



Author: Fiona H. McKay, Christina Cheng, Annemarie Wright, Jane Shill, Hugh Stephens, and Mary Uccellini
Title: Evaluating mobile phone applications for health behaviour change: A systematic review
Year: 2018
Journal: Journal of Telemedicine and Telecare
Volume: 24
Issue: 1
Pages: 22-30
URL: <http://hdl.handle.net/1959.3/456341>

Copyright: Copyright © 2016 the authors. This is the author's final peer-reviewed accepted manuscript version, hosted under the terms and conditions of the Attribution-NonCommercial 4.0 International (CC BY-NC 4.0) license. See <http://creativecommons.org/licenses/by-nc/4.0/>

This is the author's final peer-reviewed accepted manuscript version, hosted under the terms and conditions of the Attribution-NonCommercial 4.0 International (CC BY-NC 4.0) license. See <http://creativecommons.org/licenses/by-nc/4.0/>

The published version is available at: <https://doi.org/10.1177/1357633x16673538>

Evaluating mobile phone applications for health behaviour change: A systematic review

Fiona H McKay¹, Christina Cheng¹, Annemarie Wright², Jane Shill², Hugh Stephens³ and Mary Uccellini³

Abstract

Introduction: Increasing smartphones access has allowed for increasing development and use of smart phone applications (apps). Mobile health interventions have previously relied on voice or text-based short message services (SMS), however, the increasing availability and ease of use of apps has allowed for significant growth of smartphone apps that can be used for health behaviour change. This review considers the current body of knowledge relating to the evaluation of apps for health behaviour change. The aim of this review is to investigate approaches to the evaluation of health apps to identify any current best practice approaches.

Method: A systematic review was conducted. Data were collected and analysed in September 2016. Thirty-eight articles were identified and have been included in this review.

Results: Articles were published between 2011–2016, and 36 were reviews or evaluations of apps related to one or more health conditions, the remaining two reported on an investigation of the usability of health apps. Studies investigated apps relating to the following areas: alcohol, asthma, breastfeeding, cancer, depression, diabetes, general health and fitness, headaches, heart disease, HIV, hypertension, iron deficiency/anaemia, low vision, mindfulness, obesity, pain, physical activity, smoking, weight management and women's health.

Conclusion: In order to harness the potential of mobile health apps for behaviour change and health, we need better ways to assess the quality and effectiveness of apps. This review is unable to suggest a single best practice approach to evaluate mobile health apps. Few measures identified in this review included sufficient information or evaluation, leading to potentially incomplete and inaccurate information for consumers seeking the best app for their situation. This is further complicated by a lack of regulation in health promotion generally.

Keywords

Smartphone, app, apple, android, health, behaviour change, health promotion

Introduction

The growth of smartphone ownership is expected to continue to increase in the coming years, especially for those aged over 50 years.¹ Increasing smartphone access has allowed for increasing development and use of smartphone applications, commonly referred to as 'apps'. Apps are software that can be installed onto a mobile device, such as a smartphone or tablet, often for no or little cost.² Apps can be downloaded from a variety of digital marketplaces depending on the operating system of the device. The major digital marketplaces for apps are the Apple App Store (iTunes) and the Android store (Google Play), with a smaller number of apps available for other platforms including BlackBerry World and the Windows Phone Store. The number of apps created and downloaded has continued to grow since their introduction in 2008; by the middle of 2015, 100 billion apps had been downloaded from iTunes,³

while Google Play had exceeded 25 billion app downloads.⁴

In the past, mobile health interventions have typically relied on voice or text-based short message service (SMS),^{5–7} however, the increasing availability and ease of use of apps has allowed for a growth in smartphone apps that can be used for health behaviour change.^{8–10} Any person searching the iTunes or Google Play stores for a health app will see thousands of apps related to health and fitness. However, despite the growth in apps designed to assist behaviour change and to promote good

¹School of Health and Social Development, Deakin University, Australia

²VicHealth, Australia

³Dialogue Consulting, Australia

Corresponding author:

Fiona H McKay, Deakin University, Burwood Hwy, Burwood, Victoria 3125 Australia.

Email: fiona.mckay@deakin.edu.au

health, there has been little research investigating the accuracy of the health or medical advice provided, the theoretical foundations that underpin the apps, or the effectiveness of these apps in changing behaviour or promoting health.

Part of the problem in evaluating the effectiveness or accuracy of information in apps is related to limitations in the methods used and inconsistencies in the approach to research in this area.^{11,12} There have been several previous systematic reviews that have investigated other aspects of the current research around mobile phone apps. Donker and colleagues¹³ systematically investigated the effectiveness of mental health apps for mobile devices. Despite the small number of studies included in their review, they were able to conclude that apps have the potential to benefit those with poor mental health, but that more studies need to be conducted to really understand their usefulness. Payne and colleagues¹⁴ undertook a systematic review that investigated the literature around app use in behaviour change, the behaviour change features, and the usefulness of apps in health behaviour change. While they found that apps were an acceptable tool in behaviour change interventions and that consumers were comfortable taking up apps, like other systematic reviews, this study called for more research into the potential of apps to change behaviour, citing a lack of large studies. In an investigation of mobile health technology (not limited to apps) Free and colleagues¹⁵ found that text messaging services have a use in assisting patients to adhere to medication for some health conditions, but overall the findings were mixed with more research, particularly large studies, called for. The evidence around the use and usefulness of the main mobile health (mhealth) predecessor to apps, SMS health interventions, has also been systematically reviewed by Déglise and colleagues⁷ who found that in general, mobile phones and SMS (either personalised or as a bulk message) can be an appropriate way to communicate with the population about health issues.

While all of these systematic reviews provide a picture of the current evidence around the use of apps for health behaviour change or disease management, none focus on the methods used to conduct the evaluations. This current research seeks to fill this gap, first by extending the work of BinDhim and colleagues,¹⁶ through an updated search and a broader scope for inclusion, and secondly by focusing more on the actual methods for evaluation (rather than the classification undertaken by BinDhim and colleagues).¹⁶ The aim of this review is to investigate the methods used in the evaluation of health apps and to identify any current best practice approaches.

Method and approach

Six databases were systematically searched to identify relevant publications: Academic Search Complete, CINAHL Complete, E-Journal, Inspec, MEDLINE Complete and PsycINFO. Included were studies that evaluated or reviewed apps that were published between January

2008, the year the Apple App Store was launched, and April 2015. Data were collected and analysed in September 2016. The search terms consisted of three constructs. The first was related to the type of device, with the search limited to studies focusing on mobile phones, smartphones, cell phones and tablets. The second construct was related to content of the review; search terms included health, wellbeing, preventative health, smok*, nutrition, alcohol, physical activity or mental wellbeing. The third construct related to terms that would indicate an evaluation or review, and included the search terms: analys*, assess*, evaluation, review, study and behaviour. A combination of the three constructs was used to conduct the search.

The inclusion criteria comprised of studies that evaluated mobile health apps in English, evaluations or reviews of apps targeted at consumers, evaluations or reviews of apps targeted at both consumers and health professionals, and studies that evaluated the effectiveness of mobile health apps. Excluded studies included those that evaluated mobile health apps targeted only at health professionals, formative evaluations of mobile health apps, studies of apps that were not publicly or commercially available, systematic reviews of app evaluation studies (including those of BinDhim et al. and Zapata et al.),^{16,17} studies that reported primarily on the validation of any mobile health app tool,^{18–20} and studies of apps not related to health. Articles were first screened by title and abstract based on the inclusion and exclusion criteria. The full texts of selected articles were then obtained for further assessment for final inclusion.

To assess quality and bias of the included studies, a 15-item checklist (shown in Table 1) was developed based on the research questions, general systematic review appraisal tools and recommendations from BinDhim et al.¹⁶ and Zapata et al.¹⁷ The scale was binary, resulting in a maximum quality score of 15. Scores of 0–5 are considered to indicate low quality, 6–10 medium quality, and 11–15 high quality. Each article was independently reviewed by two authors.

Results

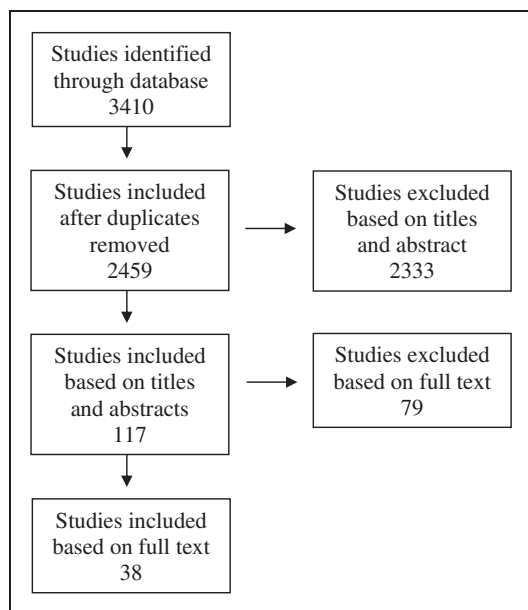
General characteristics

The database search using the key search terms resulted in a total of 2835 journal articles. After reviewing the titles and abstracts, 379 were found to be duplicates, the titles of the remaining 2459 were reviewed and 117 were found to have met the initial selection criteria (see Figure 1). After examination of the full text, 38 articles fully met the inclusion criteria and were included in the final review.

The 38 included studies were published between 2011 and 2016. Most of the studies were from English-speaking countries ($n = 22$). Of the 38 studies, 36 were reviews or evaluations of apps related to one or more health condition, the remaining two^{21,22} reported on an investigation of the usability of health apps. Studies investigated apps

Table 1. Quality and bias assessment checklist.

No.	Questions
1.	Does the study address any research question(s) or objective(s)?
2.	Does the study provide any theoretical framework for the evaluation method?
3.	Does the theoretical framework of the study include any health promotion theory?
4.	Does the study provide a timeframe of the data collection?
5.	Does the study identify the country where the search was conducted?
6.	Does the study mention that the reviewed apps were downloaded for evaluation?
7.	Does the study discuss the selection criteria for apps to be included or excluded for review?
8.	Does the study provide a clear description of the evaluation method?
9.	Are there at least two independent data extractors with a consensus procedure in place in case of disagreement?
10.	Is a list of the reviews apps provided?
11.	Does the study discuss the findings of the evaluation?
12.	Does the study look at the reviewed apps' ability to promote or enable behavioural change?
13.	Does the study discuss any limitations?
14.	Does the study provide any future recommendations in general?
15.	Does the study state any conflict of interest?

**Figure 1.** Flow diagram of selection process.

relating to the following: alcohol ($n=2$), asthma ($n=1$), breastfeeding ($n=1$), cancer ($n=4$), depression ($n=1$), diabetes ($n=3$), general health and fitness ($n=2$), headaches ($n=1$), heart disease ($n=1$), HIV ($n=1$),

hypertension ($n=1$), iron deficiency/anaemia ($n=1$), low vision ($n=1$), medication adherence ($n=1$), mindfulness ($n=1$), obesity ($n=1$), pain ($n=1$), physical activity ($n=5$), smoking ($n=5$), weight management ($n=4$) and women's health ($n=1$) (see Table 2). The median number of apps reviewed in each publication was 153 (range 3–710). The authors downloaded the apps for evaluation in only 17 of the 36 studies; the remainder reviewed the apps based solely on app store descriptions. The average quality and bias assessment score of the 36 studies that specifically reviewed apps was 10 (range 6–14), indicating medium quality. Three of the studies did not include a timeline for data collection, while only 13 included the country where the search was conducted, and only 16 studies mentioned if the apps had been downloaded for review. Most studies ($n=33$) described the method for app evaluation, however, less than half ($n=16$) provided a list of apps that were included in the review.

Evaluation methods

In their review or evaluation of apps, studies employed at least one of four main types of appraisal. Appraisals included content analysis, theory or evidence-based appraisal, usability appraisal, and/or an appraisal of effectiveness. One study,²³ looked only at the number of apps available for each health condition and was excluded from this analysis as no app evaluation was completed. Of the remaining 36 studies, 20 involved only one approach to evaluation, 16 included two approaches to evaluation, and only one incorporated three approaches to of evaluation (see Table 2).

Twenty-two of the studies conducted a content analysis.^{23–46} In these studies, authors undertook a review of key features, functions, or quality of the app, or apps were simply categorised in to the condition they addressed. The second type of appraisal related to the investigation of features that were based on theory or evidence within the app; this type of appraisal was identified in 17 studies.^{11,27,29,31,35,36,38,40,43,44,47–53} These studies investigated strategies that were promoted by the app, identified behavioural change theories incorporated into the app, or investigated the app for the inclusion of evidence, clinical outcome or health professional involvement in developing the apps. Thirdly, 10 studies investigated the usability of apps.^{11,21,28,35,37,39,41,54–56} These studies reported on app design, interface, ease of use, user engagement or user experience. The final type of app appraisal was an investigation of effectiveness, identified in four studies,^{21,22,30,43} and was used to determine if the apps can lead to behavioural change or if there was any evidence that the apps would lead to behavioural change (see Table 2).

Four types of methods were used in the evaluations; self-developed checklists, established checklists, user feedback and matched-case control design. Self-developed checklists were the most frequently used method ($n=28$). Checklists were individually developed by authors based

Table 2. Types and methods of evaluation.

Study reference	Location	Health condition	Quality score	Type	Method
Systematic reviews					
Abroms et al., 2013 ²⁷	USA	Smoking cessation	9	Content analysis Theory/evidence-based evaluation	Self-developed evaluation checklist
Arnhold et al., 2014 ²⁸	Germany	Diabetes	9	Content analysis Usability	Self-developed evaluation checklist
Azar et al., 2013 ²⁹	USA	Weight management	13	Content analysis Theory/evidence-based evaluation	Self-developed evaluation checklist
Bardus et al., 2016 ⁴⁵	USA	Weight management	10	Content analysis Quality	Established evaluation checklist
Bender et al., 2013 ³⁰	Canada	Cancer	13	Content analysis Effectiveness	Self-developed evaluation checklist
Choi et al., 2014 ³¹	South Korea	Smoking cessation	14	Content analysis Theory/evidence-based evaluation	Self-developed evaluation checklist
Chomutare et al., 2011 ³²	Norway	Diabetes	9	Content analysis	Self-developed evaluation checklist User feedback
Cohn et al., 2011 ⁴³	USA	Alcohol	10	Theory/evidence-based evaluation Effectiveness	Self-developed evaluation checklist User feedback
Conroy et al., 2014 ⁴⁴	USA	Physical activity	9	Theory/evidence-based evaluation	Established evaluation checklist
Derbyshire and Dancy, 2013 ⁵⁴	UK	Women's health	6	Usability	User feedback
Direito et al., 2014 ⁴⁷	New Zealand	Physical activity and dietary	13	Theory/evidence-based evaluation	Established evaluation checklist
Eng and Lee, 2013 ³³	USA	Diabetes and endocrinology	7	Content analysis	Self-developed evaluation checklist
García-Gómez et al., 2014 ⁵⁵	Spain	Type 2 diabetes, obesity and breast-feeding	6	Usability	User feedback
Hoepfner et al., 2016 ²⁶	USA	Smoking cessation	9	Content analysis	Self-developed evaluation checklist
Huckvale et al., 2012 ³⁴	UK	Asthma	10	Content analysis	Self-developed evaluation checklist Established evaluation checklist
Hundert et al., 2014 ³⁵	Canada	Headache	12	Content analysis Theory/evidence-based evaluation Usability	Self-developed evaluation checklist Self-developed evaluation checklist Established evaluation checklist
Kassianos et al., 2015 ³⁶	UK	Melanoma	9	Content analysis Theory/evidence-based evaluation	Self-developed evaluation checklist
Kumar et al., 2015 ³⁷	USA	Hypertension	9	Content analysis Usability	Self-developed evaluation checklist User feedback
Martínez-Pérez et al., 2013 ²³	Spain/UK	Iron-deficiency anaemia, hearing loss, migraine, low vision, asthma, diabetes mellitus, osteoarthritis and unipolar depressive disorders	7	Quantity*	Not discussed*
Martínez-Pérez et al., 2013 ⁴⁶	Spain	Heart disease	7	Content analysis	Self-developed evaluation checklist

(continued)

Table 2. Continued.

Study reference	Location	Health condition	Quality score	Type	Method
Morrissey et al., 2016 ⁵³	Ireland	Medication adherence	10	Theory/evidence-based evaluation	Established evaluation checklist
Middelweerd et al., 2014 ⁴⁸	The Netherlands	Health and fitness	13	Theory/evidence-based evaluation	Established evaluation checklist
Mobasheri et al., 2014 ³⁸	UK	Breast cancer	8	Content analysis	Self-developed evaluation checklist
Pagoto et al., 2013 ⁴⁹	USA	Weight-loss	12	Theory/evidence-based evaluation	User feedback
Pandey et al., 2013 ³⁹	USA	Cancer	7	Theory/evidence-based evaluation	Self-developed evaluation checklist
Patel et al., 2015 ²⁵	New Zealand	Weight loss and smoking cessation	6	Content analysis	User feedback
Plaza et al., 2013 ⁴⁰	Spain	Mindfulness	10	Quality	Established evaluation checklist
Reynoldson et al., 2014 ⁴¹	UK	Pain	10	Content analysis	Self-developed evaluation checklist
Robustillo Cortés et al., 2014 ⁶¹	Spain	HIV	11	Theory/evidence-based evaluation	User feedback
Sama et al., 2014 ¹¹	USA	Health and wellness	11	Content analysis	Self-developed evaluation checklist
Shen et al., 2015 ⁴²	Canada	Depression	9	Usability	Self-developed evaluation checklist
Ubhi et al., 2016 ²⁴	UK	Smoking cessation	13	Content analysis	Self-developed evaluation checklist
Vollmer Dahlke et al., 2015 ⁵⁰	USA	Cancer	12	Content analysis	Established evaluation checklist
Wearing et al., 2014 ⁵¹	USA	Paediatric obesity	11	Theory/evidence-based evaluation	Established evaluation checklist
Weaver et al., 2013 ⁵⁶	Australia	Alcohol	10	Content analysis	Self-developed evaluation checklist
Yang et al., 2015 ⁵²	USA	Physical activity	10	Usability	User feedback
Effectiveness studies					
Casey et al., 2014 ²¹	Ireland	Physical activity	High quality	Theory/evidence-based evaluation	Self-developed evaluation checklist
Kirwan et al., 2012 ²²	Australia	Physical activity	Medium quality	Usability	User feedback
			Effectiveness	Effectiveness	Matched case-control trial

on theory, established models, literature, practice guidelines or actual user evaluation that could be applied to apps. Self-developed checklists were used in 21 reviews categorised as content analysis, and 13 of the 17 reviews that used a theory or evidence-based approach (see Table 2). Bender and colleagues³⁰ and Cohn and colleagues⁴³ also used a self-developed checklist to investigate evidence indicating effectiveness.

Nine studies used an established checklist in their evaluation. Conroy et al.⁴⁴ used the Coventry, Aberdeen and London-Revised taxonomy (CALO-RE),⁵⁷ Direito et al.,⁴⁷ Middleweerd et al.,⁴⁸ and Vollmer Dahlke and colleagues⁵⁰ used the earlier Abraham and Michie⁵⁸ taxonomy of health behaviour change, while Morrissey and colleagues⁵³ used a longer taxonomy, also designed by Michie and colleagues⁵⁹ to investigate any behaviour change features. Each of these studies reported low scores on these scales, indicating that behaviour change techniques are infrequently used in health apps. Patel et al.,²⁵ and Bardus et al.⁴⁵ used the more recent Mobile App Rating Scale (MARS) designed by Stoyanov and colleagues to assess quality.²⁰ Wearing and colleagues⁵¹ employed the Flesch-Kincaid Grade Level readability formula for content analysis of paediatric obesity apps. Finally, the Nielson and Molich⁶⁰ usability heuristics were used by Hundert and colleagues³⁵ to evaluate usability of headache management apps.

User feedback was also used to evaluate apps. This method typically involved examining consumer comments and ratings downloaded from app stores or obtaining feedback from actual users through focus groups, questionnaires, or interviews. This method was mainly used for usability evaluation ($n = 12$) but was also used by Casey and colleagues²¹ to examine effectiveness. Only one study²² used a matched-case control trial method to evaluate the effectiveness of an app (see Table 2).

Recommendations

Several studies identified areas for improvement within the apps. Weaver and colleagues,⁵⁶ reviewed 48 apps that were designed to assist users to track their blood alcohol concentration (BAC). These apps were reviewed by collecting data from a field-based study measuring participants' gender, age, number of drinks consumed, hours spent drinking and their BAC level. This study found that most of the apps produced varied and unreliable results, and in a subsequent focus group discussion, participants stated that they were sceptical of the accuracy of the apps, particularly as most of the apps were difficult to use. Kumar and colleagues,³⁷ investigated hypertension apps in Google Play, and found that while 14% could transform the smartphone into a medical device to measure blood pressure or heart rate, none were approved by the US Food and Drug Administration (FDA) or validated, casting doubts on the usability and reliability of the apps. Five other studies were unable to identify any evidence, validation or health professional involvement in the

development of their reviewed apps,^{30,35,40–42} while three studies identified limited discussion of evidence-base or professional involvement in their reviews.^{36–38}

Discussion

This literature review examined 38 studies to identify a best practice approach to the review or evaluation of apps that may promote health or encourage behaviour change. The studies employed a variety of methods, and, overall, the quality of the studies was found to be medium to high. Studies employed at least one of four types of evaluation, each covering a different aspect of the mobile health apps. Half of the studies focused on one type of evaluation only; only one attempted to assess three aspects of the reviewed apps. Content analysis was the most frequently identified in studies, while evaluation of effectiveness was the least used.

One of the key objectives of this review was to identify any best practice approaches to reviewing apps that promote health or encourage behaviour change. However, as self-developed evaluation checklists were identified as the most frequently used method for evaluating health related apps, it is difficult to point to a best practice approach. None of the self-developed evaluation checklists were validated, and despite some studies indicating moderation between reviewers,^{26,27,30–32,49,51–53,61} inter-rater reliability was low in studies where it was tested,^{27,31,48,52,53} casting doubts on the reliability of the tool or the approach and therefore the findings of the study. The fact that self-developed evaluation checklists were widely used is further evidence of the current lack of validated evaluation tools to assess mobile health apps.

Of the four studies that used a pre-existing behaviour change instrument to evaluate the apps,^{44,47,48,50} three^{47,48,50} used the Taxonomy of Behaviour Change developed by Abraham and Michie,⁵⁸ while Morrissey and colleagues⁵³ used a longer taxonomy, also designed by Michie and colleagues⁵⁹ to investigate any behaviour change. This instrument has been found to be reliable in identifying behaviour change techniques in interventions. The other study⁴⁴ used the CALO-RE instrument⁵⁷ an extension of the Abraham and Michie⁵⁸ taxonomy. As it currently stands, these taxonomies are the most reliable and most commonly used instruments for assessing intervention content around behaviour change.

Two studies used the new MARS.²⁰ This scale has been designed to be a reliable, multidimensional measure for trialling, classifying, and rating the quality of mobile health apps. This review found two studies, Patel et al.²⁵ and Bardus et al.,⁴⁵ to have used this tool in their assessment of apps.

One of the challenges identified by this review was the appropriateness of the models used in the studies reviewed to evaluate the apps. Only four studies incorporated questions evaluating the specific mobile technology features in their checklist, suggesting that the other studies may not have taken the mobile environment into consideration.

This is important as current theories and checklists may be inadequate in informing mobile interventions, a consideration made by Riley and colleagues,⁶³ who have advocated for the development of new theories to capture the unique capabilities of mobile interventions. Furthermore, none of the studies examined whether the behaviour change or health promoting features incorporated specifically into the apps were appropriate for the mobile environment, or if they could be effective when used in an app. If the features are found to be ineffective, it is unlikely that the app will facilitate successful behaviour change, highlighting this an important area for future research.

Another area of concern for those interested in the behaviour change opportunities present in apps is the lack of information in any of these studies about readability, privacy or security. Previous research has highlighted the importance of privacy and security when seeking to promote behaviour change through mobile apps.^{64,65} There is some evidence suggesting that consumers believe that their right to privacy is lost when engaging in the mobile environment, yet Internet users are also often concerned about how their personal information is used.⁶⁶ The study by Chomutare and colleagues³² was the only study to acknowledge this gap in their evaluation, highlighting that their study of diabetes apps did not look at privacy or security at all. There is also an emerging body of evidence that points to readability as an important aspect of health and behaviour change technology in the online space,⁶⁷ with studies suggesting that readability of online health information is above the average reading ability of consumers.^{68,69} Concerns have been raised that information presented in these forms may be misunderstood and possibly cause harm.⁷⁰ Of the 33 studies included in this review, only Wearing and colleagues⁵¹ evaluated the readability of the app content, however, they indicated that this was included as an afterthought and not all their reviewed apps had undergone the readability test.

There are some clear limitations identified in the studies reviewed. Only 17 of the studies actually downloaded the apps for review. Two studies indicated that their self-developed evaluation checklists were not validated, and reliability could not be ascertained,^{41,47} or criteria of the questionnaire could be affected by the subjectivity of the evaluator.⁶² Abrams and colleagues²⁷ acknowledged that their guidelines were developed for a clinical setting and only assumed to be effective in a mobile app context and further research would be needed to confirm such assumptions. Middelweerd and colleagues⁴⁸ also pointed out that using Abraham and Michie's taxonomy forced evaluators to translate strategies, originally designed for other behaviour change interventions, into app features and might lead to different interpretations among researchers, resulting in the low inter-rater reliability in their study. Five studies pointed out that the behavioural change techniques or strategies found should be interpreted with caution, and that in their study they could not determine the effectiveness of these features.^{11,29,31,47,50}

Limitations

There are some limitations of this review that should also be acknowledged. While every attempt was made to ensure this literature review was comprehensive taking in all the available literature, additional articles may have been missed. However, given that the other two systematic reviews of evaluation method found during the search had included 10 or 22 studies,^{16,17} and this review found 33, the authors are confident that there is little information that is not presented here. Given the variety of different apps under study, there may also be some difficulty in making direct comparisons across the different areas of health behaviour change. However, as this is not a meta-analysis, the authors do not feel this should invalidate the findings. Three studies about specific evaluation tools were also identified, but are not presented here as they did not fit in the inclusion criteria. Two focused on usability,^{18,19} the other study was a validation of the 23-item tool, the MARS²⁰ which is focussed on engagement, functionality, aesthetics and information quality.

Conclusion

In order to harness the potential of mobile health apps for behaviour change and health, we need better ways to assess the quality and effectiveness of apps. This review is unable to suggest a single best practice approach to evaluate mobile health apps. Few measures identified in this review included sufficient information or evaluation, leading to potentially incomplete and inaccurate information for consumers seeking the best app for their situation. This is further complicated by a lack of regulation in health promotion generally.

For those seeking to complete a review of behaviour change and health promoting apps, we suggest the inclusion of three components: (a) a review of usability and functionality, (b) some critique of the apps potential to promote behaviour change, and (c) the quality of the health-related content within the apps. We were unable to find a single study or evaluation tool incorporating these three components. Such a tool would assist consumers in identifying high quality and effective health apps.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: They wish to acknowledge funding from the Victorian Health Promotion Foundation that was used to complete this work.

References

1. Deloitte. *Media consumer survey 2014*, http://landing.deloitte.com.au/rs/deloitteaus/images/Deloitte_Media_Consumer_Survey_2014.pdf?mkt_tok=3RkMMJWWf9wsRo nuK%2FPd%2B%2FhmjTEU5z16uUpWaGzi4kz2EFye%2BLIHETpodcMTcVnN77YDBceEJhgyQJxPr3CKtEN09 dxRhLgAA%3D%3D (2014, accessed 20 May 2015).
2. Australian Communication and Media Authority. Communications report 2012-3, <http://www.acma.gov.au/theACMA/Library/Corporate-library/Corporate-publications/communications-report-2012-13> (2013, accessed 29 April 2015).
3. cNet. Apple: More than 100 billion app downloads from iTunes Store, [http://www.cnet.com/news/wwdc-by-the-numbers/\(2015, accessed 28 February 2016\)](http://www.cnet.com/news/wwdc-by-the-numbers/(2015, accessed 28 February 2016)).
4. Sims G. Google Play Store vs the Apple App Store: By the numbers, <http://www.androidauthority.com/google-play-store-vs-the-apple-app-store-601836/> (2015, accessed 20 May 2015).
5. Fjeldsoe BS, Marshall AL and Miller YD. Behavior change interventions delivered by mobile telephone short-message service. *Am J Prev Med* 2009; 36(2): 165–173.
6. Chen Z-W, Fang L-Z, Chen L-Y, et al. Comparison of an SMS text messaging and phone reminder to improve attendance at a health promotion center: A randomized controlled trial. *J Zhejiang Univ Sci B* 2008; 9(1): 34–38.
7. Déglise C, Suggs LS and Odermatt P. SMS for disease control in developing countries: A systematic review of mobile health applications. *J Telemed Telecare* 2012; 18(5): 273–281.
8. Lim MS, Hocking JS, Hellard ME, et al. SMS STI: A review of the uses of mobile phone text messaging in sexual health. *Int J STD AIDS* 2008; 19(5): 287–290.
9. Luxton DD, McCann RA, Bush NE, et al. mHealth for mental health: Integrating smartphone technology in behavioral healthcare. *Prof Psychol Res Pr* 2011; 42(6): 505.
10. Rosser BA and Eccleston C. Smartphone applications for pain management. *J Telemed Telecare* 2011; 17(6): 308–312.
11. Sama PR, Eapen ZJ, Weinfurt KP, et al. An evaluation of mobile health application tools. *JMIR Mhealth Uhealth* 2014; 2(2): e19–e19.
12. Medical Technology Association of Australia. *App purchases by Australian consumers on mobile and handheld devices*, <http://ccaac.gov.au/files/2013/02/MedicalTechnologyAssociationofAustralia.pdf> (2013, accessed 29 April 2015).
13. Donker T, Petrie K, Proudfoot J, et al. Smartphones for smarter delivery of mental health programs: A systematic review. *J Med Internet Res* 2013; 15(11): e247.
14. Payne HE, Lister C, West JH, et al. Behavioral functionality of mobile apps in health interventions: A systematic review of the literature. *JMIR Mhealth Uhealth* 2015; 3(1): e20.
15. Free C, Phillips G, Galli L, et al. The effectiveness of mobile-health technology-based health behaviour change or disease management interventions for health care consumers: A systematic review. *PLoS Med* 2013; 10(1): e1001362.
16. BinDhim NF, Hawkey A and Trevena L. A systematic review of quality assessment methods for smartphone health apps. *Telemed J E Health* 2014.
17. Zapata B, Fernández-Alemán J, Idri A, et al. Empirical studies on usability of mhealth apps: A systematic literature review. *J Med Syst* 2015; 39(2): 1–19.
18. Brown W III, Po-Yin Y, Rojas M, et al. Assessment of the Health IT Usability Evaluation Model (Health-ITUEM) for evaluating mobile health (mHealth) technology. *J Biomed Inform* 2013; 46(6): 1080–1087.
19. Monkman H and Kushniruk A. A health literacy and usability heuristic evaluation of a mobile consumer health application. *Stud Health Technol Inform* 2013; 192: 724–728.
20. 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(1): e27–e27.
21. Casey M, Hayes PS, Glynn F, et al. Patients' experiences of using a smartphone application to increase physical activity: The SMART MOVE qualitative study in primary care. *Br J Gen Pract* 2014; 64(625): e500–e508.
22. Kirwan M, Duncan MJ, Vandelanotte C, et al. Using smartphone technology to monitor physical activity in the 10,000 steps program: A matched case-control trial. *J Med Internet Res* 2012; 14(2): ■■.
23. Martínez-Pérez B, de la Torre-Díez I and López-Coronado M. Mobile health applications for the most prevalent conditions by the World Health Organization: Review and analysis. *J Med Internet Res* 2013; 15(6): e120–e120.
24. Ubhi HK, Kotz D, Michie S, et al. Comparative analysis of smoking cessation smartphone applications available in 2012 versus 2014. *Addict Behav* 2016; 58: 175–181.
25. Patel R, Sulzberger L, Li G, et al. Smartphone apps for weight loss and smoking cessation: Quality ranking of 120 apps. *N Z Med J* 2015; 128: 73–76.
26. Hoepfner BB, Hoepfner SS, Seaboyer L, et al. How smart are smartphone apps for smoking cessation? a content analysis. *Nicotine Tob Res* 2016; 18(5): 1025–1031.
27. Abroms LC, Westmaas JL, Bontemps-Jones J, et al. A content analysis of popular smartphone apps for smoking cessation. *Am J Prev Med* 2013; 45(6): 732–736.
28. Arnhold M, Quade M and 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(4): e104–e104.
29. Azar KM, Lesser LI, Laing BY, et al. Mobile applications for weight management: Theory-based content analysis. *Am J Prev Med* 2013; 45(5): 583–589.
30. Bender JL, Yue RYK, To MJ, et al. A lot of action, but not in the right direction: Systematic review and content analysis of smartphone applications for the prevention, detection, and management of cancer. *J Med Internet Res* 2013; 15(12): 49–59.
31. Choi J, Noh G-Y and Park D-J. Smoking cessation apps for smartphones: Content analysis with the self-determination theory. *J Med Internet Res* 2014; 16(2): 143–156.
32. Chomutare T, Fernandez-Luque L, Arsand E, et al. 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(3): e65–e65.
33. Eng DS and Lee JM. The promise and peril of mobile health applications for diabetes and endocrinology. *Pediatr Diabetes* 2013; 14(4): 231–238.
34. Huckvale K, Car M, Morrison C, et al. Apps for asthma self-management: A systematic assessment of content and tools. *BMC Med* 2012; 10: 144.

35. Hundert AS, Huguet A, McGrath PJ, et al. Commercially available mobile phone headache diary apps: A systematic review. *JMIR Mhealth Uhealth* 2014; 2(3): e36.
36. Kassianos AP, Emery JD, Murchie P, et al. Smartphone applications for melanoma detection by community, patient and generalist clinician users: A review. *Br J Dermatol* 2015; 172(6): 1507–1518.
37. Kumar N, Khunger M, Gupta A, et al. A content analysis of smartphone-based applications for hypertension management. *J Am Soc Hypertens* 2015; 9(2): 130–136.
38. Mobasher MH, Johnston M, King D, et al. Smartphone breast applications – What’s the evidence? *Breast J* 2014; 23(5): 683–689.
39. Pandey A, Hasan S, Dubey D, et al. Smartphone apps as a source of cancer information: Changing trends in health information-seeking behavior. *J Cancer Educ* 2013; 28(1): 138–142.
40. Plaza I, Demarzo MMP, Herrera-Mercadal P, et al. Mindfulness-based mobile applications: Literature review and analysis of current features. *JMIR Mhealth Uhealth* 2013; 1(2): e24.
41. Reynoldson C, Stones C, Allsop M, et al. Assessing the quality and usability of smartphone apps for pain self-management. *Pain Med* 2014; 15(6): 898–909.
42. Shen N, Levitan M-J, Johnson A, et al. Finding a depression app: A review and content analysis of the depression app marketplace. *JMIR Mhealth Uhealth* 2015; 3(1): e16.
43. Cohn AM, Hunter-Reel D, Hagman BT, et al. Promoting behavior change from alcohol use through mobile technology: The future of ecological momentary assessment. *Alcohol Clin Exp Res* 2011; 35(12): 2209–2215.
44. Conroy DE, Yang C-H and Maher JP. Behavior change techniques in top-ranked mobile apps for physical activity. *Am J Prev Med* 2014; 46(6): 649–652.
45. Bardus M, van Beurden SB, Smith JR, et al. A review and content analysis of engagement, functionality, aesthetics, information quality, and change techniques in the most popular commercial apps for weight management. *Int J Behav Nutr Phys Act* 2016; 13: 1–9.
46. Martínez-Pérez B, de la Torre-Díez I, López-Coronado M, et al. Mobile apps in cardiology: review. *JMIR Mhealth And Uhealth* 2013; 1(2): e15–e15.
47. Direito A, Dale LP, Shields E, et al. Do physical activity and dietary smartphone applications incorporate evidence-based behaviour change techniques? *BMC Public Health* 2014; 14(1): 1656–1672.
48. Middelweerd A, Mollee J, van der Wal C, et al. Apps to promote physical activity among adults: A review and content analysis. *Int J Behav Nutr Phys Act* 2014; 11(1): 1–15.
49. Pagoto S, Schneider K, Jovic M, et al. Evidence-based strategies in weight-loss mobile apps. *Am J Prev Med* 2013; 45(5): 576–582.
50. Vollmer Dahlke D, Fair K, et al. Apps seeking theories: Results of a study on the use of health behavior change theories in cancer survivorship mobile apps. *JMIR Mhealth Uhealth* 2015; 3(1): e31.
51. Wearing JR, Nollen N, Befort C, et al. iPhone app adherence to expert-recommended guidelines for pediatric obesity prevention. *Child Obes* 2014; 10(2): 132–144.
52. Yang C-H, Maher JP and Conroy DE. Implementation of behavior change techniques in mobile applications for physical activity. *Am J Prev Med* 2015; 48(4): 452–455.
53. Morrissey EC, Corbett TK, Walsh JC, et al. Behavior change techniques in apps for medication adherence: A content analysis. *Am J Prev Med* 2016; 50(5): e143–e146.
54. Derbyshire E and Dancey D. Smartphone medical applications for women’s health: What is the evidence-base and feedback? *Int J Telemed Appl* 2013; 1–10.
55. García-Gómez JM, de la Torre-Díez I, Vicente J, et al. Analysis of mobile health applications for a broad spectrum of consumers: A user experience approach. *Health Informatics J* 2014; 20(1): 74–84.
56. Weaver ER, Horyniak DR, Jenkinson R, et al. ‘Let’s get Wasted!’ and other apps: Characteristics, acceptability, and use of alcohol-related smartphone applications. *JMIR Mhealth Uhealth* 2013; 1(1): e9.
57. Michie S, Ashford S, Snihotta FF, et al. A refined taxonomy of behaviour change techniques to help people change their physical activity and healthy eating behaviours: The CALO-RE taxonomy. *Psychol Health* 2011; 26(11): 1479–1498.
58. Abraham C and Michie S. A taxonomy of behavior change techniques used in interventions. *Health Psychol* 2008; 27(3): 379–387.
59. Michie S, Richardson M, Johnston M, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: Building an international consensus for the reporting of behavior change interventions. *Ann Behav Med* 2013; 46(1): 81–95.
60. Nielsen J and Molich R. Heuristic evaluation of user interfaces. In: Chew JC and Whiteside J (eds) *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, Seattle, WA, USA, 01–05 April 1990, pp.249–256. New York: ACM.
61. Abroms LC, Lee Westmaas J, Bontemps-Jones J, et al. A content analysis of popular smartphone apps for smoking cessation. *Am J Prev Med* 2013; 45(6): 732–736.
62. Robustillo Cortés MdlA, Cantudo Cuenca MR, et al. High quantity but limited quality in healthcare applications intended for HIV-infected patients. *Telemed J E Health* 2014; 20(8): 729–735.
63. Riley W, Rivera D, Atienza A, et al. Health behavior models in the age of mobile interventions: Are our theories up to the task? *Transl Behav Med* 2011; 1(1): 53–71.
64. Sunyaev A, Dehling T, Taylor PL, et al. Availability and quality of mobile health app privacy policies. *J Am Med Inform Assoc* 2015; 22(e1): e28–e33.
65. Mylonas A, Kastania A and Gritzalis D. Delegate the smartphone user? Security awareness in smartphone platforms. *Computers & Security* 2013; 34: 47–66.
66. Singh RI, Sumeeth M and Miller J. Evaluating the readability of privacy policies in mobile environments. *International Journal of Mobile Human Computer Interaction* 2011; 3(1): 55–74.
67. Cheng C and Dunn M. Health literacy and the Internet: A study on the readability of Australian online health information. *Aust N Z J Public Health* 2015; 39(4): 309–314.
68. Keleher H and Hagger V. Health literacy in primary health care. *Aust J Prim Health* 2007; 13(2): 24–30.
69. McCray AT. Promoting health literacy. *J Am Med Inform Assoc* 2005; 12(2): 152–163.
70. Walsh TM and Volsko TA. Readability assessment of Internet-based consumer health information. *Respir Care* 2008; 53(10): 1310–1315.