CSIRO PUBLISHING
Australian Health Review, 2020, 44, 62–82
https://doi.org/10.1071/AH18064

Going digital: a narrative overview of the effects, quality and utility of mobile apps in chronic disease self-management

Ian A. Scott[1,2,8] MBBS, FRACP, MHA, MEd, Director of Internal Medicine and Clinical Epidemiology, Associate Professor of Medicine

Paul Scuffham[3] PhD, FAHMS, Director, Centre for Applied Health Economics, Professor of Health Economics

Deepali Gupta[1,4] MPharm, Senior Smoking Cessation Pharmacist, Cardiac (HF) Pharmacist

Tanya M. Harch[5] BBus(HIM), GradCertBus(PSM), Director Strategic Partnerships

John Borchi[5] BA, BSc, PhD Chief Information Security Officer

Brent Richards[6,7] MBChB, FRACP, FCICM, Medical Director of Innovation, Director of Critical Care Research, Associate Professor of Medicine

[1]Princess Alexandra Hospital, 199 Ipswich Road, Woolloongabba, Brisbane 4102, Australia. Email: deepali.gupta@health.qld.gov.au
[2]School of Clinical Medicine, University of Queensland, 37 Trent Street, Woolloongabba, Brisbane 4102, Australia.
[3]Menzies Health Institute Queensland, Griffith University (Nathan campus), 170 Kessels Road, Nathan, Brisbane 4111, Australia. Email: p.scuffham@griffith.edu.au
[4]Queen Elizabeth II Hospital, Troughton Rd and Kessels Rd, Coopers Plains, Brisbane, 4108, Australia.
[5]eHealth Queensland, 2/315 Brunswick St, Fortitude Valley, Brisbane 4006, Australia. Email: tanya.harch@health.qld.gov.au; john.borchi@health.qld.gov.au
[6]Gold Coast University Hospital, 1 Hospital Boulevard, Southport 4215, Australia. Email: brent.richards@health.qld.gov.au
[7]Griffith University (Gold Coast campus), Parklands Drive, Southport 4215, Australia.
[8]Corresponding author. Email: ian.scott@health.qld.gov.au

Abstract

Objective. Smartphone health applications (apps) are being increasingly used to assist patients in chronic disease self-management. The effects of such apps on patient outcomes are uncertain, as are design features that maximise usability and efficacy, and the best methods for evaluating app quality and utility.

Methods. In assessing efficacy, PubMed, Cochrane Library and EMBASE were searched for systematic reviews (and single studies if no systematic review was available) published between January 2007 and January 2018 using search terms (and synonyms) of 'smartphone' and 'mobile applications', and terms for each of 11 chronic diseases: asthma, chronic obstructive lung disease (COPD), diabetes, chronic pain, serious mental health disorders, alcohol and substance addiction, heart failure, ischaemic heart disease, cancer, cognitive impairment, chronic kidney disease (CKD). With regard to design features and evaluation methods, additional reviews were sought using search terms 'design', 'quality,' 'usability', 'functionality,' 'adherence', 'evaluation' and related synonyms.

Results. Of 13 reviews and six single studies assessing efficacy, consistent evidence of benefit was seen only with apps for diabetes, as measured by decreased glycosylated haemoglobin levels (HbA1c). Some, but not all, studies showed benefit in asthma, low back pain, alcohol addiction, heart failure, ischaemic heart disease and cancer. There was no evidence of benefit in COPD, cognitive impairment or CKD. In all studies, benefits were clinically marginal and none related to morbid events or hospitalisation. Twelve design features were identified as enhancing usability. An evaluation framework comprising 32 items was formulated.

Conclusion. Evidence of clinical benefit of most available apps is very limited. Design features that enhance usability and maximise efficacy were identified. A provisional 'first-pass' evaluation framework is proposed that can help decide which apps should be endorsed by government agencies following more detailed technical assessments and which could then be recommended with confidence by clinicians to their patients.

What is known about the topic? Smartphone health apps have attracted considerable interest from patients and health managers as a means of promoting more effective self-management of chronic diseases, which leads to better health outcomes. However, most commercially available apps have never been evaluated for benefits or harms in clinical trials, and there are currently no agreed quality criteria, standards or regulations to ensure health apps are user-friendly, accurate in content, evidence based or efficacious.

What does this paper add? This paper presents a comprehensive review of evidence relating to the efficacy, usability and evaluation of apps for 11 common diseases aimed at assisting patients in self-management. Consistent evidence of benefit was only seen for diabetes apps; there was absent or conflicting evidence of benefit for apps for the remaining 10 diseases. Benefits that were detected were of marginal clinical importance, with no reporting of hard clinical end-points, such as mortality or hospitalisations. Only a minority of studies explicitly reported using behaviour change theories to underpin the app intervention. Many apps lacked design features that the literature identified as enhancing usability and potential to confer benefit. Despite a plethora of published evaluation tools, there is no universal framework that covers all relevant clinical and technical attributes. An inclusive list of evaluation criteria is proposed that may overcome this shortcoming.

What are the implications for practitioners? The number of smartphone apps will continue to grow, as will the appetite for patients and clinicians to use them in chronic disease self-management. However, the evidence to date of clinical benefit of most apps already available is very limited. Design features that enhance usability and clinical efficacy need to be considered. In making decisions about which apps should be endorsed by government agencies and recommended with confidence by clinicians to their patients, a comprehensive but workable evaluation framework needs to be used by bodies assuming the roles of setting and applying standards.

Received 6 April 2018, accepted 4 September 2018, published online 13 November 2018

Introduction

With the increasing prevalence of chronic multimorbidity, enhanced patient self-management is a priority in improving outcomes and reducing healthcare costs. With the advent of smartphones and tablets in 2007, now used by more than 80% of Australians,[1] downloadable mobile health applications (apps) have become popular as a means for enabling patients with chronic disease to participate in more effective self-management. Funders and managers of healthcare systems are increasingly interested in the potential for apps to improve chronic disease self-management and reduce hospitalisations and healthcare costs.

In 2017, the number of health apps released from iTunes and Google Play exceeded 300 000,[2] with nearly 25% dealing with disease self-management.[3] One-third of adults in the US with smartphones or tablets use health apps to achieve health behaviour goals and help with medical decision making.[4] However, although international standards exist with regard to software engineering, privacy, security and usability of mobile apps in general (e.g. International Organization for Standardization (ISO) and International Electrotechnical Commission (IEC) standards), there are currently few widely accepted criteria to ensure health apps are accurate in content, evidence based or efficacious.[5] The efficacy of most commercially available apps in improving self-management and clinical outcomes related to chronic diseases has never been evaluated in clinical trials[6] and although quality assessment tools exist for health-related websites,[7,8] no universally agreed approach exists for evaluating app clinical efficacy and safety. Involvement of users and clinical experts in development is highly variable, and adherence of apps to available clinical evidence ranges from 0% to 87%.[9] Most apps omit components that motivate patients to make lifestyle changes, instead providing information already available in paper form.[10,11] Concerns also exist about interface design, interactivity, connectivity,[12] the privacy and security of stored or transferred personal health information[13,14] and risks to patient safety from inaccurate or poorly designed apps.[15] In contrast with drugs or implantable devices, the US Food and Drug Administration (FDA)[16] and the Australian Therapeutic Goods Administration (TGA)[17] have no mandated regulatory oversight of health apps unless they are directly connected to regulated medical equipment. However, both agencies,[18,19] together with the Australian Digital Health Agency (ADHA),[20] have recently signalled their intention to focus attention on any software device, including apps, used to diagnose, prevent or manage disease.

In January 2018, the Queensland Policy and Advisory Committee on new Technology (QPACT), in partnership with Healthcare Evaluation and Assessment of Technology (HEAT) Team of Queensland Health (QH) and eHealth Queensland, established the mHealth Applications Evaluation Program (mHealth Apps Program) as a pilot initiative. The program welcomes submissions from clinicians who require funds to subject apps of their choosing to a rigorous field evaluation of quality, safety and benefit in order to receive official QH endorsement as an app that QH clinicians can recommend or 'prescribe' to their patients with confidence. The program was in rapid need of an evaluative framework and a better understanding of the efficacy and optimal design characteristics of mobile apps. The aims of this study were to: (1) assess the efficacy of currently available apps on chronic disease self-management; (2) identify design attributes that influence app usability and potential to confer benefit; and (3) review methods for evaluating app quality and utility and develop a provisional evaluative framework.

Methods

Objective no. 1: efficacy of apps in optimising chronic disease self-management

Criteria chosen by the mHealth program for app submissions (Table 1) determined the scope of literature reviews. PubMed, Cochrane Library and EMBASE were searched between January 2007 and January 2018 for systematic reviews (as defined using formal criteria[21]) of studies using search terms (and synonyms) of 'smartphone', 'mobile applications', 'mobile health' and terms for each of 11 chronic diseases, chosen for their burden of hospital utilisation,[22] namely asthma, chronic obstructive lung disease (COPD), diabetes, chronic pain, serious mental health disorders, alcohol and substance addiction, heart failure (HF), ischaemic heart disease, cancer, cognitive impairment and chronic kidney disease (CKD).

Inclusion criteria comprised systematic reviews of randomised controlled trials (RCTs) or observational studies that evaluated patient-facing interactive apps related to disease self-management, reported at least one measure of efficacy and were conducted in developed countries. Reference lists of included reviews were checked for other relevant studies. Reviews not written in English, those that analysed digital interventions not meeting our definition of a mobile app (i.e. exclusive use of text messaging, interactive voice response, desktop applications, websites) or those that examined apps used for diagnosis, screening, risk assessment, primary prevention or health promotion were excluded from the analysis.

In instances where no reviews for a particular condition could be found, individual studies were searched using the same search strategy and selection criteria, but excluding conference abstracts or case reports. Where multiple reviews dealing with the same topic were retrieved, the most recently published reviews were selected; other reviews were included if they provided additional informative data derived from thematic or subgroup analyses.

Data extracted from each review or individual study comprised: date of publication, number and type of studies, sample size, assessment of risk of bias (as defined by authors), app description, measures of efficacy, including meta-analyses of pooled data, assessment of patient adherence, usability and reference to behavioural change theory[23,24] underpinning the app, and the authors' interpretive comments on study strengths and limitations and overall results. In assessing efficacy, we made no attempt to prespecify what constituted clinically meaningful benefits in terms of absolute changes in effect measures, because this will vary according to subjective interpretations of clinicians and users and the outcomes and conditions of interest.

Table 1. Selection criteria for submissions to Queensland Policy and Advisory Committee on new Technology (QPACT) mHealth Applications Evaluation Program

TGA, Therapeutic Goods Administration

Inclusion criteria
Be a commercially available mobile health app (android or iOS platforms only)
Be patient facing (i.e. primarily for use by patients)
Assist patients to self-manage a clinical condition, or collect medical data in relation to a clinical condition
Have potential to improve patient outcomes, and reduce hospital visitation or treatment costs
Have TGA approval, should the app fall within the definition of a medical device
Exclusion criteria
Clinician facing apps (i.e. primarily used by clinicians as medical decision support)
Expressions of interest from clinicians or entities external to Queensland Health

Objective no. 2: design attributes that affect app usability and efficacy

Information about app design attributes affecting adherence, usability and potential to confer benefit were sought from the reviews retrieved, with additional reviews retrieved using app search terms that included 'quality', 'content', 'usability', 'feasibility', 'acceptability', 'functionality', 'adherence' and 'design.' In the absence of any validated usability score or taxonomy for mobile health apps, we considered all app design features that other researchers had proposed as being important on the basis of prima facie evidence.

Objective no. 3: methods for evaluating app quality and utility

Additional reviews were also sought using app search terms combined with 'quality assessment', 'evaluation' and related synonyms. Findings were used to construct a comprehensive but practical evaluation framework.

Results

Objective no. 1: efficacy of apps on chronic disease self-management

In total, 49 studies were retrieved,[11,25–72] comprising 16 reviews[11,25,29,31–39,42,51,60,64] and 33 individual studies.[26–28,30,40,41,43–50,52–59,61–63,65–72] Key findings are presented below; more detailed evidence tables are provided in Appendix 1.

Asthma

In a review of 12 RCTs of digital aids,[25] three app trials were reported; one saw a 45% increase in the number of patients with well-controlled asthma,[26] another saw no effects on several outcomes[27] and the third reported improvements in asthma-related quality of life and an 80% decrease in emergency department visits.[28]

Chronic obstructive pulmonary disease

In a review of three RCTs of digital interventions,[29] a single app trial reported no difference in COPD-related quality of life; no other outcomes were reported.[30]

Diabetes

Four reviews (one with 16 RCTs involving apps,[31] another with 12 RCTs,[32] a third with 10 RCTs[33] and a fourth with six RCTs[34]) all reported reductions in glycosylated haemoglobin levels (HbA1c) for all diabetics (mean reductions ranging from 0.48% to 0.51%) and those with type 2 diabetes (0.49–0.83%).

Other reviews[35,36] yielded mixed results, with only one-third of included studies reporting significant reductions in HbA1c.

Chronic pain

Although four reviews were retrieved assessing more than 300 apps,[11,37–39] none contained any trials of efficacy. A single app RCT showed an improvement in pain control, functionality and quality of life of patients with low back pain,[40] whereas another app involving women with widespread pain that included diaries and therapist feedback showed no effects at 11 months.[41]

Serious mental health disorders

In a review of 18 studies,[42] 14 app studies included four RCTs. Of two RCTs of depression apps, one showed no effect,[43] whereas another RCT reported improved mental health scores and reduced lost or unproductive days.[44] An RCT of an app for psychotic disorders reported 'positive effects',[45] whereas another RCT for bipolar disorder reported no effects.[46] A single-arm study of an app for schizophrenia and schizoaffective disorder was associated with significant reduction in symptoms.[47] A review of 24 app studies in children and adolescents with various mental disorders[48] included two RCTs, both showing no effects.[49,50]

Alcohol and substance addiction

In a review of five studies with three app RCTs,[51] one RCT showed increased frequency of drinking in men but not women compared with controls.[52] The second saw 1.37 fewer risky drinking days among app patients over 12 months,[53] whereas the third noted less drinking over 14 days but no effect on heavy episodic drinking.[54]

Heart failure

No reviews were found, but, of two RCTs, one using a tablet computer reported 2.2 fewer HF-related days in hospital per patient and improved HF-related quality of life and physical function over 3 months.[56] The second RCT using home telemonitoring reported no difference in hospital days and, instead, an increase in health care utilisation with more nurse visits and telephone contacts.[56]

Ischaemic heart disease

No reviews were found but, of three RCTs of apps targeting patients following acute coronary events, one reported improved medication adherence but no effect on smoking rates, physical activity or quality of life at 6 months.[57] Another showed slight weight reduction and improved quality of life at 6 weeks.[58] The third RCT showed weight loss and improved blood pressure and blood sugar control, but no change in medication adherence, smoking rates, physical activity or quality of life.[59]

Cancer

In a review of five studies[60] that reported three app trials, a controlled trial involving prostate cancer patients receiving radiotherapy reported lower levels of fatigue, nausea, insomnia, urinary symptoms and emotional dysfunction.[61] An RCT involving breast cancer patients showed no effect on physical function or quality of life,[62] whereas a pre-post study involving survivors of breast and endometrial cancer noted weight reduction but no change in physical activity or quality of life.[63]

Cognitive impairment

In a review of 24 studies,[64] two RCTs[65,66] and one controlled trial[67] of apps focused on cognitive training showed improved cognition, whereas one RCT did not.[68] Three other RCTs, one of serious games coupled with cognitive training,[69] one with a multicomponent program[70] and another of engagement,[71] showed no effects on various outcomes.

Chronic kidney disease

No reviews were found. Only one small pre-post cohort study of an app reported mean reductions in home blood pressure readings from baseline values.[72]

Objective no. 2: design attributes that affect app usability and efficacy

Eight reviews were considered relevant.[35,64,73–78] In the review of diabetes apps by Fu *et al.*[35] seven usability studies reported poor (38%) to average (80%) rates of use, due primarily to product design flaws (e.g. screen layout, system capability and reliability, and general characteristics), need for manual data entry, patients having to take more time and making more errors than expected when exporting and correcting blood sugar levels, difficulty encountered in system navigation whenever tasks required multiple steps and the absence of personalised feedback functions or social networking functions. No study considered confounders such as a patient's history of, and levels of motivation in, using digital technologies. Only 51 of 295 (17.3%) cancer apps,[73] 12 of 117 (10.3%) depression apps[74] and a minority of chronic pain apps[75] provide skill-building tools to assist in self-management. On the basis of these and other reviews[68,76–78] we identified several design features for enhancing usability (Table 2).

Objective no. 3: methods for evaluating app quality and utility

Methods for evaluating apps fall into three broad categories, as summarised in Table 3,[79] each providing a different type of evidence but with limitations. For example, some lists of criteria focus on user ratings (ease of use, reliability, aesthetics) but may exclude reference to independent expert clinician involvement or assessment.[80] Others cover development, implementation and integration, but omit evaluation of key usability attributes or privacy and security.[81] Evaluation guidelines from the Health Care Information and Management Systems Society consider efficiency, effectiveness, user satisfaction and platform optimisation, but exclude accuracy and appropriateness of app health information.[82] Such omissions pose risks to patient safety and explicit app risk assessment has been proposed.[15] Various rating scales,[83,84] scoring systems,[85,86] checklists,[87–91] toolkits,[92,93] questionnaires,[94] development guides[95] and reporting standards for app studies[96,97] all promote more transparent and objective reporting and evaluation. The existence of so many instruments suggests no single one meets all needs.

Table 2. Design features enhancing patient use of an app and ability to gain most benefit

Develop different versions of apps for different stages of disease (mild to severe)
Make the app easy to navigate and include a tutorial for using it
Engage the user by having an app that is colourful, consolidated, convenient and relevant to their needs
Intuitive interaction with appealing user interfaces
Variety of incentives to keep coming back to use it
Scripted storytelling or user-generated narrative
Visual characters and icons that empathise with users or that users can empathise with
Ideal worlds that can be explored, or real worlds that can be augmented
Educate and prepare users by having up-to-date, personalised information that informs and enlightens
Motivate the user to change behaviour by providing:
practical tips for better self-management
achievable goals of care
reminders, alerts and gamification
personalised, real-time feedback
access to expert clinical advice when and as needed
access to desired peer and social support (chat forums, social networks)
Ensure basic self-management tasks are universal functions, such as disease markers (e.g. blood glucose level in diabetes), medications, nutrition, physical exercise and bodyweight
Enable easy data entry (voice and image recognition, data transfer from other devices, built-in smartphone sensors)
Enable analysis of data and its patterns in discerning interactions between behavioural changes and disease markers
Include predictive analytics based on gathered data
Enable compatibility with different operating systems for smartphones by using packages available for different platforms
Enable data saving, transfer and sharing (from external devices, patient to medical expert, between family members)

Table 3. Various methods for evaluating apps
RCTs, randomised control trials

Method	Limitations
Content analysis: coding and interpreting qualitative, usually text-based material within the app and validating it against clinical guidelines, evidence-based protocols and behaviour change techniques	Variability in how apps have implemented guidelines, evidence-based strategies and behaviour change, requiring judgement calls on the part of the evaluator Reliance on subjective ratings of app content and features Frequent app updates with changed functionalities Content includes not only clinical guidance, but also related advertisements, industry data, software licences, privacy policies
Usability testing: ascertaining user ratings of app learnability, flexibility, satisfaction, attractiveness, consistency and error rate, by means of expert-based evaluations, observational studies or surveys of users or experimental evaluations, all of which may include laboratory testing, field-based evaluations and reviewing ratings and narrative user reviews from app marketplaces	Usability testing in laboratories may not reflect real-world usage Capturing app use in dynamic environments makes direct observations difficult in field studies Unrepresentative and small user samples Use of average user ratings may not disclose highly polarised bad or good ratings of low Ratings may be confounded by external factors, such as time to last update, app vocabulary, app description
Efficacy testing: establishing whether use of the app causes meaningful change in behaviour and clinical outcomes, with gold-standard approach being RCTs	Time and resource intense Fast moving app development with frequent updates causing lag between trial and real-world use Inability to test whether success of app-delivered intervention attributable to whole package or specific components

Several government (https://www.vichealth.vic.gov.au/media-and-resources/vichealth-apps; https://www.healthnavigator.org.nz/app-library), private (http://myhealthapps.net/; https://imedicalapps.com/#; https://practicalapps.ca/) and developer (https://www.ourmobilehealth.com/; https://www.medappcare.com/en/) websites (all accessed 13 September 2018) also exist that post apps that have been assessed using various in-house rating methods.

We contend that any evaluation method should consider clinical and technical attributes of an app that we have identified from this review as being important, while not being too general or complex or time consuming with need for training, or too specific to a particular health condition or outcome. After analysing the above reports and recent reviews,[98–101] we derived a list of criteria (Table 4) to which yes/no or scalar (Likert) responses could gauge the extent to which an app meets each criterion. Obtaining informed responses to each criterion requires retrieving relevant data using strategies listed in Table 5.[102] We suggest these criteria may suffice as a 'first-pass' assessment by panels involving clinicians in deciding whether

Table 4. Framework for evaluating app quality and utility
PHI, protected health information

What does the app actually do?
Is the app downloadable and usable on any smartphone?
What platform (android or iPhone) does it operate on?
Is it clear who this app is for (target population) and how it should be used?
Is it clear which health problem the app is designed to alleviate or what outcome it helps achieve?
Are software and hardware requirements for the app to work widely available?
Is there adequate technical and content detail provided to support adopting the app at scale?
Does the app have clinical and scientific validity?
Has its development and testing involved clinical experts?
Are the content, rules, calculators, algorithms and recommendations concordant with evidence-based clinical guidelines or systematic reviews?
Are citations or evidence links provided to substantiate content and recommendations?
Have professional organisations endorsed the app or been affiliated with its development?
How credible is the app developer?
Has the developer been fully transparent in disclosing contact information, authorship, affiliations, credentials, commercial interests (investors and shareholders), conflicts of interest and disclaimers?
Has the developer disclosed whether the app creates, receives, maintains or transmits PHI on behalf of health providers, healthcare institutions or any other third party contracted with the developer (i.e. PHI is not for exclusive use by the app user)?
Has the developer given adequate attribution for content to other bodies or individuals in terms of copyright disclosures and listing of citations?
Has the developer provided a means for users to provide feedback on the quality and usability of the app?
Has the developer sought regulatory assessment or approval for the app?
How well does the app work for its purpose?
Has software been validated and verified with regard to accuracy of clinical content and its ability to have users interrogate and correctly interpret clinical content?
Are calculators, algorithms and other interactive components accurate and have they undergone reliability testing?
Are there adequate levels of security and privacy in relation to risk as reflected in privacy policies, encryption and authentication, password protection, user controls with data sharing?
Are there adequate levels of interoperability with other devices and systems?
How well does the app engage the user and change behaviour?
Is the app free of charge, or is it affordable to the intended target audience?
Has interface design involved user participation or been informed by studies of user needs?
Has a health behaviour change theory or construct informed interface design and operation?
Is the app fast, pleasurable and easy to use (i.e. minimal training required) in routine settings?
Is the app accessible to target audiences who may have limited digital and/or health literacy, cultural barriers (language, ethnicity), visual or hearing impairment?
Does the app give the user usable answers or advice quickly?
Does the app screen look well designed and are text and images clear?
To what extent does the app provide personalised information and feedback, skills development and access to expert advice?
Has the usability and quality of the interface been adequately assessed using different metrics, namely:
user testing and satisfaction (self-report, app stars or rankings)
user engagement (number of downloads)
user learnability (amount of time and number of clicks users need to master the app)
user understanding (literacy testing and inclusion of multiple languages)
user configuration and tailoring (ability for users to customise functions and alter default settings)?
How safe, effective and cost-effective is the app in optimising health outcomes?
Have analyses sought to identify any risks to patient safety or other unintended consequences?
Are there analyses of effects on patient-important behaviours, actions or health outcomes? If so, what are the results?
Have analyses been performed that identify patient characteristics, design features or functional components that confer greater or less effectiveness?
Have analyses been undertaken that assess opportunity costs of the app versus other forms of evidence-based care for specific patient populations?

the app justifies more detailed technical assessments under a second-phase due-diligence process.

Discussion

To the best of our knowledge, this is the first review of evidence relating to the efficacy, usability and evaluation of apps that assist patients to self-manage chronic diseases. Limitations of this review are that, due to time and resource constraints, we did not search for and analyse every individual study of every app, but instead relied on systematic reviews, when these were available, some of which differed in inclusion criteria and conclusions. However, we do not feel any important study was overlooked and, in particular, RCTs with less risk of bias were diligently sought and given emphasis. Space limitations did not allow us to provide full descriptions of each app, some having multiple components, but readers can refer to the references provided for more detail. We view our framework as being a provisional 'first-pass' framework that may help select app submissions that warrant second-phase due-diligence reviews to which more detailed assessments of technical attributes relating to interoperability, security and privacy may be added.

Table 5. Strategies for retrieving relevant data in responding to evaluation criteria
NHS, National Health Service

Strategy	Limitation
1. Undertake literature reviews: search for systematic reviews or individual studies of high quality that evaluate apps in regards to content, usability, efficacy Tips: restrict search by using specific search terms and limited, recent time periods, as used in this paper under Objective no. 1	Limited published research Poor indexing of studies which makes searching time consuming Evaluations often focussed on one operating system (android or iPhone) Literature becomes quickly outdated given pace of change
2. Search app clearinghouse websites; such sites[A] can help identify app strengths and weaknesses. Tips: restrict search to those sites that are affiliated with non-commercial (academic or government) agencies and that use standardised evaluation criteria, such as NHS Health Apps (http://apps.beta.nhs.uk) or the Mobile Health Application Repository (http://nba.uth.tmc.edu/homepage/liu/mHealth/mHealth.htm)	Quality of app reviews dependent on evaluation methods used by each clearinghouse Inherent trade-off between thoroughness of review, timeliness of updates and number of apps reviewed Can be resource intensive and time consuming
3. If Strategies 1 and 2 do not provide relevant information, app stores (iTunes or Google Play) can be searched directly Tips: use search terms that are as specific as possible; for example, in finding an app that assists problem drinkers, use 'alcohol abstinence' or 'alcoholism' rather than 'alcohol'	Lack of systematic evaluation that limits accuracy App stores rely on ranking algorithms that rank the most popular or most recently released apps at a higher level than others, but may have little correlation with quality or efficacy Retrieval of many irrelevant apps because of poor indexing and desire for app stores to maximise consumer market
4. Review app descriptions, user ratings and reviews: these are often available through app stores directly or through distributors such as Amazon (we would recommend using this information only to get a description of the app and gain a sense of usability; do not use them for evaluating clinical validity, accuracy or efficacy of an app)	
5. Conduct a social media query within professional and, if available, patient networks: this allows a larger number of interested, and hopefully vetted and trusted, users to be surveyed about their impressions of an app	May oversample users with only positive experiences Does not allow standardised evaluation of validity and efficacy Patients and clinicians may have opposing views of the same app
6. Pilot the app: evaluators can download the app and test it directly, although this may require the app to be purchased and has limitations in numbers of, and levels of agreement between, evaluators	
7. Elicit feedback from patients: provide app to a representative group of patients and follow them over time with regard to clinical outcomes achieved and whether patients find the app useful, will continue to use it and would recommend it to others	

[A]Sites such as https://www.vichealth.vic.gov.au/media-and-resources/vichealth-apps, https://www.healthnavigator.org.nz/app-library, http://myhealthapps.net/; https://imedicalapps.com/#, https://practicalapps.ca/, https://digitalhealthguide.com.au, https://www.ourmobilehealth.com/ and https://www.medappcare.com/en/ (all accessed 13 September 2018) or identified by Google and PubMed searches or reference to newsletters such as mobihealthnews (http://mobihealthnews.com/, accessed 13 September 2018).

Improving the evidence base of app efficacy

The rapid pace of the technological evolution of apps and the slowness and expense of gold-standard research designs, such as RCTs, make assessments of efficacy problematic. Evaluating apps is complex given the combination of content, user, platform, links and interface attributes, and determining whether benefits are attributable to a total app package or specific components. Apps can be released, updated, modified or removed by commercial developers as studies with fixed protocols evaluating a specific app (or type of app) are underway, rendering the results potentially obsolete by the time of public release. Control groups in comparative trials may be contaminated by being exposed to similar or changing digital interventions during the course of a study. In addition, results of new clinical studies may change the evidence on which the core content and behaviour change functions contained in an app are based, such that the app itself, despite having been studied, becomes invalid. Finally, apps investigated in research settings are often not the same as those that are disseminated commercially.

Alternative study designs have been suggested that aim to overcome some of these issues, albeit with limitations. These include multiphase optimisation (MOST) strategies (factorial or fractional designs and adaptive trials),[103] *n*-of-1 studies that involve repeated measurement of many individuals over time in understanding within-person behaviour before and then after using an app, or using an app and then having it withdrawn,[104] or triangulation of 'fit-for-purpose' studies that investigate sociocultural, organisational, cognitive and other contextual determinants of effects using surveys, focus group discussions and ethnographic studies.[105]

Relationship between app use and efficacy

In many reviews, more focus is given to assessing usability than to testing efficacy. We are concerned that apps rated by many users as highly usable and that become popular may be viewed by new users as apps that predictably confer health benefits. This is akin to 'I really like and use the app so therefore it must be doing me good'. Design attributes that discourage people from using an app in the way intended will certainly compromise efficacy, but the reverse is not true. Usability is a multifaceted construct that requires app design to be informed by behaviour change theory, because information alone is insufficient to change behaviour.[106] In this review, only a minority of studies explicitly reported using behaviour change theories to underpin app design, and adherence and retention rates were frequently <80%. Although many apps make reference to behavioural theories, self-monitoring alone is often the dominant function rather than incorporating proactive skills development in self-management, which may substantially enhance efficacy.[107] Using apps pertaining to mental health and addiction problems as examples, apps need to incorporate a behavioural conditioning plan and interactive framework that include rapid access to expert help at times of crisis and that mitigate limited attention span and curtail time spent on devices by emphasising non-app-based activities.[108] Personalised information, real-time feedback and access to expert consultation when required are features highly valued by users. There is a need for a checklist of usability and functionality attributes that apps should aspire to in maximising user engagement, which this review has helped create.

Implications for app evaluation and endorsement

Lack of evidence of app efficacy, poor descriptions of processes and data sources used to develop the app and discrepancies between information generated on apps and evidence-based guidelines are key issues in evaluation.[109] Of note, apps requiring purchase are not necessarily more evidence based than free apps.[110] With regard to efficacy testing, our overview highlights several shortcomings in the current evidence base: a paucity of rigorous efficacy trials with adequate sample sizes for most conditions; inconsistency of measured outcomes, with many based on subjective patient self-report, with relatively few hard end-points (e.g. mortality, clinical events, hospitalisations); insufficient duration for assessing long-term effects; minimal reference to adherence rates and underlying behavioural change techniques; questionable fidelity of patient compliance with data input and use of algorithms; and limited generalisability given heterogeneity in populations, interventions, outcome measures and conflicting estimates of effect. The exceptions to many of these limitations are diabetes apps, for which sufficient and consistent data confirm efficacy in optimising blood sugar control, although effects on other outcomes await assessment. For all conditions, there was little focus on the needs and competencies of older patients[111] and culturally diverse groups,[112] or on cost-effectiveness, including costs of misinformation transmitted, of addressing resulting problems or of the diverse workforce required for app implementation and monitoring.[113]

Methods for evaluating the quality and utility of an app must be capable of assessing all relevant attributes of technical integrity, usability and efficacy in an explicit and transparent manner. This review suggests an evaluation framework that could be refined and tested for construct validity and internal consistency. Potential users of the framework may include the ADHA, TGA, state-based government digital health agencies and networks, app developers and any other stakeholder group who has responsibilities in health app evaluation and endorsement. In Queensland, the mHealth Apps Program, under the auspices of QPACT, HEAT, and eHealth Queensland, is being resourced to further develop the framework and its application to all future submissions from developers and interested clinicians for QPACT funding of field trials of mobile health apps within QH facilities.

Conclusion

The numbers of smartphone apps will continue to grow as will the appetite for patients and clinicians to use them in chronic disease self-management. However, the evidence to date of clinical benefit of most of the apps already available is very limited. Design features that enhance users' desire to use the app, and thereby derive most benefit, need to be considered in maximising efficacy. In making decisions about which apps should be endorsed by government agencies and recommended with confidence by clinicians to their patients, a workable evaluation framework needs to be used by those agencies that assume the role of formulating and applying app standards. The provisional framework presented in this report serves as a starting point for such work.

Competing interests

The authors declare that they have no competing interests.

Acknowledgements

There was no funding source for this research. The authors are all members of the Queensland Health mHealth Apps Working Group. The views expressed in this article should not be regarded as official policy of Queensland Health or any of its divisions, Queensland Policy and Advisory Committee on new Technology, Healthcare Evaluation and Assessment of Technology Team or eHealth Queensland. The views expressed are solely those of the authors.

References

1 Mackay M. Australian mobile phone lifestyle index. 2014. Available at: https://www.dtcollective.org.au/LiteratureRetrieve.aspx?ID=180479 [verified 22 September 2018].

2 Jahns RG. The 8 drivers and barriers that will shape the mHealth app market in the next 5 years. r2g Mobile Health Economics. 2014. Available at: https://research2guidance.com/8-drivers-and-barriers-which-will-shape-mhealth-market-in-the-next-5-years/ [verified 22 September 2018].

3 Aitken M, Lyle J. Patient adoption of mHealth: use, evidence and remaining barriers to mainstream acceptance. Parsippany, NJ: IMS Institute for Healthcare Informatics; 2015.

4 Bhuyan SS, Lu N, Chandak A, Kim H, Wyant D, Bhatt J, Kedia S, Chang CF. Use of mobile health applications for health seeking behaviour among US adults. *J Med Syst* 2016; 40: 153. doi:10.1007/s10916-016-0492-7

5 Powell AC, Landman AB, Bates DW. In search of a few good apps. *JAMA* 2014; 311: 1851–2. doi:10.1001/jama.2014.2564

6 Boulos MN, Brewer AC, Karimkhani C, Buller DB, Dellavalle RP. Mobile medical and health apps: state of the art, concerns, regulatory control and certification. *Online J Public Health Inform* 2014; 5: 229.

7 Risk A, Dzenowagis J. Review of Internet health information quality initiatives. *J Med Internet Res* 2001; 3: E28. doi:10.2196/jmir.3.4.e28

8 Gagliardi A, Jadad AR. Examination of instruments used to rate quality of health information on the internet: chronicle of a voyage with an unclear destination. *BMJ* 2002; 324: 569–73. doi:10.1136/bmj.324.7337.569

9 Subhi Y, Bube SH, Bojsen SR, Thomsen ASS, Konge L. Expert involvement and adherence to medical evidence in medical mobile phone apps: a systematic review. *JMIR Mhealth Uhealth* 2015; 3: e79. doi:10.2196/mhealth.4169

10 Goyal S, Cafazzo JA. Mobile phone apps for diabetes management: current evidence and future developments. *QJM* 2013; 106: 1067–9. doi:10.1093/qjmed/hct203

11 Wallace LS, Dhingra LK. A systematic review of smartphone applications for chronic pain available for download in the United States. *J Opioid Manag* 2014; 10: 63–8. doi:10.5055/jom.2014.0193

12 Reynoldson C, Stones C, Allsop M, Gardner P, Bennett MI, José Closs S, Jones R, Knapp P. Assessing the quality and usability of smartphone apps for pain self-management. *Pain Med* 2014; 15: 898–909. doi:10.1111/pme.12327

13 Martínez-Pérez B, De La Torre-Diez I, Lopez-Coronado M. Privacy and security in mobile phone apps: a review and recommendations. *J Med Syst* 2015; 39: 181. doi:10.1007/s10916-014-0181-3

14 Pustozerov E, von Jan U, Albrecht UV. Evaluation of mHealth applications security based on application permissions. *Stud Health Technol Inform* 2016; 226: 241–4.

15 Lewis TL, Wyatt JC. mHealth and mobile medical apps: a framework to assess risk and promote safer use. *J Med Internet Res* 2014; 16: e210. doi:10.2196/jmir.3133

16 U.S. Department of Health and Human Services, Food and Drug Administration. Mobile medical applications: guidance for industry and Food and Drug Administration staff. 2015. Available at: http://www.fda.gov/downloads/MedicalDevices/./UCM263366.pdf [verified 13 January 2018]

17 Australian Government Department of Health. Therapeutic goods administration. n. Regulation of medical software and mobile medical 'apps'. 2013. Available at: https://www.tga.gov.au/regulation-medical-software-and-mobile-medical-apps [verified 13 January 2018]

18 U.S. Department of Health and Human Services, Food and Drug Administration. Software as a medical device (SAMD): Clinical evaluation. Guidance for industry and Food and Drug Administration Staff. 2017. Available at: https://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/UCM524904.pdf [verified 21 September 2018].

19 Hau J. Update on software as a medical device (SaMD). The TGA and IMDRF perspectives. 2017. Available at: https://www.tga.gov.au/sites/default/files/update-on-software-as-a-medical-device-samd.pdf [verified 20 January 2018].

20 Australian Digital Health Agency (ADHA). Safe, seamless and secure: evolving health and care to meet the needs of modern Australia. Australia's national digital health strategy. Canberra: ADHA; 2017.

21 Moher D, Liberati A, Tetzlaff J, Altman DG, Group TP. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med* 2009; 6: e1000097. doi:10.1371/journal.pmed.1000097

22 Lozano R, Naghavi M, Foreman K, Lim S, Shibuya K, Aboyans V, Abraham J, Adair T, Aggarwal R, Ahn SY, Alvarado M, Anderson HR, Anderson LM, Andrews KG, Atkinson C, Baddour LM, Barker-Collo S, Bartels DH, Bell ML, Benjamin EJ, Bennett D, Bhalla K, Bikbov B, Bin Abdulhak A, Birbeck G, Blyth F, Bolliger I, Boufous S, Bucello C, Burch M, Burney P, Carapetis J, Chen H, Chou D, Chugh SS, Coffeng LE, Colan SD, Colquhoun S, Colson KE, Condon J, Connor MD, Cooper LT, Corriere M, Cortinovis M, de Vaccaro KC, Couser W, Cowie BC, Criqui MH, Cross M, Dabhadkar KC, Dahodwala N, De Leo D, Degenhardt L, Delossantos A, Denenberg J, Des Jarlais DC, Dharmaratne SD, Dorsey ER, Driscoll T, Duber H, Ebel B, Erwin PJ, Espindola P, Ezzati M, Feigin V, Flaxman AD, Forouzanfar MH, Fowkes FG, Franklin R, Fransen M, Freeman MK, Gabriel SE, Gakidou E, Gaspari F, Gillum RF, Gonzalez-Medina D, Halasa YA, Haring D, Harrison JE, Havmoeller R, Hay RJ, Hoen B, Hotez PJ, Hoy D, Jacobsen KH, James SL, Jasrasaria R, Jayaraman S, Johns N, Karthikeyan G, Kassebaum N, Keren A, Khoo JP, Knowlton LM, Kobusingye O, Koranteng A, Krishnamurthi R, Lipnick M, Lipshultz SE, Ohno SL, Mabweijano J, MacIntyre MF, Mallinger L, March L, Marks GB, Marks R, Matsumori A, Matzopoulos R, Mayosi BM, McAnulty JH, McDermott MM, McGrath J, Mensah GA, Merriman TR, Michaud C, Miller M, Miller TR, Mock C, Mocumbi AO, Mokdad AA, Moran A, Mulholland K, Nair MN, Naldi L, Narayan KM, Nasseri K, Norman P, O'Donnell M, Omer SB, Ortblad K, Osborne R, Ozgediz D, Pahari B, Pandian JD, Rivero AP, Padilla RP, Perez-Ruiz F, Perico N, Phillips D, Pierce K, Pope CA III, Porrini E, Pourmalek F, Raju M, Ranganathan D, Rehm JT, Rein DB, Remuzzi G, Rivara FP, Roberts T, De León FR, Rosenfeld LC, Rushton L, Sacco RL, Salomon JA, Sampson U, Sanman E, Schwebel DC, Segui-Gomez M, Shepard DS, Singh D, Singleton J, Sliwa K, Smith E, Steer A, Taylor JA, Thomas B, Tleyjeh IM, Towbin JA, Truelsen T, Undurraga EA, Venketasubramanian N, Vijayakumar L, Vos T, Wagner GR, Wang M, Wang W, Watt K, Weinstock MA, Weintraub R, Wilkinson JD, Woolf AD, Wulf S, Yeh PH, Yip P, Zabetian A, Zheng ZJ, Lopez AD, Murray CJ, AlMazroa MA, Memish ZA. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. [published erratum appears in *Lancet* 2013; 381:628] *Lancet* 2012; 380: 2095–128. doi:10.1016/S0140-6736(12)61728-0

23 Abraham C, Mitchie S. A taxonomy of behaviour change techniques used in interventions. *Health Psychol* 2008; 27: 379–87. doi:10.1037/0278-6133.27.3.379

24 Michie S, Ashford S, Sniehotta FF, Dombrowski SU, Bishop A, French DP. 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: 1479–98. doi:10.1080/08870446.2010.540664

25 Hui CY, Walton R, McKinstry B, Jackson T, Parker R, Pinnock H. The use of mobile applications to support self-management for people with asthma: a systematic review of controlled studies to identify features associated with clinical effectiveness and adherence. *J Am Med Inform Assoc* 2017; 24: 619–32.

26 Cingi C, Yorgancioglu A, Cingi CC, Oguzulgen K, Muluk NB, Ulusoy S, Orhon N, Yumru C, Gokdag D, Karakaya G, Çelebi S, Çobanoglu HB, Unlu H, Aksoy MA. The 'physician on call patient engagement trial' (POPET): measuring the impact of a mobile patient engagement application on health outcomes and quality of life in allergic rhinitis and asthma patients. *Int Forum Allergy Rhinol* 2015; 5: 487–97. doi:10.1002/alr.21468

27 Ryan D, Price D, Musgrave SD, Malhotra S, Lee AJ, Ayansina D, Sheikh A, Tarassenko L, Pagliari C, Pinnock H. Clinical and cost effectiveness of mobile phone supported self monitoring of asthma: multicentre randomised controlled trial. *BMJ* 2012; 344: e1756. doi:10.1136/bmj.e1756

28 Liu WT, Huang CD, Wang CH, Lee KY, Lin SM, Kuo HP. A mobile telephone-based interactive self-care system improves asthma control. *Eur Respir J* 2011; 37: 310–7. doi:10.1183/09031936.00000810

29 McCabe C, McCann M, Brady AM. Computer and mobile technology interventions for self-management in chronic obstructive pulmonary disease. *Cochrane Database Syst Rev* 2017; 5: CD011425.

30 Tabak M, Vollenbroek-Hutten MM, van der Valk PD, van der Palen J, Hermens HJ. A telerehabilitation intervention for patients with chronic obstructive pulmonary disease: a randomized controlled pilot trial. *Clin Rehabil* 2014; 28: 582–91. doi:10.1177/0269215513512495

31 Wu IXY, Kee JCY, Threapleton DE, Ma R, Lam VCK, Lee EKP, Wong SYS, Chung VCH. Effectiveness of smartphone technologies on glycaemic control in patients with type 2 diabetes: systematic review with meta-analysis of 17 trials. *Obes Rev* 2018; 19: 825–38. doi:10.1111/obr.12669

32 Wu Y, Yao X, Vespasiani G, Nicolucci A, Dong Y, Kwong J, Li L, Sun X, Tian H, Li S. Mobile app-based interventions to support diabetes self-management: a systematic review of randomized controlled trials to identify functions associated with glycemic efficacy. *JMIR Mhealth Uhealth* 2017; 5: e35. doi:10.2196/mhealth.6522

33 Hou C, Carter B, Hewitt J, Francisa T, Mayor S. Do mobile phone applications improve glycemic control (HbA1c) in the self-management of diabetes? A systematic review, meta-analysis, and GRADE of 14 randomized trials. *Diabetes Care* 2016; 39: 2089–95. doi:10.2337/dc16-0346

34 Cui M, Wu X, Mao J, Wang X, Nie M. T2DM self-management via smartphone applications: a systematic review and meta-analysis. *PLoS One* 2016; 11: e0166718. doi:10.1371/journal.pone.0166718

35 Fu H, McMahon SK, Gross CR, Adam TJ, Wyman JF. Usability and clinical efficacy of diabetes mobile applications for adults with type 2 diabetes: a systematic review. *Diabetes Res Clin Pract* 2017; 131: 70–81. doi:10.1016/j.diabres.2017.06.016

36 Holmen H, Wahl AK, Cvancarova Småstuen M, Ribu L. Tailored communication within mobile apps for diabetes self-management: a systematic review. *J Med Internet Res* 2017; 19: e227. doi:10.2196/jmir.7045

37 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: 557–63. doi:10.1097/AJP.0000000000000171

38 Machado GC, Pinheiro MB, Lee H, Ahmed OH, Hendrick P, Williams C, Kamper SJ. Smartphone apps for the self-management of low back pain: a systematic review. *Best Pract Res Clin Rheumatol* 2016; 30: 1098–109. doi:10.1016/j.berh.2017.04.002

39 Nicholl BI, Sandal LF, Stochkendahl MJ, McCallum M, Suresh N, Vasseljen O. Digital support interventions for the self-management of low back pain: a systematic review. *J Med Internet Res* 2017; 19: e179. doi:10.2196/jmir.7290

40 Irvine AB, Russell H, Manocchia M, Mino DE, Cox GT, Morgan R, Gau JM, Birney AJ, Ary DV. Mobile-web app to self-manage low back pain: randomized controlled trial. *J Med Internet Res* 2015; 17: e1. doi:10.2196/jmir.3130

41 Kristjánsdóttir ÓB, Fors EA, Eide E, Finset A, Stensrud TL, van Dulmen S, Wigers SH, Eide H. A smartphone-based intervention with diaries and therapist feedback to reduce catastrophizing and increase functioning in women with chronic widespread pain. Part 2: 11-month follow-up results of a randomized trial. *J Med Internet Res* 2013; 15: e72. doi:10.2196/jmir.2442

42 Batra S, Baker RA, Wang T, Forma F, DiBiasi F, Peters-Strickland T. Digital health technology for use in patients with serious mental illness: a systematic review of the literature. *Med Devices (Auckl)* 2017; 10: 237–51. doi:10.2147/MDER.S144158

43 Ly KH, Trüschel A, Jarl L, Magnusson S, Windahl T, Johansson R, Carlbring P, Andersson G. Behavioural activation versus mindfulness-based guided self-help treatment administered through a smartphone application: a randomised controlled trial. *BMJ Open* 2014; 4: e003440. doi:10.1136/bmjopen-2013-003440

44 Watts S, Mackenzie A, Thomas C, Griskaitis A, Mewton L, Williams A, Andrews G. CBT for depression: a pilot RCT comparing mobile phone vs. computer. *BMC Psychiatry* 2013; 13: 49. doi:10.1186/1471-244X-13-49

45 Forchuk C, Reiss JP, O'Regan T, Ethridge P, Donelle L, Rudnick A. Client perceptions of the Mental Health Engagement Network: a qualitative analysis of an electronic personal health record. *BMC Psychiatry* 2015; 15: 250. doi:10.1186/s12888-015-0614-7

46 Faurholt-Jepsen M, Ritz C, Frost M, Mikkelsen RL, Christensen EM, Bardram J, Vinberg M, Kessing LV. Mood instability in bipolar disorder type I versus type II: continuous daily electronic self-monitoring of illness activity using smartphones. *J Affect Disord* 2015; 186: 342–9. doi:10.1016/j.jad.2015.06.026

47 Ben-Zeev D, Brenner CJ, Begale M, Duffecy J, Mohr DC, Mueser KT. Feasibility, acceptability, and preliminary efficacy of a smartphone intervention for schizophrenia. *Schizophr Bull* 2014; 40: 1244–53. doi:10.1093/schbul/sbu033

48 Grist R, Porter J, Stallard P. Mental health mobile apps for preadolescents and adolescents: a systematic review. *J Med Internet Res* 2017; 19: e176. doi:10.2196/jmir.7332

49 Reid SC, Kauer SD, Hearps SJ, Crooke AH, Khor AS, Sanci LA, Patton GC. A mobile phone application for the assessment and management of youth mental health problems in primary care: health service outcomes from a randomised controlled trial of mobile type. *BMC Fam Pract* 2013; 14: 84. doi:10.1186/1471-2296-14-84

50 Veldhuis J. Media models matter in context: negotiated media effects of idealized body images. PhD Thesis, Vrije Universiteit Amsterdam Faculteit der Social Wetenschappen; 2014. Available at: https://research.vu.nl/en/publications/media-models-matter-in-context-negotiated-media-effects-of-ideali [verified 21 September 2018].

51 Meredith SE, Alessi SM, Petry NM. Smartphone applications to reduce alcohol consumption and help patients with alcohol use disorder: a state-of-the-art review. *Adv Health Care Technol* 2015; 1: 47–54.

52 Gajecki M, Berman AH, Sinadinovic K, Rosendahl I, Andersson C. Mobile phone brief intervention applications for risky alcohol use among university students: a randomized controlled study. *Addict Sci Clin Pract* 2014; 9: 11. doi:10.1186/1940-0640-9-11

53 Gustafson DH, McTavish FM, Chih MY, Atwood AK, Johnson RA, Boyle MG. A smartphone application to support recovery from alcoholism: a randomized clinical trial. *JAMA Psychiatry* 2014; 71: 566–72. doi:10.1001/jamapsychiatry.2013.4642

54 Witkiewitz K, Desai SA, Bowen S, Leigh BC, Kirouac M, Larimer ME. Development and evaluation of a mobile intervention for heavy drinking and smoking among college students. *Psychol Addict Behav* 2014; 28: 639–50. doi:10.1037/a0034747

55 Hägglund E, Lyngå P, Frie F, Ullman B, Persson H, Melin M, Hagerman I. Patient-centred home-based management of heart failure. Findings from a randomised clinical trial evaluating a tablet computer for self-care, quality of life and effects on knowledge. *Scand Cardiovasc J* 2015; 49: 193–9. doi:10.3109/14017431.2015.1035319

56 Vuorinen AL, Leppänen J, Kaijanranta H, Kulju M, Heliö T, van Gils M, Lähteenmäki J. Use of home telemonitoring to support multidisciplinary care of heart failure patients in Finland: randomized controlled trial. *J Med Internet Res* 2014; 16: e282. doi:10.2196/jmir.3651

57 Johnston N, Bodegard J, Jerström S, Åkesson J, Brorsson H, Alfredsson J. Effects of interactive patient smartphone support app on drug adherence and lifestyle changes in myocardial infarction patients: a randomized study. *Am Heart J* 2016; 178: 85–94. doi:10.1016/j.ahj.2016.05.005

58 Varnfield M, Karunanithi M, Lee CK, Honeyman E, Arnold D, Ding H, Smith C, Walters DL. Smartphone-based home care model improved use of cardiac rehabilitation in postmyocardial infarction patients: results from a randomised controlled trial. *Heart* 2014; 100: 1770–9. doi:10.1136/heartjnl-2014-305783

59 Blasco A, Carmona M, Fernández-Lozano I, Salvador CH, Pascual M, Sagredo PG, Somolinos R, Muñoz A, García-López F, Escudier JM, Mingo S, Toquero J, Moñivas V, González MA, Fragua JA, López-Rodríguez F, Monteagudo JL, Alonso-Pulpón L. Evaluation of a telemedicine service for the secondary prevention of coronary artery disease. *J Cardiopulm Rehabil Prev* 2012; 32: 25–31. doi:10.1097/HCR.0b013e3182343aa7

60 Rincon E, Monteiro-Guerra F, Rivera-Romero O, Dorronzoro-Zubiete E, Sanchez-Bocanegra CL, Gabarron E. Mobile phone apps for quality of life and well-being assessment in breast and prostate cancer patients: systematic review. *JMIR Mhealth Uhealth* 2017; 5: e187. doi:10.2196/mhealth.8741

61 Sundberg K, Wengström Y, Blomberg K, Hälleberg-Nyman M, Frank C, Langius-Eklöf A. Early detection and management of symptoms using an interactive smartphone application (Interaktor) during radiotherapy for prostate cancer. *Support Care Cancer* 2017; 25: 2195–204. doi:10.1007/s00520-017-3625-8

62 Uhm KE, Yoo JS, Chung SH, Lee JD, Lee I, Kim JI, Lee SK, Nam SJ, Park YH, Lee JY, Hwang JH. Effects of exercise intervention in breast cancer patients: is mobile health (mHealth) with pedometer more effective than conventional program using brochure? *Breast Cancer Res Treat* 2017; 161: 443–52. doi:10.1007/s10549-016-4065-8

63 McCarroll ML, Armbruster S, Pohle-Krauza RJ, Lyzen AM, Min S, Nash DW, Roulette GD, Andrews SJ, von Gruenigen VE. Feasibility of a lifestyle intervention for overweight/obese endometrial and breast cancer survivors using an interactive mobile application. *Gynecol Oncol* 2015; 137: 508–15. doi:10.1016/j.ygyno.2014.12.025

64 Bateman DR, Srinivas B, Emmett TW, Schleyer TK, Holden RJ, Hendrie HC, Callahan CM. Categorizing health outcomes and efficacy of mHealth apps for persons with cognitive impairment: a systematic review. *J Med Internet Res* 2017; 19: e301. doi:10.2196/jmir.7814

65 Chan SC, Chan CC, Derbie AY, Hui I, Tan DG, Pang MY, Lau SCL, Fong KNK. Chinese calligraphy writing for augmenting attentional control and working memory of older adults at risk of mild cognitive impairment: a randomized controlled trial. *J Alzheimers Dis* 2017; 58: 735–46. doi:10.3233/JAD-170024

66 Mansbach WE, Mace RA, Clark KM. The efficacy of a computer-assisted cognitive rehabilitation program for patients with mild cognitive deficits: a pilot study. *Exp Aging Res* 2017; 43: 94–104. doi:10.1080/0361073X.2017.1258256

67 Han JW, Oh K, Yoo S, Kim E, Ahn K-H, Son Y-J, Kim TH, Chi YK, Kim KW. Development of the ubiquitous spaced retrieval-based memory advancement and rehabilitation training program. *Psychiatry Investig* 2014; 11: 52–8. doi:10.4306/pi.2014.11.1.52

68 Barnes DE, Yaffe K, Belfor N, Jagust WJ, DeCarli C, Reed BR, Kramer JH. Computer-based cognitive training for mild cognitive impairment: results from a pilot randomized, controlled trial. *Alzheimer Dis Assoc Disord* 2009; 23: 205–10. doi:10.1097/WAD.0b013e31819c6137

69 Finn M, McDonald S. Computerised cognitive training for older persons with mild cognitive impairment: a pilot study using a randomised controlled trial design. *Brain Impair* 2011; 12: 187–99. doi:10.1375/brim.12.3.187

70 Hattink BJ, Meiland FJ, Overmars-Marx T, de Boer M, Ebben PW, van Blanken M, Verhaeghe S, Stalpers-Croeze I, Jedlitschka A, Flick SE, van den Leeuw J, Karkowski I, Dröes RM. The electronic, personalizable Rosetta system for dementia care: exploring the user-friendliness, usefulness and impact. *Disabil Rehabil Assist Technol* 2016; 11: 61–71. doi:10.3109/17483107.2014.932022

71 Van der Ploeg ES, Eppingstall B, O'Connor DW. Internet video chat (Skype) family conversations as a treatment of agitation in nursing home residents with dementia. *Int Psychogeriatr* 2016; 28: 697–8. doi:10.1017/S1041610215001854

72 Ong SW, Jassal SV, Miller JA, Porter EC, Cafazzo JA, Seto E. Integrating a smartphone-based self-management system into usual care of advanced CKD. *Clin J Am Soc Nephrol* 2016; 11: 1054–62. doi:10.2215/CJN.10681015

73 Bender JL, Yue RY, To MJ, Deacken L, Jadad AR. 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: e287. doi:10.2196/jmir.2661

74 Huguet A, Rao S, McGrath PJ, Wozney L, Wheaton M, Conrod J, Rozario S. A systematic review of cognitive behavioral therapy and behavioral activation apps for depression. *PLoS One* 2016; 11: e0154248. doi:10.1371/journal.pone.0154248

75 Rosser BA, Eccleston C. Smartphone applications for pain management. *J Telemed Telecare* 2011; 17: 308–12. doi:10.1258/jtt.2011.101102

76 Brzan PP, Rotman E, Pajnkihar M, Klanjsek P. Mobile applications for control and self-management of diabetes: a systematic review. *J Med Syst* 2016; 40: 210. doi:10.1007/s10916-016-0564-8

77 Zhao J, Freeman B, Li M. Can mobile phone apps influence people's health behavior change? An evidence review. *J Med Internet Res* 2016; 18: e287. doi:10.2196/jmir.5692

78 Spring B, Gotsis M, Paiva A, Spruijt-Metz D. Healthy apps: mobile devices for continuous monitoring and intervention. *IEEE Pulse* 2013; 4: 34–40. doi:10.1109/MPUL.2013.2279620

79 Jake-Schoffman DE, Silfee VJ, Waring ME, Boudreaux ED, Sadasivam RS, Mullen SP, Carey JL, Hayes RB, Ding EY, Bennett GG, Pagoto SL. Methods for evaluating the content, usability, and efficacy of commercial mobile health apps. *JMIR Mhealth Uhealth* 2017; 5: e190. doi:10.2196/mhealth.8758

80 Handel MJ. mHealth (mobile health)-using apps for health and wellness. *Explore (NY)* 2011; 7: 256–61. doi:10.1016/j.explore.2011.04.011

81 Khoja S, Durrani H, Scott RE, Sajwani A, Piryani U. Conceptual framework for development of comprehensive e-health evaluation tool. *Telemed J E Health* 2013; 19: 48–53. doi:10.1089/tmj.2012.0073

82 mHIMSS App Usability Work Group. Selecting a mobile app: evaluating the usability of medical applications. Chicago, IL: Health Care Information Management Systems Society; 2012. Available at: https://www.himss.org/selecting-mobile-app-evaluating-usability-medical-applications-0 [verified 21 September 2018]

83 Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile apps rating scale: a new tool for assessing the quality of health mobile apps. *JMIR Mhealth Uhealth* 2015; 3: e27. doi:10.2196/mhealth.3422

84 Jin M, Kim J. Development and evaluation of an evaluation tool for healthcare smartphone applications. *Telemed J E Health* 2015; 21: 831–7. doi:10.1089/tmj.2014.0151

85 Chyjek K, Farag S, Chen KT. Rating pregnancy wheel applications using the APPLICATIONS scoring system. *Obstet Gynecol* 2015; 125: 1478–83. doi:10.1097/AOG.0000000000000842

86 Baumel A, Faber K, Mathur N, Kane JM, Muench F. Enlight: a comprehensive quality and therapeutic potential evaluation tool for mobile and web-based ehealth interventions. *J Med Internet Res* 2017; 19: e82. doi:10.2196/jmir.7270

87 McNiel P, McAthtur EC. Evaluating health mobile apps: information literacy in undergraduate and graduate nursing courses. *J Nurs Educ* 2016; 55: 480–2. doi:10.3928/01484834-20160715-12

88 Meurk C, Leung J, Hall W, Head BW, Whiteford H. Establishing and governing e-mental health care in Australia: a systematic review of challenges and a call for policy-focused research. *J Med Internet Res* 2016; 18: e10. doi:10.2196/jmir.4827

89 Maheu MM, Pulier ML, Roy S. Finding, evaluating, and using smartphone applications. In: Koocher GP, Norcross JC, Breene BA, editors. Psychologists' desk reference. 3rd edn. New York: Oxford University Press; 2013. pp. 704–8.

90 Wyatt JC, Thimbleby H, Rastall P, Hoogewerf J, Wooldridge D, Williams J. What makes a good clinical app? Introducing the RCP Health Informatics Unit checklist. *Clin Med (Lond)* 2015; 15: 519–21. doi:10.7861/clinmedicine.15-6-519

91 Anderson K, Burford O, Emmerton L. App Chronic disease checklist: protocol to evaluate mobile apps for chronic disease self-management. *JMIR Res Protoc* 2016; 5: e204. doi:10.2196/resprot.6194

92 Maheu MM, Nicolucci V, Pulier ML, Wall KM, Frye TJ, Hudlick E. The Interactive Mobile App Review Toolkit (IMART): a clinical practice-oriented system. [published erratum appears in: *J Technol Behav Sci* 2017; 2 (3-4):211] *J Technol Behav Sci* 2017; 1: 3–15. doi:10.1007/s41347-016-0005-z

93 The Royal Australian College of General Practitioners (RACGP). mHealth in general practice: a toolkit for effective and secure use of mobile technology. Melbourne: RACGP; 2016.

94 Yasini M, Beranger J, Desmarais P, Perez L, Marchand G. mHealth quality: aA process to seal the qualified mobile health apps. *Stud Health Technol Inform* 2016; 228: 205–9.

95 Dialogue Consulting. Guidelines for developing healthy living apps. Melbourne: VicHealth; 2015.

96 Eysenbach G. CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of web-based and mobile health interventions. *J Med Internet Res* 2011; 13: e126. doi:10.2196/jmir.1923

97 Agarwal S, LeFevre AE, Lee J, L'Engle K, Mehl G, Sinha C, Labrique A. Guidelines for reporting of health interventions using mobile phones: mobile health (mHealth) evidence reporting and assessment (mERA) checklist. *BMJ* 2016; 352: i1174. doi:10.1136/bmj.i1174

98 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. doi:10.1089/tmj.2014.0088

99 Chan S, Torous J, Hinton L, Yellowlees P. Towards a framework for evaluating mobile mental health apps. *Telemed J E Health* 2015; 21: 1038–41. doi:10.1089/tmj.2015.0002

100 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: 1051–9. doi:10.1016/j.amepre.2016.07.009

101 McKay FH, Cheng C, Wright A, Shill J, Stephens H, Uccellini M. Evaluating mobile phone applications for health behaviour change: a systematic review. *J Telemed Telecare* 2018; 24: 22–30. doi:10.1177/1357633X16673538

102 Boudreaux ED, Waring ME, Hayes RB, Sadasivam RS, Mullen S, Pagoto S. Evaluating and selecting mobile health apps: strategies for healthcare providers and healthcare organizations. *Transl Behav Med* 2014; 4: 363–71. doi:10.1007/s13142-014-0293-9

103 Collins LM, Murphy SA, Strecher V. The multiphase optimisation strategy (MOST) and the sequential multiple assignment randomiised trial (SMART): new methods for more potent eHealth interventions. *Am J Prev Med* 2007; 32: S112–8. doi:10.1016/j.amepre.2007.01.022

104 McDonald S, Quinn F, Vieira R, O'Brien N, White M, Johnston DW, Sniehotta FF. The state of the art and future opportunities for using longitudinal n-of-1 methods in health behaviour research: a systematic literature overview. *Health Psychol Rev* 2017; 11: 307–23. doi:10.1080/17437199.2017.1316672

105 Kaplan B. Evaluating informatics applications – some alternative approaches: theory, social interactionism, and call for methodological pluralism. *Int J Med Inform* 2001; 64: 39–56. doi:10.1016/S1386-5056(01)00184-8

106 Michie S, Johnston M. Theories and techniques of behaviour change: developing a cumulative science of behaviour change. *Health Psychol Rev* 2012; 6: 1–6. doi:10.1080/17437199.2012.654964

107 Payne HE, Lister C, West JH, Bernhardt JM. Behavioral functionality of mobile apps in health interventions: a systematic review of the literature. *JMIR Mhealth Uhealth* 2015; 3: e20. doi:10.2196/mhealth.3335

108 Bakker D, Kazantzis N, Rickwood D, Rickard N. Mental health smartphone apps: review and evidence-based recommendations for future developments. *JMIR Ment Health* 2016; 3: e7. doi:10.2196/mental.4984

109 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. doi:10.2196/jmir.1874

110 Pagoto S, Schneider K, Jojic M, DeBiasse M, Mann D. Evidence-based strategies in weight-loss mobile apps. *Am J Prev Med* 2013; 45: 576–82. doi:10.1016/j.amepre.2013.04.025

111 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. doi:10.2196/jmir.2968

112 Sarkar U, Gourley GI, Lyles CR, Tieu L, Clarity C, Newmark L, Singh K, Bates DW. Usability of commercially available mobile applications for diverse patients. *J Gen Intern Med* 2016; 31: 1417–26. doi:10.1007/s11606-016-3771-6

113 Roess A. The promise, growth, and reality of mobile health – another data-free zone. *N Engl J Med* 2017; 377: 2010–1. doi:10.1056/NEJMp1713180

Appendix 1. Evidence table for systematic reviews and individual trials of apps for chronic disease management

3D, three dimensional; A-CHESS, Addiction-Comprehensive Health Enhancement Support System; ACS, acute coronary syndrome; BDI, Beck's Depression Inventory; BMI, body mass index; BP, blood pressure; BSL, blood sugar level; CAP-CR, Care Assessment Platform of cardiac rehabilitation; CCQ, Clinical COPD Questionnaire; COPD, chronic obstructive pulmonary disease; DBP, diastolic blood pressure; EBAC, estimated blood alcohol concentration; ED, emergency department; EMAs, ecological momentary assessments; EQ5D, EuroQol 5D version; ESRD, end-stage renal disease; FACT-G, Functional Assessment of Cancer G version; GP, general practitioner; HADS, Hospital Anxiety and Depression Scale; HF, heart failure; HIS, home intervention system; HRQoL, health-related quality of life; IQR, interquartile range; IRR, incidence rate ratio; KDS, Kessler Psychological Distress Scale; LDL-C, low-density lipoprotein cholesterol; LVEF, left ventricular ejection fraction; m app, mobile health application; MD, mean difference; MI, myocardial infarction; NYHA, New York Heart Association; PANSS, positive and negative symptom scale; PEFR, peak expiratory flow rate; PHQ, Patient Health Questionnaire; QoL, quality of life; RCT, randomised controlled trial; RR, relative risk; SBP, systolic blood pressure; SDS, self-directed support; SF, Short Form; SGRQ, St George Respiratory Questionnaire; SMD, standardised mean difference; SR, systematic review; T1DM, type 1 diabetes mellitus; T2DM, type 2 diabetes mellitus; TCR, traditional, centre-based cardiac rehabilitation; TMG, telemonitoring group; WEL, Weight Efficacy Lifestyle Questionnaire

Reference	Type of studies, no. participants, risk of bias	App description	Measures of clinical efficacy	Self-efficacy adherence, contextual factors	Authors' interpretive comments
Asthma					
Hui *et al.*[25]	SR of 12 RCTs: 3 RCTs specific to m apps, $n = 513$, high to unclear risk of bias	Included RCTs incorporated 10 features grouped into 7 categories: education, monitoring/electronic diary, action plans, medication reminders/prompts, facilitating professional support, raising patient awareness of asthma control	In one RCT ($n = 136$),[26] more app patients achieved a well-controlled asthma score than the control group (49% vs 27%; $P < 0.05$). In a second RCT ($n = 288$),[27] no between-group difference in asthma control, QoL, asthma exacerbations, ED attendances, admissions, unscheduled GP consultations, or steroid courses In the third RCT ($n = 89$),[28] use of the app resulted in higher asthma-related QoL scores at 6-months follow-up (MD 5.50 (95% CI 1.48–9.52) for the physical component score of the SF-12 questionnaire; MD 6.00 (95% CI 2.51–9.49) for the mental component), improved lung function (PEFR) at 4, 5 and 6 months (MD (95% CI) 27.80 (4.51–51.09), 31.40 (8.51–54.29) and 39.20 (16.58–61.82), respectively) and reduced ED visits (OR 0.20; 95% CI 0.04–0.99)	No studies explicitly reported adoption of and adherence to the technology system No reference to behavioural change theories	Mobile apps, incorporating an action plan and other self-monitoring features, may be an effective option for supporting self-management Insufficient evidence to identify the important application features that attract and encourage patients to continue using it May have missed some contemporary studies Methodological limitations relating to full app descriptions, and included trials of interventions that included multiple features apart from apps
COPD					
McCabe *et al.*[29]	SR of 3 RCTs: 1 RCT specific to m apps, $n = 30$, high risk of bias	RCT[30] ($n = 30$) comprising 2 modules: activity coach and web portal for recording symptoms and activity levels Daily diary completion on web portal triggered decision support system in cases of exacerbation Activity coach comprised 3D accelerometer and smartphone with Bluetooth, both worn by participant	No difference in health related quality of life (CCQ and SGRQ) up to 6 months Improvement in daily step count (by 1000 steps) up to 4 months Effects on ED visits, hospitalisations, functional capacity, lung function or smoking cessation not reported	Self-efficacy (as measured by the COPD Self-Efficacy Scale or any validated instrument) was not reported Participant engagement not sustained after 6 months	No firm conclusions of efficacy can be drawn Study highlighted need for sustained engagement with the app over time Limited evidence suggests app is not harmful and may be more beneficial in those with interest in using technology

Diabetes					
Wu *et al.*[31]	SR of 17 RCTs, $n = 2225$, low risk of bias	16 trials involved interventions that actively engaged patients and included one or more of the following components: collecting biodata for healthcare providers, giving patient feedback from providers, supported drug titration (including self-titration), monitoring drug, diet, physical exercise compliance, goal setting, self-monitoring, reminders for medication intake or checking parameters of blood glucose and/or patient education	Significant reduction in HbA1c (pooled weighted MD –0.51%; 95% CI –0.71%, –0.30%; $P < 0.001$) favouring apps Pooled weighted MD was –0.83% in patients with T2DM <8.5 years and –0.22% in patients with T2DM ≥8.5 years, with a significant subgroup difference ($P = 0.007$). No subgroup differences were found among different follow-up durations, trial locations, patient age, healthcare provider contract time, baseline BMI and baseline HbA1c Very few hypoglycaemic events were reported	No measures of self-efficacy or adherence No reference to behavioural change theory	Evidence of publication bias Key longer-term outcomes, such as mortality, major stroke, loss of vision, amputation and ESRD, were not considered in included trials Six studies did not report baseline characteristics (e.g. baseline duration of T2DM), and patient compliance Only six trials had follow-up duration >6 months Unable to explore effect of different sex, race and background intervention (e.g. insulin vs oral medication) because of insufficient reporting among included trials Unable to determine which components are more effective
Wu *et al.*[32]	SR of 12 RCTs, $n = 974$, low or unclear risk of bias	Apps contained modules relating to medication management, lifestyle modification, complication prevention (smoking cessation and hypoglycaemia prevention) and psychosocial care	Meta-analysis of pooled data showed apps associated with significant reduction in HbA1c (MD 0.48%; 95% CI 0.19–0.77%) without excess adverse events Larger HbA1c reductions noted among patients with T2DM vs T1DM (MD (95% CI) 0.67% (0.30–1.03%) vs 0.36% (0.08–0.81%), respectively)	No measures of self-efficacy or adherence Greater HbA1c reductions noted with complication prevention module vs no prevention module (MD (95% CI) 1.31% (0.66–1.96%) vs 0.38% (0.09–0.67%); $P = 0.01$) and structured display vs no structured display (MD (95% CI) 0.69% (0.32–1.06%) vs 0.16% (0.16–0.48%); $P = 0.03$) Functions of clinical decision-making, personalised feedback, lifestyle modification or general education had no impact on HbA1c.	Exploratory and observational nature of subgroup analyses and possible misclassification precludes firm conclusions about modular efficacies and risks Funnel plot asymmetry suggests risk of publication bias
Fu *et al.*[35]	SR of 20 articles included: 7 unique usability studies from 8 reports, 12 unique clinical effectiveness studies (6 RCTs, 6 pre-post studies), high risk of bias	Most studies used a diabetes app in conjunction with a website, healthcare provider feedback or Bluetooth-enabled devices such as glucometers and blood pressure monitors Study reports lacked details of the mobile devices used	Clinical effectiveness: App use in all studies decreased HbA1c, ranging from 0.15 to 1.9% from baseline	Five studies did not report intervention adherence rates Most studies did not report name of the diabetes app and the study device (e.g. smartphone and tablet) Only one study reported a theoretical basis (Information-Motivation-Behavioural Skills Model); two studies applied the theoretical framework for an education class and disease management counselling Reductions in HbA1c were only statistically significant in four studies where the app design provided the greatest interactive features: instant app feedback messages and alert reminders from an algorithm database website Involvement of primary care provider or a dietician were also covariates	Limitations of clinical effectiveness studies: use of diabetes apps in combination with other interventions posed risk of confounding; only two studies reported power analysis; eight studies did not use any form of blinding Findings strongly suggest that efforts to improve user satisfaction, incorporate established principles of health behaviour change, and match apps to user characteristics will increase therapeutic impact

(*continued next page*)

Appendix 1. (*continued*)

Reference	Type of studies, no. participants, risk of bias	App description	Measures of clinical efficacy	Self-efficacy adherence, contextual factors	Authors' interpretive comments
Holmen *et al.*[36]	SR of 6 trials, $n = 431$, unclear risk of bias	Most apps had reminders, feedback messaging functions based on BSL and BP trends, and warnings when hypoglycaemic episodes occurred Feedback used was either automatic or manual; 3 apps alerted patients when they missed readings Tracking and imputation of medication, levels of physical activity and weight were functions in 3 apps	HbA1c was reported in 4 trials, with 2 reporting significant decrease in apps groups (–0.4% to –1.95%), 2 trials reporting no change between groups Change in blood pressure was reported inconsistently, with 1 trial reporting significant reduction in mean daytime ambulatory SBP in the intervention group; 3 trials reported no significant change in either SBP or DBP between groups; 2 trials did not report BP changes One trial reported significant increase in depressive symptoms using the HADS in the intervention group No trial reported lifestyle measures such as physical activity or dietary habits	Theoretical foundation and reference to clinical guidelines was largely lacking No significant differences between groups in diabetes knowledge Various usability and satisfaction scales were reported using non-validated and non-comparable questionnaires with inconsistent results	Overall methodological quality was low, with small samples, weak designs, lack of detailed description of comparison groups, poor reporting of study details
Hou *et al.* [33]	SR of 14 RCTs, $n = 1360$, medium to high risk of bias	Apps provided personalised feedback on self-monitoring data, such as BSL, food intake, physical activity In 8 apps, BSL was automatically transferred and other data manually entered 3 apps provided feedback when needed (e.g. patient data considered abnormal); in others, frequency of feedback ranged from once a week to once every 3 months	All 10 studies of T2DM reported a reduction in HbA1c in participants using an app, with a median reduction of 0.55% (range 0.15–1.87%) After pooling, mean HbA1c reduction was 0.49% (95% CI 0.30–0.68; $P = 0.01$) No HbA1c reductions seen in T1DM Subgroup analyses showed effects did not differ significantly by follow-up duration, mean diabetes duration, mean age, number of self-monitoring tasks or type of feedback	Behavioural mechanisms were not considered Subgroup analysis by follow-up duration suggested effect of diabetes apps on BSL control may attenuate over time, due possibly to lack of user friendliness, perceived additional benefits and use of gamification elements, resulting in lack of efficacy following use Younger than older patients more likely to benefit from app use, because more amenable to new technologies and more familiar with use of mobile phones	Publication bias cannot be ruled out Patient-important outcomes were not considered Some of the effect attributed to apps could be explained by healthcare providers No clear definition of diabetes apps, with authors defining their interventions in different ways
Cui *et al.*[34]	SR of 6 RCTs, $n = 1022$, low risk of bias	All apps comprised four parts including a mobile/smartphone with self-management apps, measuring devices, patients who uploaded the data to the apps and providers who analysed the data and provided feedback	Overall effect: MD in HbA1c –0.40% (95% CI –0.69, –0.11%; $P = 0.007$) with SMD of –0.40% (95% CI –0.69, –0.10%; $P = 0.008$) Subgroup analysis showed similar effect with –0.33% (95% CI –0.59 to –0.06%; $P = 0.02$) in MD and –0.38% (95% CI –0.71 to –0.05%; $P = 0.02$) in SMD in studies where patients' baseline HbA1c levels were <8.0% No significant reduction in HbA1c in those trials where app had no feedback function Patients with milder disease with baseline HbA1c <8% benefited more than those with higher baseline HbA1c No effects of app interventions on BP, serum lipids, weight, medication changes; one trial showed significantly increased exercise	No reference to self-efficacy or adherence No reference to behaviour change theory Increased accessibility to educational resources and self-management strategies, more frequent physical and emotional symptom tracking and increased access to peer support were seen as main strengths of apps in optimising self-management	Substantial heterogeneity between the studies was the main limitation of this meta-analysis Insufficient information to separate individual patients with better and worse glycaemic control

Chronic pain					
Irvine *et al.*[40]	RCT, $n = 597$, low risk of bias	Patients randomised to: mobile–web intervention called 'FitBack' intervention, alternative care group that received 8 emails urging participants to link to six Internet resources for back pain, and control group FitBack intervention designed to encourage users to adopt appropriate pain-prevention behaviours, tracking them against self-reported pain level during brief repeat interactions Users received weekly emails with gain-framed pain self-care messages and prompts to return to the FitBack program to track pain and self-care activities	At 4 months, control group 1.7-fold more likely to report back pain than FitBack Group; alternative care group 1.6-fold more likely to report back pain at 4-month follow-up Improved scores for pain control, QoL, work productivity, functionality	High rates of user satisfaction and patient activation; 96–98% understood the program Reference to social cognitive theory and theory of planned behaviour	Cannot gauge importance of email reminders on results, which potentially could influence response rate Only prompted treatment group if they did not open the first message, which may have biased the response rate Cannot verify participants provided accurate information on eligibility criteria, surveys, and 4-month follow-up period was limited Unable to tease out which aspects of FitBack were most effective Cannot determine whether social desirability bias may have influenced responses to assessment items
Kristjánsdóttir *et al.*[41]	RCT, $n = 140$, low risk of bias	After 4-week in-patient rehabilitation program, subjects randomised to smartphone intervention comprising 1 face-to-face session and 4 weeks of written communication via a smartphone Participants received 3 smartphone diary entries daily to support their awareness of and reflection on pain-related thoughts, feelings and activities Registered diaries immediately available to a therapist who submitted personalised written feedback daily based on cognitive behavioural principles Both intervention and control groups given access to a non-interactive website after discharge to promote constructive self-managemen	At 11-months, favourable between-group differences seen after intervention and at 5-month follow-up on catastrophising, acceptance, functioning and symptom level no longer evident More improvement in catastrophising scores during follow-up period in the intervention than control group (mean (± s.d.) –2.36 ± 8.41 vs 0.40 ± 7.20, respectively; $P = 0.045$)	Withdrawal rate of 30% Of those who completed the study, 86% agreed somewhat or totally that participation was useful No measure of app adherence Main theoretical framework based on cognitive behavioural fear-avoidance model and cognitive behavioural theory, and comprised, more specifically, elements from the acceptance and commitment therapy	Multicomponent intervention and unable to tease out which had most effect High withdrawal rate and response rate to assessment questionnaires <70% at follow-up Selection bias may also have had an impact (i.e. those with positive long-term effects may have elected not to participate in the study)

(*continued next page*)

Appendix 1. (*continued*)

Reference	Type of studies, no. participants, risk of bias	App description	Measures of clinical efficacy	Self-efficacy adherence, contextual factors	Authors' interpretive comments
Serious mental health disorders					
Batra *et al.*[42]	SR of 18 studies, 14 studies specific to m apps, 4 RCTs ($n=244$), 10 observational studies ($n=429$), high risk of bias	Apps collected patient data in real time In 5 studies, apps collected EMAs; in 3 studies, apps used for daily mood and symptom monitoring; in 1 study, app used skill assessment scale for assessing functional capacity; in 4 studies, app used to deliver psychosocial or behavioural therapy Frequently used format was completion of self-reports or questionnaires when prompted by an auditory signal generated by the app Questionnaires aimed to assess domains of mood, sleep, activity and symptoms	RCT[43] ($n=81$), app for depression delivering behavioural and mindfulness treatments had no effects on outcomes or recovery rates; subgroup analysis indicated behavioural activation had more favourable effect for patients with higher severity of depression, whereas mindfulness program more effective for mildly depressed patients RCT[44] ($n=35$): app delivering cognitive behavioural therapy in major depression improved PHQ-9, BDI-II and K-10 scores, with SDS scores showing significant reduction in number of days lost or unproductive Cohort study[45] ($n=95$): app delivering illness monitoring and management functions for mood or psychotic disorders was 'perceived to have positive outcomes on lives of participants' with no specific measures reported Cohort study[46] ($n=33$): app that monitored disease with feedback of clinical assessments in bipolar disorder was able to differentiate between Type 1 and Type 2 disorder, but no effect on outcomes Cohort study[47] ($n=32$): app for schizophrenia or schizoaffective disorder that prompted patients to complete thrice-daily mood assessments and offered advice based on results was associated with significant reduction in symptoms at study end in PANSS total, PANSS positive, PANSS general psychopathology and BDI-II and BDI-II scores All 10 observational studies except one pre-post study focused on disease monitoring only with no education, self-care or therapeutic interventions	Adherence rates varied between 68% and 100% Overall, studies indicated high feasibility and acceptability, but reference to behavioural theory or user-centred approach to app design was lacking in most studies	Apps seem to be feasible for short-term use in patients with serious mental illness; long-term effectiveness and adherence data from naturalistic studies will help demonstrate their usefulness and facilitate their adoption and integration into the mental health care systems Many studies involved patients with only mild to moderate illness severity, and most apps tested in pilot studies with small populations, with lack of standardised evaluation and reporting Patient characteristics (sex, education level) differed from one another and may not be representative of intended user populations
Grist *et al.*[48]	SR of 24 studies (2 RCTs, 22 observational studies) of 15 apps used in children and adolescents <18 years of age	Apps mostly focused on self-monitoring of mood and other symptoms across a range of disorders (depression, eating disorders being most prevalent), with skills training, goal setting and coping tactics also included in most	Only 2 studies reported efficacy outcome data RCT[49] ($n=11$): app for any mental health problem showed no effects on depression, anxiety, stress RCT[50] ($n=206$):app for eating disorders showed no effect on self-esteem or body satisfaction	Most studies indicated acceptability was generally positive with average to high ratings for ease of use, satisfaction, and usability Adherence rates ranged from 68% to 83% No reference to behavioural theory or principles	Methodological concerns about quality of research, small samples sizes, poor reporting of demographic data, recruitment of patients with mild to moderate severity of disease, uncertainty around validity of diagnosis, and no reference to parent or guardian involvement

Alcohol and substance abuse					
Meredith *et al.*[51]	SR of 5 studies: 3 RCTs specific form apps, $n = 1729$, low risk of bias	RCT[52] ($n = 1286$): randomising participants to Group 1 (app that offered real-time eBAC calculation), Group 2 (web-based app offering real-time eBAC calculation with planning and follow-up functions) and Group 3 (control) RCT[53] ($n = 349$): patients receiving residential treatment randomised to usual treatment or usual treatment plus an app with the A-CHESS, an application designed to improve continuing care for alcohol use disorders RCT[54] ($n = 94$): randomised to receive mobile app with 31 different self-help modules for 14 days, app with assessments (without intervention modules) for 14 days or minimal assessments by mobile phone	Per-protocol analyses at 7 weeks revealed only one significant time × group interaction, where Group 1 participants increased the frequency of their drinking occasions compared with controls ($P = 0.001$) Secondary analyses by gender showed a significant difference among men in Group 1 for frequency of drinking occasions per week ($P = 0.001$), but not among women After 8 months of the intervention and 4 months of follow-up, patients in the app group reported significantly fewer risky drinking days (>4 or >3 standard drinks per day for men or women respectively) than control patients, with a mean of 1.39 vs 2.75 days (MD 1.37; 95% CI 0.46–2.27; $P = 0.003$). At 4 weeks, among those randomised to the intervention app, receiving more modules of the intervention was significantly associated with a lower likelihood of any drinking during the 14-day assessment period, but no effect on reducing heavy episodic drinking	Self-reported app use was higher in Group 1 than Group 2 (74% vs 41% respectively) Attrition was 23–39%, higher among heavier drinkers and highest in Group 2 Apps were designed with reference to Theory of Planned Behaviour Theoretical basis of A-CHESS is self-determination theory, which posits that meeting 3 needs contributes to an individual's adaptive functioning: being perceived as competent, feeling related to others, and feeling internally motivated and not coerced in one's actions During the 8-month intervention period, patients randomised to the A-CHESS group used the system, on average, 41.1% of days (mean no. days of use, 100.2; median 103.0) and viewed a mean number of 1967 pages (median 1745 pages) Of the 170 patients who received A-CHESS, 122 (71.8%) pressed the panic button at least once By the end of month 4, ~80% of A-CHESS participants were still using the app No reference to behaviour change theory	Apps studied using eBAC calculation did not affect alcohol consumption among university students and one app may have led to a negative effect among men Future research should: (1) explore ways to increase user retention; (2) include apps facilitating technical manipulation for evaluation of added components; (3) explore the effects of adapting app content to possible gender differences; and (4) offer additional interventions to high-risk users Patients in the treatment group received a smartphone whereas those in the control group did not, and the application included a weekly self-assessment, possibly producing an assessment effect and more counsellor contact than for those in the control group The study involved only patient self-report, without urine testing, and each survey asked about drinking only in the past 30 days, which does not capture a complete picture of each patient's drinking and can underestimate or overestimated drinking behaviour Multifeatured smartphone application may have significant benefit to patients in continuing care for alcohol use disorders
Heart failure					
Hägglund *et al.*[55]	RCT, $n = 82$, low risk of bias	Recently discharged patients randomised to new HIS consisting of specialised software, a tablet computer wirelessly connected to a weight scale, containing information about HF and lifestyle advice according to current guidelines, present dose of diuretic, changes in patient-measured weight and HRQoL over time versus control group receiving usual care	After 3 months, intervention group showed fewer HF-related days in hospital (1.3 vs 3.5 days per patient; risk ratio 0.38; 95% CI 0.31–0.46; $P < 0.05$) and significant improvement in self-care (median (IQR) scale score 17 (13–22) vs 21 (17–25); $P < 0.05$), QoL as measured with the Kansas City Cardiomyopathy Questionnaire (median (IQR) score 65.1 (38.5–83.3) vs 52.1 (41.1–64.1); $P < 0.05$) and physical limitation (median (IQR) score 54.2 (37.7–83.3) vs 45.8 (25.0–54.2)l $P < 0.05$). No difference in knowledge scores	No reference to adherence or behaviour change theory	

(*continued next page*)

Appendix 1. (*continued*)

Reference	Type of studies, no. participants, risk of bias	App description	Measures of clinical efficacy	Self-efficacy adherence, contextual factors	Authors' interpretive comments
Vuorinen *et al.*[56]	RCT, $n = 94$, low risk of bias	Patients with LVEF <35% and NYHA functional class ≥2 were randomised to telemonitoring group who measured their bodyweight, BP and pulse, and answered symptom-related questions on a weekly basis, reporting their values to an HF nurse using a mobile phone app and who contacted patient with instructions if necessary, versus a control group that received multidisciplinary treatment according to standard practices	No difference was found in the number of HF-related hospital days (primary outcome) Intervention group used more healthcare resources due to increased number of nurse visits (IRR = 1.73, $P < 0.001$), spent more time at nurse reception (MD 48.7 min, $P < 0.001$) and greater number of telephone contacts with nurse (IRR = 3.82, $P < 0.001$ for nurse-induced contacts; IRR = 1.63, $P = 0.049$ for patient-induced contacts) Significantly more medication changes in intervention group ($P = 0.042$ for medication increase; $P = 0.026$ for medication decrease) No statistically significant differences in patients' clinical health status or self-care behaviour	Adherence, calculated as a proportion of weekly submitted self-measurements, was close to 90% 95% of patients found making and reporting measurements useful; automatic feedback received after sending measurements useful for 91% 91% of patients felt feedback motivated them to take measurements and report them regularly No reference to behavioural change theory	Intervention significantly increased nurse workload by increasing the number of reception visits and the number of telephone contacts Failure to reduce HF-related hospital days attributed to: control group receiving high standard interactive care; all patients lived a short distance from health services and were able to visit the clinic easily without great effort; study population relatively young; medications for all patients optimised during baseline visits; short follow-up period
Ischaemic heart disease					
Johnstone *et al.*[57]	RCT ($n = 174$), unclear risk of bias	Post-MI patients randomised to interactive patient support tool (active group) or a simplified tool (control group) in addition to usual post-MI care	At 6 months, greater patient-registered drug adherence achieved in active versus control group (non-adherence score: 16.6 vs 22.8; $P = 0.025$), with score defined as a combination of adherence failure events (2 missed doses registered in 7-day cycles) and treatment gaps (4 consecutive missed doses) No significant change in risk factors, smoking cessation, increased physical activity or QoL	Patient satisfaction significantly higher in active versus control group (system usability score 87.3 vs 78.1; $P = 0.001$)	No blinding of either observer or patient Number of pill counts was limited; objective validation of self-registered drug use was not possible; limited pill counts made correlation between e-diary registration and actual missed dose impossible to assess Study not powered to show differences in secondary end-points Prerequisite of patient inclusion was ownership of smartphone, skewing population to younger MI patients more likely to be early adopters of smartphones
Varnfield *et al.*[58]	RCT, $n = 120$, high risk of bias	Post-MI patients randomised to smartphone-based home service delivery (CAP-CR), including health and exercise monitoring, motivational and educational material delivery, and weekly mentoring consultations), or TCR	Both groups showed significant improvements in 6-min walk test from baseline to 6 weeks (TCR: 537 ± 86 to 584 ± 99 m; CAP-CR: 510 ± 77 to 570 ± 80 m), which was maintained at 6 months CAP-CR showed slight weight reduction (89 ± 20 to 88 ± 21 kg) and demonstrated significant improvements in emotional state (K10: median (IQR) 14.6 (13.4–16.0) to 12.6 (11.5–13.8)) and QoL (EQ5D index: median (IQR) 0.84 (0.8–0.9) to 0.92 (0.9–1.0)) at 6 weeks	CAP-CR had significantly higher uptake (80% vs 62%), adherence (94% vs 68%) and completion (80% vs 47%) rates than TCR ($P < 0.05$) CAP-CR reduced waiting time from referral to commencing cardiac rehabilitation by 2 weeks on average Difficulty using information technology tools was listed as a CAP-CR dropout reason (7%)	Limitations of small sample size and low power for functional capacity outcomes, particularly due to considerable dropouts Study focused only on patients referred to cardiac rehabilitation after MI and did not address all patients eligible for cardiac rehabilitation Willingness to be randomised to specific treatment strategy did impact on trial recruitment No blinding of patients to different treatment modes

Blasco *et al.*[59]	RCT, $n=203$, low risk of bias	Post-ACS patients with at least 1 risk factor randomised to TMG where patients sent, through mobile phones, using structured questionnaire, weight, heart rate, and BP weekly, and capillary plasma lipid profile and glucose monthly, which cardiologist accessed through a web interface and sent recommendations via SMS, versus control group of TCR	TMG patients were significantly more likely (RR 1.4; 95% CI 1.1–1.7) to experience improvement in cardiovascular risk factor profile than control patients (69.6% vs 50.5%; $P=0.010$) Significantly more TMG patients achieved treatment goals for BP (62.1% vs 42.9%; $P=0.012$) and, among diabetics, HbA1c (86.4% vs 54.2%; $P=0.018$) BMI was significantly lower in TMG (change in BMI –0.77 vs –0.29 kg/m^2; $P=0.005$) No differences in medication adherence rates, physical activity, smoking cessation, LDL-C, or QoL scores	98% completed more than 50% of questionnaire sessions (mean 89.2 ± 16.0% sessions completed) Only 0.5 messages per patient missed due to mobile phone being turned off No reference to behavioural change theory	Patients were not blinded to intervention and could have shared telemonitoring data with family physicians Small sample with duration of follow-up only 12 months Study not designed to analyse clinical events
Cancer					
Rincon *et al.*[60]	SR of 5 studies, 3 studies offering treatment, high risk of bias	Non-RCT[61] ($n=130$ patients receiving radiotherapy for prostate cancer): intervention group reported symptoms daily via app during treatment and 3 weeks after to a clinic nurse who viewed the report via web interface and, in case of an alert, contacted patient by telephone to discuss and manage symptoms; the EORTC QLQ-C30 and its module PR25 and the Sense of Coherence questionnaire were administered at three time points in both groups RCT[62] ($n=356$): all patients instructed to perform a 12-week regimen of aerobic and resistance exercise; intervention group received pedometer and newly developed app to provide information and monitor prescribed exercises; control group received exercise brochure Before–after study[63] ($n=50$ endometrial and breast cancer survivors): exercise and nutrition counselling using app into which entry and exit quality of life (FACT-G) and WEL	Intervention group rated significantly lower levels of fatigue and nausea at end of radiotherapy, and had significantly less burden in emotional functioning, insomnia and urinary-related symptoms at treatment end and at 3 months In multivariate analyses, with education and sense of coherence as covariates, intervention group still significantly rated emotional functioning ($P=0.007$), insomnia ($P=0.017$), and urinary-related symptoms ($P=0.008$) better than control group Physical function, physical activity and QoL scores were significantly improved regardless of the intervention method, and changes were not significantly different between groups Significant reductions in before and after groups in weight (105.0 ± 21.8 vs 98.6 ± 22.5 kg), BMI (34.9 ± 8.7 vs 33.9 ± 8.4 kg/m^2), waist circumference (108.1 ± 14.9 vs 103.7 ± 15.1 cm; $P<0.001$ for all) and WEL score (99.38 ± 41.8 vs 120.19 ± 47.1; $P=0.043$) No significant differences in FACT-G, macronutrient consumption and physical activity patterns	No reference to quality certification or user comments regarding its use Mean Likert scale response for overall satisfaction with app was 4.27/5	Limitations relate to: (1) no cancer-focused apps are being used in studies involving cancer patients; (2) many of the studies had small samples; (3) studies were without rigorous design based on RCTs; (4) studies were not free to the user; (5) no theoretical framework reported; (6) usability and accessibility issues were not assessed; (7) little evidence on patient satisfaction level; (8) little assessment of QoL and/or well-being assessment relevant to oncological settings

(*continued next page*)

Appendix 1. (*continued*)

Reference	Type of studies, no. participants, risk of bias	App description	Measures of clinical efficacy	Self-efficacy adherence, contextual factors	Authors' interpretive comments
		data, along with anthropometrics, daily food intake, and physical activity were entered			
Cognitive impairment					
Bateman *et al.*[64]	SR of 24 studies, 7 RCTs, 2 non-RCTs, 15 pre-post pilot studies, $n = 576$, high risk of bias	App functions categorised as cognitive training and serious games, wandering and wayfinding, reminiscence therapy, prompts and multicomponent interventions, engagement interventions, exercise interventions Common quantitative health outcomes included cognition, function, mood and QoL	14 studies (58%) reported efficacy outcomes For cognitive training, 3 RCTs and one controlled trial showed improvement in cognition, whereas one RCT did not RCT[65] of serious games coupled with cognitive training showed no improvement in cognition, anxiety or depression, although visual attention was improved RCT[68] of multicomponent interventions showed no effect on cognition, autonomy, on caregiver needs RCT[69] of engagement showed no effects on agitation Pilot studies of various interventions were mostly negative studies[66,67,70,71]	Large number of studies did not take end users' (people with dementia or caregivers) input into consideration during the development of apps, creating potential mismatch between proposed solution and participant needs No reference to behavioural change theory	Small samples and poor study designs do not allow for any firm conclusions as to efficacy, although cognitive training seems to exert an effect Need for greater agreement on study design and health outcome measures
Chronic kidney disease					
Ong *et al.*[72]	Cohort study, $n = 47$	App targeted four behavioural elements: monitoring BP, medication management, symptom assessment and tracking laboratory results Prebuilt customisable algorithms provided real-time personalised patient feedback and alerts to providers when predefined treatment thresholds were crossed or critical changes occurred	Statistically significant mean reductions in home BP readings between baseline and exit (SBP, 23.4 mmHg (95% CI 25.0–21.8 mmHg); DBP, 22.1 mmHg (95% CI 22.9–21.2 mmHg)); 27% with normal clinic BP readings had newly identified masked hypertension 127 medication discrepancies were identified; 59% were medication errors that required an intervention to prevent harm	User adherence was high (80% performed >80% of recommended assessments) and sustained In exit interviews, patients indicated feeling more confident and in control of their condition; clinicians perceived patients to be better informed and more engaged Reference to Chronic Care Model as underpinning behavioural change	App developed using an iterative, user-centred design method to ensure an exceptional user experience Automatic transfer of BP readings circumvented the pitfalls of manual entry, and user exit interviews suggested that healthcare providers had more confidence in the values transmitted Only mobile application for out-patient use that supports patient-driven medication reconciliation with a feedback loop to correct errors Study limited by the lack of a comparison group to control for secular trends and possibility of selection bias