

# SUITABILITY OF CURRENT EVALUATION FRAMEWORKS FOR USE IN THE HEALTH TECHNOLOGY ASSESSMENT OF MOBILE MEDICAL APPLICATIONS: A SYSTEMATIC REVIEW

Magdalena Ruth Moshi

School of Public Health, University of Adelaide

[magdalena.moshi@adelaide.edu.au](mailto:magdalena.moshi@adelaide.edu.au)

Rebecca Tooher

School of Public Health, University of Adelaide

Tracy Merlin

Adelaide Health Technology (AHTA), School of Public Health, The University of Adelaide

**Objectives:** To identify and appraise existing evaluation frameworks for mobile medical applications (MMA) and determine their suitability for use in health technology assessment (HTA) of these technologies.

**Methods:** Systematic searches were conducted of seven bibliographic databases to identify literature published between 2008 and 2016 on MMA evaluation frameworks. Frameworks were eligible if they were used to evaluate at least one of the HTA domains of effectiveness, safety, and/or cost and cost-effectiveness of an MMA. After inclusion, the frameworks were reviewed to determine the number and extent to which other elements of an HTA were addressed by the framework.

**Results:** A total of forty-five frameworks were identified that assessed MMAs. All frameworks assessed whether the app was effective. Of the thirty-four frameworks that examined safety, only seven overtly evaluated potential harms from the MMA (e.g., the impact of inaccurate information). Only one framework explicitly considered a comparator. Technology specific domains were sporadically addressed.

**Conclusion:** None of the evaluation frameworks could be used, unaltered, to guide the HTA of MMAs. To use these frameworks in HTA they would need to identify relevant comparators, improve assessments of harms and consider the ongoing effect of software updates on the safety and effectiveness of MMAs. Attention should also be paid to ethical issues, such as data privacy, and technology specific characteristics. **Implications:** Existing MMA evaluation frameworks are not suitable for use in HTA. Further research is needed before an MMA evaluation framework can be developed that will adequately inform policy makers.

**Keywords:** Technology assessment, Health policy/standards, Mobile health, Mobile applications

Mobile health (mHealth) has the potential to change health systems and how care is delivered (1). One form of mHealth is mobile medical applications (MMAs) also known as ‘apps’. These are a type of software available for mobile platforms (e.g., smartphone, tablet, smartwatch) (1). In a medical context, MMAs may be used by patients to self-manage and/or screen medical conditions, rather than presenting at hospitals or clinics for additional appointments. MMAs may also allow for medical practitioners and/or allied health workers to remotely monitor, screen and manage their patients (2;3).

A potential barrier to the successful integration of MMAs into health systems is that many come at a cost to the patient, or require in-app purchases, which some patients are unable to afford. While some MMAs may have a negligible costs—and thus will not warrant public funding—others may require subscriptions or

come with accessories, such as wearables and implantable devices. Furthermore, medical practitioners and allied health workers that use MMA-based services during a clinical encounter are often unable to claim reimbursement for the interpretation of MMA output or for treatment guided by MMA results.

Health management organizations (HMO) in the United States have reimbursed some MMAs since 2013 (4). Similarly, since 2014, private health insurers have reimbursed specific MMAs in Germany (5;6). It is unclear how these apps were selected for reimbursement, although this may have depended on whether the MMA was approved by the relevant regulatory authority (e.g., the United States Food and Drug Administration [FDA]).

Countries with tax funded universal healthcare like Australia and Great Britain currently do not reimburse the use of MMAs. However, the National Institute for Health Care Excellence (NICE) in Britain is currently investigating ways to assess MMAs and provide guidance on their use (7;8). If the use of MMAs becomes routine in clinical consultations, MMA-guided care will need to be formally assessed.

The authors thank David Tamblyn for assisting in the trialing and testing of the standardization tool. Magdalena Moshi is a recipient of an *Australian Government Research Training Program Scholarship*.

This systematic review is part of a larger research project to develop or adapt an evaluation framework for MMAs and determine the feasibility of a reimbursement pathway for MMAs in Australia. The aim of our review was to identify and appraise existing evaluation frameworks for MMAs and determine their suitability for use in health technology assessment (HTA). In this context an evaluation framework was defined as a method for determining an MMA's effectiveness, safety and/or cost, cost-effectiveness.

## METHODS

### Literature Search

We searched PubMed (MEDLINE), EMBASE, CINAHL, PsychINFO, The Cochrane Library, Compendex, and Business Source Complete between January 1, 2008 (when the first publicly accessible online application store opened) and October 31, 2016 (9). We used a broad search strategy including terms for MMAs (e.g., mHealth app\*, telehealth app\*), mobile platforms (e.g., cellular phone, mobile device) and evaluation (e.g., criteri\*, apprais\*). Grey literature sources were also searched to identify any relevant material that may have not been identified through the database search. The full search strategy is given in the Supplementary Materials.

### Study Eligibility Criteria

Papers were selected for inclusion if they met the predetermined eligibility criteria. The population of interest were participants aged 18 years or over that used an MMA. The intervention of interest was an MMA evaluation framework. This included frameworks that assessed all mHealth apps, as MMAs are a subset of these. MMAs were defined as mobile apps (including accompanying accessories or attachments) available on various platforms (smartphone, tablets, smart watches, etc.) that have a therapeutic or diagnostic intended purpose. Framework(s) aimed solely at assessing pregnancy, health promotion, or disease prevention apps (e.g., medication management, smoking cessation, and weight management) were excluded as the apps' intended purpose was not diagnostic or therapeutic. The outcomes of interest were the core HTA evaluation domains of effectiveness, safety, and/or cost, cost-effectiveness. There was no comparator as the aim of the systematic review was not to determine the effectiveness of these evaluation frameworks but, rather, to identify the HTA domains that they address. Only frameworks available in English were included. Frameworks that were duplicated in several articles were collated and reported as a single record.

### Study Selection

Two reviewers (M.M. and T.M.) screened the literature separately and applied the inclusion criteria. M.M. reviewed all title and abstracts retrieved from the searches, while T.M.

assessed 10 percent. The full text articles were screened against the inclusion criteria by M.M. Any articles in which M.M. was unsure of eligibility, were discussed with T.M. and a consensus decision made. The reference lists of included papers were perused to identify any additional relevant references.

### Data Extraction

The data extracted from the papers included: Author and dates of publication, source affiliation, country of origin, name of framework, study design, description of framework, intended audience/user, type of MMA, framework scoring system, and HTA domains addressed. The included papers were not critically appraised for study quality as this was a methodological systematic review.

### Framework Assessment

A checklist was created to act as a tool to standardize data extraction. Using the checklist, each framework was assessed to determine if it included any of nine traditional HTA domains; six were core domains considered essential for a full HTA: current use of the technology; description and technical characteristics; effectiveness; safety; cost and cost-effectiveness; organizational aspects, and three were optional domains: legal aspects; ethical aspects; social aspects (10;11). The checklist was trialed and tested by an HTA expert and M.M. and was found to have reasonable inter-rater reliability (Kappa = 0.77).

## RESULTS

The systematic searches retrieved 12,690 citations. An additional twelve papers were identified from grey literature sources. Six additional frameworks were identified through the perusing of the included publications' reference lists. Three frameworks were excluded as the information provided was not in a usable form. An evidence base of forty-six papers met the inclusion criteria, and two of these papers published on the same framework. Thus, forty-five frameworks were identified that assessed whether MMAs are safe, effective and/or cost-effective. Figure 1 illustrates the complete study selection process. Tables 1 and 2 provide details on all forty-five MMA evaluation frameworks.

### Overview of Frameworks

All of the included frameworks addressed mHealth applications, with 73 percent ( $n = 33$ ) explicitly assessing MMAs. Most of the frameworks that evaluated MMAs were sourced from universities. The remaining 27 percent ( $n = 12$ ) were developed by private organizations, institutes, medical schools, or governmental organizations.

The frameworks originated from three geographical regions. Most (49 percent,  $n = 22$ ) came from North America, while

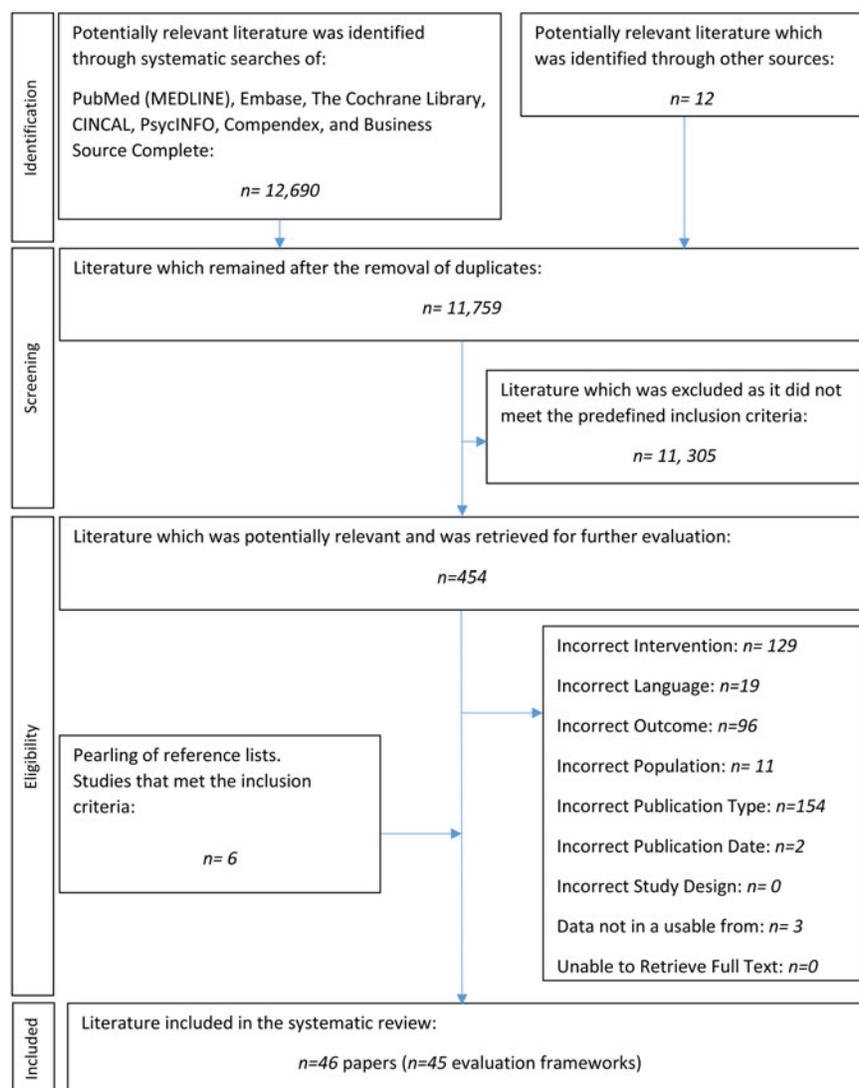


Fig. 1. PRISMA flow-chart of literature selection.

Europe contributed 40 percent ( $n = 18$ ) and the remaining 11 percent ( $n = 5$ ) originated from the Asia-Pacific region.

The frameworks that assessed MMAs came in a variety of different formats. Some of these formats included questionnaires, data extraction criteria, flow charts, and varying types of lists. Due to the variety of formats, it was difficult to categorize the frameworks into types. Less than half of the frameworks had a scoring system (Table 1).

#### Intended Audience

The included frameworks had different intended audiences or purposes, such as: MMA developers ( $n = 3$ ), used for quality assurance for user protection ( $n = 13$ ), for patients ( $n = 17$ ), or used for quality assurance in a research setting ( $n = 18$ ). All of these frameworks assessed the HTA domains concerning the current use of the technology, description and technical characteristics, effectiveness, safety, and ethical aspects.

#### Intended Health Condition

Most of the frameworks were aimed at evaluating MMAs that focused on the treatment, management or diagnosis of chronic health conditions. Diabetes and mental health were the most commonly addressed conditions with 13 percent ( $n = 6$ ) of the frameworks focused on MMAs for each condition. Cancer and pain were the next most frequently addressed conditions (7 percent,  $n = 3$ , each). Three frameworks (7 percent) were aimed at MMAs managing sexuality transmitted infections, with two of these focused specifically on the human immunodeficiency virus (HIV).

#### Critical Appraisal of the Evidence-Base Underpinning the MMA

The credibility of the information included in an MMA was assessed by fifteen (33 percent) frameworks, including through: level of evidence or grade of the recommendation (12;13); assessment in a randomized controlled trial (13–15);

Table 1. Description of Frameworks

Author, date	Intended audience	Type of mobile application	Source affiliation <sup>1</sup>	Scoring system in framework	Country of origin <sup>1</sup>
Albrecht, Von Jan & Pramann (20), 2013	Patients	MMA	Institute	No	Norway
Anxiety and Depression Association of America (ADAA) (38), 2016	Quality assurance for user protection	MMA	Private organization	Yes	USA
Arnhold, Quade & Kirch (57), 2014	Quality assurance for research setting	MMA	University	Yes	Germany
Aungst et al. (32), 2014	Health professionals	MMA	University	No	USA
Basilico et al. (41), 2016	Quality assurance for research setting	MMA	University	Yes	Italy
Beatty, Fukuoka & Whooley (15), 2013	Patients, health professionals	MMA	University	No	USA
BinDhim et al. (58), 2015	Regulators	mHealth	University	Yes	Australia
Brooks et al. (39), 2015	Patients	MMA	University	No	USA
Chan et al. (18), 2015	Patients, health professionals	MMA	University	No	USA
Chomutare et al. (59), 2011	Quality assurance for research setting	MMA	University	No	Norway
Demidowich et al. (60), 2012	Quality assurance for research setting	MMA	Medical school	Yes	USA
Drincic et al. (16), 2016	Patients	MMA	University	No	USA
Fairburn & Rothwell (33), 2015	Quality assurance for research setting	MMA	University	No	UK
Ferrero-Alvarez-Rementeria (36), 2013	All stakeholders in mHealth	mHealth	Governmental organization	No	Spain
Gautham, Iyengar, & Johnson, C. W. (30), 2015	Health professionals	MMA	University	No	UK
Gibbs et al. (25), 2016	Quality assurance for research setting	MMA	University	No	UK
Grundy et al. (26), 2016	Patients, health professionals, and app developers	mHealth	University	Yes	Australia
Hacking Medicine Institute (HMi) (34), 2016	Quality assurance for user protection	MMA	Institute	Yes	USA
Hoppe, Cade & Carter, (61) (2016)	Patients	MMA	University	Yes	UK
Huckvale et al. (27), 2015	Quality assurance for research setting	MMA	University	Yes	UK
IMS Institute for Healthcare Informatics (62), 2013	Quality assurance for user protection	mHealth	Institute	Yes	USA
Jin & Kim (63), 2015	Health professionals	mHealth	University	Yes	Republic of Korea
Kassianos et al. (21), 2015	Quality assurance for research setting	mHealth	University	No	UK
Laloo et al. (47), 2015	Quality assurance for research setting	MMA	University	Yes	Canada
Lee et al. (42), 2015	Patients, health professionals, and app developers	MMA	University	No	New Zealand
Martinez-Perez et al. (28;29), 2013/2015	Patients, health professionals	MMA	University	Yes	Spain
McMillan et al. (19), 2016	Quality assurance for research setting	MMA	University	Yes	UK
mHIMSS App Usability Work Group (31), 2012	Health professionals	mHealth	Private organization	No	USA
Mobasherhi et al. (64), 2014	Quality assurance for research setting	MMA	University	No	UK
Murfin (24), 2013	Health professionals	mHealth	University	No	USA
Pandey et al. (22), 2013	Quality assurance for research setting	MMA	University	No	USA
Portelli & Eldred (17), 2016	Quality assurance for research setting	MMA	University	Yes	UK
Powell et al. (14), 2016	Patients, health professionals	MMA	Private organization	Yes	USA
Psyber Guide (65)	Quality assurance for user protection	MMA	Private organization	Yes	USA
Reynoldson et al. (43), (2014)	Patients	MMA	University	Yes	UK
Robustillo Cortes et al. (12), 2014	Patients	MMA	University	No	Spain
Schnall et al. (44), 2015	Quality assurance for research setting	MMA	University	Yes	USA
Shah, Jonassiant & Castro (66), 2014	Patients	MMA	University	Yes	USA
Shaia et al. (45), 2016	Patients	MMA	Medical school	Yes	USA
Shen et al. (46), 2015	Quality assurance for research setting	MMA	University	No	Canada
Singh et al. (37), 2016	Quality assurance for research setting	MMA	University	No	USA
Social Wellth (67), 2016	Quality assurance for research setting	mHealth	Private organization	Yes	USA

Table 1. Continued

Author, date	Intended audience	Type of mobile application	Source affiliation <sup>1</sup>	Scoring system in framework	Country of origin <sup>1</sup>
Stoyanov et al. (13), 2015	Quality assurance for research setting	MMA	University	Yes	Australia
Walsworth (40), 2012	Patients	mHealth	University	No	USA
Yasini & Marchard (23), 2015	Quality assurance for user protection	mHealth	Private organization	No	France

**Explanatory note:** <sup>1</sup>Based on first author affiliation.

study design (i.e., clinical trial, controlled trial) (16;17); improved health outcomes (18); sample size, intervention fidelity and evaluation design (19); and publication in peer-reviewed journals (20–23). Powell et al. (14), Beatty et al. (15), and Stoyanov et al. (13) specifically question if the MMA was clinically tested using RCTs. Stoyanov et al. (13) asked whether the MMA has been verified through trialing/testing, ranking the responses with the lowest being no RCTs and the highest being multiple RCTs. One framework considered whether a systematic review or meta-analysis had been conducted about the MMA and topic area (24). Furthermore, twenty-seven (60 percent) of the included frameworks asked about the MMA's source of information.

#### HTA Domains

More than half of the assessed frameworks included the following HTA domains: MMA effectiveness; description of technical characteristics; safety; current use of the technology; and ethical aspects (Figure 2). Five frameworks assessed six domains (19;25–29), whereas two frameworks only assessed a single domain (Table 2) (30;31). The average number of domains addressed was four ( $\bar{x} = 3.9$ ).

#### Core HTA Domains

**Effectiveness.** Every framework assessed the effectiveness of MMAs in some capacity. Eleven frameworks (24 percent) evaluated user satisfaction. Thirty (67 percent) frameworks evaluated the technical efficacy of MMAs. Beatty et al. (15), Drincic et al. (16), and McMillan et al. (19) appraised efficacy of the applications, but did not provide any further detail of what they meant; it could be interpreted as both therapeutic and/or diagnostic effectiveness. Only one framework explicitly considered comparative effectiveness (32). The framework by Aungst et al. (32) asked whether an MMA already exists for the current reference condition (clinical tool used in practice).

#### Investigative MMAs

The diagnostic accuracy of MMAs was assessed by 29 percent ( $n = 13$ ) of the frameworks. Martinez-Perez et al. (28;29) reviewed the accuracy of an MMA's calculations. Stoyanov

et al. (13), Powell et al. (14), Fairburn and Rothwell (33), Gibbs et al. (25), and Murfin (24), all assessed the accuracy or specificity of the information given in the MMAs, while Hacking Medical Institute (HMi) (34) reviewed the MMA's clinical credibility. Powell et al. (14) included four response options which rated how the MMA was designed to improve a specific condition, whereas Gibbs et al. (25) ranked the accuracy of the information also using four options. None of the thirteen frameworks assessed subsequent changes in patient or decision-making management associated with use of the MMA (a necessary domain for determining the effectiveness of investigative interventions) (35). Furthermore, none of the included frameworks assessed the clinical utility of MMAs, that is, the health impacts of an MMA that provides diagnostic, information.

#### Therapeutic MMAs

Therapeutic effectiveness was assessed by 71 percent ( $n = 32$ ) of the frameworks. Three frameworks addressed primary patient-relevant outcomes including quality of life and mortality (16;18;19), whereas, 25/32 (78 percent) made provision for the reporting of surrogate outcomes (e.g., physiological, biochemical, and/or behavior change parameters); for example, a diabetes management MMA that could log glucose (HbA1C) readings, or an HIV management app that could track T-cell counts.

#### Safety

Safety was addressed in thirty-four (76 percent) frameworks with twenty-seven (79 percent) of these assessing the source of the information used by the MMA, and three appraising how the information sources were selected. Only seven frameworks evaluated the harms of the app itself (e.g., adverse events) (14;18;19;26;32;36;37). Six frameworks addressed whether the MMA had been trialed or tested and whether safety concerns had been identified during the process (15;16;20;22;38;39).

#### Cost, Cost-Effectiveness

Only one framework assessed the cost-effectiveness domain by asking whether a health economic evaluation had been conducted (40). However, this domain was partially addressed by

**Table 2.** HTA Domains Addressed by Each Framework

Author, date	Health technology assessment (HTA) domains								
	Domains addressed in a full HTA						Auxiliary domains addressed in a full HTA		
	Current use of the technology	Characteristics of the technology	Effectiveness	Safety	Cost, cost-effectiveness	Organisational aspects	Legal aspects	Ethical aspects	Social aspects
Albrecht, Von Jan & Pramann (20), 2013	✓	✓	✓	✓	X	X	X	✓	X
Anxiety and Depression Association of America (ADAA) (38), 2016	X	✓	✓	✓	X	X	X	X	X
Arnhold, Quade & Kirch (57), 2014	X	X	✓	X	X	X	X	✓	X
Aungst et al. (32), 2014	X	X	✓	✓	X	✓	X	X	X
Basilico et al. (41), 2016	✓	✓	✓	✓	~	X	X	X	X
Beatty, Fukuoka & Whooley (15), 2013	X	✓	✓	✓	~	X	X	X	✓
BinDhim et al. (58), 2015	✓	✓	✓	X	X	X	X	X	X
Brooks et al. (39), 2015	X	X	✓	✓	X	X	X	X	X
Chan et al. (18), 2015	X	✓	✓	✓	X	✓	X	✓	X
Chomutare et al. (59), 2011	X	✓	✓	X	X	X	X	X	X
Demidowich et al. (60), 2012	X	✓	✓	X	X	X	X	X	X
Drincic et al. (16), 2016	✓	✓	✓	✓	X	X	X	X	X
Fairburn & Rothwell (33), 2015	✓	X	✓	✓	X	X	X	X	X
Ferrero-Alvarez-Rementería (36), 2013	✓	✓	✓	✓	X	X	X	✓	X
Gautham, Iyengar, & Johnson (30), 2015	X	X	✓	X	X	X	X	X	X
Gibbs et al. (25), 2016	✓	✓	✓	✓	~	X	✓	✓	X
Grundy et al. (26), 2016	✓	✓	✓	✓	X	X	✓	✓	X
Hacking Medicine Institute (HMi) (34), 2016	✓	✓	✓	✓	X	X	X	✓	X
Hoppe, Cade & Carter, (61) (2016)	✓	✓	✓	X	X	X	X	X	X
Huckvale et al. (27), 2015	✓	✓	✓	✓	~	X	✓	✓	X
IMS Institute for Healthcare Informatics (62), 2013	X	✓	✓	X	X	X	X	X	X
Jin & Kim (63), 2015	X	X	✓	✓	X	X	X	✓	X
Kassianos et al. (21), 2015	X	✓	✓	✓	X	X	X	X	X
Laloo et al. (47), 2015	X	X	✓	✓	X	X	X	X	✓
Lee et al. (42), 2015	✓	✓	✓	✓	~	X	X	✓	X
Martinez-Perez et al. (28;29), 2013/2015	X	✓	✓	✓	X	✓	X	✓	✓
McMillan et al. (19), 2016	✓	✓	✓	✓	X	X	X	✓	✓
mHIMSS App Usability Work Group (31), 2012	X	X	✓	X	X	X	X	X	X
Mobasheri et al. (64), 2014	✓	✓	✓	✓	X	X	X	✓	X
Murfin (24), 2013	X	✓	✓	✓	X	X	X	✓	X
Pandey et al. (22), 2013	✓	X	✓	✓	~	X	X	✓	X
Portelli & Eldred (17), 2016	X	✓	✓	✓	X	X	X	X	✓
Powell et al. (14), 2016	✓	✓	✓	✓	X	X	X	✓	X

Table 2. Continued

Author, date	Health technology assessment (HTA) domains								
	Domains addressed in a full HTA						Auxiliary domains addressed in a full HTA		
	Current use of the technology	Characteristics of the technology	Effectiveness	Safety	Cost, cost-effectiveness	Organisational aspects	Legal aspects	Ethical aspects	Social aspects
Psyber Guide (65),	✓	✓	✓	✓	✗	✗	✗	✓	✗
Reynoldson et al. (43), (2014)	✓	✓	✓	✓	~	✗	✗	✓	✗
Robustillo Cortes et al. (12), 2014	✗	✓	✓	✓	✗	✗	✓	✓	✗
Schnall et al. (44), 2015	✓	✓	✓	✗	~	✗	✗	✗	✗
Shah, Jonassiant & Castro (66), 2014	✗	✓	✓	✗	✗	✗	✗	✗	✗
Shaia et al. (45), 2016	✓	✓	✓	✓	~	✗	✗	✓	✗
Shen et al. (46), 2015	✓	✓	✓	✓	~	✗	✗	✓	✗
Singh et al. (37), 2016	✗	✓	✓	✓	✗	✗	✗	✓	✓
Social Wellth (67), 2016	✓	✓	✓	✗	✗	✗	✗	✓	✗
Stoyanvo et al. (13), 2015	✓	✓	✓	✓	~	✗	✗	✓	✗
Walsworth (40), 2012	✓	✗	✓	✓	✓	✗	✗	✗	✗
Yasini & Marchard (23), 2015	✓	✓	✓	✓	✗	✗	✗	✗	✗

Note. ✓ Domain was addressed; ~ Domain was partially addressed; ✗ Domain was not addressed. HTA, health technology assessment.

11 (24 percent) frameworks that reviewed the cost of MMAs in terms of the price to download the application or to undertake in-app purchases (13;15;22;25;27;41–46).

#### Current use of the Technology

The current use of the technology was assessed by 25 (55 percent) of the frameworks. Seventeen assessed (68 percent) usage of the MMA (e.g., rates, use, trends), sixteen (64 percent) assessed the intended user population and fifteen (60

percent) considered the intended purpose of the app (e.g., diagnosis, management, or treatment).

#### Description and Technical Characteristics

Technical characteristics of MMAs were assessed by 78 percent ( $n = 35$ ) of the frameworks. The type of device (e.g., mobile platform, operating systems, software versions) was evaluated in twenty-three (67 percent) frameworks and nineteen (54 percent) evaluated whether experts were consulted during the

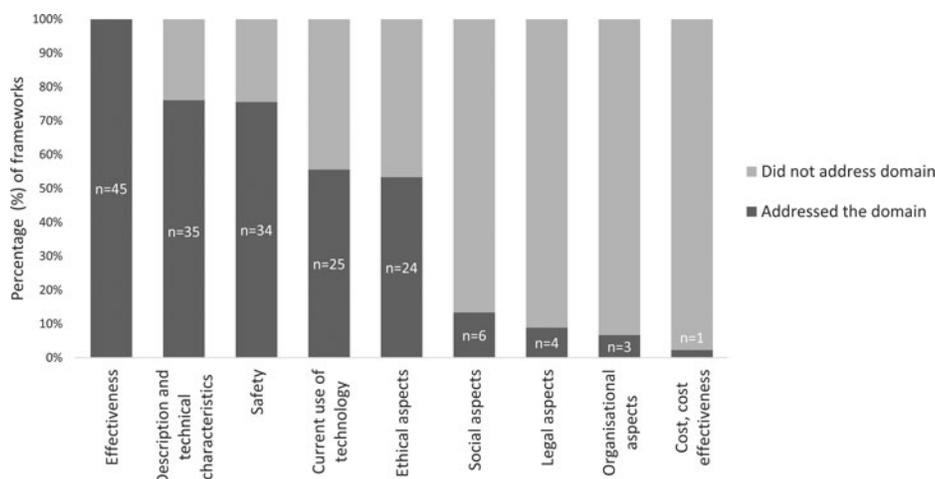


Fig. 2. The proportion of frameworks that address each HTA domain.

development of the app. Eleven (31 percent) assessed whether the MMA had communicative capabilities (e.g., communication with personal health records, communication with electronic health records, and healthcare provider-patient communication), and eight (23 percent) considered whether the MMA had personalization capabilities.

#### Organizational Aspects

Only three (7 percent) of the included frameworks assessed whether the MMA would have organizational implications. Two of the frameworks recorded if any training was needed to use the application and if adopting the MMA would alter the usage of existing services (18;28;29). One framework assessed whether the MMA would alter the daily practices of clinicians (32).

#### Optional HTA Domains

**Legal Aspects.** Four (9 percent) of the identified frameworks assessed the legal implications of MMAs (12;25–27). Three of these determined whether there were legal implications by asking whether the MMA had a disclaimer concerning clinical accountability (12;25;26). Two of the frameworks required consideration of the possibility of copyright infringement (26;27).

**Ethical Aspects.** Ethical considerations were examined by twenty-four (53 percent) frameworks. Of these, eighteen (75 percent) recorded whether the MMA had a privacy policy (although only four considered the individual content of the privacy policy); eighteen (75 percent) evaluated patient confidentiality provisions in the app; and fourteen (58 percent) assessed conflicts of interest (e.g., affiliation, funding, third party sponsorship). Four (17 percent) frameworks appraised equity (e.g., socioeconomic status, disability, language, and age) (13;18;27;43), and an additional four (17 percent) frameworks assessed MMA accessibility (e.g., geographical location) (19;27–29;42).

**Social Aspects.** Six (13 percent) frameworks assessed how the MMA provides social support to the users (15;17;28;29;37;47), for example, whether the MMA provides psychosocial support, if the MMA can provide support through social media, or if access to social support is facilitated.

## DISCUSSION

None of the included frameworks could be used “off the shelf” to evaluate MMAs in a full HTA requiring assessment across all six core HTA domains. Frameworks by Grundy et al. (26), HMi (34), Huckvale et al. (27), Martinez-Perez et al. (28;29), and McMillan et al. (19) all assessed six HTA domains, but none of these addressed all of the six core domains. Ethical, social, and legal considerations are frequently omitted in typical HTAs. However, we found that, for MMA specific

evaluations, ethical issues were often addressed, whereas organization of care, and cost and cost-effectiveness domains, together with legal considerations, were the least likely to be addressed (11).

#### Safety

Nearly a quarter of the evaluation frameworks did not assess safety in any capacity. Only five (16 percent) frameworks explicitly considered the MMA’s ability to cause harm or adverse events. None of the frameworks explicitly assessed the comparative safety of the MMA with reference to other MMAs or current clinical practice without use of an MMA. It is possible that evaluators of MMAs do not find safety as important a concern as the effectiveness this technology. MMAs with attachments (such as glucometers, oximeters, or electrocardiogram leads) that have the potential to physically harm, may be more obvious candidates for safety assessment, rather than the individual MMA itself.

A further concern regards the source of information on which the assessment of safety was based. We found that only one quarter of frameworks checked this factor. Lack of attention to information sources is problematic because of the potential harms caused by misinformation. The International Medical Device Regulators Forum (IMDRF)(48) states that the greatest risks and benefits posed by software which acts as a medical device (SaMD), such as an MMA, relates to its output and how it impacts on a patient’s clinical management or other healthcare related decisions, not from direct contact with the device itself. Apps which utilize poor/weak evidence bases could present a range of clinical harms. For example, chronically ill patients using medication incorrectly due to inaccurate feedback from the MMA; rehabilitation patients doing inappropriate exercises; or, potentially more seriously, the long-term consequences to health of receiving a false negative diagnosis from an investigational MMA. However, both the physical harm and risks associated with misinformation are of interest in an HTA and may affect subsequent policy decisions, regarding access to, or reimbursement of MMAs.

#### Effectiveness

Normally to assess the effectiveness of an intervention in an HTA, the results of the intervention are compared with current practice or an existing intervention. However, only one of forty-five frameworks considered in this systematic review referred to the availability of a comparator MMA or to the current management of the condition without the MMA. Without a comparator identified, it is impossible to adequately assess the effectiveness of an MMA or conduct a full HTA that could inform policy decision making.

**Investigative MMAs.** The safety and effectiveness of an investigative medical service can be determined through direct or linked

evidence (35;49–51). None of the included frameworks appeared to use a direct evidence approach to evaluate an MMA. Frameworks did address the diagnostic accuracy of an MMA; however, none linked this to subsequent changes in management or healthcare decision making. In any case, those frameworks that assessed the diagnostic accuracy of MMAs did not collect sufficient information to enable a full assessment.

**Therapeutic MMAs.** In the evaluation of a therapeutic medical service safety and effectiveness can be determined through direct randomized trials (preferred), indirect comparisons of randomized trials, or nonrandomized trials, and observational studies. The purpose of this evidence is to identify the best available clinical evidence for the primary indication relative to the main comparator (50;51). None of the included frameworks adequately addressed the key elements evaluated to demonstrate the therapeutic effectiveness of an MMA, with only seven frameworks considering the quality of the evidence base (such as whether clinical trials were considered or what health outcomes eventuated from use of the MMA).

#### Cost and Cost-effectiveness and Organizational Issues

The frameworks did not consider the impact that the direct costs of an MMA would have on the current health system, or the potential effect on other medical services or devices (Table 2). Only one framework, by Walsworth (40), addressed cost and cost-effectiveness and it only assessed if the value of the MMA justified the cost. Formal economic evaluations of the value for money associated with the use of the MMA was not required by any of the identified evaluation frameworks. It may be that the cost impact of MMAs is considered to be trivial and, therefore, cannot justify a formal economic evaluation; the cost of some MMAs is small. However, costs do not just relate to the unit price of the MMA but also to downstream costs associated with behavior affected by the MMA.

Although no single framework addressed all of the information necessary for an HTA of an MMA, there were elements considered across the frameworks that if combined could produce a comprehensive evaluation framework. Technology-specific characteristics are particularly relevant.

#### Technology Specific Considerations

There are several technology specific considerations that may need to be addressed when conducting an HTA on an MMA.

The first is a requirement to assess ethical aspects, specifically data privacy. Over half of the frameworks identified had assessed ethical issues concerned with MMAs. Connectivity to the Internet, networks, and other devices through a portable handheld device (i.e., smartphone or tablet) is a unique vulnerability of apps that are used for medical purposes. If the MMA is jeopardized (e.g., hacked or viruses), there is the potential to

compromise sensitive personal health information (52;53). The IMDRF (54) regards security concerns relating to the privacy and confidentiality of data (of an SaMD) as safety concerns. The accessibility, availability, and integrity of the device output are crucial for patient treatment and diagnoses (54). A further concern is that companies have been known to sell consumer data (55). The FDA (52;53) has attempted to address cybersecurity concerns by publishing pre- and postmarket guidance documents which provide recommendations for the management of cyber threats to medical devices.

Second, compatibility and connectivity concerns are important for the evaluation of MMAs. MMA performance may vary between different platforms (i.e., smartphone, tablet, or smart-watch), with different operating systems (i.e., Android versus iOS), and for different generations of the same device (i.e., iPhone 5 versus iPhone 6). Additionally, the impact of software updates must also be allowed for, as MMAs are a dynamic technology which is constantly changing. One update that makes an incremental change to the MMA may not alter its intended purpose. However, multiple subsequent incremental updates may change the intended purpose of the MMA (56). As highlighted by the IMDRF (54), if not managed systematically, any modification (e.g., updates) to the software throughout its lifecycle, including maintenance, poses a risk to the patient. A full HTA that is used to inform policy decisions regarding an MMA may need to assess these compatibility and connectivity concerns to ensure that the app is consistent across various platforms, operating systems, and devices, as well as identify when software modifications such as updates, should trigger reassessment of the MMA.

#### Key Components of an MMA Evaluation Framework

MMA evaluation frameworks intended to appraise apps for HTA purposes should include: consideration of a comparator; a complete assessment of safety and harms from misinformation; a more detailed evaluation of ethical issues such as equity and secure management of confidential data; a consideration of the impact of software updates on the safety and effectiveness of the MMA. It is difficult to determine from this systematic review whether social, legal, and organizational aspects, or the cost and cost-effectiveness of MMAs should be evaluated. It would be helpful to identify indicators that could trigger an assessment of these factors. More research is needed to determine the concepts that should be included in an MMA evaluation framework for HTA purposes and what structure the framework could take. For example, would the structure of such a framework follow the HTA domains or use another categorization method which is more suitable to address the unique challenges presented by MMAs (e.g., development quality, information security, technical considerations).

The second stage of our research project is to conduct interviews with MMA developers, health professionals and policy

makers to identify factors important in the use, assessment and reimbursement of MMAs.

## LIMITATIONS

As with any systematic review there were some limitations with the research. There is a risk of publication bias, although we attempted to limit this by conducting grey literature searches and including all frameworks in the review that met the selection criteria. Another possible limitation was that the checklist we created to standardize and identify which HTA domains the frameworks addressed, could have been idiosyncratic. The tool was pilot tested by an HTA expert and found to have fair inter-rater reliability. The use of the core HTA domains to assess the MMA evaluation frameworks may have limited the concepts identified. To address this, we have also highlighted several technology-specific considerations that would need to be included in a HTA of MMAs.

In conclusion, none of the forty-five identified frameworks could be used, unaltered, to assess an MMA in a full HTA to inform a policy decision. While several of the identified MMA evaluation frameworks addressed up to six of the HTA domains, there was a lack of detail that would be required to undertake a full HTA. To adapt these frameworks for use in the HTA of MMAs there would need to be greater consideration of the comparator, and a fuller assessment of the harms associated with MMAs. Our results also indicate that an HTA of an MMA should pay particular attention to the ethical issues associated with the technology, in particular to the secure handling of confidential data. The impact of MMA updates on overall conclusions of safety and effectiveness would also need consideration.

## Policy Implications

This research has various policy implications. First, there is a need to develop an MMA evaluation framework that is compatible with HTA and addresses all of the relevant policy concerns. Further information is needed from developers and users of apps about the technology-specific characteristics of MMAs that would need to be addressed in a HTA evaluation framework to inform policy decisions on MMAs. Second, due to technology specific considerations, such as the app development cycle, varying platforms, and cybersecurity risks, regulatory and reimbursement authorities may need to work collaboratively with each other if MMAs are to be safely integrated into clinical practice and healthcare delivery.

## SUPPLEMENTARY MATERIAL

The supplementary material for this article can be found at <https://doi.org/10.1017/S026646231800051X>

Supplementary Materials:

<https://doi.org/10.1017/S026646231800051X>

## CONFLICTS OF INTEREST

The authors have nothing to declare.

## REFERENCES

- Powell AC, Landman AB, Bates DW. In search of a few good apps. *JAMA*. 2014;311:1851-1852.
- Royal Australian College of General Practice. 1.3 m-Health: Royal Australian College of General Practice; [cited April 6, 2016]. <http://www.racgp.org.au/digital-business-kit/m-health/> (accessed August 6, 2018).
- Laranjo L, Lau A, Oldenburg B, et al. *mHealth technologies for chronic disease prevention and management*. An evidence check review brokered by the Sax Institute for Healthdirect Australia Sax Institute. Ultimo, NSW: Sax Institute for Healthdirect Australia; 2015.
- Dolan B. WellDoc's BlueStar secures first mobile health reimbursement: *Mobihealthnews*; 2013 [cited April 29, 2016]. <http://mobihealthnews.com/23026/welldocs-bluestar-secures-first-mobile-health-reimbursement> (accessed August 6, 2018).
- Paris V, Devaus M, Wei L. Health systems institutional characteristics: A survey of 29 OECD countries. OECD Health Working Papers No 50. OECD Publishing 2010.
- Dolan B. Caterna offers prescribable, reimbursed eye-strengthening gaming app in Germany: *Mobihealthnews*; 2014 [cited April 29, 2016]. <http://mobihealthnews.com/31549/caterna-offers-prescribable-reimbursed-eye-strengthening-gaming-app-in-germany> (accessed August 6, 2018).
- Clifford R. NICE to bring HTA processes to mobile health apps: PM Group Worldwide Limited 2017 [cited May 1, 2017]. [http://www.pmlive.com/pharma\\_news/nice\\_to\\_bring\\_hta\\_processes\\_to\\_mobile\\_health\\_apps\\_1185722](http://www.pmlive.com/pharma_news/nice_to_bring_hta_processes_to_mobile_health_apps_1185722) (accessed August 6, 2018).
- Mulryne J, Clemence B. NICE to introduce HTA-style app assessment process from March 2017: EU Regulatory and Litigation; 2017 [cited May 1, 2017]. <http://www.digitalhealthdownload.com/2017/02/nice-introduce-hta-style-app-assessment-process-march-2017/> (accessed August 6, 2018).
- Donker T, Petrie K, Proudfoot J, Clarke J, Birch MR, Christensen H. Smartphones for smarter delivery of mental health programs: A systematic review. *J Med Internet Res*. 2013;15:e247.
- Busse R, Orvain J, Velasco M, et al. Best practice in undertaking and reporting health technology assessments - Working group 4 report. *Int J Technol Assess Health Care*. 2002;18:361-422.
- Merlin T, Tamblin D, Ellery B, Group IQA. What's in a name? Developing definitions for common health technology assessment product types of the International Network of Agencies for Health Technology Assessment (inahta). *Int J Technol Assess Health Care*. 2014;30:430-437.
- Robustillo Cortés Me, Cantudo Cuenca MR, Morillo Verdugo R, Calvo Cidoncha E. High quantity but limited quality in healthcare applications intended for HIV-infected patients. *Telemed J E Health*. 2014;20:729-735.
- Stoyanov SR, Hides L, Kavanagh DJ, et al. Mobile app rating scale: A new tool for assessing the quality of health mobile apps. *JMIR Mhealth Uhealth*. 2015;3:e27.
- Powell AC, Torous J, Chan S, et al. Interrater Reliability of mHealth app rating measures: Analysis of top depression and smoking cessation apps. *JMIR Mhealth Uhealth*. 2016;4:e15.
- Beatty AL, Fukuoka Y, Whooley MA. Using mobile technology for cardiac rehabilitation: A review and framework for development and evaluation. *J Am Heart Assoc*. 2013;2:e000568.
- Drincic A, Prahalad P, Greenwood D, Klonoff DC. Evidence-based mobile medical applications in diabetes. *Endocrinol Metab Clin North Am*. 2016;45:943-965.
- Portelli P, Eldred C. A quality review of smartphone applications for the management of pain. *Br J Pain*. 2016;10:135-140.

18. Chan S, Torous J, Hinton L, Yellowlees P. Towards a framework for evaluating mobile mental health apps. *Telemed J E Health*. 2015;21:1038-1041.
19. McMillan B, Hickey E, Patel MG, Mitchell C. Quality assessment of a sample of mobile app-based health behavior change interventions using a tool based on the National Institute of Health and Care Excellence behavior change guidance. *Patient Educ Couns*. 2016;99:429-435.
20. Albrecht UV, Von Jan U, Pramann O. Standard reporting for medical apps. In: Mantas J, Hasman A, editors. *Informatics, management and technology in healthcare*. 190: IOC Press; 2013:201-203.
21. Kassianos AP, Emery JD, Murchie P, Walter FM. Smartphone applications for melanoma detection by community, patient and generalist clinician users: A review. *Br J Dermatol*. 2015;172:1507-1518.
22. Pandey A, Hasan S, Dubey D, Sarangi S. Smartphone apps as a source of cancer information: Changing trends in health information-seeking behavior. *J Cancer Educ*. 2013;28:138-142.
23. Yasini M, Marchand G. Mobile health applications, in the absence of an authentic regulation, does the usability score correlate with a better medical reliability? *Stud Health Technol Inform*. 2015;216:127-131.
24. Murfin M. Know your apps: An evidence-based approach to evaluation of mobile clinical applications. *J Physician Assist Educ*. 2013;24:38-40.
25. Gibbs J, Gkatzidou V, Tickle L, et al. 'Can you recommend any good STI apps?' A review of content, accuracy and comprehensiveness of current mobile medical applications for STIs and related genital infections. *Sex Transm Infect*. 2017;93:234-235.
26. Grundy QH, Wang Z, Bero LA. Challenges in assessing mobile health app quality: A systematic review of prevalent and innovative methods. *Am J Prev Med*. 2016;51:1051-1059.
27. Huckvale K, Morrison C, Ouyang J, Ghaghda A, Car J. The evolution of mobile apps for asthma: An updated systematic assessment of content and tools. *BMC Med*. 2015;13:58.
28. Martínez-Pérez B, De La Torre-Díez I, Candelas-Plasencia S, López-Coronado M. Development and evaluation of tools for measuring the quality of experience (QoE) in mHealth applications. *J Med Syst*. 2013;37:9976.
29. Martínez-Pérez B, de la Torre-Díez I, López-Coronado M. Experiences and results of applying tools for assessing the quality of a mHealth app named Heartkeeper. *J Med Syst*. 2015;39:142.
30. Gautham M, Iyengar MS, Johnson CW. Mobile phone-based clinical guidance for rural health providers in India. *Health Informatics J*. 2015;21:253-266.
31. mHIMSS App Usability Work Group. Selecting a mobile app: Evaluating the usability of medical applications. Chicago, IL: Healthcare Information and Management Systems Society; 2012.
32. Aungst TD, Clauson KA, Misra S, Lewis TL, Husain I. How to identify, assess and utilise mobile medical applications in clinical practice. *Int J Clin Pract*. 2014;68:155-162.
33. Fairburn CG, Rothwell ER. Apps and eating disorders: A systematic clinical appraisal. *Int J Eat Disord*. 2015;48:1038-1046.
34. Hacking Medicine Institute. RANKED Hacking Medicine Institute; [cited February 27, 2017]. <http://www.rankedhealth.com/approach/> (accessed August 8, 2018).
35. Merlin T, Lehman S, Hiller JE, Ryan P. The "linked evidence approach" to assess medical tests: A critical analysis. *Int J Technol Assess Health Care*. 2013;29:343-350.
36. Ferrero-Alvarez-Rementeria J, Santana-Lopez V, Escobar-Ubreva A, Vazquez-Vazquez M. Quality and safety strategy for mobile health applications: A certification programme. *Eur J ePractice*. 2013;21:14-26.
37. Singh K, Drouin K, Newmark LP, et al. Developing a framework for evaluating the patient engagement, quality, and safety of mobile health applications. *Issue brief (Commonw Fund)*. 2016;5:1-11.
38. Anxiety and Depression Association of America. Mental health apps: Anxiety and depression association of America; 2016 [cited March 5, 2017]. <https://www.adaa.org/finding-help/mobile-apps>.
39. Brooks GC, Vittinghoff E, Iyer S, et al. Accuracy and usability of a self-administered 6-minute walk test smartphone application. *Circ Heart Fail*. 2015;8:905-913.
40. Walsworth DT. Medical apps: Making your mobile device a medical device. *Fam Pract Manag*. 2012;19:10-13.
41. Basilico A, Marcegaglia S, Bonacina S, Pincioli F. Advising patients on selecting trustful apps for diabetes self-care. *Comput Biol Med*. 2016;71:86-96.
42. Lee H, Sullivan SJ, Schneiders AG, et al. Smartphone and tablet apps for concussion road warriors (team clinicians): A systematic review for practical users. *Br J Sports Med*. 2015;49:499-505.
43. Reynoldson C, Stones C, Allsop M, et al. Assessing the quality and usability of smartphone apps for pain self-management. *Pain Med*. 2014;15:898-909.
44. Schnall R, Mosley JP, Iribarren SJ, et al. Comparison of a user-centered design, self-management app to existing mHealth apps for persons living with HIV. *JMIR Mhealth Uhealth*. 2015;3:e91.
45. Shaia KL, Farag S, Chyjek K, Knopman J, Chen KT. An evaluation of mobile applications for reproductive endocrinology and infertility providers. *Telemed J E Health*. 2017;23:254-258.
46. Shen N, Levitan MJ, Johnson A, et al. Finding a depression app: A review and content analysis of the depression app marketplace. *JMIR Mhealth Uhealth*. 2015;3:e16.
47. Lalloo C, Jibb LA, Rivera J, Agarwal A, Stinson JN. "There's a pain app for that": Review of patient-targeted smartphone applications for pain management. *Clin J Pain*. 2015;31:557-563.
48. International Medical Device Regulators Forum. *Draft: Software as a Medical Device (SaMD): Clinical evaluation*. International Medical Device Regulators Forum; 2016.
49. Medical Services Advisory Committee (MSAC). *Guidelines for the assessment of diagnostic technologies*. Canberra, ACT: Commonwealth of Australia 2005.
50. Medical Services Advisory Committee. *Technical guidelines for preparing assessment reports for the Medical Services Advisory Committee – Medical Service Type: Therapeutic*. Australian Government; 2016.
51. Medical Services Advisory Committee. *Technical guidelines for preparing assessment reports for the Medical Services Advisory Committee – Service Type: Investigative*. Australian Government; 2016.
52. U.S. Food and Drug Administration. Cybersecurity: U.S. Department of Health and Human Services; 2017 [cited June 6, 2017]. <https://www.fda.gov/MedicalDevices/DigitalHealth/ucm373213.htm> (accessed August 8, 2018).
53. U.S. Food and Drug Administration. *FDA fact sheet. The FDA's roll in medical device cybersecurity*. Washington, DC: U.S. Department of Health and Human Services; 2017.
54. International Medical Device Regulators Forum. "Software as a medical device": Possible framework for risk categorization and corresponding considerations. International Medical Device Regulators Forum; 2014.
55. Forbrukerradet. *Consumer protection in fitness wearables*. Norway: Forbrukerradet; 2016.
56. O'Mealy PL, ed. *Update on software as a medical device (SaMD)*. ARCS Scientific Congress 2016; Canberra: Australian Government.
57. Arnhold M, Quade M, Kirch W. Mobile applications for diabetics: A systematic review and expert-based usability evaluation considering the special requirements of diabetes patients age 50 years or older. *J Med Internet Res*. 2014;16:e104.
58. BinDhim NF, Hawkey A, Trevena L. A systematic review of quality assessment methods for smartphone health apps. *Telemed J E Health*. 2015;21:97-104.

59. Chomutare T, Fernandez-Luque L, Arsand E, Hartvigsen G. Features of mobile diabetes applications: Review of the literature and analysis of current applications compared against evidence-based guidelines. *J Med Internet Res*. 2011;13:e65.
60. Demidowich AP, Lu K, Tamler R, Bloomgarden Z. An evaluation of diabetes self-management applications for Android smartphones. *J Telemed Telecare*. 2012;18:235-238.
61. Hoppe CD, Cade JE, Carter M. An evaluation of diabetes targeted apps for Android smartphone in relation to behaviour change techniques. *J Hum Nutr Diet*. 2017;30:326-338.
62. IMS Institute for Healthcare Informatics. *Patient apps for improved healthcare from novelty to mainstream*. IMS Institute for Healthcare Informatics; 2013.
63. Jin M, Kim J. Development and evaluation of an evaluation tool for healthcare smartphone applications. *Telemed J E Health*. 2015;21:831-837.
64. Mobasheri MH, Johnston M, King D, Leff D, Thiruchelvam P, Darzi A. Smartphone breast applications - What's the evidence? *Breast*. 2014;23:683-689.
65. PsyberGuide. PsyberGuide Ratings- Explanation: PsyberGuide; [cited March 5, 2017]. <http://psyberguide.org/psyberguide-ratings-explanation/>.
66. Shah N, Jonassaint J, De Castro L. Patients welcome the sickle cell disease mobile application to record symptoms via technology (SMART). *Hemoglobin*. 2014;38:99-103.
67. Social Wealth. Our approach: Social wealth. 2016 [cited February 27, 2017]. <http://www.happtique.com/>.