](#)]
9. Lai E, Calderwood A, Doros G, Fix OJ. The Boston Bowel Preparation Scale: a valid and reliable instrument for colonoscopy-oriented research. *Gastroint* 2009;69. [doi: [10.1016/j.gie.2008.05.057.The](#)]

10. Kahlid-de Bakker C, Jonkers D, Smits K, Mesters I, Masclee A, Stockbrügger R. Participation in colorectal cancer screening trials after first-time invitation. *Endoscopy* 2011;43:1059-1086.
11. Rosenfeld G, Krygier D, Enns RA, Singham J, Wiesinger H, Bressler B. The impact of patient education on the quality of inpatient bowel preparation for colonoscopy. *Can J Gastroenterol* 2010 Sep;24(9):543-546 [FREE Full text] [Medline: 21152458]
12. Lorenzo-Zuniga V, Moreno DVV, Marin I, Barbera M, Boix J. Improving the quality of colonoscopy bowel preparation using a smart phone application: a randomized trial. *Dig Endosc* 2015;27(5).

## Abbreviations

**BBPS:** Boston bowel preparation scale

**ColoprAPP:** Colonoscopy Preparation App

**NRS:** numeric rating scale

*Edited by G Eysenbach; submitted 16.03.17; peer-reviewed by K Cato; accepted 14.07.17; published 20.09.17.*

*Please cite as:*

*Walter B, Schmid R, von Delius S*

*A Smartphone App for Improvement of Colonoscopy Preparation (ColoprAPP): Development and Feasibility Study*

*JMIR Mhealth Uhealth* 2017;5(9):e138

URL: <http://mhealth.jmir.org/2017/9/e138/>

doi: [10.2196/mhealth.7703](https://doi.org/10.2196/mhealth.7703)

PMID: [28931498](https://pubmed.ncbi.nlm.nih.gov/28931498/)

©Benjamin Walter, Roland Schmid, Stefan von Delius. Originally published in JMIR Mhealth and Uhealth (<http://mhealth.jmir.org>), 20.09.2017. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mhealth and uhealth, is properly cited. The complete bibliographic information, a link to the original publication on <http://mhealth.jmir.org/>, as well as this copyright and license information must be included.

Original Paper

# A User-Centered Approach: Understanding Client and Caregiver Needs and Preferences in the Development of mHealth Apps for Self-Management

Roxanna M Bendixen<sup>1</sup>, OTR, PhD; Andrea D Fairman<sup>2</sup>, OTR, PhD; Meredith Karavolis<sup>3</sup>, BHSc; Carly Sullivan<sup>3</sup>, BHSc; Bambang Parmanto<sup>4</sup>, PhD

<sup>1</sup>School of Health and Rehabilitation Sciences, Department of Occupational Therapy, University of Pittsburgh, Pittsburgh, PA, United States

<sup>2</sup>School of Health and Rehabilitation Sciences, Department of Occupational Therapy, MGH Institute of Health Professions, Boston, MA, United States

<sup>3</sup>Department of Occupational Therapy, Duquesne University, Pittsburgh, PA, United States

<sup>4</sup>School of Health and Rehabilitation Sciences, Health and Information Management, University of Pittsburgh, Pittsburgh, PA, United States

**Corresponding Author:**

Roxanna M Bendixen, OTR, PhD  
School of Health and Rehabilitation Sciences  
Department of Occupational Therapy  
University of Pittsburgh  
5020 Forbes Tower  
Pittsburgh, PA, 15260  
United States  
Phone: 1 412 383 6603  
Fax: 1 412 383 6613  
Email: [bendixen@pitt.edu](mailto:bendixen@pitt.edu)

## Abstract

**Background:** Many adolescents and young adults with chronic illness or disability often fail to develop the self-management skills necessary to independently handle medical and self-management routines. In light of these needs, we are developing iMHere 2.0 (Interactive Mobile Health and Rehabilitation), a mobile health (mHealth) system to support a self-management program.

**Objective:** Our objective was to gather data from persons with brain and spinal cord anomalies (BSA) and their caregivers to better understand how mHealth would be most helpful in supporting them to proactively manage daily self-care routines and to access medical care as needed. The specific purpose was not only to gather feedback and to gain increased insight into the design of the new version of iMHere, but also to gather perspectives of new groups, namely adolescents as young as 12 years and their parents and/or caregivers.

**Methods:** Our project employed focus group sessions and surveys to collect data from participants with BSA, as well as their caregivers. A total of six focus group sessions were conducted on four separate occasions until the data gathered reached saturation. The objectives of our focus group sessions were to better understand ways to develop mHealth systems to support self-management, to promote independence, to motivate long-term system use, and to prevent medical problems that lead to hospitalizations and emergency room visits for youth and young adults with BSA.

**Results:** A total of 16 youth and young adults with BSA and 11 caregivers participated in the sessions. Within and among our groups, the following five overarching themes emerged from the data: (1) make it easy, (2) engage, (3) educate and prepare, (4) motivate and support, and (5) personalize. Participants shared their perspectives and detailed information about mHealth apps that would be important for independence in self-care and self-management.

**Conclusions:** Our findings suggest that most individuals keep their mobile phones with them at all times and typically use a mobile phone for social media, music, photos, and texting. Our qualitative analysis indicates that youth and young adults with BSA, as well as their caregivers, acknowledge the importance of being actively engaged in developing and using mHealth apps that monitor and manage their health care needs. Information gleaned from these focus group sessions and surveys have provided data to refine the iMHere 2.0 mHealth prototype platform that we have developed.

(*JMIR Mhealth Uhealth* 2017;5(9):e141) doi:[10.2196/mhealth.7136](https://doi.org/10.2196/mhealth.7136)

**KEYWORDS**

mobile health; telemedicine; self-care; adolescence; spina bifida; cerebral palsy; spinal cord injury

## Introduction

### Background

Self-management is a gradual process beginning in childhood, with increasing responsibility during adolescence until one reaches his or her maximum degree of independence in adulthood [1]. Self-management skills and behaviors change over time depending on life circumstances and phases of disease and, therefore, require support strategies tailored to meet these changing needs [2]. Self-management support is the assistance given to people with chronic conditions, which enables them to manage their health on a day-to-day basis. Self-management support can aid and inspire people to learn more about their conditions and to take an active role in their health care [3]. Typically, self-management skills progress during the adolescent and young adult years, which is a developmentally appropriate time in one's life to seek separation from parents and gain full independence with regard to self-management [4,5]. Unfortunately, many adolescents and young adults with chronic illness or disability often fail to develop the self-management skills necessary to independently handle medical and self-management routines [6-8]. For example, individuals with developmental brain and spinal cord anomalies (BSA), including persons with spina bifida (SB), congenital hydrocephalus, and cerebral palsy, often incur preventable secondary conditions such as pressure ulcers, urinary tract infections, and sepsis [9-11]. These complicated conditions occur most frequently in the youngest age groups and account for more than 30% of hospitalizations [10]. However, proactive management of primary and secondary conditions can have a positive effect on medical outcomes and rates of hospitalization [12]. Currently, mobile health (mHealth) interventions are becoming increasingly important for reducing health care costs, encouraging proactive self-management skills, and improving well-being [13-18].

In light of the growing need of persons with BSA, we developed iMHere (Interactive Mobile Health and Rehabilitation), an mHealth system to support a self-management program [19]. The initial version of iMHere 1.0 connected adults with disabilities to clinicians and was implemented to support individuals with SB [20] and spinal cord injury (SCI). iMHere 1.0 was developed on an Android platform and did not support iPhone or other non-Android users. Previous surveys, usability testing, and qualitative feedback were gathered from persons with SB and SCI to improve upon the functionality and user interface of iMHere 1.0 [21]. In the process of gathering these data, it was determined that a new version of the system was needed to allow for a cross-platform design to ensure that the system could be used on the iPhone operating system, Android, and other platforms. This design also allowed us to create a new system that is more appealing for sustained use by a younger population and to integrate caregivers and community support. The newer version of iMHere 2.0, currently in development, will have a cross-platform design and will incorporate support for adults and youth with disabilities and chronic conditions, clinical personnel, caregivers, and community support.

### Objectives

The specific purpose of our study was to employ a user-centered approach to gather increased feedback and to gain insight into the design of the new version of iMHere. We also sought to gather perspectives of new populations, namely adolescents as young as 12 years and their paid or unpaid caregivers. Our project employed surveys and focus group sessions with participants with BSA, as well as their caregivers. We chose the focus group method of data collection, which is a technique involving the use of in-depth group interviews to facilitate the generation of deep, rich, and diverse data through the social interaction of the group participants [22,23]. Our focus group objectives were to better understand ways to develop mHealth systems to support self-management, to promote independence, to motivate long-term system use, and to prevent medical problems that lead to hospitalizations and emergency room visits for youth and young adults with BSA. This iterative user-centered design occurred within the development phase of iMHere 2.0 to ensure that the final mHealth system would fulfill the users' desired functionality and meet their needs.

## Methods

### Participants

Participant (youth and young adults with BSA) and caregiver recruitment for our qualitative focus group sessions occurred through collaboration with human services agencies, including the Spina Bifida Association of Western Pennsylvania (SBAWP), the Community Living and Support Services (CLASS), research registries, and personal referrals. The study received approval from the institutional review board (IRB) of the University of Pittsburgh (Protocol # PRO14120443). iMHere 2.0 is specifically being developed for youth and young adults with BSA, as well as their caregivers. Our recruitment efforts were intended to target youth and caregivers who met the inclusion criteria, but adults with BSA also participated in the study. We sought to achieve a wide range of perspectives and ideas from our focus group discussions through recruiting and enrolling participants with variations in age, gender, living environment, and disease severity. The IRB-approved flyers were emailed to potential participants from our collaborating agencies and posted in the community. Individuals were requested to contact the study coordinator if they met the criteria and were interested in participating in a focus group session. Before participating in the study, written informed consent was obtained from all adult participants and at least one parent or guardian for minor participants. Written assent was obtained from minor participants whenever possible; dissent from all minors was respected (Multimedia Appendices 1 and 2). Inclusion criteria required a diagnosis of BSA or the caregiver of an individual with BSA, the ability to understand and speak English, and willingness to participate in a one-time only focus group interview lasting up to 120 min. Participants received US \$25 remuneration for participation, reimbursement for parking (if applicable), and a free meal. Focus group sessions were

facilitated by the first and second authors (experienced focus group moderators) between May and October 2015 in central locations most convenient to our participants. Focus group sessions comprised participants and research team members only. Apart from the moderators, a notetaker was present to observe and document nonverbal interactions, verbal exchanges, and general content of the group discussions.

### Focus Group Sessions

A PowerPoint presentation developed by the moderators was used to organize and guide the focus groups. The PowerPoint slides also served as a means to provide screenshots of the existing version of the iMHere app in development for participants to view. Guided focus group sessions lasted approximately 120 min and were audiorecorded and professionally transcribed. Each focus group session began with a broad definition of mHealth technology and self-management and a request to participants to individually discuss their experiences with using mobile devices and apps ([Multimedia Appendices 3 and 4](#)). Our research objective was to better understand ways to develop mHealth systems to support self-management for youth and their caregivers; therefore, guided focus group questions centered on areas such as (1) examples of how participants have used mHealth apps in the past; (2) what apps have they found beneficial or unfavorable and why?; (3) what resources participants may need to better support their health?; (4) feedback on mHealth designs to help motivate and engage users; and (5) overall expectations for technology to support their wellness and prevent medical complications. Focus group sessions ended with the collection of demographic data, along with a follow-up survey requesting additional comments regarding the mHealth functions, importance of the functions for promoting self-management and independence, as well as personal experience with technology. Following each focus group session, the research team met for a debriefing session to share information and to discuss observations and potential ways to improve the focus group process. Field notes and debriefing sessions supplemented the participants' oral discussions and enabled a richer analysis of the data [22,23].

### Data Analyses

Final datasets included participant demographics, follow-up surveys, transcribed focus group sessions, as well as fieldwork observations and debriefing sessions. Throughout the research process, the primary author used constant comparison techniques, whereby each focus group session, along with the corresponding fieldwork observations and debriefing sessions, was compared with the previous data and not considered independently. This enabled data to be treated as a whole in

lieu of fragmenting the data. Constant comparison also enabled the initial identification of emerging patterns and themes [24].

The use of framework analysis, as described by Ritchie and Spencer [25], provided clear steps to assist in the management and analyses of the data. Initial key patterns within the datasets were first identified by the primary author through a rigorous approach of familiarization through reading and re-reading each transcript on numerous occasions, along with information from the questionnaires, fieldwork observations, and debriefing sessions [26-28]. This was achieved through hand coding datasets line-by-line by marking the data with varied colored highlighter pens to identify important quotes, initial ideas, and concepts arising from the data. Two research assistants, who were not present during the focus group sessions, thoroughly reviewed the datasets independent of the first author and each other, with direction from the second author. Research assistants used cutting and sorting as a formal way of identifying and organizing important quotes. Following the independent identification of ideas and concepts within the datasets, researchers and research assistants came together on three occasions to discuss their findings as a group and to categorize initial codes. Researchers conducted focused coding to eliminate, combine, and divide categories, thereby allowing for the development of patterns within the data. Patterns were then synthesized and reduced to themes [25,26,28,29]. Data relevant to each theme were fully reviewed and discussed by the research team. Themes were refined and rephrased to more clearly describe the *story* being told by our participants and their caregivers. Agreement among the independent reviewers provided confidence that we had identified appropriate themes. Our central research objective shaped our framework and final interpretation [26,29]. Six in-depth focus group sessions were sufficient to achieve data saturation and to enable the development of meaningful themes and useful interpretations [30-32].

## Results

A total of six focus group sessions were conducted on four separate occasions. Initially, 19 youth and young adults with BSA demonstrated interest in participating in focus group sessions; 3 did not show up for the session and hence could not provide assent, leaving a total of 16 youth and young adults with BSA in our focus groups. A total of 15 caregivers (paid and unpaid) contacted the study coordinator and discussed participation. Of these 15, 2 caregivers canceled because of an illness, and 2 caregivers did not show up, leaving a total of 11 caregivers in our focus group sessions ([Table 1](#)).



**Table 1.** Demographic characteristics of focus group participants.

Characteristic	Participants (n=16)	Caregivers (n=11)
<b>Gender</b>		
Female	12	5
Male	4	6
<b>Age range, in years</b>		
12-17	3	0
18-24	3	0
25-34	3	0
35-44	0	0
45-54	5	1
55-65	2	10
Range	15-62	45-64
Mean (SD)	34.7 (17.4)	51.1(6.6)
<b>Race</b>		
White	15	11
Asian	1	0
<b>Ethnicity</b>		
Hispanic	1	0
Non-Hispanic	15	11
<b>Education</b>		
Some high school	4	0
High school graduate	3	0
Some college	1	1
Trade/Technical/Vocational	0	2
Associate's degree	0	1
Bachelor's degree	7	3
Master's degree	1	3
Doctoral degree	0	1
<b>Marital status</b>		
Single	13	2
In a committed relationship	3	7
Divorced/Separated	0	2
<b>Residence</b>		
Urban	7	3
Suburban/Rural	9	8

Within and among our groups, five overarching themes emerged from the data, which are as follows: (1) make it easy, (2) engage, (3) educate and prepare, (4) motivate and support, and (5) personalize. The *make it easy* theme describes why mobile phone users typically refrain from or give up using apps. The *engage* theme illustrates the importance of having an app that is colorful, consolidated, convenient, and helpful. The *educate and prepare* theme presents our participants' and caregivers'

views on the importance of having individualized information to inform and educate. The *support* theme is focused on the participants' consistent descriptions of their desired social needs and personal feedback. Peer and social support was an important and recurring theme throughout our focus group sessions. Finally, the *personalize* theme provides us with thoughtful and innovative ideas to inspire users to adopt the mHealth app for independence and self-management of their personal health and

wellness. Table 2 presents examples of the initial concepts which created our codes, the themes that developed from these codes, and the percentage of total participants (youth and young adults,

as well as caregivers) who discussed these ideas and concepts throughout our focus group sessions.

**Table 2.** Examples of codes, developed themes, and prevalence of overall participant/caregiver contribution.

Examples of codes	Themes	Participants (n=16), n (%)	Caregivers (n=11), n (%)
Don't use; Too difficult; Need to update; Difficult to type; Too many levels to get through; Too time-consuming; Too slow; Difficult to read—text to voice	Make it easy	13 (81)	10 (91)
Fun reminders; Free movies/Downloads/Games; Educational challenges; Rewards; Colorful and fun; Trivia of the day; Jokes; Entertaining	Engage	14 (87)	10 (91)
Look up information; Learn about my disease; Help to manage stress, etc; Guidance—understanding; Learn from others; Searchable information; Accurate and reliable; Updated information; Resources in the community; Teach basic skills	Educate and prepare	16 (100)	11 (100)
Social engagement; Share; Feedback from others; Peer support; Supportive quote; Social media focused on the disease; Thumbs up; Notification when meet goals; Provide assistance	Motivate and support	16 (100)	10 (91)
Tailor to individual needs; Customizable information; Individualized tracking, view progress; Health reminders; Personalize coaching; Healthy tips; Customizable avatar; Rankings—compare to others; Personal point system	Personalize	12 (75)	11 (100)

It was evident throughout our focus group sessions with both youth and young adult participants as well as caregivers that mobile phone use is extremely prevalent and is an integral part of their daily lives. As one of our participants voiced, “I would feel naked if I didn't have my iPhone with me.”

### Make it Easy

#### Participants (Youth and Young Adults With BSA)

Our participants described the difficulty they face when attempting to download and set up a new app on a phone, demonstrating that accessibility is salient in the minds of our respondents. Participants requested options that were “easy to navigate,” “hands free,” “easy to use and reliable,” “fewer keystrokes,” “fast,” “bigger buttons,” and “colorful so that things stand out.” Certain apps they had attempted to use were too difficult and time-consuming to set up. Participants discussed their frustrations with apps needing to be updated or new versions installed, with one young participant summing up their exasperation by stating: “...So I just said ‘forget it’.”

#### Caregivers

Many of our caregivers were sensitive to the difficulties of typing on small phones, setting up apps, and accessing information. One parent stated:

*If you think about it, you could make it so it's a picture rather than needing to read a word. You could make it so that it talks to you rather than needing to read the word.*

Caregivers were just as vocal as our young participants about apps that required too long to program, that continuously updated, or that required the input of information more than one time rather than reusing information. Examples included excessive time and effort to input information before an app could even be used. As one caregiver remarked:

*I mean, I know I don't have that great of patience, but just keep it simple, right?*

### Engage

#### Participants (Youth and Young Adults With BSA)

Participants discussed apps that they found engaging and interesting and noted that being “convenient, consolidated, and helpful” was appealing to them but also requested bright, colorful, and vivid images. Participants spoke of the ways that an app might “get my attention, if I was busy.” One young adult who occasionally uses basic alarms and reminders stated:

*In truthfulness, I don't always take its reminder. I go on. It's just that you're busy in the day, and things happen. You go on to the next thing. Then it rings for 5 minutes. It'll stop. Then it'll be over. Then I forget about it.*

Ideas discussed focused on the use of “fun” reminders such as visually appealing animations, a specific person's voice or person laughing, or the use of one's favorite music. One of our participants suggested that when an alarm or reminder pops up, add a “joke or trivia of the day, famous quotes or something fun; make someone smile.”

#### Caregivers

For many of our caregivers, the features that were suggested to promote engagement focused on ensuring that the mHealth app would serve a useful, relevant purpose while also being “fun and interactive.” Caregivers echoed participants' suggestions that if the app “brought value to my daily life, was user-friendly, and entertaining...,” it would be used more and more throughout the day.

## Educate and Prepare

### Participants (Youth and Young Adults With BSA)

Very few of our participants reported using health-related apps on their phones but admitted the benefits of having access to educational resources that were relevant and conducive to their health needs. Current apps on the market were not always pertinent or informative for our particular participants. One young participant stated:

*I have the condition of Spina bifida. To have an app that would [easily] go into different sites that focus exactly on that condition so that I can be updated on the newest things that are coming up in terms of medications, and treatments, and that type of thing. That would be cool.*

Another participant stated:

*I know there're certain kinds of shunts and stuff. I'm not really aware of what kind of shunt I have, so maybe something that would tell me about that.*

Assistance in tracking personal health issues and providing education on health trends was a recurring theme with our participants. Many of the young adult participants specifically mentioned how worthwhile it would be to have personal health-related information on an app, "all in one place" that they could easily access and present to their doctor. Doctors' appointments can be rushed and the ability to have complete health information or health-related questions available in an organized fashion may increase efficiency in the limited time available.

### Caregivers

Our caregivers also felt strongly about the provision of educational resources and wanted assurance that the information was "searchable with keywords" and "customizable to that individual." Caregivers requested automaticity in the ability to "continually update when new guidelines emerge and get a message about the update." The ability to receive text messages for updates, reminders, and special notices was considered beneficial by our caregivers, who did not want to consistently log back into an app throughout the day to check on important issues. Of great importance was the caregiver's ability to oversee what educational resources their children had access to. One mother noted:

*The parents need input on the information. I can see them thinking they know it all based off what came across the screen. The information should supplement the family, not replace the family, right?*

For caregivers of our younger participants, educational requests that were focused on basic skills were suggested. Interestingly, caregivers were seeking information on safety in daily living skills such as the safest way to get in and out of a shower or information on safety within the community such as how to cross the street or efficiently scan the environment. One caregiver talked about the importance of education by stating:

*...physical community resources. Accessibility. Places a family can go in the community like parks and things like that, that would be great.*

Although caregiver requests for education regarding the diagnosis and basic living skills were common, a number of caregivers also discussed the importance of information regarding insurance and the purchase of medical supplies. The potential of an app that could provide information on where supplies could be purchased for the best price or lowest insurance co-pay, such as a Web-based versus a medical supply store, appeared to be of great value to caregivers. One caregiver suggested:

*An app that would help the consumer like me do a little comparative shopping when we're buying supplies, that kind of thing.*

## Motivate and Support

### Participants (Youth and Young Adults With BSA)

Most of our participants reported using their phones for social media, especially Facebook or Snapchat, and keeping in touch with social acquaintances and family through texting. Social support and feedback was the most consistent and prevalent theme in our youth and young adult-based focus group sessions. Sharing, connecting, and communicating with friends and family to help achieve personal goals was a topic of discussion throughout. One participant mentioned that it would be beneficial to have support for stressful days:

*Yes. Deep breaths or just remind me to bring your head back to what you're doing now instead of worrying about what's happening later.*

Another young participant who echoed the importance of support noted:

*...just talking to someone who's going through the same thing that I am and seeing how they handle it and what they do to calm themselves down and stay sane.*

Interestingly, the ability to motivate others through the meeting of one's goals was also a frequent topic. One participant said:

*I think if you were to give informational tips...that you could share with your friends and family, like healthy tips based on what milestones you've reached through using the app—being able to share how you were successful so that they can also be successful.*

Participants talked at length about social connection. The ability to share goals and milestones with friends and family and to receive feedback or advice on those goals and milestones was a consistent theme. Participants varied in their suggestions on feedback from something as simple as "a little emoticon that's active or that shows an expression or energy level as you're meeting the goals" to more direct feedback such as "Great job using that app for your medications!"

### Caregivers

Caregivers were invested in having the information they needed so that they were fully able to provide support and feedback to

their children/clients. According to a caregiver, such information should include:

*...a report regarding missed medications or not responding to flags, poor health reports.*

A few caregivers were focused on medication management and wanted the ability to use an app “to scan the barcode on the medication bottle,” or for information such as prescribed dosage, side effects, or potential interactions among medications prescribed. Caregivers were also invested in providing feedback to their children when they were “doing a great job on using the app.” One caregiver requested:

*maybe we could have a summary of the data and what they can be praised for, like “Hey great job on keeping your schedule this week”; making it positive for them.*

## Personalize

### Participants (Youth and Young Adults With BSA)

An essential component for any health-related app or device that is focused on personal health and wellness is to “know the user.” The ability to access user-relevant information and set personal goals was a discussion point with our young participants as well as the ability to add or remove information. One participant noted:

*...add information or take away information because health and different things come and go with our history. It should be customizable for that individual person.*

Our participants were also acutely interested in visually following their progress on individualized goals and being able to share that process with others. Tracking progress from start to finish through the use of reports or spreadsheets appeared to be extremely useful for the participants. The ability to compare personal outcomes over time and compete with themselves was exceptionally appealing to our young participants. One participant stated:

*So be able to see your progress in compiled form so you can judge where you are and compare it to where you wanna be.*

A novel idea for personalization and to provide feedback on meeting health-related goals was the use of a customizable avatar—an avatar that matures, develops, and changes as the client learns more, achieves more goals, and becomes more independent in their self-management skills. One of the participants said:

*You can have a little avatar—that would be cool*

Additionally, incorporating monetary and tangible rewards for meeting goals was also relevant to our participants. One participant stated:

*Somehow the points that you would earn for doing healthy things then could be converted into a gift certificate or something you really want.*

Ideas for material rewards included a downloaded movie or music (with less focus on games), money or gift cards, candy, and restaurant or store gift certificates.

## Caregivers

Caregivers understood the importance of providing specific feedback on personal goal attainment and the importance of rewarding “small steps.” Most of our caregivers focused on the use of a point system for rewards and recommended a system that would allow participants to achieve certain points and redeem them with toys, food, certificates, or even stickers.

Finally, a few of our caregivers acknowledged they didn’t “always push for independence; I feel like I’ve held her back some.” Referring to a personalized app focused on self-management and self-care, one of the caregivers noted:

*...would be a great idea to help lift some of that pressure off us.*

Another caregiver stated:

*Using something tailored like this can help them start to gain some of that confidence in being able to do things for themselves.*

## Discussion

### Principal Findings

Focus groups were held with youth and young adults with BSA and their paid or unpaid caregivers in our attempt to better understand how mHealth would be helpful to support them in proactively managing daily self-management routines and accessing medical care as needed [12,15]. Participants shared their personal views and provided detailed information about mHealth apps that would be important for independence in self-care and self-management. Our objective was to use the data gleaned from these focus group sessions to refine the mHealth prototype platform that we have developed, called iMHere 2.0, an app with a suite of modules that can be personalized and tailored to an individual’s needs.

Our findings suggest that most individuals keep their mobile phones with them at all times and typically use a mobile phone for social media, music, photos, and texting. This is consistent with a previous study on the use of cellphone in the general population [33]. This is also good news because the “always carry, always on” connectivity can be harnessed by mHealth apps such as iMHere to support self-management.

On the basis of participant and caregiver comments, a few mHealth-related apps were typically used. Some participants discussed the occasional use of mHealth apps, such as alarms for medical reminders or the use of diet or fitness apps, but consistent or daily use was rarely occurring. In fact, many of our participants voiced how they often ignored basic alarms and suggested the use of something more appealing such as their favorite music, a loved one’s voice, or even a joke of the day when they responded to their reminders. Participants also reported concerns about apps taking too long to program and being difficult and time-consuming to use, along with their frustration with continuous updates.



Recommendations from our participants focused on supporting personal health and wellness and included reminders for attending doctors' appointments, taking medications, and ordering medical supplies, which is consistent with past research [13,15,17]. For our young adult group, suggestions for increasing independence included education on their diagnosis and medical management, as well as stress reduction, time management, and social relationships. One unique feature of iMHere 2.0 is the ability to deliver care according to various clinical practice guidelines. Providing care guidelines that are relevant, customizable, continually updated, and easy to access through the use of basic keywords was an essential component for caregivers. Additionally, tracking health trends and receiving information based on those trends was a common topic of discussion. The ability to receive basic guidance or advice for self-management based on the information being collected through the mHealth app seemed to increase a sense of independence for the participants. The security of knowing that a clinician or provider was available if such self-management strategies were not successful was essential. An additional benefit included allowing their doctor to have access to these records and to view trends before a health care appointment.

The ability to develop personal self-management goals and to visually track progress toward those goals was also a function that appeared relevant to our focus group participants. Moreover, personalized coaching was an important component to assist with goal attainment and to provide affirmation and feedback. Interestingly, the greatest motivator for long-term use of the mHealth system was not tangible (ie, money, games, and free movie), instead it was the ability to engage with peers and family and share experiences regarding goal attainment and milestones. Sharing their successes so that others may learn about and acknowledge their accomplishments appeared to be exceptionally motivating. Social support and social recognition was consistently discussed as a reinforcer for the development of self-management skills.

### Limitations

It's important to note that there are limitations in our focus group methods. As recruitment strategies were mainly carried out through our existing collaborations with human services agencies and personal referrals, participants not involved with these agencies may not have been fully represented in the focus

group sessions. Therefore, our participants' experiences may not be universally representative of all youth/young adults and caregivers in the BSA community. In addition, our inability to recruit minority participants limits our findings. It is also important to mention that our focus group sessions were held in closed meeting rooms, and not in natural environments, which may have changed the behavior of some of our participants. As our focus group sessions were one-time only with each participant, data collection techniques only included fieldwork observations, follow-up surveys, and debriefing sessions with the focus group moderators. Each of the authors was directly involved in all focus group sessions and analyses of the data, which may elicit bias into the study and may have shaped the interpretation and understanding of these data.

### Conclusions

These focus group sessions informed refinements in the iMHere 2.0 system. For example, we are "making it easy" through automatic updates, accessibility preferences, and simplicity in the navigation of the system. We are "engaging" youth, young adults, and caregivers through color coding and special icons within iMHere 2.0 and using a variety of personally selected avatars and sounds for alarms. We are "educating and preparing" through the incorporation of resources that are inclusive of medical and psychosocial elements of the diagnosis and that are searchable and customizable for each individual. We are "motivating and supporting" through the ability to text within the system and to receive feedback and encouragement on goals and self-management tasks. We are "personalizing" through customizing the app for each user's needs, tracking progress on individual goals, preparing reports on health care trends, and providing a rewards center for earning points when using the app to manage individualized health care needs.

Our qualitative analysis indicates that youth and young adults with BSA, as well as their caregivers, acknowledge the importance of being actively engaged in developing and using mHealth apps that monitor and manage their health care needs. Future research will continue to refine iMHere 2.0 through usability studies, which will lead to a longitudinal clinical trial. Through the use of focus groups, our findings will allow us to move forward with quantitative procedures that are meaningful and appropriate for our population.

### Acknowledgments

We wish to thank the families who participated in the focus group sessions; without the engagement of these families in the research process, this project could not move forward. We also wish to acknowledge the research team members who helped coordinate and run each of the focus group sessions, including Dr Andi Saptono, I Made Agus Setiawan, Zara Ambadar, Delhari DeAlmeira, and Zakiy Alfikri. The contents of this publication were supported, in part, by the grants from the National Institute on Disability, Independent Living, and Rehabilitation Research (NIDILRR; grant numbers 90DP0064 and 90RE5018), which is a center within the Administration for Community Living (ACL), US Department of Health and Human Services (HHS). The contents of this publication do not necessarily represent the policy of NIDILRR, ACL, and HHS, and you should not assume endorsement by the federal government.

### Conflicts of Interest

None declared.



## Multimedia Appendix 1

Informed consent form - youth.

[[PDF File \(Adobe PDF File\), 124KB - mhealth\\_v5i9e141\\_app1.pdf](#)]

## Multimedia Appendix 2

Informed Consent Form - caregivers.

[[PDF File \(Adobe PDF File\), 117KB - mhealth\\_v5i9e141\\_app2.pdf](#)]

## Multimedia Appendix 3

Focus group guide - youth.

[[PDF File \(Adobe PDF File\), 59KB - mhealth\\_v5i9e141\\_app3.pdf](#)]

## Multimedia Appendix 4

Focus Group Guide - caregivers.

[[PDF File \(Adobe PDF File\), 62KB - mhealth\\_v5i9e141\\_app4.pdf](#)]

## References

1. Lorig KR, Holman H. Self-management education: history, definition, outcomes, and mechanisms. *Ann Behav Med* 2003 Aug;26(1):1-7. [Medline: [12867348](#)]
2. Audulv Å. The over time development of chronic illness self-management patterns: a longitudinal qualitative study. *BMC Public Health* 2013 May 07;13:452 [FREE Full text] [doi: [10.1186/1471-2458-13-452](#)] [Medline: [23647658](#)]
3. Dicianno BE, Parmanto B, Fairman AD, Crytzer TM, Yu DX, Pramana G, et al. Perspectives on the evolution of mobile (mHealth) technologies and application to rehabilitation. *Phys Ther* 2015 Mar;95(3):397-405. [doi: [10.2522/ptj.20130534](#)] [Medline: [24925075](#)]
4. Binks JA, Barden WS, Burke TA, Young NL. What do we really know about the transition to adult-centered health care? a focus on cerebral palsy and spina bifida. *Arch Phys Med Rehabil* 2007 Aug;88(8):1064-1073. [doi: [10.1016/j.apmr.2007.04.018](#)] [Medline: [17678671](#)]
5. Kieckhefer GM, Trahms CM. Supporting development of children with chronic conditions: from compliance toward shared management. *Pediatr Nurs* 2000;26(4):354-363. [Medline: [12026469](#)]
6. Mechling LC. Assistive technology as a self-management tool for prompting students with intellectual disabilities to initiate and complete daily tasks. *Education and Training in Developmental Disabilities* 2007;42(3):252-269 [FREE Full text]
7. Lindsay S, Kingsnorth S, McDougall C, Keating H. A systematic review of self-management interventions for children and youth with physical disabilities. *Disabil Rehabil* 2014;36(4):276-288 [FREE Full text] [doi: [10.3109/09638288.2013.785605](#)] [Medline: [23614359](#)]
8. Sawyer SM, Aroni RA. Self-management in adolescents with chronic illness. What does it mean and how can it be achieved? *Med J Aust* 2005 Oct 17;183(8):405-409. [Medline: [16225444](#)]
9. Dicianno BE, Fairman AD, Juengst SB, Braun PG, Zabel TA. Using the spina bifida life course model in clinical practice: an interdisciplinary approach. *Pediatr Clin North Am* 2010 Aug;57(4):945-957. [doi: [10.1016/j.pcl.2010.07.014](#)] [Medline: [20883884](#)]
10. Kinne S, Patrick DL, Doyle DL. Prevalence of secondary conditions among people with disabilities. *Am J Public Health* 2004 Mar;94(3):443-445. [doi: [10.2105/AJPH.94.3.443](#)]
11. Verhoef M, Barf HA, Post MW, van Asbeck FW, Gooskens RH, Prevo AJ. Secondary impairments in young adults with spina bifida. *Dev Med Child Neurol* 2004 Jun;46(6):420-427 [FREE Full text] [Medline: [15174535](#)]
12. Dicianno B, Peele P, Lovelace J, Fairman A, Smyers D, Halgas M, et al. Amga. University of Pittsburgh Medical Center Case Study: University of Pittsburgh; 2012. Best practices in managing patients with multiple chronic conditions URL: <http://www.amga.org/PI/Collabs/MPMCC/Compendiums/UniversityofPittsburgh.pdf> [accessed 2017-08-30] [WebCite Cache ID 6t6UxjDag]
13. Free C, Phillips G, Galli L, Watson L, Felix L, Edwards P, et al. The effectiveness of mobile-health technology-based health behaviour change or disease management interventions for health care consumers: a systematic review. *PLoS Med* 2013 Jan;10(1):e1001362 [FREE Full text] [doi: [10.1371/journal.pmed.1001362](#)] [Medline: [23349621](#)]
14. Kiasnja P, Pratt W. Healthcare in the pocket: mapping the space of mobile-phone health interventions. *J Biomed Inform* 2012;45(1):184-198. [doi: [10.1016/j.jbi.2011.08.017](#)]
15. Fiordelli M, Diviani N, Schulz PJ. Mapping mHealth research: a decade of evolution. *J Med Internet Res* 2013 May 21;15(5):e95 [FREE Full text] [doi: [10.2196/jmir.2430](#)] [Medline: [23697600](#)]

16. Free C, Phillips G, Felix L, Galli L, Patel V, Edwards P. The effectiveness of M-health technologies for improving health and health services: a systematic review protocol. BMC Res Notes 2010;3:250 [FREE Full text] [doi: [10.1186/1756-0500-3-250](https://doi.org/10.1186/1756-0500-3-250)] [Medline: [20925916](https://pubmed.ncbi.nlm.nih.gov/20925916/)]
17. Free C, Phillips G, Watson L, Galli L, Felix L, Edwards P, et al. The effectiveness of mobile-health technologies to improve health care service delivery processes: a systematic review and meta-analysis. PLoS Med 2013 Jan;10(1):e1001363 [FREE Full text] [doi: [10.1371/journal.pmed.1001363](https://doi.org/10.1371/journal.pmed.1001363)] [Medline: [23458994](https://pubmed.ncbi.nlm.nih.gov/23458994/)]
18. de Jongh T, Gurol-Urganci I, Vodopivec-Jamsek V, Car J, Atun R. Mobile phone messaging for facilitating self-management of long-term illnesses. Cochrane Database Syst Rev 2012(12):CD007459. [doi: [10.1002/14651858.CD007459.pub2](https://doi.org/10.1002/14651858.CD007459.pub2)] [Medline: [23235644](https://pubmed.ncbi.nlm.nih.gov/23235644/)]
19. Parmanto B, Pramana G, Yu DX, Fairman AD, Dicianno BE, McCue MP. iMHere: a novel mhealth system for supporting self-care in management of complex and chronic conditions. JMIR Mhealth Uhealth 2013 Jul 11;1(2):e10 [FREE Full text] [doi: [10.2196/mhealth.2391](https://doi.org/10.2196/mhealth.2391)] [Medline: [25100682](https://pubmed.ncbi.nlm.nih.gov/25100682/)]
20. Dicianno BE, Fairman AD, McCue M, Parmanto B, Yih E, McCoy A, et al. Feasibility of using mobile health to promote self-management in Spina Bifida. Am J Phys Med Rehabil 2016 Jun;95(6):425-437. [doi: [10.1097/PHM.0000000000000400](https://doi.org/10.1097/PHM.0000000000000400)] [Medline: [26488144](https://pubmed.ncbi.nlm.nih.gov/26488144/)]
21. Fairman AD, Dicianno BE, Datt N, Garver A, Parmanto B, McCue M. Outcomes of clinicians, caregivers, family members and adults with Spina Bifida regarding receptivity to use of the iMHere mhealth solution to promote wellness. Int J Telerehabil 2013 Jun;5(1):3-16 [FREE Full text] [doi: [10.5195/ijt.2013.6116](https://doi.org/10.5195/ijt.2013.6116)] [Medline: [25945209](https://pubmed.ncbi.nlm.nih.gov/25945209/)]
22. Rabiee F. Focus-group interview and data analysis. Proc Nutr Soc 2007;63(4):655-660. [doi: [10.1079/PNS2004399](https://doi.org/10.1079/PNS2004399)] [Medline: [15831139](https://pubmed.ncbi.nlm.nih.gov/15831139/)]
23. Kitzinger J. The methodology of focus groups: the importance of interactions between research participants. Sociol Health Illn 1994;16:103-121. [doi: [10.1111/1467-9566.ep11347023](https://doi.org/10.1111/1467-9566.ep11347023)]
24. Anderson C. Presenting and evaluating qualitative research. Am J Pharm Educ 2010 Sep;74(8):1-7. [doi: [10.5688/aj7408141](https://doi.org/10.5688/aj7408141)]
25. Bryman A, Burgess RH. Qualitative data analysis for applied policy research. In: Analyzing qualitative data. London: Routledge; 1994:305-329.
26. Boeije H. Analysis in Qualitative Research. Thousand Oaks, CA: Sage Publications Ltd; 2009.
27. Shenton AK. Strategies for ensuring trustworthiness in qualitative research projects. EFI 2004 Jul 19;22(2):63-75. [doi: [10.3233/EFI-2004-22201](https://doi.org/10.3233/EFI-2004-22201)]
28. Sandelowski M. Qualitative analysis: what it is and how to begin. Res Nurs Health 1995 Aug;18(4):371-375. [doi: [10.1002/nur.4770180411](https://doi.org/10.1002/nur.4770180411)]
29. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: [10.1191/1478088706qp063oa](https://doi.org/10.1191/1478088706qp063oa)]
30. Guest G. How many interviews are enough?: an experiment with data saturation and variability. Field Methods 2006 Feb 01;18(1):59-82. [doi: [10.1177/1525822X05279903](https://doi.org/10.1177/1525822X05279903)]
31. Patton M. Qualitative research and evaluation methods. Thousand Oaks, CA: Sage Publications; 2002.
32. Rainie L, Zickuhr K. Pew Research Center. Washington, DC: Pew Research Center; 2015. Americans' views on mobile etiquette URL: <http://www.pewinternet.org/2015/08/26/americans-views-on-mobile-etiquette/> [accessed 2017-08-07] [WebCite Cache ID 6sXZcbuw4]
33. Krueger R, Casey M. Focus groups: a practical guide for applied research. Thousand Oaks, CA: Sage Publications; 2000.

## Abbreviations

**ACL:** Administration for Community Living  
**BSA:** brain and spinal cord anomalies  
**CLASS:** Community Living and Support Services  
**HHS:** US Department of Health and Human Services  
**iMHere:** interactive Mobile Health and Rehabilitation  
**IRB:** institutional review board  
**mHealth:** mobile health  
**NIDILRR:** National Institute on Disability, Independent Living, and Rehabilitation Research  
**SB:** spina bifida  
**SBAWP:** Spina Bifida Association of Western Pennsylvania  
**SCI:** spinal cord injury

*Edited by G Eysenbach; submitted 09.12.16; peer-reviewed by M Losa Iglesias, K Becerro de Bengoa Losa, A Athanasiou; comments to author 22.01.17; revised version received 27.02.17; accepted 15.07.17; published 26.09.17.*

*Please cite as:*

*Bendixen RM, Fairman AD, Karavolis M, Sullivan C, Parmanto B*

*A User-Centered Approach: Understanding Client and Caregiver Needs and Preferences in the Development of mHealth Apps for Self-Management*

*JMIR Mhealth Uhealth 2017;5(9):e141*

URL: <http://mhealth.jmir.org/2017/9/e141/>

doi: [10.2196/mhealth.7136](https://doi.org/10.2196/mhealth.7136)

PMID: [28951378](https://pubmed.ncbi.nlm.nih.gov/28951378/)

©Roxanna M Bendixen, Andrea D Fairman, Meredith Karavolis, Carly Sullivan, Bambang Parmanto. Originally published in JMIR Mhealth and Uhealth (<http://mhealth.jmir.org>), 26.09.2017. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mhealth and uhealth, is properly cited. The complete bibliographic information, a link to the original publication on <http://mhealth.jmir.org/>, as well as this copyright and license information must be included.

Original Paper

# Mobile Augmented Reality as a Feature for Self-Oriented, Blended Learning in Medicine: Randomized Controlled Trial

Christoph Noll<sup>1\*</sup>; Ute von Jan<sup>1\*</sup>, Dr rer biol hum; Ulrike Raap<sup>2\*</sup>, Dr med; Urs-Vito Albrecht<sup>1\*</sup>, Dr med, MPH

<sup>1</sup>PL Reichertz Institute for Medical Informatics, Hannover Medical School, Hannover, Germany

<sup>2</sup>Universitätsklinik für Dermatologie und Allergologie, Klinikum Oldenburg AöR, Oldenburg, Germany

\* all authors contributed equally

**Corresponding Author:**

Urs-Vito Albrecht, Dr med, MPH  
PL Reichertz Institute for Medical Informatics  
Hannover Medical School  
Carl-Neuberg-Str. 1  
Hannover, 30625  
Germany  
Phone: 49 511 532 ext 3508  
Fax: 49 511 532 2517  
Email: [albrecht.urs-vito@mh-hannover.de](mailto:albrecht.urs-vito@mh-hannover.de)

## Abstract

**Background:** Advantages of mobile Augmented Reality (mAR) application-based learning versus textbook-based learning were already shown in a previous study. However, it was unclear whether the augmented reality (AR) component was responsible for the success of the self-developed app or whether this was attributable to the novelty of using mobile technology for learning.

**Objective:** The study's aim was to test the hypothesis whether there is no difference in learning success between learners who employed the mobile AR component and those who learned without it to determine possible effects of mAR. Also, we were interested in potential emotional effects of using this technology.

**Methods:** Forty-four medical students (male: 25, female: 19, mean age: 22.25 years, standard deviation [SD]: 3.33 years) participated in this study. Baseline emotional status was evaluated using the Profile of Mood States (POMS) questionnaire. Dermatological knowledge was ascertained using a single choice (SC) test (10 questions). The students were randomly assigned to learn 45 min with either a mobile learning method with mAR (group A) or without AR (group B). Afterwards, both groups were again asked to complete the previous questionnaires. AttrakDiff 2 questionnaires were used to evaluate the perceived usability as well as pragmatic and hedonic qualities. For capturing longer term effects, after 14 days, all participants were again asked to complete the SC questionnaire. All evaluations were anonymous, and descriptive statistics were calculated. For hypothesis testing, an unpaired signed-rank test was applied.

**Results:** For the SC tests, there were only minor differences, with both groups gaining knowledge (average improvement group A: 3.59 [SD 1.48]; group B: 3.86 [SD 1.51]). Differences between both groups were statistically insignificant (exact Mann Whitney U,  $U=173.5$ ;  $P=.10$ ;  $r=.247$ ). However, in the follow-up SC test after 14 days, group A had retained more knowledge (average decrease of the number of correct answers group A: 0.33 [SD 1.62]; group B: 1.14 [SD 1.30]). For both groups, descriptively, there were only small variations regarding emotional involvement, and learning experiences also differed little, with both groups rating the app similar for its stimulating effect.

**Conclusions:** We were unable to show significant effects for mAR on the immediate learning success of the mobile learning setting. However, the similar level of stimulation being noted for both groups is inconsistent with the previous assumption of the success of mAR-based approach being solely attributable to the excitement of using mobile technology, independent of mAR; the mAR group showed some indications for a better long-term retention of knowledge. Further studies are needed to examine this aspect.

**Trial Registration:** German Clinical Trials Register (DRKS): 00012980; [http://www.drks.de/drks\\_web/navigate.do?navigationId=trial.HTML&TRIAL\\_ID=DRKS00012980](http://www.drks.de/drks_web/navigate.do?navigationId=trial.HTML&TRIAL_ID=DRKS00012980) (Archived by WebCite at <http://www.webcitation.org/6tCWom2Jb>).

(JMIR Mhealth Uhealth 2017;5(9):e139) doi:[10.2196/mhealth.7943](https://doi.org/10.2196/mhealth.7943)

**KEYWORDS**

problem-based learning; cellular phone; education; medical; mHealth

## Introduction

According to authors such as Johnson et al [1] and Kroeker [2], augmented reality (AR) will become one of the major user interfaces of the 21st century. AR allows real and virtual objects to coexist and interact in the same space and time [3]. Using AR, virtual information can be interwoven with reality, which leads to an augmentation of the physical environment. Thanks to the ready and still growing availability of smartphones and tablets and their ever-increasing processing power, AR can now be used in a mobile manner (ie, mobile Augmented Reality [mAR]) as well. Whereas previously, AR was mainly of relevance for entertainment, marketing, or video games, it is now also entering the challenging field of teaching and training. One significant benefit of mAR for learning is the ease of modeling objects and presenting them to learners in real-world settings, so that they can get a clear idea about what they are to learn [4], and there are various studies evaluating the effects this technology has on the learning process for various user groups and settings [4-7].

In preparatory work done at Hannover Medical School, there was already an initial investigation into the possible uses of mAR for teaching and learning in a medical education setting [6]. For this purpose, a *mobile Augmented Reality blended learning environment* (mARble) app was built, which was then evaluated in comparison with conventional learning material (textbook), specifically with respect to its learning efficiency [6]. Despite the low number of cases (n=10) for that pilot study, it was possible to show positive activation for those participants who had been learning with the mAR app, and when checking the participants' knowledge gain, the mAR group performed significantly better than those who had learned with the conventional textbook material [6].

However, it remained unclear whether this activation had to be attributed to using a different medium and its exciting novelty. Initially, it was unclear to what extent mAR had actually contributed to the learning success, a problem also mentioned by Radu [4] when contemplating the effects of different media—with entirely different means of presentation—on learning. With this study, we wanted to address this issue.

## Methods

### The Learning Environment mARble

The iPhone operating system (iOS, Apple Inc)-based app mARble-Derma (mARble-dermatology) was developed at the Peter L. Reichertz Institute for Medical Informatics of Hannover Medical School, in collaboration with Ulrike Raap, formerly of the Clinic for Dermatology and Allergy at Hannover Medical

School, and her team at the clinic. It provides users with learning content organized in the form of digital flashcards. Using paper-based markers that can be placed on the skin of users, the app employs AR to recall content linked to the markers, overlay it on images of the environment if desired, and to thus add an entirely new level of information [6]. The app's code and its content are kept separately. Via an extensible markup language-based file format, content can easily be edited or added without changing the code [6,8].

### Learning Material

The subject of dermatology was chosen for the study, as dermatology is a specialty where visual information is of high relevance when it comes to diagnosing various skin conditions, making it ideal for AR-based scenarios. The lecturer for dermatology selected altogether five relevant topics (malignant melanoma, basal cell carcinoma, psoriasis vulgaris, bullous pemphigoid, and atopic dermatitis) from the learning catalog. The learning material for the selected topics was adapted from relevant literature [9], as well as the course material normally provided to students by the department. In close collaboration with the lecturer, it was then integrated into the app. All images originated in the department and were professionally produced for teaching purposes.

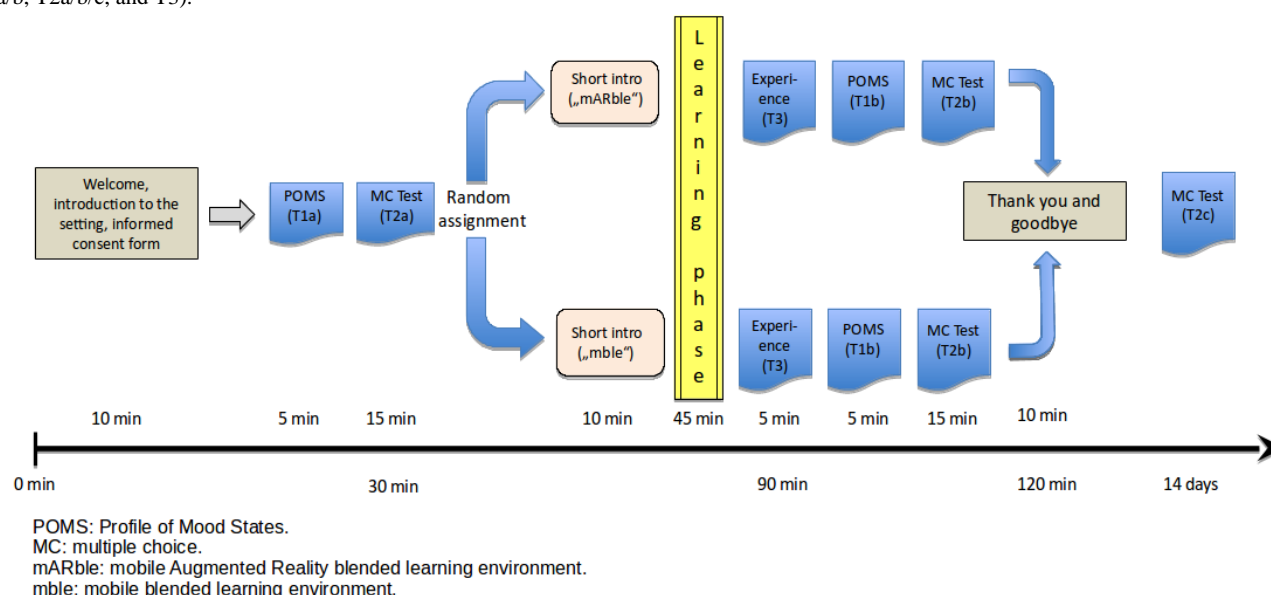
### Fine-Tuning the Content: Selecting the AR Markers and Their Corresponding Content

For selecting a suitable subset out of the available markers and flashcards, a randomized single-blinded questionnaire was employed. For each of the available markers, this questionnaire contained images that had been acquired by overlaying the respective finding onto the skin of a test subject using the app. These images were then rated by 16 doctors (9 junior doctors and 7 dermatologists) working at the clinic for Dermatology and Allergy of Hannover Medical School. For each image, the doctors were asked to give a free text answer stating their diagnosis. A subsequent analysis of interrater reliability [10] led to the aforementioned selection from originally 10 markers and 6 subject areas. With one exception, only markers that were correctly recognized and had shown an interrater reliability of at least 60% were included. The marker and the subject area for “atopic dermatitis” (item 2) were included despite poor reliability (46% [6/13]); whereas location is an important aspect when diagnosing this condition, it could not adequately be deduced from the presented image. It is to be expected that with a more carefully chosen view better depicting the location, the association of the presented image with the correct diagnosis would have been more reliable, as the term “eczema,” which also covers “atopic dermatitis,” was often used to describe the depicted finding. Altogether, eight markers from five subject areas were finally included (Table 1).



**Table 1.**

Item	N	Correct answers	Specialist		Incorrect answers	Specialist		Correct answers, % (n/N)
			yes	no		yes	no	
Item 1 melanoma metastases	14	9	4	5	5	2	3	64 (9/14)
Item 2 atopic dermatitis	13	6	2	4	7	4	3	46 (6/13)
Item 3 psoriasis vulgaris, single spot	16	14	6	8	2	1	1	88 (14/16)
Item 4 basal cell carcinoma, pigmented	16	3	3	0	13	4	9	19 (3/16)
Item 5 nodular melanoma	15	10	4	6	5	3	2	67 (10/15)
Item 6 basal cell carcinoma	16	16	7	9	0	0	0	100 (16/16)
Item 7 superficial spreading melanoma	14	9	2	7	5	3	2	64 (9/14)
Item 8 bullous pemphigoid	15	14	5	9	1	1	0	93 (14/15)
Item 9 urticaria	10	5	2	3	5	1	4	50 (5/10)
Item 10 psoriasis vulgaris, multiple spots	13	8	3	5	5	1	4	62 (8/13)

**Figure 1.** Schedule and tests that were performed. Throughout the text, individual steps or tests are referenced with the labels shown in this figure (T1a/b, T2a/b/c, and T3).

## Objective

The hypothesis to be tested in the study was that there is no significant difference in the score of correct answers (learning success) between learners who have access to mAR and those who do not. In addition, it was of interest whether there were indicators hinting at better long-term retention of acquired knowledge for those who had learned with mAR. We were also interested in whether the emotional involvement seen in the prestudy could be reproduced.

## Study

The study was conducted with approval by the institutional review board of Hannover Medical School, study number 1823-2013, amended 2014. For this study, it was decided to use the design of a two-arm, prospective randomized trial. There were two study groups, both of which were equipped with smart devices (iOS-based smartphones and tablets, specifically iPads, iPad Mini tablets, iPhones 4, or iPhones 5) with preinstalled copies of the mobile learning environment. For both groups,

the software was identical, with the exception of the mAR functionality, which was only provided to one group (Figure 1).

## Sample Size Calculation

Experiences from our previous study [6] had shown that recruiting students for extracurricular activities such as participation in a study is extremely difficult. We therefore decided to take a conservative approach in our calculations, leading to a reasonable (and realistically obtainable) number of participants while still keeping the power at an acceptable level. On the basis of our previous results [6], the sample size required for Mann Whitney *U* testing (unpaired rank sum, two-sided, effect size  $d=0.73$ , Laplace distribution, minimum power of .8) was calculated with G\*Power 3.1 (Heinrich-Heine-Universität Düsseldorf) [11,12], leading to 21 individuals per group (altogether 42 participants). However, we chose to recruit 2 additional candidates to be able to compensate for spontaneous dropouts, at least for the initial part of the study.

**Table 2.** Descriptive statistics of the participants (N=44).

Group	Gender	n	Age in years, mean (SD <sup>a</sup> )	iPhone owners	Other smartphone or tablet
A	Female	9	21.45 (2.39)	9	14
	Male	13			
B	Female	16	23.05 (3.97)	16	8
	Male	6			

<sup>a</sup>SD: standard deviation.

### Study Population

A total of 44 third-year medical students (25 females, 19 males, mean age=22.25 years [SD 3.33]) were included in the study. None of them had previously finished the dermatology module (Table 2).

### Implementation

After all the participants had given their consent to being included in the study, they were given a brief introduction into the study's topic and its schedule. Following this initial introduction, the participants were allocated to the two study groups by letting them choose a random envelope containing information about their assignment to one of the two groups, their individual study ID, and the questionnaires used in the study. These envelopes had been prepared by the study team beforehand. These were sealed, with no labeling or other discernible markings on the outside that could have provided a hint as to their specific content, and they were also mixed before being presented to the participants. Before the students opened their chosen envelopes, it was not possible for either the students or the study team to determine which group assignment was given by the envelopes' contents.

To assess the initial emotional status of the students, the participants were asked to fill out a German version of the "Profile of Mood States" (POMS) questionnaire [13,14]. As shown in Figure 1, they were given 5 min for answering this questionnaire (T1a). To obtain a baseline about their knowledge regarding the subject areas, they were also asked to answer a single choice (SC) test consisting of 10 questions (T2a), for which they were given 15 min.

Whereas the setting was otherwise identical, group A learned with mAR and group B without the mAR component. Both groups were led into two different rooms where they were again given a brief introduction, this time into the basic operation of the app mARble (Figure 2). The students were then equipped with mobile devices (one per individual) on which the app had been preinstalled. They also received headphones for individual use. The participants were then allowed to study using the app for a time span of 45 min (Figure 1). Group B simply used the flashcard-based material containing textual information as well as corresponding images (Figure 2). Members of group A were given the opportunity to use the additional markers (Figure 2), for example, to place them on their own bodies, to view the corresponding findings overlaid on their skin, and to quickly access the same textual as well as image data as group B. All participants learned at their own pace. For both groups, members

of the study team were present to quietly observe the learning process and to be able to react to potential technical problems.

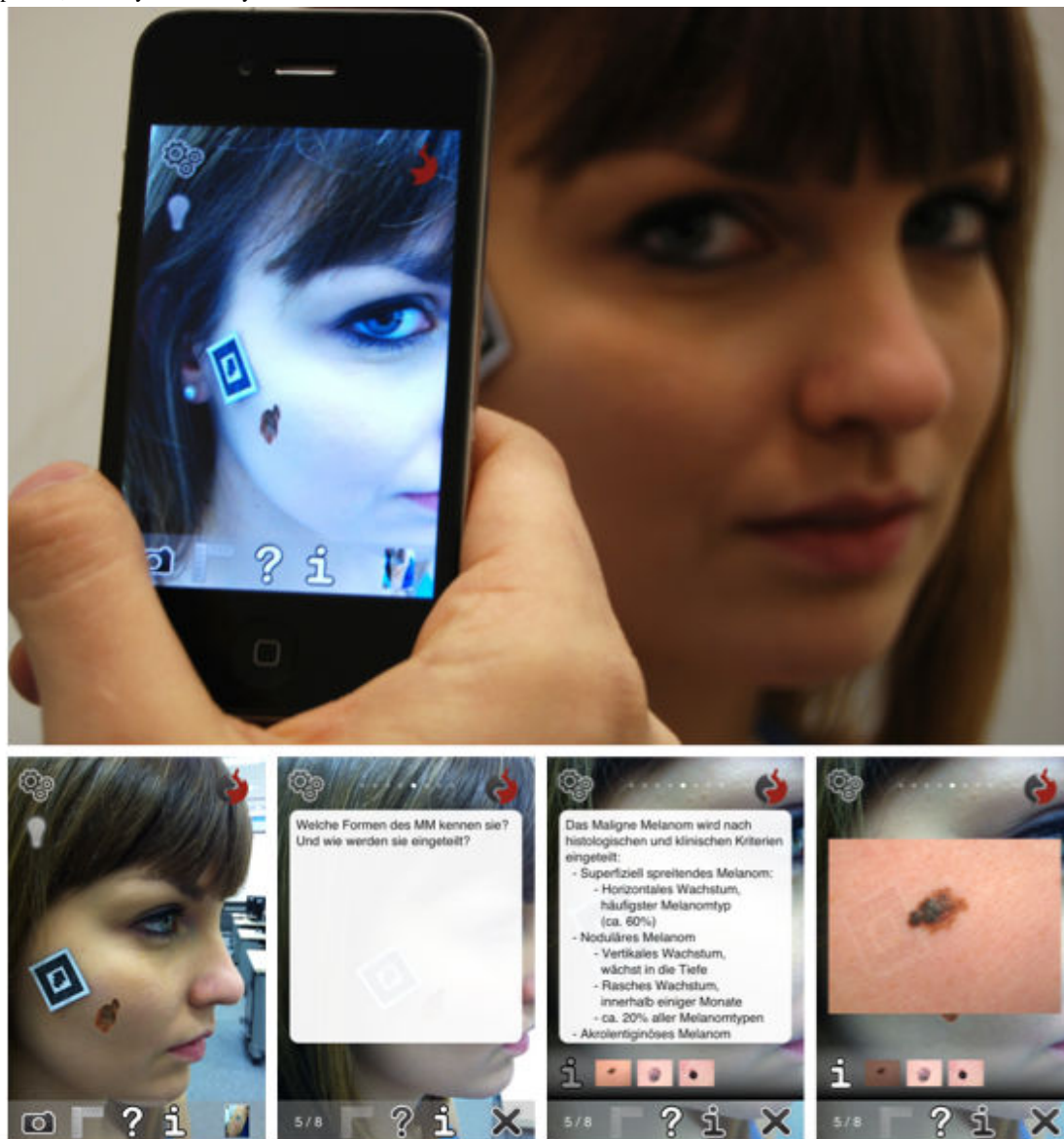
The control group B was provided with the (content-wise) same app as group A, but the members of this group were not given any markers that they could have used to trigger the mAR-based functionalities of the app. They were only told about how they could access the provided content (flashcards) using the app's navigation menu (Figure 2). During the learning phase, the participants of group B were allowed to learn at their own pace, without any interaction with other members of their group, and to take notes on paper if they wished to do so. Following the learning phase, the participants were asked to complete a questionnaire (AttrakDiff 2, T3) covering user experience-related aspects of what they had just experienced [15]. They were also asked to once again fill out the POMS questionnaire (T1b) about their emotional status. For filling out both questionnaires, they were given 10 min. Finally, to determine how much they had learned, they were once again asked to answer the SC test consisting of 10 questions, with the questions being presented in a random order (T2b).

Similar to group B, group A was briefed about using the app, the included flashcards, as well as attachments. Additionally, they were familiarized with using the paper-based markers serving as triggers for the mAR-based functions of the app. The participants were asked to use all of the provided eight markers for the five subject areas by placing them on their own skin and to also utilize the markers for "help" and "contact." For two subject areas, multiple markers were available. There were three markers for "malignant melanoma" and two markers for "psoriasis vulgaris." Similar to group B, for the study phase of 45 min duration, group A was asked to study individually, without any interaction with other members of their group, and taking paper-based notes was also allowed. After finishing the learning phase, again, similar to group B, they were also asked to complete the user experience (AttrakDiff 2, T3) and POMS questionnaires (T1b), as well as the 10-question SC test (T2b) about the five subject areas they had learned.

After reminding the participants about the Web-based follow-up survey (T2c), planned for 14 days after the day the study had taken place, everyone was thanked and the study group B were informed that they would get an opportunity to experience the full functionality of mARble, including mAR, at a later date if they so desired.

For the follow-up survey (T2c), the participants were invited via an email that contained an individual link, leading them to a Web-based version with the same questionnaire as before, consisting of 10 questions presented in a random order.

**Figure 2.** Upper row: Using mobile Augmented Reality (mAR), a malignant melanoma is simulated on the cheek of a student. Lower row: Screenshots taken within the mobile Augmented Reality blended learning environment (mARble) app. Left to right: Overlaid image, question side of a flashcard, answer side with links to additional image material, and a presentation of said image material. For the control group B, the app was provided without the mAR component, and they were solely able to access the flashcard-based information.



## Evaluation Tools

### *Emotional Involvement (T1a+T1b): POMS Questionnaire, German Version*

Similar to the previous study by Albrecht et al [6,8], before and immediately after the learning phase, the emotional status of the students was evaluated based on the POMS questionnaire [13]. It was applied in its German, slightly modified version, as described by Biel et al [14]. This questionnaire contains 35 adjectives that can be divided into groups associated with four different emotional states, that is, fatigue–inertia (14 items), vigor–activity (7 items), tension–anxiety (7 items), and depression–dejection (7 items). Ratings are assigned based on a 7-point rating scale representing the experienced intensity (ranging from “not at all” to “very strongly”).

### *Learning Success: Single Choice Tests (T2a, T2b, and T2c)*

The learning outcome was evaluated by means of the aforementioned paper-based SC tests (single choice answers) consisting of 10 questions. There were 88 test forms, with questions and answers being presented in a random order. For the follow-up survey, a Web-based questionnaire was used, which participants were able to access using their participant ID as well as a password they had received at the beginning of the study. As the participant IDs had been randomly assigned to the students—the IDs and corresponding passwords were noted on a slip of paper in the envelope the students had chosen themselves at the beginning—it was not possible to identify individual students.

The questions employed for testing closely followed the methodology also used in official exams for medical students as they are compiled by the German Institute for Medical and

Pharmaceutical Examination Questions (Institut für medizinische und pharmazeutische Prüfungsfragen). The questions' language and content were adapted to reflect the material provided in the lecture notes available for the dermatology and allergy class at Hannover Medical School, and they were checked for correctness and solvability by the module's lecturer. The content provided by the app was also checked with respect to whether it was adequate for solving the test questions and whether it was presented in a manner that made it possible to go through all of this content within the given time frame of 45 min.

On the basis of the tests conducted before and after the learning phase, the learning efficiency (T2a, T2b, and T2c) for both groups was evaluated descriptively using the calculated mean values and corresponding SD. For hypothesis testing, a nonparametrical signed-rank test for unpaired samples was conducted (exact Mann Whitney *U* test) with Statistical Package for the Social Sciences (SPSS) version 24 (IBM Corp). All questionnaires were included in this evaluation, and each of them had been fully completed. For the follow-up survey, only those questionnaires were included in the evaluation that had been completed in the time span between the start of the follow-up period (after 14 days) up to 8 days later. Missing questionnaires were replaced by the mean values calculated for the respective group.

### **Learning Experience (T3): AttrakDiff 2**

Isleifsdóttir et al [16] describe "user experience" as an important factor to consider when designing software. In their previous, preliminary evaluation of mARble, Albrecht et al [6] employed the AttrakDiff 2 questionnaire as described by Hassenzahl et al [15,17,18] to evaluate this aspect of the app. It uses altogether 28 questions, covering four different aspects (pragmatic quality, PQ; hedonic quality (HQ)-stimulation, HQ-S; hedonic quality-identification, HQ-I; and attractiveness, ATT), with 7 questions per group. For each item, semantic differentials are used, with opposite adjectives (eg, "good-bad" and "confusing-clear") being placed at the poles of a 7-point Likert scale. In the work presented here, using a similar setting, the AttrakDiff 2 questionnaire was again used to evaluate the app's ATT, as well as its hedonic and pragmatic qualities.

For each of the 28 attributes included in the questionnaire, mean values as well as corresponding SDs were calculated for the ratings given by the participants. For each dimension, average ratings were calculated and plotted for clarity. The values for PQ (on the x-axis) were plotted against those obtained for HQ (aggregated from the values obtained for hedonic stimulation and hedonic identification, placed on the y-axis). By including the corresponding confidence intervals into the plot, rectangles are shown that allow asserting to what extent the user experiences between both groups differ or overlap.

### **Analyzing User Behavior Based on Log Files Recorded on the Devices**

To provide insights into how the participants had learned, the usage of markers as well as the included flashcards were tracked via the logging functionality integrated into the app. The recorded data included the date and time at which a marker or flashcard had been used, the type of the event (marker in focus,

flashcard being invoked), the title of the marker or flashcard being used, as well as the duration of the event in seconds. As there were multiple flashcards per subject, for the flashcards, a numeric identifier was recorded as well. It was also noted whether the answer or question side of the flashcard had been displayed.

For all participants, the log files recorded during the learning phase were transferred to a central database. How long the provided flashcard content had been utilized (median values and interquartile range [IQR]) was then calculated for each group, in aggregated form as well as per flashcard (stratified for questions and answers) and per participant. For group A, median values and IQRs for the markers were calculated as well.

## **Results**

### **Item Analysis: Single Choice Tests (T2a, T2b, and T2c)**

The three SC tests were subjected to an item analysis to determine their difficulty and selectivity. For both groups, for each of the questions in a test, a difficulty index *p* was calculated with the following formula:

$$p = N_C / N$$

(*N<sub>C</sub>*=number of participants with a correct answer, *N*=number of participants in the group). A selectivity index *r* as point-biserial correlation (*r<sub>p.bis</sub>*) was calculated for each test as well.

For the pretest T2a, *p* was .7682 for group A and .7782 for group B. Thus, initially, the overall difficulty for both groups was almost identical, despite differences on a per-question level, which, however, is to be expected to be able to discriminate between high and low performing participants [19]. Overall, over the course of the study, *p* rose for both groups, denoting decreasing difficulty. Directly after the initial learning phase, *p* was .8400 for group A and .8555 for group B. At the time of the final follow-up test, there were again only minor negligible differences between both groups with *p*=.8667 (group A) and *p*=.8650 (group B).

### **Learning Success: Single Choice Test (T2a, T2b, and T2c)**

Immediately after the learning phase (post 1, T2b), as well as after 2 weeks (post 2, T2c), both groups showed improvements compared with their initial level of knowledge (baseline, T2a). Although there were only minor differences between both groups immediately following the learning phase, with the average number of correctly answered questions rising by 3.59 (SD 1.48) for group A and 3.86 (SD 1.51) for group B (difference 2.7% between both groups), the differences between the two groups were statistically insignificant (exact Mann Whitney *U*, *U*=173.5, *P*=.10, *r*=.247).

Descriptively, at the time of the final test after 2 weeks (Table 3 and Figure 3), both groups did not do as well as before. However, those who had learned with mAR (group A) made an average of 8.1% fewer errors compared with those who had learned without the benefits of mAR (group B).



**Table 3.** Results (number of correctly answered questions) and changes for the single choice tests administered during the study.

Phase	Group A (mARble <sup>a</sup> )		Group B (mble <sup>b</sup> )	
	Mean (SD <sup>c</sup> )	Change to the previous phase	Mean (SD)	Change to the previous phase
Pre (T2a)	3.41 (1.33)	-	3.91 (1.90)	-
Post 1 (T2b)	7.00 (1.48)	+3.59	7.77 (1.51)	+ 3.86
Post 2 (T2c)	6.67 (1.62)	-0.33	6.63 (1.30)	-1.14
Total (T2a to T2c)	-	+3.26	-	+2.72

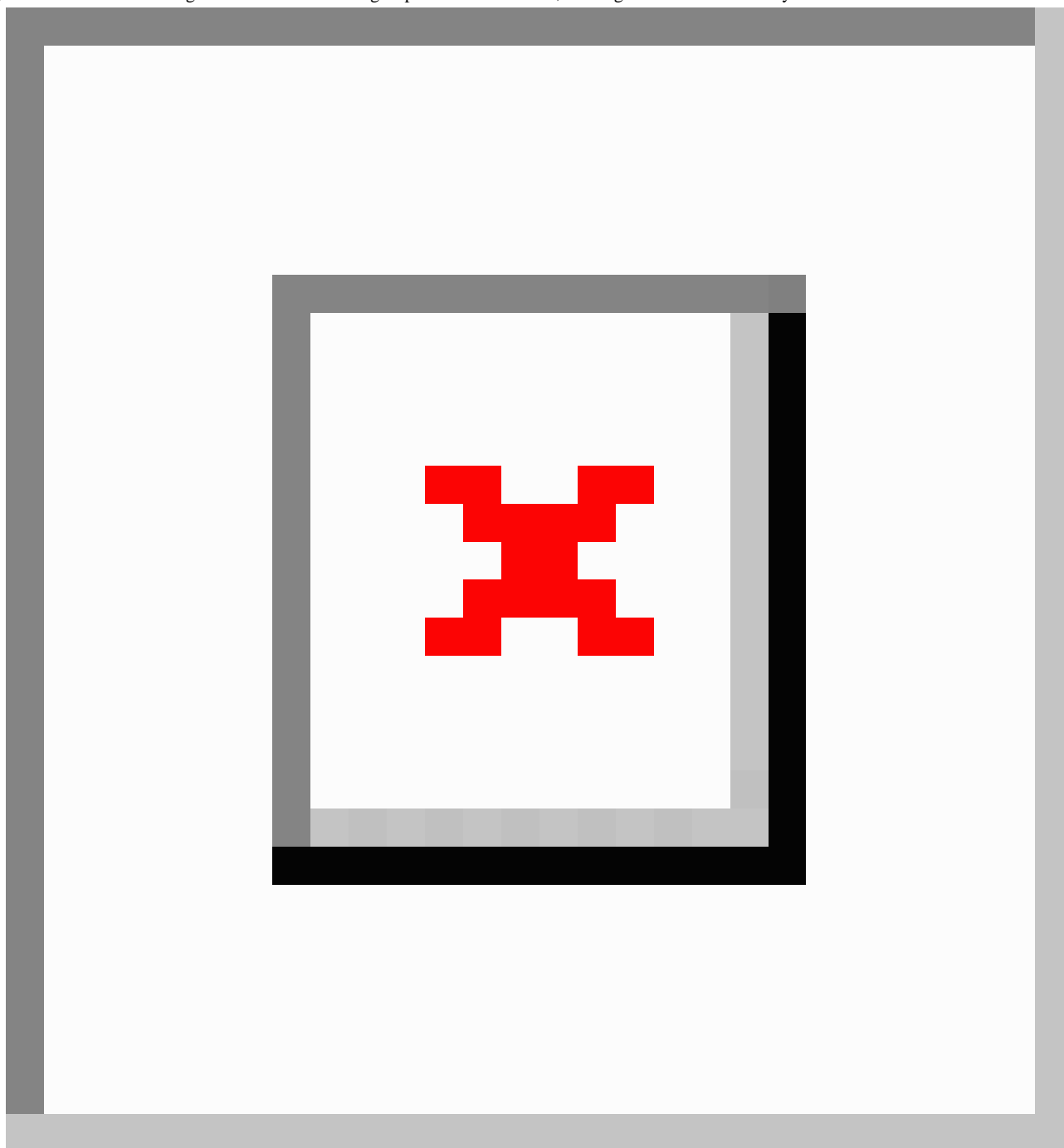
<sup>a</sup>mARble: mobile Augmented Reality blended learning environment.

<sup>b</sup>mble: mobile blended learning environment.

<sup>c</sup>SD: standard deviation.



**Figure 3.** Results of the single choice tests for both groups. For the third test, missing values are denoted by m.



### Evaluating App Usage Based on Log Files Recorded on the Devices

For both groups, utilization periods for the question as well as answer cards differed (Tables 4 and 5). With a total time of 42,977 s of using the flashcards (usage times for questions and answers summarized), group A used considerably less time than group B (59.816 s, see Table 4). For group A, the median usage time per flashcard was 45 s for questions and 370 s for answer cards. Altogether, each member of group A had used question cards for a median of 311 s (IQR 236 s) and answer cards for a median of 1587.5 s (IQR 503 s). For group B, the median use

amounted to 71 s for each question and 245 s per answer card. Again, looking at median values, each member of group B had used the question cards for 534 s (IQR 265 s), and answers were viewed for 2094 s (IQR 874 s), both time spans again being longer than those of group A (Table 5). The lower utilization times recorded for group A can be explained by the additional effort required by interacting with the markers (in-focus time for the markers, sum for all participants: 3603 s, median time per participant 156 s, IQR 85 s; see Tables 4 and 5). Also, considerable time was spent on selecting the desired markers, placing them on the skin, focusing on them with the device's camera etc (12,820 s, see Table 4 and Figure 4).

**Table 4.** Combined utilization times in seconds, including in-focus time spans for the markers, interaction time span, and presentation times for both question and answer parts of the flashcards, stratified by group (group A with mobile Augmented Reality [mAR], group B without mAR functionality, both n=22).

Utilization of specific parts of the application	Group A time, s (%)	Group B time, s (%)
In-focus time span for the markers, n (%)	3603 (6.07)	-
Other interactions with the markers, n (%)	12,820 <sup>a</sup> (21.58)	-
Presentation time: Questions, n (%)	8177 (13.77)	11,992 (20.05)
Presentation time: Answers, n (%)	34,800 (58.59)	47,824 (79.95)
Total learning time, N	59,400 (100.00)	59,816 (100.00)

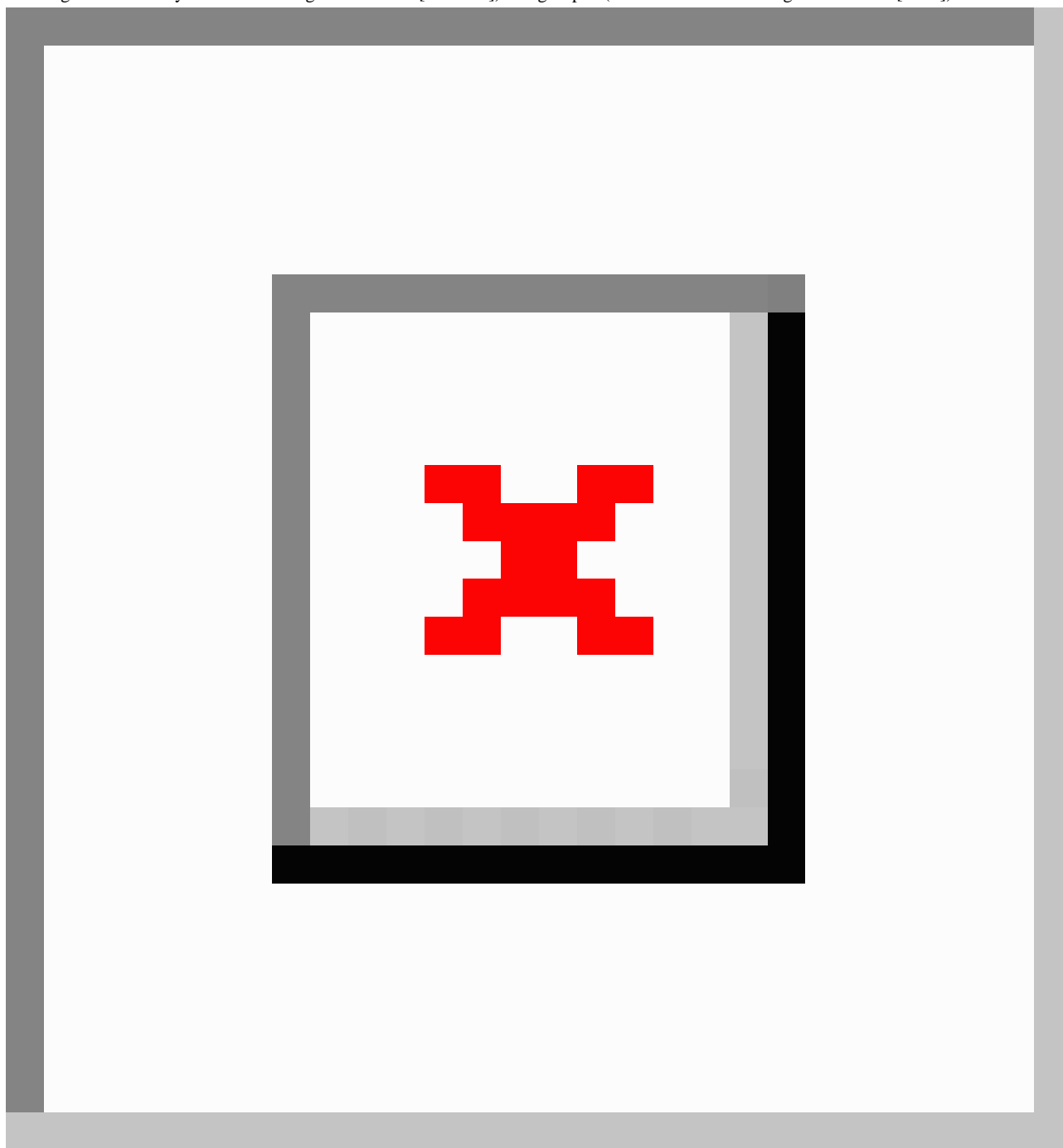
<sup>a</sup>This value was calculated rather than measured.

**Table 5.** Utilization times in seconds (median values and interquartile ranges) for the markers (in-focus time span), marker interaction, and flashcards (questions and answers) stratified by group (group A with mobile Augmented Reality [mAR], group B without mAR functionality, both n=22).

Utilization of specific parts of the application	Group A time (s)		Group B time (s)	
	Median	IQR <sup>a</sup>	Median	IQR
Markers: In-focus time per participant	156.0	85.0	-	-
Presentation time per question/participant	45.0	71.0	71.0	84.5
Total presentation time per participant (questions only)	311.0	236.0	534.0	265.0
Presentation time per answer/participant	245.0	276.0	370.0	311.0
Total presentation time per participant (answers only)	1587.5	503.0	2094.0	874.0

<sup>a</sup>IQR: interquartile range.

**Figure 4.** Graphical representation of the distribution of time spent on interacting with the markers as well as the flashcards, stratified for group A (mobile Augmented Reality blended learning environment [mARble]) and group B (mobile blended learning environment [mble]).



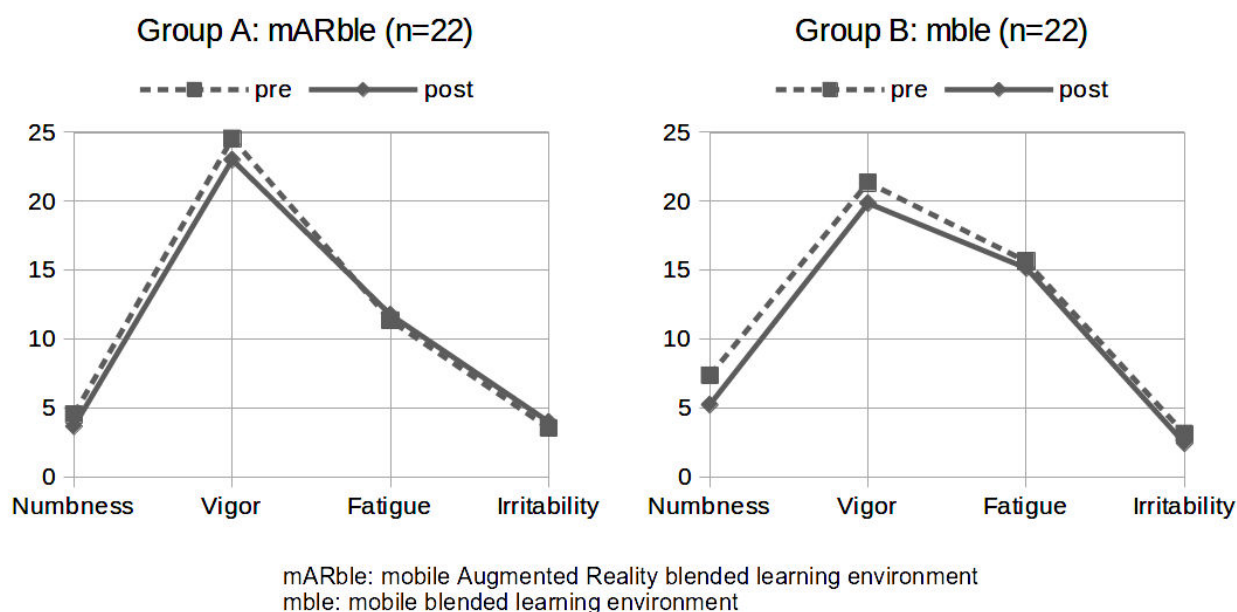
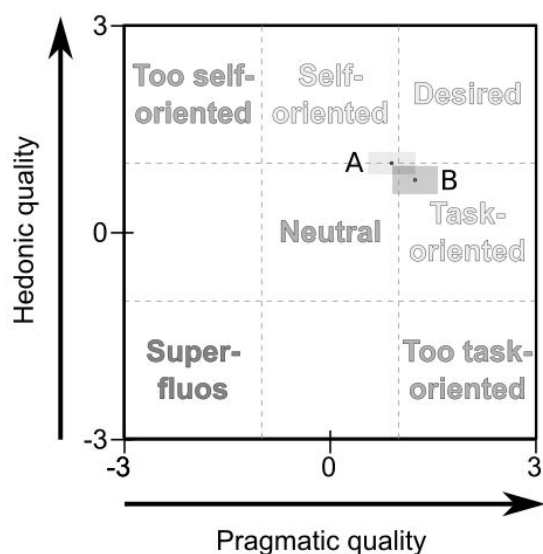
#### Emotional Involvement (T1a+T1b): POMS Questionnaire, German Version

For the two groups, the results of the POMS tests applied before and after the learning phase with the aim of determining whether there were any changes in the participant's emotional status did not show significant differences with respect to the evaluated qualities (see Table 6 and Figure 5). Descriptively, differences were seen for the two dimensions of "irritability" and

"numbness," whereas for both groups, "fatigue" did not change as much. For "vigor," the decrease was almost equal for both groups (decrease for group A: 1.54, for group B: 1.5). For group B, "numbness" decreased by 2.11, from 7.36 (SD 8.54) to 5.25 (SD 7.56). This decrease was larger than for group A, where "numbness" had only been reduced by 0.87, with an initial value of 4.55 (SD 4.78), which changed to 3.68 (SD 4.52) after the learning phase. For "irritability," there was a slight increase for group A and a slight decrease for group B.

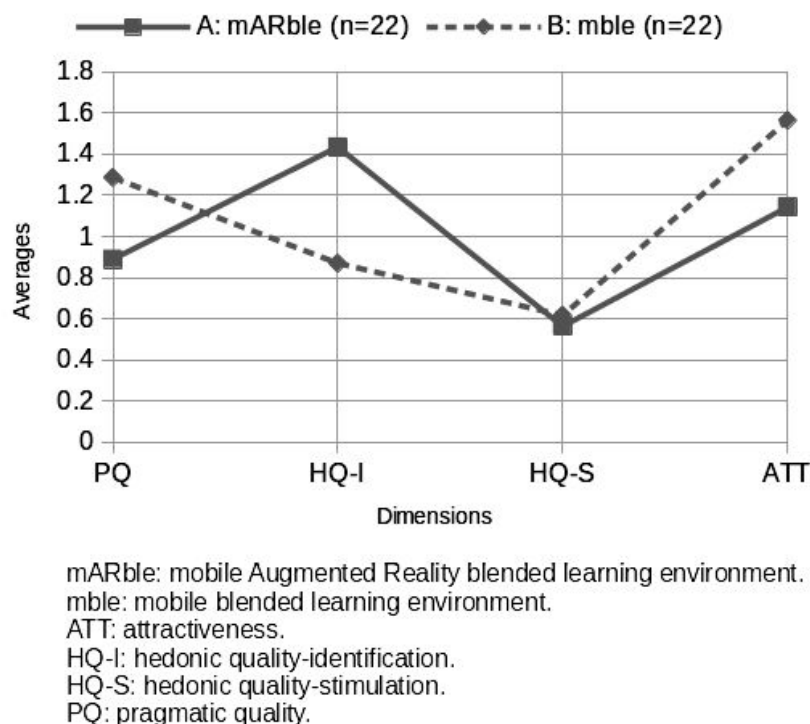
**Table 6.** Aggregated values for numbness, vigor, fatigue, and irritability for groups A and B (both n=22).

Group	Phase	Dimensions			
		Numbness, mean (SD) <sup>a</sup>	Vigor, mean (SD)	Fatigue, mean (SD)	Irritability, mean (SD)
A	Pre	4.55 (4.78)	24.55 (5.47)	11.36 (5.11)	3.55 (4.70)
	Post	3.68 (4.52)	23.01 (5.02)	11.77 (6.79)	4.00 (4.35)
B	Pre	7.36 (8.54)	21.36 (6.69)	15.64 (8.64)	3.14 (3.51)
	Post	5.25 (7.56)	19.86 (6.39)	15.15 (9.47)	2.45 (3.19)

<sup>a</sup>SD: standard deviation.**Figure 5.** Aggregated values for irritability, fatigue, vigor, and numbness.**Figure 6.** Left: Portfolio with average values of the dimensions pragmatic quality (PQ) and hedonic quality (HQ) and the respective confidence rectangles of A (mobile Augmented Reality blended learning environment [mARble]) and B (mobile blended learning environment [mble]), modified following Hassenzahl et al. Right: Corresponding values.

Pragmatic quality (PQ)		
Group	Mean	CI
A	0.890	± 0.339
B	1.286	± 0.340

Hedonic quality (HQ)		
Group	Mean	CI
A	1.000	± 0.171
B	0.744	± 0.215

**Figure 7.** Average values calculated for the four qualities (both groups).

### Learning Experience (T3): AttrakDiff 2

The learning experience was rated positively by all participants, independent of whether they had learned with or without the mAR component, with only marginal differences (descriptive) between both groups (Figures 6 and 7, Table 7). Nevertheless, as the confidence rectangles for both groups overlap (Figure 6; [15]), this is statistically insignificant [18]. However, AR-based learning was rated better with respect to HQ, and there was also

an emphasis on “self-orientation,” which can be attributed to the greater degree of self-centeredness (HQ-I) calculated for this group. In contrast, for group B, ratings emphasized the PQ of the learning experience, mirroring its perceived task-orientation. Differences between the average values calculated for PQ and HQ (aggregated from HQ-I and HQ-S) are negligible. Both groups gave similar ratings for stimulation (HQ-S), with the app without mAR being rated slightly more attractive (group A: 1.143, group B: 1.564).

**Table 7.** Aggregated values calculated for the four qualities covered by AttrakDiff 2: pragmatic quality (PQ), identification (HQ-I), stimulation (HQ-S), and attractiveness (ATT) for groups A and B (both n=22).

Group	PQ <sup>a</sup> mean (SD <sup>b</sup> )	HQ-I <sup>c</sup> mean (SD)	HQ-S <sup>d</sup> mean (SD)	ATT <sup>e</sup> mean (SD)
A	0.890 (0.76)	1.435 (0.51)	0.564 (0.57)	1.143 (0.60)
B	1.286 (0.77)	0.870 (0.78)	0.617 (0.66)	1.564 (0.65)

<sup>a</sup>PQ: pragmatic quality.

<sup>b</sup>SD: standard deviation.

<sup>c</sup>HQ-I: hedonic quality-identification.

<sup>d</sup>HQ-S: hedonic quality-stimulation.

<sup>e</sup>ATT: attractiveness.

## Discussion

### Ascertaining the Effects of mAR

The basic suitability of the mAR-based concept for teaching purposes had already been evaluated in a previous study, where a comparison between conventional learning (using textbooks) and app-based learning was presented, which also included mAR [6]. At that time, a clear advantage of the app-based

approach versus textbook-based learning was shown. However, it was unclear whether the positive effects that had been noted could in fact be attributed to the AR component. There was also a suspicion that the learning medium itself, that is, the excitement of using a mobile phone or tablet personal computer (PC), might already have influenced the results [6]. On the contrary, in this study, with the learning scenarios and presentation of the learning content being identical (multimedia-supported flashcards presented on mobile phones



and tablet PCs) with the exception of the mAR component, it was possible to examine the influence of the mAR component on both learning success and learning experience.

### Principal Findings

Surprisingly, the test scores showed an almost identical increase in the average number of correct answers for both groups (pre to post 1, average improvement for group A: 3.59 [SD 1.48], group B: 3.86 [SD 1.51]; exact Mann Whitney  $U$ ,  $U=173.5$ ;  $P=.10$ ;  $r=.247$ ). Therefore, simply attributing the learning success to the mAR component seems implausible. In comparison with our previous study, whether the greater increase in knowledge is simply because of the use of mobile technologies in general rather than the influence of the mAR component (with its seemingly small contribution shown here), warrants further scrutiny and needs to be considered in future work. However, indications—albeit small—of possible long-term effects may be of interest; at the time of the follow-up test 14 days later, the average number of correct answers only decreased by 0.33 (SD 1.62) for group A but by 1.14 (SD 1.30) for group B, which had not had access to the mAR component of the app while learning. Unfortunately, the dropout rate at T2c (Figure 1) was too high to permit a more confident assertion, but it may be reasonably assumed that the mAR component contributes to committing what is learned to long-term memory, and this is indeed an interesting subject to be examined in later studies. On a side note, we do not believe that repeat testing—that is, using the same tests for T2a, T2b, and T2c (Figure 1)—had a significant influence on the results. During the course of the study, none of the students were provided with either their test scores or the correct answers to the presented questions, which would have given them the opportunity to improve their results. They were only able to base their answers on the provided study material, and if any of the participants had cheated or memorized the answers based on the previously administered tests, we would have expected a more significant increase of their knowledge.

Particularly noteworthy was that group A, learning with mAR support, spent obviously much less time on using the flashcard-based content (identical for both groups) than their counterparts in group B (group A: 1587.5 s [IQR 503 s], group B: 2094 s [IQR 874 s]). Group A spent a significant amount of the allocated time on interacting with the markers, which amounted to a total of 3603 s for all participants (median marker usage per participant: 156.0s, IQR 85.0, also see Table 5). Whereas for the missing 12,820 s, there was no hard evidence proving additional marker usage in the log files (see Table 4), there were observations by the principal investigator who was present during the learning phase that there had indeed been significant mAR related interaction which—for technical reasons—had not been recorded by the app. This included time spent on searching for the desired markers, placing the markers on the skin, trying to focus on the markers, etc, which can certainly be rated as marker-related use of the app. It is up for speculation whether there is an effect of AR and interaction on the learning success that might have effects on better committing knowledge to long-term memory. Future study designs need to consider this aspect carefully. However, some indications for a potentially positive impact of interactive components on the

learning process and commitment of knowledge to long-term memory can be found in literature.

In comparison to other technology-supported learning techniques, there are several mentions of potentially positive as well as negative effects of AR on the learning process [4]. In the past, there were fears that with AR demanding a higher level of focus from learners than, for example, simple multimedia supported learning modules—and possibly requiring more attention for technical aspects—AR might in fact distract students from the presented content [20]. However, we do not believe this to be true, as nowadays, when implemented in a mobile manner, on devices users are familiar with, many of the complexities previously attributed to AR are much less of an issue. This was also corroborated by observations we made during the study, where none of the participants of the mARble group indicated problems with handling the application. In fact, there were early mentions of AR and its playful aspects possibly decreasing cognitive load [21], encouraging students to be creative, to explore the provided content, and to make exciting discoveries on their own, thereby also improving learner's motivation.

The directed attention required when using AR is often also described as beneficial. AR's ability to direct its users' attention to the relevant content, effectively highlighting important content [4,22], as well as the ability to physically enact a learning experience or at least interact with the content, may lead to enhanced memory encoding and better retention of what is being learned [4]. There are also indications that this physical interaction may activate kinesthetic schemas [23], which may also have a positive influence on the learning outcome and help with transferring acquired knowledge from working memory (with relatively low-capacity) to (high-capacity) long-term memory [24].

The learning experience for both groups was evaluated based on the method described by Hassenzahl et al [15,17,18]. Descriptively, mAR was rated more self-oriented, which was because of higher average values in the hedonic domain and smaller average values for pragmatic qualities in comparison with mobile blended learning environment (mble). Nevertheless, as the confidence rectangles for both groups overlap (see Figure 6), this is statistically insignificant [17]. In detail, both systems were rated as similarly stimulating (see Figure 7), which is consistent with ratings for mARble in the previously conducted study [6]. Thus, the stimulating effect is probably rather attributable to the app and the devices it runs on rather than to the mAR component. With respect to a possible self-oriented perception of the AR-based learning experience, the intense (and time-consuming) engagement with the mAR component may be an explanation. However, this hypothesis needs to be further corroborated by additional studies.

In contrast, there were no significant differences between both groups in the emotional realm, as evaluated by the POMS questionnaire. For “numbness,” “vigor,” “fatigue,” and “irritability,” there were only marginal differences in the ratings of both groups (see Table 6 and Figure 5).

## Limitations

As indicated, the study design was adapted according to the general difficulty of recruiting students. The highly streamlined and demanding curriculum medical students have to deal with does not give them much room for participating in activities that they perceive as further reducing their spare time. With this kept in mind, we were forced to make a compromise to the study design by calculating the sample size with a power of 0.8. For the future, for disciplines where visual content plays an important role in medical education, we will therefore aim at integrating our approach into the curriculum, thus also giving us access to a larger number of (potential) study participants.

There is room for debate about whether the random allocation of female and male participants to the two groups, which lead to a rather heterogeneous sample for both, had any influence on the results of the SC tests.

With respect to the markers, based on the chosen technical approach, it was impossible to record usage times other than those that were caused by the markers being in the camera's focus, defined as the time span from recognition of a marker to a flashcard being displayed. Other efforts required for making use of the markers, leading up to them being in focus (selection of the desired markers, placing them on the skin, and trying to focus the camera) were not logged. In follow-up studies, a way for recording the time to fulfill these tasks needs to be found. Finally, assessing emotional involvement solely based on the

POMS questionnaire is less than ideal, and care should be taken to identify an instrument better suited to evaluating the self-oriented character of the mAR-based approach.

## Conclusions

Using mobile technologies for learning purposes integrated into a multimedia-based concept, for example, with a flashcard-based approach similar to the one presented here, can be an effective approach that is at least equivalent to conventional ways of learning, if not better [6]. In this study, isolated indications for the actual impact of mAR on learning success could not be found. The effect described in the previous study may be attributable to the impact of other mobile design aspects rather than the mAR component. Larger-scale evaluations seem advisable for providing final evidence. However, whereas both groups of students obtained similar results regarding learning success, compared with their counterparts, the mARble group spent a significant part of their allocated learning time on AR-related interactions instead of on the flashcards providing textual information, pointing to the potential benefits of mAR on knowledge retention. The (descriptive) indications we found for mAR's potentially positive influence on committing knowledge to long-term memory also point in this direction. Finally, the presented work also found indications pointing to the self-oriented character of mAR-based learning but unfortunately with a lack of significance. Whether—and if so, how—this contributes to the learning process also needs to be investigated in future studies.

## Acknowledgments

The authors would like to thank Bernhard Häussermann and Felix Strübing for technical support during the development of mARble, Anke Mittelstädt and Enno Rubner for assistance with the photography, and all students for their participation in this trial. No external funding was obtained to conduct the study. The manuscript was checked against the Consolidated Standards of Reporting Trials of Electronic and Mobile HEalth Applications and onLine TeleHealth checklist [25].

## Conflicts of Interest

None declared.

## Multimedia Appendix 1

Completed CONSORT checklist.

[[PDF File \(Adobe PDF File\), 62KB - mhealth\\_v5i9e139\\_app1.pdf](#)]

## References

1. Johnson L, Smith R, Levine A, Haywood K. The 2010 Horizon Report: Australia-New Zealand Edition. Austin, Texas: The New Media Consortium; 2010.
2. Kroecker KL. Mainstreaming augmented reality. *Commun ACM* 2010 Jul 01;53(7):19-21. [doi: [10.1145/1785414.1785422](#)]
3. Azuma RT. A survey of augmented reality. *Presence-Teleop VIRT* 1997 Aug;6(4):355-385. [doi: [10.1162/pres.1997.6.4.355](#)]
4. Radu I. Why should my students use AR? A comparative review of the educational impacts of augmented-reality. 2012 Presented at: 2012 IEEE International Symposium on Mixed and Augmented Reality (ISMAR); 2012; Atlanta, Georgia, USA p. 313-314. [doi: [10.1109/ISMAR.2012.6402590](#)]
5. Figueiredo M, Gomes J, Gomes C, Lopes J. Augmented reality tools and learning practice in mobile-learning. In: International Conference on Universal Access in Human-Computer Interaction. Universal Access to Information and Knowledge. Proceedings, Part II. Heidelberg, New York, Dordrecht, London: Springer Cham; 2014 Presented at: 8th International Conference, UAHCI 2014; June 22-27, 2014; Heraklion, Crete, Greece p. 301-312. [doi: [10.1007/978-3-319-07440-5\\_28](#)]

6. Albrecht UV, Folta-Schoofs K, Behrends M, von Jan U. Effects of mobile augmented reality learning compared to textbook learning on medical students: randomized controlled pilot study. *J Med Internet Res* 2013;15(8):e182 [FREE Full text] [doi: [10.2196/jmir.2497](https://doi.org/10.2196/jmir.2497)] [Medline: [23963306](https://pubmed.ncbi.nlm.nih.gov/23963306/)]
7. Ferrer-Torregrosa J, Jiménez-Rodríguez M, Torralba-Estelles J, Garzón-Farínós F, Pérez-Bermejo M, Fernández-Ehrling N. Distance learning icts and flipped classroom in the anatomy learning: comparative study of the use of augmented reality, video and notes. *BMC Med Educ* 2016 Sep 01;16(1):230 [FREE Full text] [doi: [10.1186/s12909-016-0757-3](https://doi.org/10.1186/s12909-016-0757-3)] [Medline: [27581521](https://pubmed.ncbi.nlm.nih.gov/27581521/)]
8. Albrecht UV, Behrends M, Matthies HK, von Jan U. Medical students experience the mobile augmented reality blended learning environment (mARble®): an attractive concept for the net generation? In: Sampson DG, Isaias P, Ifenthaler D, Spector JM, editors. *Ubiquitous and Mobile Learning in the Digital Age*. New York, NY: Springer; 2013:109-113.
9. Moll I, Jung EG, Augustin M. *Duale Reihe Dermatologie*. Stuttgart: Thieme; 2010.
10. Wirtz M, Caspar F. Beurteilerübereinstimmung und Beurteilerreliabilität: Methoden zur Bestimmung und Verbesserung der Zuverlässigkeit von Einschätzungen mittels Kategoriensystemen und Ratingskalen. Göttingen: Hogrefe; 2002.
11. Faul F, Erdfelder E, Lang AG, Buchner A. G\*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav Res Methods* 2007 May;39(2):175-191. [Medline: [17695343](https://pubmed.ncbi.nlm.nih.gov/17695343/)]
12. Faul F, Erdfelder E, Buchner A, Lang AG. Statistical power analyses using G\*Power 3.1: tests for correlation and regression analyses. *Behav Res Methods* 2009 Nov;41(4):1149-1160. [doi: [10.3758/BRM.41.4.1149](https://doi.org/10.3758/BRM.41.4.1149)] [Medline: [19897823](https://pubmed.ncbi.nlm.nih.gov/19897823/)]
13. McNair DM, Lorr M, Droppleman LF. *EITS Manual for the Profile of Mood States*. San Diego, CA: Educational and Industrial Testing Service; 1971.
14. Biehl B, Dangel S, Reiser A. POMS: Profile of Mood States - deutsche Fassung. In: *Internationale Skalen für Psychiatrie*. Weinheim: Collegium Internationale Psychiatrica Scalarum (CIPS); 1981.
15. Hassenzahl M, Burmester M, Koller F. AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität. In: Szwillus G, Ziegler J, editors. *Mensch & Computer 2003*. Wiesbaden: Vieweg+Teubner Verlag; 2003:187-196.
16. Isleifsdóttir J, Lárusdóttir M. Measuring the user experience of task oriented software. In: Law ELC, Bevan N, Christou G, Springett M, Lárusdóttir M, editors. *Proceedings of the International Workshop on Meaningful Measures: Valid Useful User Experience Measurement (VUUM)*. Toulouse, France: Institute of Research in Informatics of Toulouse (IRIT); 2008:97-102.
17. Hassenzahl M. The interplay of beauty, goodness, and usability in interactive products. *Hum Comput Interact* 2004 Dec 1;19(4):319-349. [doi: [10.1207/s15327051hci1904\\_2](https://doi.org/10.1207/s15327051hci1904_2)]
18. Hassenzahl M, Burmester M, Koller F. Der User Experience (UX) auf der Spur: Zum Einsatz von [www.attrakdiff.de](http://www.attrakdiff.de). In: Brau H, Diefenbach S, Hassenzahl M, Koller F, Peissner M, Röse K, editors. *Usability Professionals 2008*. Stuttgart: Fraunhofer IRB Verlag; 2008:78-82.
19. Kuni H, Becker P. Multiple Choice als Numerus clausus: (4) Prüfung nicht bestanden: die Kandidaten oder die Fragen? [German]. Marburg: Universität Marburg URL: <http://staff-www.uni-marburg.de/~kunih/all-doc/mcalsnc4.pdf> [accessed 2017-04-28] [WebCite Cache ID 6q3jyyuaK]
20. Chen YC. Peer learning in an AR-based learning environment. 2008 Presented at: 16th International Conference on Computers in Education; 2008; Taipei, Taiwan p. 291-295.
21. Kaufmann H, Dünser A. Summary of usability evaluations of an educational augmented reality application. Berlin, Heidelberg: Springer; 2007 Presented at: ICVR'07 Proceedings of the 2nd international conference on Virtual reality; July 22 - 27, 2007; Beijing, China p. 660-669. [doi: [10.1007/978-3-540-73335-5\\_71](https://doi.org/10.1007/978-3-540-73335-5_71)]
22. Nischelwitzer A, Lenz FJ, Searle G, Holzinger A. Some aspects of the development of low-cost augmented reality learning environments as examples for future interfaces in technology enhanced learning. In: Stephanidis C, editor. *Universal Access in Human-Computer Interaction Applications and Services*. Berlin, Heidelberg: Springer; 2007:728-737.
23. Abrahamson D, Trninic D. Toward an embodied-interaction design framework for mathematical concepts. In: *Proceedings of the 10th International Conference on Interaction Design and Children*. New York, NY, USA: ACM; 2011 Presented at: IDC '11; June 20 - 23, 2011; Ann Arbor, Michigan p. 1-10. [doi: [10.1145/1999030.1999031](https://doi.org/10.1145/1999030.1999031)]
24. Shirazi A, Behzadan AH. Design and assessment of a mobile augmented reality-based information delivery tool for construction and civil engineering curriculum. *J Prof Issues Eng Educ Pract* 2015 Jul;141(3). [doi: [10.1061/\(ASCE\)EI.1943-5541.0000229](https://doi.org/10.1061/(ASCE)EI.1943-5541.0000229)]
25. Eysenbach G, CONSORT-EHEALTH GROUP. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. *J Med Internet Res* 2011;13(4):e126 [FREE Full text] [doi: [10.2196/jmir.1923](https://doi.org/10.2196/jmir.1923)] [Medline: [22209829](https://pubmed.ncbi.nlm.nih.gov/22209829/)]

## Abbreviations

**AR:** augmented reality  
**ATT:** attractiveness  
**HQ:** hedonic quality

**HQ-I:** hedonic quality-identification  
**HQ-S:** hedonic quality-stimulation  
**iOS:** iPhone operating system  
**IQR:** interquartile range  
**mAR:** mobile Augmented Reality  
**mARble:** mobile Augmented Reality blended learning environment  
**mble:** mobile blended learning environment  
**PC:** personal computer  
**POMS:** Profile of Mood States Questionnaire  
**PQ:** pragmatic quality  
**SC:** single choice  
**SD:** standard deviation

*Edited by G Eysenbach; submitted 29.04.17; peer-reviewed by A Oliveira, R Grainger, M Lang; comments to author 27.07.17; revised version received 06.08.17; accepted 09.08.17; published 14.09.17.*

Please cite as:

Noll C, von Jan U, Raap U, Albrecht UV

Mobile Augmented Reality as a Feature for Self-Oriented, Blended Learning in Medicine: Randomized Controlled Trial

JMIR Mhealth Uhealth 2017;5(9):e139

URL: <http://mhealth.jmir.org/2017/9/e139/>

doi: [10.2196/mhealth.7943](https://doi.org/10.2196/mhealth.7943)

PMID: [28912113](https://pubmed.ncbi.nlm.nih.gov/28912113/)

©Christoph Noll, Ute von Jan, Ulrike Raap, Urs-Vito Albrecht. Originally published in JMIR Mhealth and Uhealth (<http://mhealth.jmir.org>), 14.09.2017. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mhealth and uhealth, is properly cited. The complete bibliographic information, a link to the original publication on <http://mhealth.jmir.org/>, as well as this copyright and license information must be included.

Original Paper

# Social Communication Coaching Smartglasses: Well Tolerated in a Diverse Sample of Children and Adults With Autism

Neha U Keshav<sup>1</sup>, PhD; Joseph P Salisbury<sup>1</sup>, PhD; Arshya Vahabzadeh<sup>1,2</sup>, MD; Ned T Sahin<sup>1,3</sup>, PhD

<sup>1</sup>Brain Power, Cambridge, MA, United States

<sup>2</sup>Psychiatry Academy, Massachusetts General Hospital, Boston, MA, United States

<sup>3</sup>Department of Psychology, Harvard University, Cambridge, MA, United States

**Corresponding Author:**

Ned T Sahin, PhD

Brain Power

1 Broadway 14th Floor

Cambridge, MA,

United States

Phone: 1 617 331 8401

Email: [sahin@post.harvard.edu](mailto:sahin@post.harvard.edu)

## Abstract

**Background:** Augmented reality (AR) smartglasses are an emerging technology that is under investigation as a social communication aid for children and adults with autism spectrum disorder (ASD) and as a research tool to aid with digital phenotyping. Tolerability of this wearable technology in people with ASD is an important area for research, especially as these individuals may experience sensory, cognitive, and attentional challenges.

**Objective:** The aim of this study was to assess the tolerability and usability of a novel smartglasses system that has been designed as a social communication aid for children and adults with autism (the Brain Power Autism System [BPAS]). BPAS runs on Google Glass Explorer Edition and other smartglasses, uses both AR and affective artificial intelligence, and helps users learn key social and emotional skills.

**Methods:** A total of 21 children and adults with ASD across a spectrum of severity used BPAS for a coaching session. The user's tolerability to the smartglasses, user being able to wear the smartglasses for 1 minute (initial tolerability threshold), and user being able to wear the smartglasses for the entire duration of the coaching session (whole session tolerability threshold) were determined through caregiver report.

**Results:** Of 21 users, 19 (91%) demonstrated tolerability on all 3 measures. Caregivers reported 21 out of 21 users (100%) as tolerating the experience, while study staff found only 19 out of 21 users managed to demonstrate initial tolerability (91%). Of the 19 users who demonstrated initial tolerability, all 19 (100%) were able to use the smartglasses for the entire session (whole session tolerability threshold). Caregivers reported that 19 out of 21 users (91%) successfully used BPAS, and users surpassed caregiver expectations in 15 of 21 cases (71%). Users who could communicate reported BPAS as being comfortable (94%).

**Conclusions:** This preliminary report suggests that BPAS is well tolerated and usable to a diverse age- and severity-range of people with ASD. This is encouraging as these devices are being developed as assistive technologies for people with ASD. Further research should focus on improving smartglasses design and exploring their efficacy in helping with social communication in children and adults with ASD.

(JMIR Mhealth Uhealth 2017;5(9):e140) doi:[10.2196/mhealth.8534](https://doi.org/10.2196/mhealth.8534)

**KEYWORDS**

autism; tech; digital health; smartglasses; augmented reality; autism spectrum disorder; technology; medtech; education

## Introduction

Modern smartglasses are small head-mounted displays that integrate a range of sensors that can capture video, audio, and movement data. Smartglasses can deliver an augmented reality

(AR) experience, where the user can see virtual objects overlaid on top of their real-world view as they look through the optical display. Smartglasses delivering AR are believed to have considerable potential as educational and health care tools, and



an increasingly wide range of smartglasses are available for developers and consumers [1,2].

A wide range of assistive technologies have been developed for autism spectrum disorder (ASD) including smartphone and tablet apps, computer programs, social robots, and virtual reality [3]. There have been encouraging findings about the positive impact of such technologies, yet many children and adults with ASD continue to have considerable unmet educational and health care needs. Interest has been growing in the use of AR as a teaching tool for children and adults with ASD, and understanding how people with ASD experience and are affected by head-mounted displays remain key questions that face the field [4]. An AR experience can be delivered on a variety of different platforms including smartphones, tablets, stationary displays, and on “heads-up” smartglasses. Much of the current AR research has been on AR delivered through handheld/“heads-down” devices [5-7]. Studies have demonstrated that AR delivered on smartphones, tablets, and desktop computers may help people with ASD with their attention [6], emotion recognition [8], ability to notice social cues [5], social skills [9], ability to engage in pretend play [10], and even as a navigation aid for planning trips [7]. However, using AR is not a risk-free endeavor; children using smartphone-based AR have developed postural and grip strain in addition to experiencing falls [11], and smartphone-based AR games can lead to injury through distraction, with resultant major trauma already being reported [12].

Smartglasses may offer several advantages when compared to smartphone and tablet devices and have been described as the platform of the future for AR [13]. Use of smartglasses may be less distracting and may require less cognitive workload than smartphones [14,15]. By looking through smartglasses, users can continue to look heads-up at the environment around them and also remain hands-free because smartglasses are head-worn [16]. These advantages may enable users to continue to observe the social world around them, something that is considerably impacted when using a smartphone [17]. Additionally, smartglasses allow users to keep their hands unoccupied, making it easier to use them in nonverbal communication and/or academic and occupational activities, which are particularly pertinent considerations for children and adults with ASD who demonstrate impairment in social communication [18]. To our knowledge, we have published the first report of the feasibility of using AR smartglasses to provide social and cognitive coaching in children with ASD [16].

Research is required to determine the tolerability of AR smartglasses given that ASD is accompanied by a range of sensory, behavioral, and cognitive challenges that may make wearing such devices difficult. Many people with ASD have sensory sensitivities, and they may struggle to wear conventional prescription glasses [19], brush their teeth, or comb their hair [20,21]. Smartglasses often have a similar form factor to prescription glasses, and in the case of Glass Explorer Edition (formerly known as Google Glass), may weigh the same as typical pair of prescription lenses and frame [22]. Unlike prescription glasses, smartglasses produce additional sensory stimuli in the form of visual input via their optical displays and audio via their speakers. It is therefore important to study how

people with ASD respond to and tolerate wearing such devices. With the exception of conventional prescription glasses, there are only rare occasions when one would need to “wear” a face-mounted object. In this regard, wearing smartglasses may be a particularly novel experience, with few daily life comparators. This is an important consideration because people with ASD can exhibit considerable distress when exposed to unfamiliar situations, changes in routine, or changes in environment [18]. Despite the abovementioned concerns, there continues to be a dearth of research into AR smartglasses for people with ASD. The authors have found that many clinicians, educators, and people from the ASD community have expressed doubt as to whether children and adults with ASD would tolerate wearing AR smartglasses. This has led to the commonly encountered question: *but will they even wear it?* This is not surprising given that wearing conventional glasses has been highlighted as a major challenge by prominent ASD charities [23].

The importance of understanding how people with ASD will respond to such devices is heightened by the potential benefits of conducting research with smartglasses. Smartglasses, like smartphones, contain myriad sensors, such as an accelerometer and camera, and are able to collect video, audio, movement, physiologic, and user interaction data [24]. These quantitative data can be collected and analyzed to undertake digital phenotyping, and more importantly, to help support research efforts to help subtype highly clinically heterogeneous behavioral conditions such as ASD [25].

To explore the tolerability of AR smartglasses, we studied whether children and adults with ASD were able to tolerate wearing the Brain Power Autism System (BPAS), novel social communication coaching smartglasses that use AR and emotional artificial intelligence [16]. BPAS has undergone feasibility [16], acceptability [26], safety [27], and clinical impact studies [28]. BPAS is based on a highly modified version of Google Glass Explorer Edition and Glass Enterprise Edition (both overseen by X Development LLC, formerly known as Google X).

## Methods

The methods and procedures of this study were approved by Asentral, Inc, Institutional Review Board, an affiliate of the Commonwealth of Massachusetts Department of Public Health.

### User Recruitment

A sequential sample of 21 children and adults with clinically diagnosed ASD were recruited from a database of individuals who completed a Web-based signup form expressing interest in participating in smartglasses research. Individuals represented a demographically and clinically diverse group comprising different ages, genders, verbal abilities, and level of functioning (Table 1). Written consent was obtained from the legal guardians of children and from cognitively able adults. Children aged 7 to 17 years provided written assent when possible. In this report, every user was accompanied by a parent or other caregiver during the session, and users and caregivers could ask for the session to stop at any time and for any reason. All users

completed the Social Communication Questionnaire (SCQ) so we could document their level of social communication impairment [29]. The SCQ score demonstrates that the user sample represented a wide range of social communication impairment.

### Data Collection Procedure

Users and caregivers were given an introductory explanation and demonstration of BPAS smartglasses. Users were then given the chance to wear the smartglasses (Figure 1), aided as needed by study staff and their caregivers for correct initial placement (Figure 2).

The user's tolerability to the smartglasses was determined through caregiver report, the user's ability to wear the smartglasses for 1 minute (initial tolerability threshold), and the user's ability to wear the smartglasses for the entire duration of the coaching session (whole session tolerability threshold). The initial tolerability threshold provides a rapid understanding of how well a user would respond to the physical form factor of the smartglasses, an important consideration given the unique set of sensory and cognitive challenges of each user. The whole session tolerability threshold represents how well the user

tolerates wearing the smartglasses but also represents their use of the coaching apps as they undertake a series of structured activities with their caregiver in a session lasting between 1 and 1.5 hours. At the end of the session, caregivers could rate how well they felt the user tolerated using BPAS through a 5-point Likert scale (1=very low, 5=very high). A tolerability rating of low or very low was deemed to be a negative indication of tolerability, while neutral, high, or very high caregiver ratings were noted as an indication of tolerability.

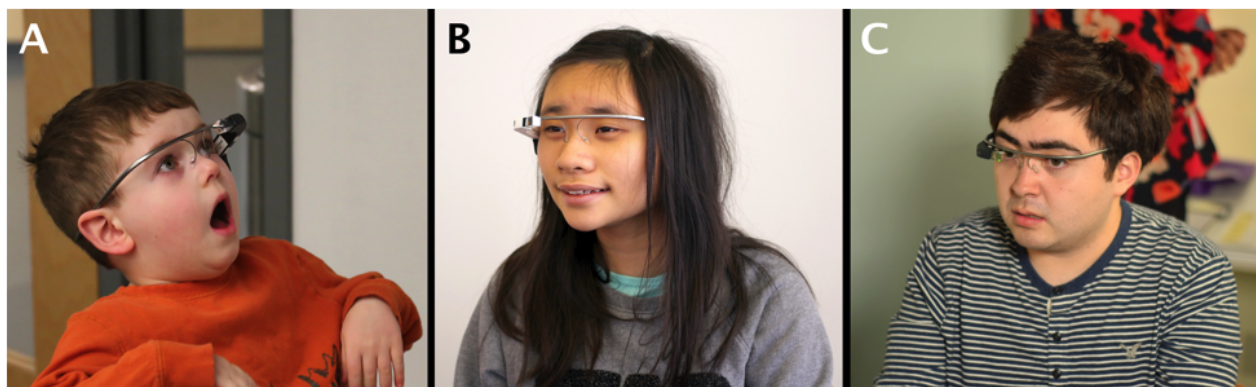
Caregivers were also asked to use a 5-point Likert scale to rate if they felt the user was able to successfully use BPAS with their assistance (1=strongly disagree, 5=strongly agree) and whether they felt the user responded more positively to BPAS smartglasses than they had expected (1=strongly disagree, 5=strongly agree). For these responses, a higher standard had to be set compared to tolerability: a rating of agree/strongly agree (4 or 5) was determined to be a positive response for each of these questions. Users who could communicate verbally with their caregiver or study staff were asked to rate how comfortable the smartglasses were. Both caregivers and users were able to provide additional feedback to any question in the interviews.

**Table 1.** Demographics of users (N=21).

Characteristic	Mean (SD) range or n (%) <sup>a</sup>
Age, years, mean (SD) range	11.9 (4.9) 4.4-21.5
<b>Gender, n (%)</b>	
Male	19 (91)
Female	2 (10)
<b>Verbal, n (%)</b>	
Yes	19 (91)
No	2 (10)
Social Communication Questionnaire score, mean (SD) range	18.5 (6.1) 6-28

<sup>a</sup>Percentages can equal more than 100 due to rounding.

**Figure 1.** Three users with autism wearing the Brain Power Autism System and using its socioemotional coaching apps. Pictures used with user/caregiver permission.



**Figure 2.** Caregiver assisting user in wearing Brain Power Autism System smartglasses during testing (A, B). User demonstrated tolerability on all 3 measures and was witnessed to spontaneously hug his caregiver during use of social communication app use (C, D). Pictures used with user/caregiver permission.



## Results

A total of 19 out of 21 users (91%) demonstrated tolerability on all 3 measures (caregiver report, initial tolerability threshold, and whole session tolerability threshold; [Table 2](#)). Of the 19 users who managed to pass the initial tolerability threshold (19/21, 91%), all went on to use BPAS for the entire coaching session, passing the whole session tolerability threshold (19/19, 100%). Two users, both nonverbal, did not pass the initial 1-minute tolerability threshold as they would not continue to

wear the smartglasses once placed. These users were both nonverbal and were aged 5.5 and 5.75 years with SCQ scores of 25 and 28. Users who were verbal and able to answer questions (18/21) rated the smartglasses as being comfortable to use (17/18, 94%; [Table 3](#)). A majority of caregivers felt users responded more positively to the smartglasses than they had expected (15/21, 71%). A number of caregivers provided additional feedback, suggesting that users may benefit from extended and/or repeated orientation and introduction sessions with BPAS. The results are graphically represented in [Figure 3](#).

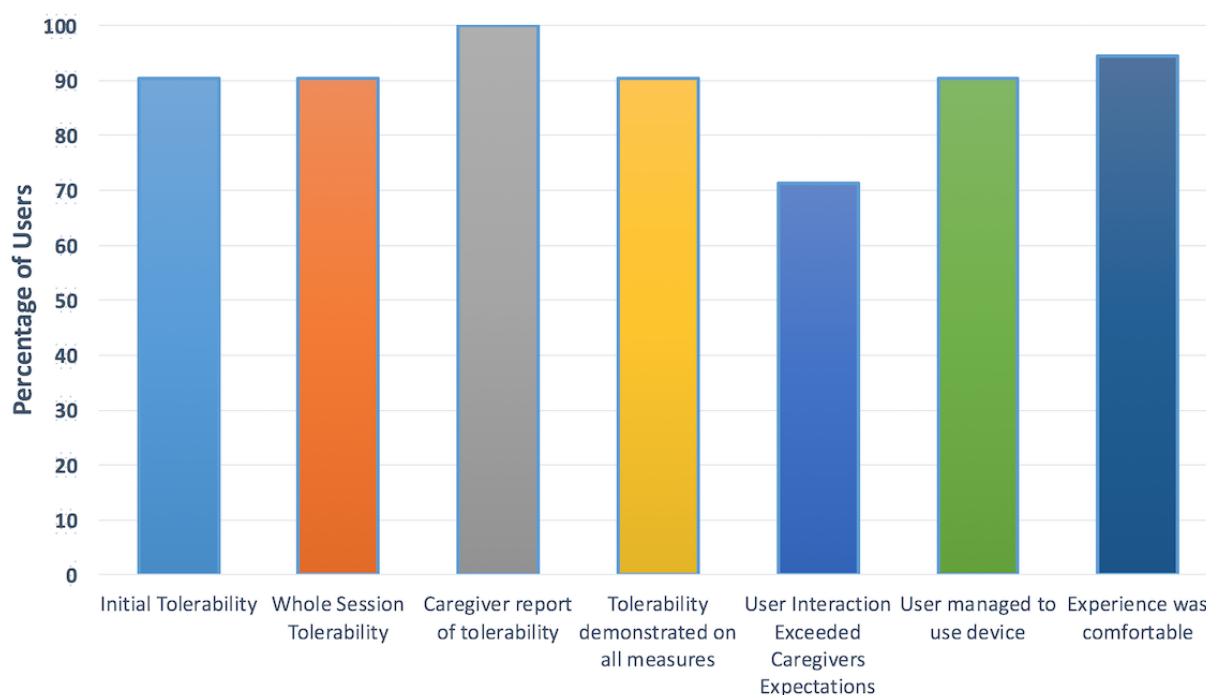
**Table 2.** Tolerability report of Brain Power Autism System (N=21).

Tolerability measures	n (%)
<b>Initial tolerability threshold (worn for at least 1 minute of continuous use)</b>	
Yes	19 (91)
No	2 (10)
<b>Whole session tolerability threshold (worn until session completion)</b>	
Yes	19 (91)
No	2 (10)
<b>Caregiver report of tolerability</b>	
Yes	21 (100)
No	0 (0)
<b>Users demonstrating tolerability on all measures</b>	
Yes	19 (91)
No	2 (10)

**Table 3.** User experience report (N=21).

User experience responses	n (%)
<b>User reported that experience was comfortable (only users who were able to answer questions: n =18)</b>	
Yes	17 (94)
Mixed response/sometimes	1 (6)
No	0 (0)
<b>Caregiver reported that user managed to use the device with assistance</b>	
Yes	19 (91)
Mixed response/sometimes	0 (0)
No	2 (10)
<b>Caregiver reported user responded more positively to the device than expected</b>	
Yes	15 (71)
Mixed response/sometimes	6 (29)
No	0 (0)



**Figure 3.** Summary of Brain Power Autism System tolerability and experience.

## Discussion

Children and adults with ASD have considerable unmet educational and behavioral health needs, and technology-aided solutions may provide a scalable and effective tool to help address these demands for resources. While AR smartglasses have been designed as a social communication aid for people with ASD [16], there is only limited research to understand how acceptable and tolerable this technology is to these individuals [16]. Our preliminary study shows that a diverse range of children and adults with ASD can tolerate wearing and using BPAS, AR smartglasses designed to function as a social communication aid for people with ASD.

Tolerability was demonstrated across all 3 of our measures in 91% of users. Every user who demonstrated initial 1-minute tolerability managed to continue to demonstrate tolerability for the entire coaching session that ran between 1 and 1.5 hours. This suggests that the tolerability of users with ASD to such smartglasses can be accurately predicted based on their ability to initially use the device over a relatively short amount of time. The 2 users who did not tolerate wearing the device were both nonverbal, younger in age, and had greater social impairment as highlighted by their higher SCQ scores. Further investigation is warranted to determine how to improve tolerability in younger children with ASD who have greater language and social communication challenges.

Our data help to answer our initial question: *but will they even wear it?* In our experience, this is one of the most common questions that parents, educators, and researchers ask us when BPAS is shown to them. These data show that children and adults with ASD can not only wear smartglasses for relatively

lengthy durations of time but are able to use them and describe the experience as comfortable. How the users interacted with BPAS surpassed the expectations of their caregivers in most cases. We did find that both nonverbal users struggled to wear the smartglasses and were unable to pass the initial tolerability threshold. Based on feedback from caregivers, a more gradual introduction and orientation process to the smartglasses may have been useful in these cases.

Additionally, given that sensor-rich smartglasses are quantitative data gathering tools, it is helpful to know that they can be worn for such durations in people with behaviorally heterogeneous conditions that could benefit from digital phenotyping and subtyping.

While our results are promising, there are a number of limitations. Although this work is, to our knowledge, the first report of the tolerability of smartglasses as a social communication aid in people with ASD, our sample size is moderate (N=21). Additionally, given the customized nature of BPAS, generalizability may be limited in the case of other smartglasses, different smartglasses software apps, and even an unmodified Google Glass device.

More longitudinal research would be useful to determine whether the tolerability that we have observed continues to last after repeated coaching sessions. Understanding the effects of repeated sessions over a longer duration of time is important, as many training and educational programs for people with ASD involve repeated sessions over a long period of time. Further research is required to investigate the efficacy of AR smartglasses in ASD, but the tolerability and usability of such devices does not appear to be a substantial barrier to their use.



## Authors' Contributions

NS is the inventor of the Brain Power Autism System. NK, JS, AV, and NS designed and undertook the intervention. The authors all contributed to the writing of this technology report.

## Conflicts of Interest

This report was supported by Brain Power, a neurotechnology company developing a range of artificially intelligent wearable technologies. Brain Power has engineering and technical partnerships with major technology companies and also receives funding support from federal and congressional sources.

## References

1. Elder S, Vakaloudis A. A technical evaluation of devices for smart glasses applications. 2015 Presented at: Internet Technologies and Applications (ITA); 2015; Wrexham.
2. Elder S, Vakaloudis A. Towards uniformity for smart glasses devices: an assessment of function as the driver for standardisation. 2015 Presented at: Technology and Society (ISTAS), 2015 IEEE International Symposium; 2015; Dublin.
3. Grynspan O, Weiss P, Perez-Diaz F, Gal E. Innovative technology-based interventions for autism spectrum disorders: a meta-analysis. *Autism* 2013 Oct 03;18(4):346-361. [doi: [10.1177/1362361313476767](https://doi.org/10.1177/1362361313476767)]
4. Newbutt N, Sung C, Kuo H, Leahy M. The acceptance, challenges, and future applications of wearable technology and virtual reality to support people with autism spectrum disorders. In: *Recent Advances in Technologies for Inclusive Well-Being*. New York: Springer; 2017:221-241.
5. Chen C, Lee I, Lin L. Augmented reality-based video-modeling storybook of nonverbal facial cues for children with autism spectrum disorder to improve their perceptions and judgments of facial expressions and emotions. *Comput Hum Behav* 2016 Feb;55:477-485. [doi: [10.1016/j.chb.2015.09.033](https://doi.org/10.1016/j.chb.2015.09.033)]
6. Escobedo L, Tentori M, Quintana E, Favela J, Garcia-Rosas D. Using augmented reality to help children with autism stay focused. *IEEE Pervasive Comput* 2014 Jan;13(1):38-46. [doi: [10.1109/MPRV.2014.19](https://doi.org/10.1109/MPRV.2014.19)]
7. McMahon D, Cihak D, Wright R. Augmented reality as a navigation tool to employment opportunities for postsecondary education students with intellectual disabilities and autism. *J Res Technol Educ* 2015 Jul 31;47(3):157-172. [doi: [10.1080/15391523.2015.1047698](https://doi.org/10.1080/15391523.2015.1047698)]
8. Chen C, Lee I, Lin L. Augmented reality-based self-facial modeling to promote the emotional expression and social skills of adolescents with autism spectrum disorders. *Res Dev Disabil* 2014 Nov 08;36:396-403. [doi: [10.1016/j.ridd.2014.10.015](https://doi.org/10.1016/j.ridd.2014.10.015)] [Medline: [25462499](https://pubmed.ncbi.nlm.nih.gov/25462499/)]
9. Tentori M, Hayes G. Designing for interaction immediacy to enhance social skills of children with autism. 2010 Presented at: Proceedings of the 12th ACM international conference on Ubiquitous computing; 2010; Copenhagen.
10. Bai Z, Blackwell A, Coulouris G. Using augmented reality to elicit pretend play for children with autism. *IEEE T Vis Comput Gr* 2015;21(5):598-610.
11. Radu I, Guzdial K, Avram S. An Observational Coding Scheme for Detecting Children's Usability Problems in Augmented Reality. 2017 Presented at: Proceedings of the 2017 Conference on Interaction Design and Children; Stanford; 2017.
12. Franchina M, Sinkar S, Ham B, Lam GC. A blinding eye injury caused by chasing Pokemon. *Med J Aust* 2017 May 15;206(9):384. [Medline: [28490297](https://pubmed.ncbi.nlm.nih.gov/28490297/)]
13. Monkman H, Kushniruk A. A see through future: augmented reality and health information systems. *Stud Health Technol Inform* 2015;208:281-285. [Medline: [25676988](https://pubmed.ncbi.nlm.nih.gov/25676988/)]
14. He J, Choi W, McCarley JS, Chaparro BS, Wang C. Texting while driving using Google Glass: promising but not distraction-free. *Accid Anal Prev* 2015 Aug;81:218-229. [doi: [10.1016/j.aap.2015.03.033](https://doi.org/10.1016/j.aap.2015.03.033)] [Medline: [26024837](https://pubmed.ncbi.nlm.nih.gov/26024837/)]
15. He J, Ellis J, Choi W, Wang P. Driving while reading using Google glass versus using a smart phone: which is more distracting to driving performance. 2015 Presented at: 11th Annual Symposium on Graduate Research and Scholarly Projects; 2015; Wichita.
16. Liu R, Salisbury JP, Vahabzadeh A, Sahin NT. Feasibility of an autism-focused augmented reality smartglasses system for social communication and behavioral coaching. *Front Pediatr* 2017;5:145 [FREE Full text] [doi: [10.3389/fped.2017.00145](https://doi.org/10.3389/fped.2017.00145)] [Medline: [28695116](https://pubmed.ncbi.nlm.nih.gov/28695116/)]
17. Chen P, Pai C. Smartphone gaming is associated with pedestrians? Head-turning performances: an observational study of street-crossing behaviours at uncontrolled intersection in Taipei [in press]. *Int J Sustainable Transport* 2017.
18. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders (DSM-5)*. Arlington: American Psychiatric Publishing; 2013.
19. DeLeon IG, Hagopian LP, Rodriguez-Catter V, Bowman LG, Long ES, Boelter EW. Increasing wearing of prescription glasses in individuals with mental retardation. *J Appl Behav Anal* 2008;41(1):137-142 [FREE Full text] [Medline: [18468288](https://pubmed.ncbi.nlm.nih.gov/18468288/)]
20. Smith MD, Belcher R. Teaching life skills to adults disabled by autism. *J Autism Dev Disord* 1985 Jun;15(2):163-175. [Medline: [3997743](https://pubmed.ncbi.nlm.nih.gov/3997743/)]
21. Matson JL, Taras ME, Sevin JA, Love SR, Fridley D. Teaching self-help skills to autistic and mentally retarded children. *Res Dev Disabil* 1990;11(4):361-378. [Medline: [2270319](https://pubmed.ncbi.nlm.nih.gov/2270319/)]

22. Walsh G. The weight of spectacle frames and the area of their nose pads. *Ophthalmic Physiol Opt* 2010 Jul;30(4):402-404. [doi: [10.1111/j.1475-1313.2010.00755.x](https://doi.org/10.1111/j.1475-1313.2010.00755.x)] [Medline: [20629962](#)]
23. Speaks A. Autism challenge for our experts: kid needs glasses; won't tolerate anything on face.: Autism Speaks; 2012. URL: <https://www.autismspeaks.org/blog/2015/05/29/autism-challenge-kid-needs-glasses-wont-tolerate-anything-face> [accessed 2017-09-06] [WebCite Cache ID 6tHexh1sd]
24. Hernandez J, Li Y, Rehag J, Picard R. Physiological parameter estimation using a head-mounted wearable device. *Bioglass*; 2014 Presented at: Wireless Mobile Communication and Healthcare (Mobihealth), 2014 EAI 4th International Conference on Bioglass; 2014; Athens.
25. Onnela J, Rauch SL. Harnessing smartphone-based digital phenotyping to enhance behavioral and mental health. *Neuropsychopharmacology* 2016 Dec;41(7):1691-1696 [FREE Full text] [doi: [10.1038/npp.2016.7](https://doi.org/10.1038/npp.2016.7)] [Medline: [26818126](#)]
26. Sahin NT, Keshav NU, Salisbury JP, Vahabzadeh A. Cool enough for school: second version of Google Glass rated by children facing challenges to social integration as desirable to wear at school. *bioRxiv* 2017. [doi: [10.1101/171033](https://doi.org/10.1101/171033)]
27. Sahin NT, Keshav NU, Salisbury JP, Vahabzadeh A. An augmented reality social communication aid for children and adults with autism: user and caregiver report of safety and lack of negative effects. *bioRxiv* 2017. [doi: [10.1101/164335](https://doi.org/10.1101/164335)]
28. Vahabzadeh A, Keshav N, Salisbury J, Sahin N. Preliminary report on the impact of smartglasses-based behavioral and social communication aid on hyperactivity in children and adults with autism. *bioRxiv* 2017. [doi: [10.1101/165514](https://doi.org/10.1101/165514)]
29. Rutter M, Bailey A, Lord C. *The Social Communication Questionnaire: Manual*. Torrance: Western Psychological Services; 2003.

## Abbreviations

**AR:** augmented reality

**ASD:** autism spectrum disorder

**BPAS:** Brain Power Autism System

**SCQ:** Social Communication Questionnaire

*Edited by G Eysenbach; submitted 22.07.17; peer-reviewed by B Bie; comments to author 15.08.17; revised version received 18.08.17; accepted 19.08.17; published 21.09.17.*

*Please cite as:*

Keshav NU, Salisbury JP, Vahabzadeh A, Sahin NT

*Social Communication Coaching Smartglasses: Well Tolerated in a Diverse Sample of Children and Adults With Autism*

*JMIR Mhealth Uhealth* 2017;5(9):e140

URL: <http://mhealth.jmir.org/2017/9/e140/>

doi: [10.2196/mhealth.8534](https://doi.org/10.2196/mhealth.8534)

PMID: [28935618](https://pubmed.ncbi.nlm.nih.gov/28935618/)

©Neha U. Keshav, Joseph P. Salisbury, Arshya Vahabzadeh, Ned T. Sahin. Originally published in JMIR Mhealth and Uhealth (<http://mhealth.jmir.org>), 21.09.2017. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mhealth and uhealth, is properly cited. The complete bibliographic information, a link to the original publication on <http://mhealth.jmir.org/>, as well as this copyright and license information must be included.

---

Publisher:  
JMIR Publications  
130 Queens Quay East.  
Toronto, ON, M5A 3Y5  
Phone: (+1) 416-583-2040  
Email: [support@jmir.org](mailto:support@jmir.org)

---

<https://www.jmirpublications.com/>