Original Article

Healthc Inform Res. 2018 July;24(3):207-226.
https://doi.org/10.4258/hir.2018.24.3.207
pISSN 2093-3681 • eISSN 2093-369X

HIR
Healthcare Informatics Research

Effectiveness of Mobile Health Application Use to Improve Health Behavior Changes: A Systematic Review of Randomized Controlled Trials

Myeunghee Han, MS, RN[1], Eunjoo Lee, PhD, RN[2]

[1]School of Nursing, University of Maryland, Baltimore, MD, USA; [2]College of Nursing, Research Institute of Nursing Science, Kyungpook National University, Daegu, Korea

Objectives: The purpose of this study was to examine the effectiveness of mobile health applications in changing health-related behaviors and clinical health outcomes. **Methods:** A systematic review was conducted in this study. We conducted a comprehensive bibliographic search of articles on health behavior changes related to the use of mobile health applications in peer-reviewed journals published between January 1, 2000 and May 31, 2017. We used databases including CHINAHL, Ovid-Medline, EMBASE, and PubMed. The risk of bias assessment of the retrieved articles was examined using the Scottish Intercollegiate Guidelines Network. **Results:** A total of 20 articles met the inclusion criteria. Sixteen among 20 studies reported that applications have a positive impact on the targeted health behaviors or clinical health outcomes. In addition, most of the studies, which examined the satisfaction of participants, showed health app users have a statistically significant higher satisfaction. **Conclusions:** Despite the high risk of bias, such as selection, performance, and detection, this systematic review found that the use of mobile health applications has a positive impact on health-related behaviors and clinical health outcomes. Application users were more satisfied with using mobile health applications to manage their health in comparison to users of conventional care.

Keywords: Health Behavior, Mobile Applications, Smartphone, Review, Mobile Health

I. Introduction

The global mobile health (mHealth) application (app) market has been growing at a tremendous rate, and it is expected to continue to flourish [1]. These mHealth apps provide quick and easy access, transfer, and tracking of health information as well as interactive displays and interventions that can allow users to be highly engaged in promoting health outcomes and changing health-related behaviors [2]. Thus, health-related apps have a great potential to aid a wide range of target audiences with a variety of health issues [3].

Despite the evolution and widespread use of these mHealth apps, the factors involved in smartphone and health app use and their effectiveness are not yet fully understood and the field of research related to mHealth apps is still in a nascent

Submitted: March 15, 2018
Revised: May 7, 2018
Accepted: May 30, 2018

Corresponding Author
Eunjoo Lee, PhD, RN
Department of Nursing, College of Nursing, Research Institute of Nursing Science, Kyungpook National University, 680, Gukchaebosang-ro, Jung-gu, Daegu 41944, Korea. Tel: +82-53-420-4934, E-mail: jewelee@knu.ac.kr

stage [4]. Kitsiou et al. [5] mentioned that a wide range of mHealth apps have not been strictly evaluated. For this reason, most consumers use mHealth apps without any concrete information about their effectiveness or harm and evidence of the effectiveness of mHealth has been inconclusive and not fully understood [4].

Thus, research determining the effectiveness of health apps is urgently needed [2]. This study aimed to demonstrate the effectiveness of mHealth apps in changing health-related behaviors and clinical health outcomes through a systematic review of randomized controlled trials (RCTs).

II. Methods

1. Search Strategy

We searched the electronic literature of RCTs published from January 1, 2000 to May 31, 2017, using four databases: the Cumulative Index to Nursing and Allied Health Literature (CINAHL), PubMed, Excerpta Medica dataBASE (EMBASE), and Ovid Medline.

A university librarian was consulted who was a subject expert in the field of teaching and learning of systematic review. Searches used the following medical subject headings terms and keywords in various combination. We derived three broad themes that were then combined using the Boolean operator 'AND'. The first theme 'mobile' was created using the Boolean operator 'OR' to combine text words (mobile*, OR smartphone*). The second theme 'application' was created using the Boolean operator 'OR' to combine text words (app*, OR application*). The third theme 'health behaviors', was created using the Boolean operator 'OR' to combine text words (health behaviors*, OR health behaviors change*, OR behaviors change*).

In this study, to secure the quality of this systematic review, the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines were used [6]. This is a tool developed to verify the quality of evidence obtained during systematic reviews.

2. Study Selection

Two investigators independently reviewed the titles first and then examined the abstracts. Data extraction was conducted by one reviewer (MH Han) and was rechecked for by another reviewer to confirm the accuracy (EJ Lee). The same investigators read and screened the full texts to make the final decision. The reasons for inclusion and exclusion were recorded.

We included articles with the following characteristics: (1) published in English, (2) published during the period from 2000 to 2017, (3) results related to changes in health behaviors, (4) RCTs designed for app-based interventions to improve any health-related behaviors. The exclusion criteria were being other kinds of study than RCTs; qualitative studies; books; conference proceedings; reviews; dissertations; protocols; or studies examining text messages, Web, emails, Twitter, social network services, or personal digital assistant-based health interventions. We also excluded studies lacking behavior change indicators or outcomes, not using apps as the primary intervention tools; or focusing primarily on app design and development. Conference abstracts, protocol papers, reviews, editorials, and commentary were also excluded.

References that clearly did not meet all criteria were excluded. Full-text articles that appeared to be relevant were retrieved and independently assessed by two reviewers. Disagreements were resolved through a meeting. The initial search revealed 1,247 articles: 66 in CINAHL, 481 in Ovid-Medline, 626 in EMBASE, and 74 in PubMed. Following the PRISMA guidelines (Figure 1), we removed duplicates and screened the titles and abstracts, which narrowed the list down to 57 relevant articles. Two investigators reviewed these 57 articles, 37 of which conducted other interventions with apps, compared two app-functions, examined protocols, or had unclear outcomes. Thus, finally, 20 articles were included for this review.

3. Data Collection and Analysis

From the 20 included articles, the following information was retrieved and analyzed: first author, year of publication, country, study design, themes, participants' character, sample size, mean age, intervention tool, follow-up duration, intervention characteristics, outcome measurements, as well as reported outcomes and significant levels. The search was broad with no limited target health-related behaviors in the search strategy. In this study, apps were considered effective if statistically significant results of health-related behavior changes were reported for them.

4. Risk of Bias Assessment

Risk of bias assessment for included studies was conducted by two authors using a modified version of the Scottish Intercollegiate Guideline Network (SIGN) checklist for RCT [7]. Specifically, the SIGN checklists were applied to grade the level of evidence of each study. The evaluation items are divided into 7 categories as follows: random sequence generation (selection bias), allocation concealment (selection

 https://doi.org/10.4258/hir.2018.24.3.207

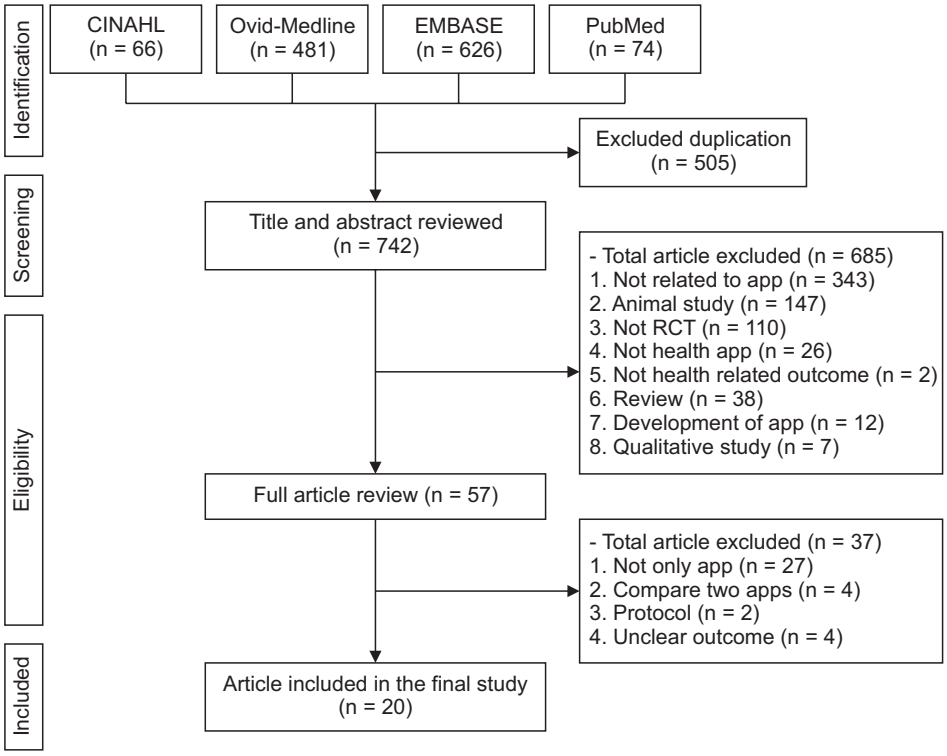

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram.

bias), blinding of participants and personnel (performance bias), blinding of outcome assessment (detection bias), incomplete outcome data (attrition bias), selective reporting (reporting bias), and other bias. Each domain was classified as having a low, high, or unclear risk of bias. This assessment was conducted by two researchers, and disagreements were resolved by discussion.

III. Results

1. Characteristics of Included Studies

The RCTs included in this review were published between 2014 and 2017. Most (n = 16) were simply RCTs; the rest were an open-labelled RCT, an unmasked RCT, a cluster RCT, and a single-blinded parallel 3-arm pilot cluster RCT. The longest study duration was 8 months [8].

Five studies had a large number of participants [8-12]. The cluster RCT study had the largest sample size (n = 1,192) [11]. Two studies had a moderate number of participants [13,14]. The other 13 studies had a small number of participants. All 20 studies had between 80% and 100% retention rates; 18 (90%) studies achieved high (80–100%) retention rates in the intervention group, and only two (10%) studies [8,14] had a moderate retention rate.

The 20 selected studies were analyzed in this systematic review, and the following 16 themes related to health behaviors were created: physical activity (4), alcoholism (1), dietary change (1) adherence of medication or therapy (2), preparation of clinical procedure (1), PTSD management (1), weight loss (2), prenatal education and engagement (1), adherence to follow-up clinic appointments (1), improvement of CPR skill performance (2), suicide prevention (1), prevention of CHD (1), smoking cessation (1), and knowledge improvement of pap testing (1). The comprehensive characteristics of included articles are summarized in Table 1.

2. Risk of Bias of Selected Studies

The evaluation of risk of bias for all 20 studies was conducted using the SIGN checklist for RCTs. The results were summarized using the risk of bias table of RevMan 5.3 software. A total of 15 studies properly reported random sequence generation. Only one article did not mention random sequence generation [15].

For allocation concealment, only 6 studies explicitly mentioned that allocation was concealed [8,10,16-18]. However, 8 studies did not discuss allocation concealment adequately. Participants were blinded in 4 studies [16,17,19,20]. However, due to the traits of mHealth apps, some studies could not be conducted with perfect blinding. For the remaining 16 studies, either they were not blinded or information on blinding was not clearly provided in the reporting.

Among the studies in this review, three reported the blinding of outcome assessment [16,17,19]. For reporting bias, 9 studies had a low risk of bias, and 11 were evaluated as un-

Table 1. Summary characteristics of articles included

Study	Year	Country	Study design	Themes	Participants	Group (N, Age)	Intervention	Follow-up	Features	Outcome measurement	Outcomes
Glynn et al. [19]	2014	West of Ireland	Open-label RCT	Physical activity	Patient referred by their primary care health professional or self-referred	-Intervention group (n = 37, 46 ± 11 years) -Control group (n = 40, 42 ± 11 years)	Smartphone app (SMART MOVE)	2 mo	-Automatic feedback -Tracking of step count and calories burnt -Graphic display of step-count history -Goal setting	-Mean daily step count -Improvement of daily step count in the intervention group	-Mean of daily step count -Intervention group (n = 4,365) -Control group (n = 5,138) -Mean improvement in the intervention group (n = 1,029; 95% CI, 213–1843; $p = 0.009$)
Gustafson et al. [8]	2014	USA	Unmasked RCT	Alcoholism	Those who met the criteria for DSM-IV alcohol dependence	-Intervention group (n = 170, 38.3 ± 9.5 years) -Control group (n = 179, 38.4 ± 11.2 years)	Smartphone app (A-CHESS)	8 mo	-Education (audio-guided relaxation) -Interactive features (counselor discuss with users)	-Count the fewer risk of drinking days	-Fewer risk of drinking day: intervention group (1.39), control group (2.75) (Mean difference = 1.37; 95% CI, 0.46–2.27; $p = 0.003$)
Ipjian and Johnston [25]	2017	USA	RCT	Dietary change	Healthy adults	-Intervention group (n = 15, 35.5 ± 14.9 years) -Control group: Journal group (n = 15, 33.3 ± 16.8 years)	Smartphone app (MyFitnessPal)	1 mo	-Received feedback on sodium content of foods -Entering the data (daily food and beverage entry to monitor dietary sodium levels) -Instruction and educational materials	-24-hour sodium excretion -Satisfaction	-24-hour sodium excretion: intervention group (-838 ± 1,093 mg/day), control group (236 ± 1,333 mg/day) ($p = 0.010$) -Intervention group reported significantly greater satisfaction than the journal group ($p = 0.010$)

Table 1. Continued 1

Study	Year	Country	Study design	Themes	Participants	Group (N, Age)	Intervention	Follow-up	Features	Outcome measurement	Outcomes
Hammonds et al. [15]	2015	USA	RCT	Adherence to medication	College students who had a current prescription for an antidepressant and regularly used a smartphone device	-Intervention group (n = 30, 20.3 ± 4.0 years) -Control group (n = 27, 20.9 ± 4.7 years)	Smartphone app	1 mo	-Reminder (participant entered the prescribed information regarding dosing; participants were asked to use the medication reminder app to indicate when they had taken their medication by responding to the message received)	-Percent adherence	-Participants in the treatment group were 3.5 times more likely to adhere to their medication regimen than those in the control group (95% CI, 0.945–12.966; X^2 [1, n = 40] = 3.64; p = 0.057; ϕ = 0.30).
Kang et al. [12]	2016	China	RCT	Preparation before procedure (bowel preparation)	Colonoscopy outpatients	-Intervention group (n = 387, 44.4 ± 13.2 years) -Control group (education of standard form) (n = 383, 45.5 ± 13.0 years)	Smartphone app (WeChat)	6 mo	-Social media instruction	-Level of adequate bowel preparation (Ottawa score<6).	-A higher proportion of patients in the intervention group had adequate bowel preparation than in the control group (82.2% vs. 69.5%, p < 0.001). -A higher proportion of patients in the intervention group had cecal intubation and have adenomas than in the control group (97.2% vs. 93.2%, p = 0.014; 18.6% vs. 12.0%, p = 0.012).

Table 1. Continued 2

Study	Year	Country	Study design	Themes	Participants	Group (N, Age)	Intervention	Follow-up	Features	Outcome measurement	Outcomes
Kuhn et al. [13]	2017	USA	RCT	Self-management of PTSD symptoms	PTSD	-Intervention group (n = 62, 39.43 ± 15.16 years) -Control group: Waitlist condition (n = 58, 39.12 ± 14.08 years)	Smartphone app (PTSD coach)	3 mo	-Instruction/Education: Offer sound psychoeducational information and evidence-based cognitive behavior tools	-PTSD symptoms -Depression symptoms -Psychosocial functioning	-Intervention group participants had significantly greater improvement in PTSD symptoms (p = 0.035), depression symptoms (p = 0.005), and psychosocial functioning (p = 0.007) than waitlist participants. -A greater proportion of PTSD coach participants achieved clinically significant PTSD symptom improvement (p = 0.018) than waitlist participants.
Laing et al. [10]	2014	USA	RCT	Loss weight	Overweight primary care patients	-Participants (n = 212, 43.3 ± 14.3 years): Intervention group (n = 107) and control group (usual care; n = 105)	Smartphone app (MyFitnessPal)	6 mo	-Self-monitoring -Goal setting -Feedback	-Weight loss at 6 months -SBP -Satisfaction with the app	-Weight change: No difference between groups (-0.30 kg [95% CI, -1.50–0.95; p = 0.63]) -Systolic pressure: No difference between groups (-1.7 mmHg; 95% CI, -7.1–3.8; p = 0.55) -Most users reported satisfaction.

Table 1. Continued 3

Study	Year	Country	Study design	Themes	Participants	Group (N, Age)	Intervention	Follow-up	Features	Outcome measurement	Outcomes
Leoford et al. [14]	2016	USA	RCT	Prenatal education and engagement	Pregnant mothers	-Intervention group (n = 65, 43.3 ± 14.3 years) -Control group: Notebook (n = 62, 29.29 ± 4.80 years)	Smartphone app	5 mo	-Entering the data (record information: pregnancy experience—weight, blood pressure, experience between prenatal appointments)	-The frequency of recording information	-Intervention group: Record information about pregnancy more frequently (p = 0.04), developed greater patient activation (p = 0.02) than control group. -No difference was detected in interpersonal clinical communication.
Lee et al. [1]	2017	USA	RCT	Neck exercise (pain manage)	Office workers	-Participants (43.3 ± 14.3 years): Intervention group (n = 11) and control group (Brochure; n = 9)	Smartphone app (MyFitnessPal)	2 mo	-Education of neck exercise	-VAS for pain intensity -NDI for functional disability -MVFS/MVES -Level of exercise adherence	-VAS & NDI: In the intervention group, VAS (p = 0.003) and NDI (p = 0.005) improved significantly after the app-based neck exercise. -MVES (p = 0.013), improved significantly -In the intervention group, the level of exercise adherence was high.

Table 1. Continued 4

Study	Year	Country	Study design	Themes	Participants	Group (N, Age)	Intervention	Follow-up	Features	Outcome measurement	Outcomes
Lyu et al. [18]	2016	China	RCT	Clinical follow-up with discharged patients	Patients diagnosed with head and neck tumor	-Intervention group (n = 53, 61.0 ± 13.0 years) -Control group: Telephone (TFU) (n = 46, 61.5 ± 9.3 years)	Smartphone app (Wechat)	6 mo	-Entering the data	-Time consumption for follow-up delivery -Total economic cost -Lost to follow-up rate -Satisfaction with methods	-Time consumption for follow-up delivery: WFU group (23.36 ± 6.16 minutes) was significantly shorter than that in the TFU group (42.89 ± 7.15 minutes) ($p < 0.001$). -Total economic cost in WFU group (RMB 90 Yuan) was much lower than in the TFU group, 9.80% (5/51). -Lost follow-up rate: Intervention group (7.02%, 4/57) and TFU group (9.80%, 5/51) -Satisfaction: Intervention group (94.34%, 50/53) and TFU group (80.43%, 37/46) (95% CI, 0.057–0.067; $p = 0.034$)

Table 1. Continued 5

Study	Year	Country	Study design	Themes	Participants	Group (N, Age)	Intervention	Follow-up	Features	Outcome measurement	Outcomes
Nord et al. [11]	2016	Sweden	Cluster RCT	CPR	Students (13 years)	-Intervention group (n = 549, 13 years) -Control group: DVD-based CPR training (n = 575, 13 years)	Smartphone app	1 session (30 minutes)	-Training/Education	-CPR skills -Willingness to act -Improvement of compression depth	-The DVD-based group was superior to the app-based group in CPR skills: a total score of 36 (33–38) vs. 33 (30–36) directly after training ($p < 0.001$) and 33 (30–36) and 31 (28–34) at 6 months ($p < 0.001$). -If a friend suffered cardiac arrest, 78% (DVD) vs. 75% (app) would do compression and ventilations, whereas only 31% (DVD) vs. 32% (app) would perform standard CPR if the victim was a stranger. -At 6 months, the DVD group performed significantly better in 8 out of 12 CPR skill components. Both groups improved compression depth from baseline to follow-up.

Table 1. Continued 6

Study	Year	Country	Study design	Themes	Participants	Group (N, Age)	Intervention	Follow-up	Features	Outcome measurement	Outcomes
Park [27]	2014	Korea	RCT	CPR	Those who agreed to complete CPR curriculum	-Intervention group (n = 33, 61.0 ± 13.0 years) -Control group: Traditional group (n = 31, 61.5 ± 9.3 years)	Smartphone app	1 session	-Education (CPR curriculum)	-The proper compression depth -The proper chest compression -Awareness of chest compression	-Proper compression depth (mm) was shown in the traditional group (53.77) compared to the smartphone group (48.35) ($p < 0.01$). -The proper chest compression (%) was formed suitably ($p < 0.05$) in the traditional group (73.96%) more than smartphone group (60.51%). -The traditional group (3.83 points) had higher awareness of chest compression accuracy ($p < 0.001$) than the smartphone group (2.32 points).

Table 1. Continued 7

Study	Year	Country	Study design	Themes	Participants	Group (N, Age)	Intervention	Follow-up	Features	Outcome measurement	Outcomes
Paul et al. [26]	2016	United Kingdom	RCT	Physical activity	People with stroke	-Intervention group (n = 15, 56.3 ± 8.7 years) -Control group: Usual care (n = 8, 55.3 ± 12.6 years)	Smartphone app (STARFISH)	6 wk	-Goal setting -Visualization on physical activity -Receive a real-time feedback on their physical activity	-Daily step count -Sedentary time -Heart rate -Blood pressure -Body mass index -Fatigue -Daily activity -Ten-meter walk test -Quality of life scale -General well-being index	-Average daily step count increased by 39.3% (4,158–5,791 steps/day) in the intervention group and reduced by 20.2% (3,694–2,947 steps/day) in the control group (p = 0.005 for group-time interaction). -There were no significant for the other outcome measures.
Perera et al. [28]	2014	New Zealand	RCT	Adherence to therapy	People undergoing ART	-Intervention group (n = 17) -Control group: Standard group (n = 11)	Smartphone app	3 months	-Reminder: 24-hour medication clock -A personalized disease state simulation of immune activity	-Self-reported adherence -Pharmacy dispensing record -HIV viral load	-Participants in the intervention group showed a significantly higher level of self-reported adherence to ART at 3 months (p = 0.03) and decreased viral load (p = 0.023) compared to individuals using the standard version.

Table 1. Continued 8

Study	Year	Country	Study design	Themes	Participants	Group (N, Age)	Intervention	Follow-up	Features	Outcome measurement	Outcomes
Tighe et al. [20]	2017	Australia	RCT	Suicide prevention	Indigenous Australian youth	-Intervention group (n = 31, 27.48 ± 9.54 years) -Control group: Waitlisted (n = 31, 24.97 ± 6.28 years)	Smartphone app (ibobbly)	6 wk	-Education (delivered acceptance-based therapy)	-Depressive Symptom Inventory-Suicidality Subscale (DSI-SS) -Patient Health Questionnaire (PHQ-9) -Kessler Psychological Distress Scale (K-10) -Barratt Impulsive Scale (BIS-11)	-DSI-SS were significant in the ibobbly arm (t = 2.40; df = 58.1; p = 0.0195); these differences were not significant compared with the waitlist arm (t = 1.05; df = 57.8; p = 0.2962). -Participants in the ibobbly group showed substantial and statistically significant reductions in PHQ-9 and K10 scores compared with waitlist. -No differences were observed in impulsivity between intervention and control groups.

https://doi.org/10.4258/hir.2018.24.3.207

Table 1. Continued 9

Study	Year	Country	Study design	Themes	Participants	Group (N, Age)	Intervention	Follow-up	Features	Outcome measurement	Outcomes
Van Reijen et al. [9]	2017	New Zealand	RCT	Neuromuscular training	Athletes	-Intervention group (n = 110, 37.6 ± 13.1 years) -Control group: Booklet (n = 110, 38.1 ± 13.7 years)	Smartphone app	2 mo	-Training (exercise program)	-Compliance with the exercise program -Incidence density of self-reported recurrent ankle sprains	-The mean compliance to the exercise scheme was 73.3% (95% CI, 67.7–78.1) in the App group compared with 76.7% (95% CI, 71.9–82.3) in the Booklet group. No significant difference in compliance was found between groups. -The incidence densities of self-reported time-loss recurrences were not significantly different between groups (HR = 3.07; 95% CI, 0.62–15.20).
Zhang et al. [29]	2017	Singapore	RCT	CHD prevention	Working people	-Intervention group (n = 40) -Control group: Health promotion website only (n = 40)	Smartphone app (Care4Heart)	6 weeks	-Education (heart disease prevention [SBCHDP] program)	-Awareness of CHD -Knowledge of CHD -Blood cholesterol control	-Participants in the intervention group had a better awareness of CHD ($X^2 = 6.486$, $p = 0.039$). -A better overall CHD knowledge level ($t = 3.171$, $p = 0.002$) -Better behavior concerning blood cholesterol control ($X^2 = 4.54$, $p = 0.033$) than participants in the control group.

Table 1. Continued 10

Study	Year	Country	Study design	Themes	Participants	Group (N, Age)	Intervention	Follow-up	Features	Outcome measurement	Outcomes
Carter et al. [16]	2013	New Zealand	RCT	Loss weight	Overweight volunteer	-Intervention group (n = 43, 41.2 ± 8.5 years) -Control group: Website group (n = 40, 42.5 ± 8.3 years) and Paper diary group (n = 42, 41.9 ± 10.6 years)	Smartphone app (My Meal Mate)	6 mo	-Goal setting -Self-monitoring of diet and activity -Feedback via weekly text message	-Self-monitoring -Mean weight change -BMI change -Change in body fat	-Self-monitoring declined over time in all groups. -Mean weight change at 6 months was −4.6 kg (95% CI, −6.2 to −3.0) in the smartphone app group, −2.9 kg (95% CI, −4.7 to −1.1) in the diary group, and −1.3 kg (95% CI, −2.7 to 0.1) in the website group. -BMI change at 6 months was −1.6 kg/m^2 (95% CI, −2.2 to −1.1) in the smartphone group, −1.0 kg/m^2 (95% CI, −1.6 to −0.4) in the diary group, and −0.5 kg/m^2 (95% CI, −0.9 to 0.0) in the website group. -Change in body fat was −1.3% (95% CI, −1.7 to −0.8) in the smartphone group, −0.9% (95% CI, −1.5 to −0.4) in the diary group, and −0.5% (95% CI, −0.9 to 0.0) in the website group.

Table 1. Continued 11

Study	Year	Country	Study design	Themes	Participants	Group (N, Age)	Intervention	Follow-up	Features	Outcome measurement	Outcomes
Cheung et al. [17]	2015	Hong Kong	Single blinded, parallel, 3-arm pilot cluster RCT	Smoking cessation	Quitters who had stopped smoking recently	-Intervention group (n = 54, 40.4 ± 10.4 years) -Control group: Facebook online social group (n = 40, 37.6 ± 8.0 years) and Booklet group (n = 54, 42.7 ± 10.4 years)	Smartphone app (Whats App)	2 mo	-Group discussion -Training for smoking cessation	-2- and 6-month relapse rates (the proportion of participants who smoked at least 5 cigarettes in 3 consecutive days)	-The WhatsApp group (17%, 7/42) reported more relapse than the control group (42.6%, 23/54) at 2-month (OR = 0.27; 95% CI, 0.10–0.71) and 6-month (40.5%, 17/42 vs. 61.1%, 33/54; OR = 0.43; 95% CI, 0.19–0.99) follow-ups.
Christensen [30]	2014	New England	RCT	Enhance knowledge of pap testing	Women attending a university in New England	-Intervention group (n = 37, 20.41 years) -Control group: Standard pamphlet on pap testing (n = 37, 20.81 years)	Smartphone app (Digital health education application on Pap testing)	2 mo	-Education	-Knowledge of pap testing	-Pap testing knowledge: Intervention group (5.26 ± 1.66 to 10.78 ± 1.51), control group (5.70 ± 1.61 to 8.92 ± 1.88) (p < 0.001); Knowledge scores on the posttest increased significantly in both groups but was significantly higher in the intervention group.

RCT: randomized controlled trial, DSM-IV: Diagnostic and Statistical Manual of Mental Disorders 4th edition, PTST: posttraumatic stress disorder, SBP: systolic blood pressure, VAS: Visual Analogue Scale, NDI: Neck Disability Index, MVFS: maximal voluntary flexion strength, MVES: maximal voluntary extension strength, CPR: cardiopulmonary resuscitation, ART: antiretroviral therapy, HIV: human immunodeficiency virus, CHD: Coronary heart disease, BMI: body mass index.

clear on the presence of bias. A summary of Cochrane's risk of bias table is presented in Figure 2.

3. Content Characteristics of Apps

Some app characteristics of contents were categorized according to the behavior change technique taxonomy by Abraham and Michie [21] that is, providing information, planning (goal setting), reminding, providing feedback, or monitoring. Furthermore, additional app characteristics, such as entering data, education/training, and communication were derived from this study. Ten apps have multiple functions to manage health-related behaviors. The most common function of mHealth apps is providing the opportunity for education or training.

All participants in the control groups underwent standardized or usual care. Table 2 shows a summary of app characteristics.

IV. Discussion

These days, mHealth apps seem to be ubiquitous, and the body of research indicating their effectiveness has been growing rapidly. However, evidence for the effectiveness of mHealth apps has been uncertain, and much remains unknown in terms of health-related behavior changes and clinical results.

In this study, 20 RCTs were included to evaluate the effectiveness of mHealth apps for health-related behavior change. Seventeen studies among 20 showed a positive contribution to the enhancement of health-related behaviors. This result is similar to other previous evidence reviews [3]. Therefore, using mHealth apps could be an effective strategy to improve outcomes of users along with the high popularity of smartphone use in the everyday lives of users. However, three studies did not show positive effects of mHealth apps on health-related behavior changes. Laing et al. [10] found that

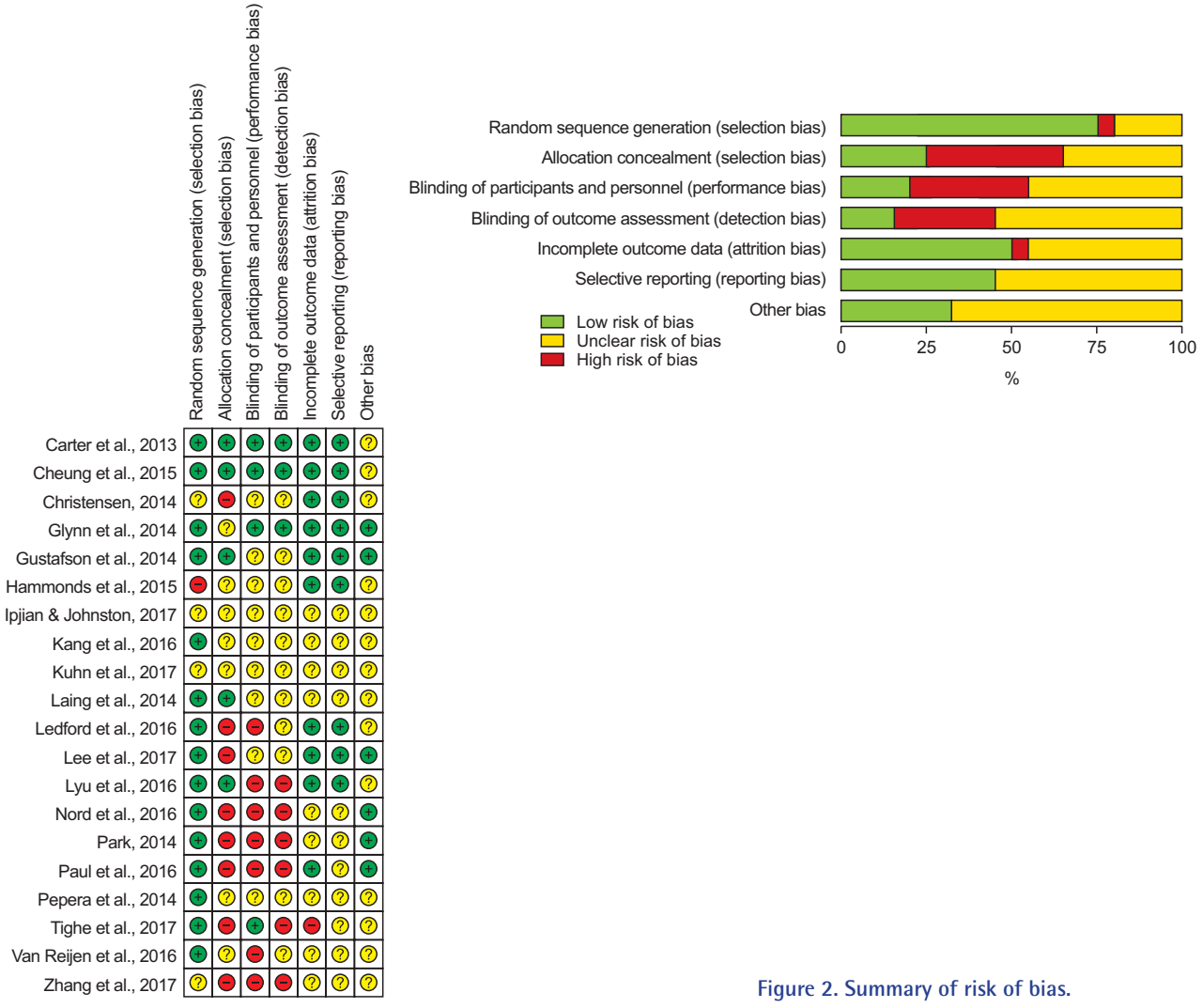

Figure 2. Summary of risk of bias.

https://doi.org/10.4258/hir.2018.24.3.207

Table 2. Summary characteristics of application contents

Study	Providing information	Planning (goal setting)	Remind	Feedback	Entering the data	Education/ Training	Communication
Glynn et al. [19]		√		√	√		
Gustafson et al. [8]						√	√
Ipjian and Johnston [25]	√			√	√	√	
Hammonds et al. [15]			√		√		
Kang et al. [12]						√	
Kuhn et al. [13]	√					√	
Laing et al. [10]		√		√	√		
Leoford et al. [14]					√		
Lee et al. [1]						√	
Lyu et al. [18]					√		√
Nord et al. [11]						√	
Park [27]						√	
Paul et al. [26]		√		√			
Perera et al. [28]			√				
Tighe et al. [20]						√	
Van Reijen et al. [9]						√	
Zhang et al. [29]						√	
Carter et al. [16]		√		√	√		
Cheung et al. [17]						√	√
Christensen [30]						√	

there was no difference in weight loss between intervention and control groups. In addition, Nord et al. [11] reported that a DVD-based CPR education group had better performance than an app group. Tighe et al. [20] also reported that using an app had no effect on decreasing depression symptoms and impulsivity behaviors. However, it is difficult to assess the effectiveness of apps based on the results of a single study, and more studies with controlled research design are needed.

This study differs from other reviews in that we only included RCTs to examine the effectiveness of mHealth apps, because RCTs are considered the 'gold standard' in evaluating the effects of intervention and provide a valuable source of evidence in research and treated as a powerful experimental tool to examine the effectiveness of intervention [22]. Therefore, this systematic review, which only analyzed the results of RCTs, has provided more reliable evidence for the effectiveness of smartphone health apps.

More than half of the reviewed studies had small samples (<60). In addition, 11 reported that the duration of intervention was less than 2 months. According to Man-Son-Hing et al. [23], trials with larger samples and longer intervention durations or follow-up times are more reliable to appraise the effectiveness of intervention. Based on the results of this review, to demonstrate a certain effect using mHealth apps for health-related behavior changes, more research with long intervention durations and large samples is needed.

According to Zhao et al. [3], the retention rate is defined as the proportion of participants who remain to complete a study. The Cochrane Handbook for Systematic Reviews of Interventions reported that studies with retention rates over 80% are classified as having low attrition, and studies with retention between 60% and 79% are classified as having moderate attrition [24]. Eighteen studies achieved high (>80%) retention rates in the intervention group. In this study, over half of the studies had a moderate-high retention rate. It can be assumed that the reasons for the high retention rate were the high feasibility and acceptability of app use in users' everyday lives. Thus, mHealth apps could be effectively adopted for users to improve health-related outcomes by managing and supporting health-related behaviors of users.

In this review, some studies [8,10,13,15-19,25,26] considered multiple function apps, such as entering data and

providing feedback, education, and reminders. In contrast, other studies [1,9,11,12,14,20,27-30] considered apps that have only one function; most of these apps had an education function. One study reported that having multiple app functions is much better to manage health status and to improve health-related behaviors [4]. However, this result might not be concrete because applying apps to the different types of situations, health behaviors, and participants can yield different results and effectiveness. For example, Laing et al. [10] conducted a study regarding an app with multiple functions for weight loss, and the effect was not significant. However, Zhang et al. [29] reported a significant result in improvement of coronary heart disease (CHD) knowledge and awareness using a single-function app. This suggests that many unconditional features of apps might not work properly, and a customized app is needed that fits the purpose and intent of the user. In this aspect, mHealth apps could provide individualized information via feedback to users and benefit them.

With respect to the risk of bias of included studies, the categories of selection bias (allocation concealment), performance bias (blinding of participants and researchers), and detection bias (blinding of outcome assessment) indicated high risk of bias. Therefore, to enhance the quality of studies and ensure low risk of bias, researchers should consider rigorous study design and reporting. Based on the results of this review, further studies using meta-analysis are needed to identify the effects of mHealth apps with specific outcomes. In addition, the effectiveness of the apps should be verified with the effects or risks being used with verification.

In this review, there were several limitations. First, we used broad key words, such as 'health', 'behaviors', 'smartphone', and 'mobile'. For this reason, many articles related to the use of mHealth apps on specific diseases or health conditions—for instance, diabetes, hypertension, or asthma—might not have been included in this study. Second, in this review, only RCTs were included to analyze the effectiveness of mHealth apps. However, some of the RCTs did not fully follow the form of the RCT or applied a modified form of RCT. However, well-structured RCT is needed to verify mHealth app effectiveness. Third, almost all of the studies considered in this review were conducted in developed countries. Hence, it is difficult to generalize our results to developing countries. Fourth, we did not conduct a meta-analysis because of interventions with different kinds of mHealth apps. However, with the results of meta-analysis, the effectiveness of mHealth apps can be verified more clearly.

This systematic review was conducted to examine the effectiveness of mHealth apps to lead to changes in their targeted health-related behavior. This study summarized the characteristics and changes in targeted health outcomes. To our knowledge, there has been no previous systematic review of RCTs for identifying the effectiveness of mHealth apps in improving health-related behaviors. Similar to previous studies, this systematic review also found that the use of mHealth apps has a positive impact on health-related behaviors, such as physical activity, diet change, adherence to medication or therapy, and knowledge enhancement related to clinical procedures. Moreover, most apps seem to promote better clinical health outcomes. Most app users are satisfied with the use of mHealth apps to manage their health in comparison to users of conventional care. Although most studies analyzed indicated statistically significant effects to improve health, more RCTs with larger samples and longer applied interventions are still needed to confirm the effectiveness of mHealth apps. To assess the efficacy of mHealth apps in greater detail, further research is needed.

Conflict of Interest

No potential conflict of interest relevant to this article was reported.

References

1. Lee J, Lee M, Lim T, Kim T, Kim S, Suh D, et al. Effectiveness of an application-based neck exercise as a pain management tool for office workers with chronic neck pain and functional disability: a pilot randomized trial. Eur J Integr Med 2017;12:87-92.
2. Boudreaux ED, Waring ME, Hayes RB, Sadasivam RS, Mullen S, Pagoto S. Evaluating and selecting mobile health apps: strategies for healthcare providers and healthcare organizations. Transl Behav Med 2014;4(4):363-71.
3. Zhao J, Freeman B, Li M. Can mobile phone apps influence people's health behavior change? An evidence review. J Med Internet Res 2016;18(11):e287.
4. 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;10(1):e1001363.
5. Kitsiou S, Pare G, Jaana M, Gerber B. Effectiveness of mHealth interventions for patients with diabetes: an overview of systematic reviews. PLoS One

 https://doi.org/10.4258/hir.2018.24.3.207

2017;12(3):e0173160.
6. Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Syst Rev 2015;4:1.
7. Scottish Intercollegiate Guidelines Network [Internet]. Edinburgh: Scottish Intercollegiate Guidelines Network, Healthcare Improvement Scotland; c2018 [cited at 2018 Jul 1]. Available from: http://www.sign.ac.uk/.
8. Gustafson DH, McTavish FM, Chih MY, Atwood AK, Johnson RA, Boyle MG, et al. A smartphone application to support recovery from alcoholism: a randomized clinical trial. JAMA Psychiatry 2014;71(5):566-72.
9. Van Reijen M, Vriend I, Zuidema V, van Mechelen W, Verhagen EA. The "Strengthen your ankle" program to prevent recurrent injuries: a randomized controlled trial aimed at long-term effectiveness. J Sci Med Sport 2017;20(6):549-554.
10. Laing BY, Mangione CM, Tseng CH, Leng M, Vaisberg E, Mahida M, et al. Effectiveness of a smartphone application for weight loss compared with usual care in overweight primary care patients: a randomized, controlled trial. Ann Intern Med 2014;161(10 Suppl):S5-12.
11. Nord A, Svensson L, Hult H, Kreitz-Sandberg S, Nilsson L. Effect of mobile application-based versus DVD-based CPR training on students' practical CPR skills and willingness to act: a cluster randomised study. BMJ Open 2016;6(4):e010717.
12. Kang X, Zhao L, Leung F, Luo H, Wang L, Wu J, et al. Delivery of instructions via mobile social media app increases quality of bowel preparation. Clin Gastroenterol Hepatol 2016;14(3):429-435.e3.
13. Kuhn E, Kanuri N, Hoffman JE, Garvert DW, Ruzek JI, Taylor CB. A randomized controlled trial of a smartphone app for posttraumatic stress disorder symptoms. J Consult Clin Psychol 2017;85(3):267-73.
14. Ledford CJW, Canzona MR, Cafferty LA, Hodge JA. Mobile application as a prenatal education and engagement tool: a randomized controlled pilot. Patient Educ Couns 2016;99(4):578-82.
15. Hammonds T, Rickert K, Goldstein C, Gathright E, Gilmore S, Derflinger B, et al. Adherence to antidepressant medications: a randomized controlled trial of medication reminding in college students. J Am Coll Health 2015;63(3):204-8.
16. Carter MC, Burley VJ, Nykjaer C, Cade JE. Adherence to a smartphone application for weight loss compared to website and paper diary: pilot randomized controlled trial. J Med Internet Res 2013;15(4):e32.
17. Cheung YT, Chan CH, Lai CK, Chan WF, Wang MP, Li HC, et al. Using WhatsApp and Facebook online social groups for smoking relapse prevention for recent quitters: a pilot pragmatic cluster randomized controlled trial. J Med Internet Res 2015;17(10):e238.
18. Lyu KX, Zhao J, Wang B, Xiong GX, Yang WQ, Liu QH, et al. Smartphone application WeChat for clinical follow-up of discharged patients with head and neck tumors: a randomized controlled trial. Chin Med J (Engl) 2016;129(23):2816-23.
19. Glynn LG, Hayes PS, Casey M, Glynn F, Alvarez-Iglesias A, Newell J, et al. Effectiveness of a smartphone application to promote physical activity in primary care: the SMART MOVE randomised controlled trial. Br J Gen Pract 2014;64(624):e384-91.
20. Tighe J, Shand F, Ridani R, Mackinnon A, De La Mata N, Christensen H. Ibobbly mobile health intervention for suicide prevention in Australian Indigenous youth: a pilot randomised controlled trial. BMJ Open 2017;7(1):e013518.
21. Abraham C, Michie S. A taxonomy of behavior change techniques used in interventions. Health Psychol 2008;27(3):379-87.
22. Sullivan GM. Getting off the "gold standard": randomized controlled trials and education research. J Grad Med Educ 2011;3(3):285-9.
23. Man-Son-Hing M, Laupacis A, O'Rourke K, Molnar FJ, Mahon J, Chan KB, et al. Determination of the clinical importance of study results. J Gen Intern Med 2002;17(6):469-76.
24. Linke SE, Gallo LC, Norman GJ. Attrition and adherence rates of sustained vs. intermittent exercise interventions. Ann Behav Med 2011;42(2):197-209.
25. Ipjian ML, Johnston CS. Smartphone technology facilitates dietary change in healthy adults. Nutrition 2017;33:343-7.
26. Paul L, Wyke S, Brewster S, Sattar N, Gill JM, Alexander G, et al. Increasing physical activity in stroke survivors using STARFISH, an interactive mobile phone application: a pilot study. Top Stroke Rehabil 2016;23(3):170-7.
27. Park SS. Comparison of chest compression quality between the modified chest compression method with the use of smartphone application and the standardized traditional chest compression method during CPR. Technol Health Care 2014;22(3):351-8.
28. Perera AI, Thomas MG, Moore JO, Faasse K, Petrie KJ. Effect of a smartphone application incorporating

personalized health-related imagery on adherence to antiretroviral therapy: a randomized clinical trial. AIDS Patient Care STDS 2014;28(11):579-86.
29. Zhang H, Jiang Y, Nguyen HD, Poo DC, Wang W. The effect of a smartphone-based coronary heart disease prevention (SBCHDP) programme on awareness and knowledge of CHD, stress, and cardiac-related lifestyle behaviours among the working population in Singapore: a pilot randomised controlled trial. Health Qual Life Outcomes 2017;15(1):49.
30. Christensen S. Evaluation of a nurse-designed mobile health education application to enhance knowledge of Pap testing. Creat Nurs 2014;20(2):137-43.

https://doi.org/10.4258/hir.2018.24.3.207