Royal College
of Physicians

Setting higher standards

Rapid literature review for National Information Board workstream 1.2

Final report

version 0.97h

August 2015

Edited by Jeremy Wyatt, based on contributions by Leo Chant, Jan Hoogewerf, Tom Lewis, Shiva Sathanandam, Harold Thimbleby and Darren Wooldridge

Owner	Jeremy Wyatt		
Author(s)	Leo Chant Jan Hoogewerf Tom Lewis Shiva Sathanandam Harold Thimbleby Darren Wooldridge Jeremy Wyatt	**Version**	0.96 for review
Version Date	15-7-2015	**Status**	Draft

Document Control

Version	Date	Reviewed By	Amendment History
0.93	18-6-15		
0.94	24-6-15		
0.95	14-7-15		
0.96	15-7-15		

Contents

Executive summary

Our literature review covers digital healthcare tools including apps or web services for the public or patients, apps for clinicians and personal health records used by both parties. The main aspects considered include factors determining user acceptance and uptake rates, usability and accuracy of the technology, and the technology's impact on clinical outcomes and usage of healthcare resources. Potential assessment methods, several "digital healthcare myths" and privacy issues are also discussed.

Patient and public use of apps: Summary of evidence on acceptability, usability and impact (section 3)

There is some evidence from surveys about the factors that make apps more acceptable to patients. There is also some evidence about usage rates for apps in specific areas such as diabetes, but there is no study that allows us to predict the likely influence of the proposed national app assessment programme on patient / public app download or usage rates.

The usability of apps for patients remains very variable, with many examples of poor usability. In addition, there are several studies showing poor accuracy of apps that attempt to make a diagnosis (eg. of melanoma from images of pigmented skin lesions) or predict risks from clinical features (eg. of cardiac risk prediction for patients).

There is some evidence from randomised trials (RCTs) about the positive impact of apps on patient decisions and actions, including behaviour change. For example, one RCT showed significantly increased weight loss over 6 months in users of the MyMealMate app. There is also some evidence of the modest benefit of patients using apps on clinical outcomes, eg. on HbA1C in diabetics. However, there is no rigorous evidence about the impact of patients using apps on their utilisation of health service resources.

Further research is needed on:

- The impact of patients using apps on their utilisation of health service resources.

- The likely influence of the proposed national app assessment programme on patient / public app download or usage rates

- How to improve the usability and accuracy of apps intended for public use in potentially safety critical scenarios

Clinician use of apps: Summary of evidence on acceptability, usability and impact (section 4)

There is some concern about the accuracy of apps intended to be used by clinicians, for example for drug dose conversion or calculation of risk scores. This may be explained by the low rates of engaging clinicians in the design and testing of apps. The recent move by the Royal College of Physicians to remind clinicians that they should only use CE marked apps for such purposes and the proposed national assessment process, emphasizing testing of the accuracy of such apps before use by clinicians, is likely to cause app developers to engage more clinicians in the development and testing process.

We have obtained early results of a recent internet survey of app use by 1104 physician respondents to the RCP Panel survey (panel size 2658, response rate 42%). Overall, 54% of respondents use apps to support their clinical work (especially to look up guidelines for risk assessment and to inform

prescribing); and rated these as either "Very important" (29%) or Essential" (13%) to their clinical work. The biggest concerns about clinical apps are about the quality of the underlying evidence base on which the app is based (43%) and the accuracy of information generated (43%); information governance is also a concern (25%). The majority of respondents (73%) felt that the RCP should be responsible for recommending apps that are of high quality and safe to use; many respondents also expected specialist societies (69%) or the NHS (50%) to play a role in app recommendation.

However, despite these promising results, there is clear evidence that the design and usability of apps for clinicians – even those that have passed CE marking – fail to take account of well-known human factors issues. We give several examples for well-known CE marked apps, with detailed critiques of screen shots.

We were unable to locate any studies of the impact of app use by clinicians on the quality and safety of care nor on efficiency / resource utilisation. More studies are therefore needed on:

- The factors that influence app use by clinicians
- The accuracy and reliability of apps intended for use by clinicians
- The impact of app usage by clinicians on patient outcomes and healthcare resource utilisation.

Personal health records (PHRs): Summary of evidence on acceptability and impact (section 5)

The evidence on patient decisions and actions indicates better adherence to medical advice, enhanced empowerment and better communication with the use of PHRs. This evidence is largely based on single studies of acceptable quality. Problems with web portal, technology availability and poor socio-economic and educational status are seen as barriers.

The evidence on behavioural change is more robust and includes large systematic reviews across different chronic conditions. Significant positive behavioural changes, e.g. smoking cessation, increased physical activity, mood improvement in cancer patients and better health promotion are seen with the use of PHRs. The effects are largely seen in the short-term, with studies reporting significantly decreasing effects with time.

The largest volume of evidence on PHRs is on patient clinical outcomes. Various positive outcomes have been outlined - including significant reduction in HbA1c levels in Diabetic patients, improvement in daily functional effect and moderate improvements in quality of life. Although occasional studies report no effect, the vast majority are positive, especially for HbA1c reduction. The reports on other outcomes such as cholesterol improvement and blood pressure control have not shown any significant benefits.

Evidence on the cost of running PHRs / cost effectiveness as well as the decisions and actions taken by clinicians is scarce and we are unable to provide conclusive statements on these measures.

Younger patients were found to be more likely to access PHRs and many studies have reported the access of laboratory test results as the most useful feature of the PHRs. Patient satisfaction was generally high across the majority of the studies and individual studies have identified a list of barriers to usage. These frequently included poor technology access, elderly and low educational status.

The utilisation of healthcare services has often been reported as a positive outcome with small individual studies showing evidence of reduction in admission rates and emergency visits. However a large US based study has found that there could be significant increase in patient visits, telephone calls and hospitalisations.

More evidence is needed on:

- The impact of PHR use on patient outcomes and resource utilisation in the UK

- Which PHR functions are most useful and contribute to clinical and efficiency impacts

- How to encourage more clinicians to engage in patient centred PHRs, eg. to respond to patient messages (in a US study, only 14% of health professionals used the patient PHR at least once a day).

Summary of other findings and insights

To support the health and care system in developing robust methods for assuring quality and assessing apps and related products, we provide:

- A table listing nine potential methods for assuring the quality of apps and web delivered services, the likely implications for the health and care system of adopting each of these, and the pros and cons of each (our work, no published evidence – section 6.1)

- The results of a recent survey of EU stakeholders on their preferences about app assessment and quality improvement methods (section 6.1)

- A taxonomy or ontology of apps etc. that we developed to identify those factors most useful to the organisation running an assessment process, and to the users of assessed products (section 6.2)

- Our proposals for a risk-related evaluation strategy (section 6.3)

- Our analysis of a number of "digital healthcare myths" (section 6.4)

- A summary (section 7.2) of how the results of the literature review and other evidence can inform 11 key features of the proposed NIB 1.2 app assessment process

- We also comment on the variable quality of empirical studies in digital healthcare, and give suggestions for how these might be improved (section 7.3)

- A list of 9 key unanswered research questions that could form the basis of a specific NIHR funding call (section 7.4)

- A bibliography of over 150 literature references (see reference section) as well as 30 relevant web links in the text

1. Introduction and background to the review

1.1. The variable quality of health apps, personal health records and related tools

Digital healthcare or eHealth offers great potential to promote greater public engagement in health, improve efficiency and open up new care pathways and models of care [Wyatt 2005]. A huge number and variety of apps and web-delivered services such as personal health records (PHRs) are now available to patients, the public and health professionals to promote health, prevent illness, promote self-management and support health and clinical care. These digital health tools cover everything from "serious" games to records, administrative applications and clinical decision support. In England there is an NHS health apps library (http://apps.nhs.uk/), which includes apps which have been reviewed by the NHS to ensure they are clinically safe and relevant to people living in England. The apps are also rated by users.

It is widely agreed, however, that the quality of these apps and services is very variable, sometimes dangerously so. For example, industry, academic and policy delegates at a recent EU stakeholder meeting on eHealth in Riga were asked to vote on the main issue regarding the quality and safety of lifestyle and wellbeing apps [source: http://ec.europa.eu/digital-agenda/en/news/mhealth-green-paper-next-steps]. The results are shown in the graph.
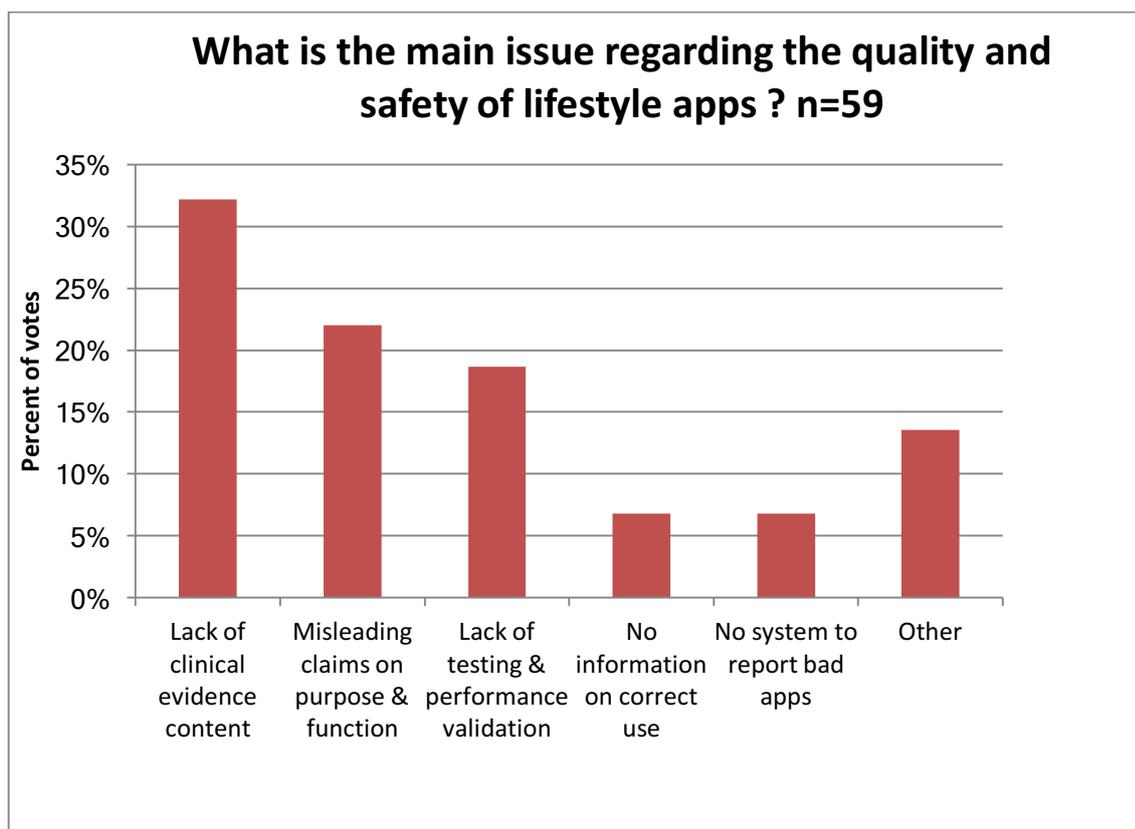
Another issue is the sheer volume and exponential growth of lifestyle and health apps, meaning that it is practically impossible to assess each and every app [van Velsen 2013]. As a result of these concerns about app quality and also disquiet over the low uptake of digital health tools for other reasons, including low awareness and lack of patient and professional incentives, the National Information Board (NIB) workstream 1.2 is working on the development of a process for assessing these, initially focussing on apps.

1.2. Review aims and scope

The literature review was commissioned by the Health and Social Care Information Centre (HSCIC) in collaboration with NHS England, the National Institute for Health and Care Excellence (NICE) and Public Health England (PHE) as part of the National Information Board (NIB) workstream 1.2: providing citizens with access to an endorsed set of NHS and social care apps. Despite rapid development in new technologies and approaches it is perceived that health and social care are lagging behind other domains. The aim of this review is to inform the work of the NIB 1.2 core group in developing a practical, sustainable and robust method for assessing or endorsing apps and related products. For the purposes of this report, a mobile app means any software application created for or used on a mobile/computer device for lifestyle, medical or other health-related purposes.

The aim of the study as set out in the original brief was to answer the question "What does the literature tell us about the current adoption and future potential benefits of health and Wellbeing Applications and Personal Health Records to enable patient centred care and reduction in the cost of care delivery?"

In summary the scope is:

- Technologies of interest & user groups:

    Apps used by the public and patients (Section 3) and apps used by health professionals (Section 4)
    Web based services for patients (eg. Big White Wall)
    Personal electronic health records for patients (Section 5)
- Aspects of interest:

    o Acceptance, adoption & usage rate; enablers of and barriers to this (Sections 3.1, 4.1, 5.1 and 5.2)
    o Usability, and factors that determine this (Sections 3.2 and 4.2)
    o Benefits and impacts of use for users and the health system (Sections 3.4, 4.3 and 5.4)
    o Safety issues and privacy
    o Additional issues, eg. The pros and cons of linking apps to electronic health records (Section 6).
- Focus areas of relevance as set in Appendix 3

- Methods to measure and improve the quality and impact of digital health tools, with pros and cons of these (Section 6.1)

The study was constrained to a ten week time period between 20 March and 23 May 2015. The study was also limited to 70 researcher and research assistant days of effort.

2. Methods

2.1. Methods used to search & synthesise the literature

We carried out a rapid literature review and synthesis of the best available evidence to explore the above aspects for each of the technologies of interest using the following methods (details given in Appendix 4):

1. Selection of appropriate research databases to use for the study

2. Developed appropriate initial search terms for the population, intervention, outcome and study methods

3. Carried out iterative testing and refinement of the initial search terms in-house

4. Disseminated the search results to out-worker team for review and feedback on search terms

5. Mapped the number and type of studies found in tables listing aspects of the technology versus disease focus areas for each of the 3 relevant combinations of technology and user (apps etc. for the public and patients; apps etc. for clinicians; PHRs for patients)

6. Retrieved PDFs of the systematic reviews (SRs) and other high quality articles relevant to each aspect (for some questions / aspects, may not always be Randomised Controlled Trials (RCTs) – see table in Appendix 4) for further appraisal and summarisation

7. Where necessary, reviewed reference lists in retrieved articles and carried out further searches (eg. using PubMed's Clinical Queries)

8. Summarised in an accessible form the methods and results of the most recent or relevant SRs and/or selected high quality articles, where these added useful insights.

In addition, several web searches were carried out to supplement literature searches or to obtain material unlikely to be published (eg. annual uptake of apps by diabetics).

2.1.1 Types of studies sought

For this review we needed to answer a variety of questions about the technologies ranging from the attributes that appear to guide user choice to usage rates, usability, accuracy and impact on users or health systems. We therefore matched the type of study sought to the question, as recommended by Sackett & Wennberg [1997]. The table below shows this.

Review question	Preferred study: a systematic review of the following study types, or a single well conducted, large study
What is the need for or requirement for the technology?	Surveys, qualitative studies (ie. interviews, focus groups etc.); formal systems analysis / business process modelling work
What are the user attitudes to or perceptions of the technology?	Surveys, qualitative studies
What are the usage rates for the technology?	Analysis of log file data; *possibly* surveys of eligible users

Review question	Preferred study: a systematic review of the following study types, or a single well conducted, large study
Is the technology usable?	Formal usability studies; user centred design workshops; task analysis, eye tracking studies etc.; checklist-based assessment of the system using a reputable checklist of desirable system attributes
Is there risk of cyberdivide for this technology?	Assessment of technology usability / usage rates by different age or sensory limitation groups in the four focus areas
Is the app / PHR advice or output accurate?	Comparison of the advice or output against a robust gold standard obtained for a prospective cohort or using simulated cases, preferably with data entered by typical users
What is the quality of the data captured or shared?	Analysis of the accuracy & completeness of data against a reliable gold standard source
What are the benefits or impact of the technology on clinical outcomes, knowledge about disease, self-efficacy (empowerment), drug adherence, health related behaviours, NHS resource usage, etc.?	Randomised trial; *possibly* a controlled before-after or interrupted time series study if no RCTs. Exclude simple before after and other study types.
What are the safety implications or risks posed by the technology?	Analysis of adverse incidents or near misses using root cause analysis; analysis of the accuracy of system output or advice against a reliable gold standard
What are the privacy risks?	Checklist based assessment of privacy risks / threats and controls
Does the technology show value for money, is it cost effective?	Formal cost effectiveness, cost utility or cost consequence analysis

Note that for questions about the impact of a technology, we consider RCTs to be the gold standard [Liu & Wyatt 2011]. Liu and Wyatt [2011] provide a comprehensive deconstruction of the nine main arguments against RCTs, with 115 references and a table listing RCTs of atypical interventions such as service dogs, intercessory prayer and educational visits. RCTs are increasingly being adopted by international funders outside healthcare such as the Gates Foundation and DfID, who support RCTs on interventions such as modest payments to families in low & middle income countries.

Also note that more reliable results are always obtained from well conducted systematic reviews than from individual studies, and that it is possible to carry out SRs of any type of study design including qualitative studies, not just RCTs. We located several SRs of qualitative studies, surveys etc. in our searches. Sackett DL, Wennberg JE. Choosing the best research design for each question **BMJ** 1997; 315: 1636

3. Review results on apps and web services used by patients and the public

3.1. Patient acceptability and adoption rates of apps

Acceptability is only the first stage of technology adoption. If the idea of the technology is not acceptable, people will not sign up for, let alone use the technology. This section reviews studies of app acceptability; the next section discusses studies of adoption, which for apps consists of downloading the app and then using it often and for long enough to make a difference to the user's health.

3.1.1. Acceptability of apps and factors that influence this

Generic issues influencing the acceptability of apps identified from the literature include:

- Physical considerations
  - There is a need to consider the physical limitations of mobile devices and their utility. Issues such as screen size and usability have been identified as important factors for elderly patients.

- Software considerations
  - There is a need to consider the software limitations of mobile apps including usability issues, theories of behaviour change, the fact that patients often use multiple apps (eg. glucose app *and* diet management app). There can be information overload or problems with poorly designed displays (Wyatt 1998) which pose issues for clinicians.

- Financial barriers
  - There is a need to consider the business case for use of apps to promote widespread adoption including evidence for remuneration and cost-effectiveness of integrated apps.
  - Consideration should be given to the cost of mobile apps for patients and integration into existing CCG models eg. Big White Wall.

- Infrastructure limitations
  - Currently, apps do not integrate well with existing EHRs/NHS IT systems and therefore any information stored within an app often needs to be manually transferred to a system.
  - Issues related to management of 'Big Data'.

- Education issues/digital literacy
  - It is important to consider the 'digital divide', in particular, the individual needs of an ageing population with regards to smartphones and mobile technology. This includes ensuring elderly users understand how to use the technology effectively and appropriately.
  - There is a need to consider how to effectively educate clinicians and other healthcare professionals as to how to optimise use of mobile apps in healthcare.

- Lack of adequate regulation
  - The regulatory stance adopted by the Medicines and Healthcare Products Regulatory Agency (MHRA) only covers a very small subsection of all health apps available leaving

many potentially unsafe apps in the public domain. Please see the assessment / quality improvement section for more information on this (Lewis and Wyatt, 2014).

- o Patients have expressed numerous privacy concerns regarding mobile apps.

- Patient Safety

  - o Numerous safety issues have been highlighted in health apps used by clinicians. This includes apps that provide out of date information, incorrect algorithms and inaccurate drug dosage information. Please see app quality section for more information on this area.

- Social considerations

  - o Consideration needs to be given to the impact that patient consumer groups can have with regards to promoting specific apps. Engagement with these groups could positively inform app developers. Surveys of patient consumer groups have shown that app selection amongst patients is guided and influenced by social media, word of mouth, and local recommendations.

3.1.2. Adoption rates: condition specific examples

Diabetes apps:

The growth in the rate of publishing diabetes apps has been exponential. In 2009 there were 60 apps available, in 2011 there were 260 and by 2013 there were 650 [Arnhold 2014]. 54% of the apps offered a single function, usually documentation of factors relevant to diabetes (diet, exercise, medication) or results of blood sugar testing. The graph shows this, and also a trend towards fewer free apps. The price of two thirds of the paid-for apps lay in the price range 1 cent to $3 US. There is a suggestion of a drop off in the recent diabetes app publication rate.
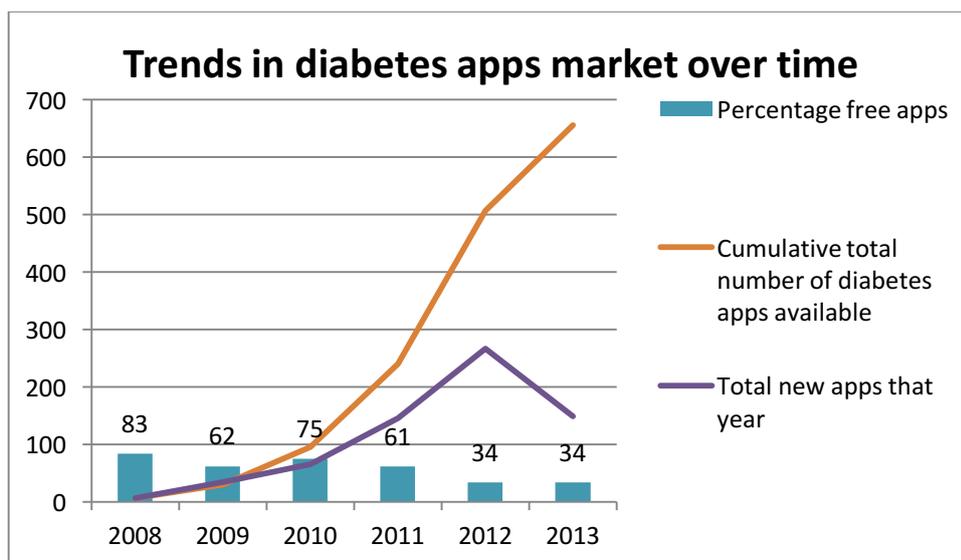
Approximately 1100 diabetes apps are now available for major operating systems, an increase from the 700 in 2013. One study [Research2guidance 2013] estimated in 2013 that 1.2% of smartphone

owners with diabetes were using a mobile app to help them manage their condition. This is expected to rise to approximately 8% by 2018 (14ompute. 24 million globally). We suspect that many survey respondents had T2 diabetes rather than the younger patients with T1, for whom rates seem likely to be be higher.

Global penetration of diabetes app users within the target group

There are at least two acceptability issues for apps identified from various sources in the literature that are specific to diabetes:

* Lack of automated data entry / smartphone compatible blood glucose monitoring devices

* The need for multiple types of apps to manage the condition eg carb counter, glycaemic control

Diabetes App usage rates

So far we have found few studies of app usage rates, and our attempt to obtain this data from NHS Choices failed as they do not collect it. A survey of 639 patients with diabetes (40% with type 1, 50% with type 2, 10% parents or carers of people with diabetes) by DiabetesMine in 2013 on the frequency of app usage by diabetics who responded to the survey is shown in the graph:

Frequency with which diabetics use smartphone software or apps

- Never
- Often - once a month or more
- Sometimes - every few months
- Only before a doctor's appointment
- Rarely - a few times a year
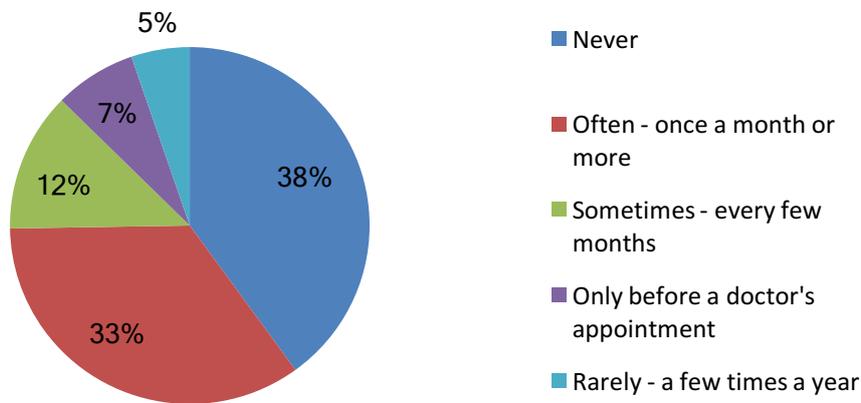
It is notable that 38% of survey respondents did not use apps at all. This is much lower than the 94% of diabetics in the Research2 study 2013 who did not use apps, this is probably because the DiabetesMine survey was carried out on an internet site. The next graph shows the types of app used by the diabetics who were app users:

Type of app used by diabetics who use apps
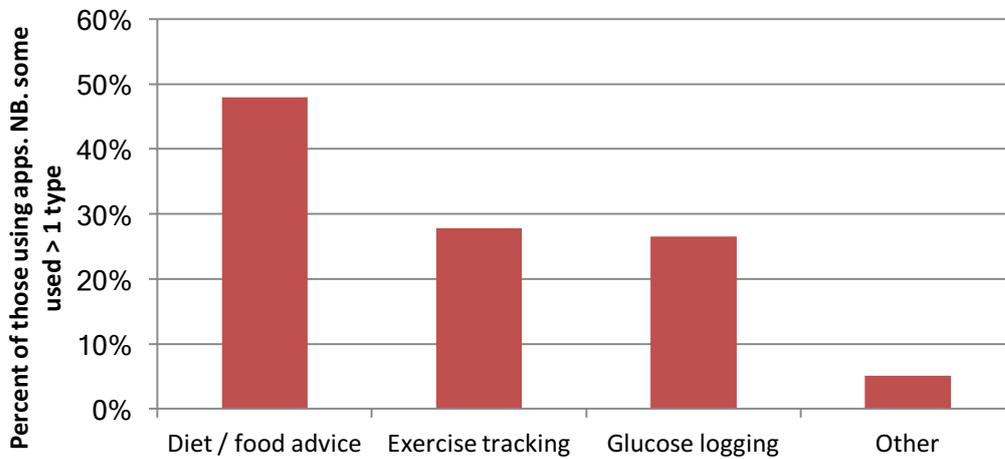
Note that some diabetics used more than one kind of apps, so the percentages add up to more than 100%.

Smoking cessation apps

Usage issues specific to smoking cessation apps include:

- Smoking cessation apps that follow evidence based guidelines and theories are often least downloaded [Abroms 2011, Abroms 2013]. A further study corroborates that many patients do not check the credibility of information before downloading.

- While not about smoking cessation *per se*, one study evaluated various digital media for patients with chronic obstructive pulmonary disease (COPD – a smoking related disease) and identified the netbook as the preferred device, due to its good controllability, fast response time, and large screen size [Cheung 2013].

- Many users download multiple smoking cessation apps however often do not continue to engage with these apps [BinDhim et al 2014].

Asthma apps

One search in 2011 identified 84 apps for asthma which increased to 241 in 2013 [http://www.acutemedicine.org.uk/wp-content/uploads/2014/10/6.2-How-Good-are-Medical-Apps-for-Patient-Care.pdf].

Dementia apps

Analysis of dementia apps in the literature identifies two main usage scenarios [Sposaro 2010]:
- Mobile apps to support carers
- Mobile apps to support family and patients with dementia eg medication adherence apps, support groups, patient education

Apps specifically designed for patients with dementia include: patient education, apps to help coordinate care, location tracking apps. Apps not specifically designed for patients with dementia but which may be of value include brain training games, music, relaxation, and memory activities.

No high quality evidence was found to identify how many or how often mobile apps to support patients with dementia are being used by either patients or caregivers.

Mood Monitoring apps

One survey of 100 psychiatric outpatients showed that 97% of patients reported owning a phone and 72% reported that their phone was a smartphone. Patients in all age groups indicated greater than 50% interest in using a mobile application on a daily basis to monitor their mental health condition [Torous 2014]. Interest in utilising mobile applications to help track and monitor mental health conditions varies with 67% in the aforementioned study versus 76% in an Australian study [Proudfoot 2010].

A further study by the Torous et al surveyed 320 psychiatric outpatients from four clinics across a geographically and socioeconomically diverse population and found an overall patient interest in utilising smartphones to monitor symptoms was 70.6% (226/320) [Torous 2014].

No surveys have been identified that measure how many apps are used or how often they are used to support mood monitoring by patients.

3.1.3. General factors that influence patient choice of apps

A pilot study of 250 worldwide patient and consumer groups by MyHealthApps in 2013 identified 5 key themes [https://alexwyke.files.wordpress.com/2013/10/health-app-white-paper-to-go.pdf]:

- Give people more control over their condition, or keep them healthy (35%)

- Be easy to use (26%)

- Be trustworthy (25%)

- Allow networking with other people like them, or with people who understand them (11%)

- Be able to be used regularly (3%).

These results may be skewed by the fact that 81% of respondents came from the UK. A second more globally representative study by PatientView in 2014 [PatientView/Health 2.0/Ticbiomed] surveyed 1130 disparate patient and consumer groups including groups for patients with a long term condition (or carers) to identify the qualities patients seek in health apps, these can be seen in the table below.

Provide trustworthy, accurate information	69%
Be easy to use/simple/well-designed	66%
Provide guarantees that my personal data is secure	62%
Be free	56%
Contain no advertisements	51%
Work effectively and consistently over time	44%
Not be expensive to buy, and provide value for money	28%
Allow me to network with other people important to me	26%
Be packed with detail (I don't mind complex apps)	23%

The same study also identified a range of common reasons why people may *not* use health apps:
- 37% Sheer number is confusing

- 32% Not sure if they will help

- 31% Prefer face-to-face contact with doctor/nurse

- 30% Unsure about which apps are relevant to patient

- 27% Unsure as to who produces the apps

- 27% App not used by my doctor/nurse

- 17% Too expensive

- 8% Not trustworthy/reliable

- 3% Not in my language.

Five studies reported on perceived barriers, whereas no studies reported on perceived facilitators of the use of interactive, Web-based interventions. Perceived barriers were typically of a technical nature including:

- problems with Internet connection

- slow loading of website

- security concerns

- discomfort with using the computer or Internet

- problems with related hardware (eg, PDA, monitor).

In a systematic review (SR) of 18 studies of web interventions to support people with long term conditions [Kuijpers et al, 2013], ten studies were identified that described general users' experiences, for example, satisfaction scores and a judgment of intervention content. In general, patient satisfaction was high. The personalised nature of the interventions was often cited by participants as being important. In one study [Ross et al 2004], nurses and physicians reported that their workload did not increase as a result of the intervention.

It also seems likely that social media or word of mouth recommendations will also influence choice of apps, but we found no studies mentioning this.

3.2. Usability of apps designed for patients and the public

According to the International Standard 9241 (see references for more details), usability is a combination of three key factors: user satisfaction, effectiveness and efficiency. Usability is seen within a process of iterative design: systems are not usable as such, but they can be made more usable. ISO 9241 thus emphasizes usability as a process of continual improvement.

3.2.1. ISO 9241: The standard on usability and its misunderstandings

ISO 9241 as a whole covers most aspects of usability, from accessibility to forms and haptic interaction. ISO 9241-210 in particular is the part of the standard specifying human-centred design. The standard strongly argues that to design usable systems, the user must be at the centre of the entire design process from initial conception onwards, that the process is iterative, and must involve a design team that includes multidisciplinary skills and perspectives. ISO 9241-210 spells out the advantages of human-centred design:

    a)   increasing the productivity of users and the operational efficiency of organizations;

    b)   being easier to understand and use, thus reducing training and support costs;

    c)   increasing usability for people with a wider range of capabilities and thus increasing accessibility;

    d)   improving user experience;

    e)   reducing discomfort and stress;

    f)   providing a competitive advantage, for example by improving brand image;

g)   contributing towards sustainability objectives.

ISO 9241 is the gold standard for benchmarking the usability of apps and the literature on app development and evaluation. Unfortunately, in contrast to ISO 9241, the literature reviewed shows widespread use of what might be called "DIY usability," a term we introduce for the popular simplification of usability.

Consistent with what undergraduate textbooks on usability [e.g., Shneiderman, *et al*, 2013] teach, DIY usability causes problems for several key reasons:

- *Everybody has an opinion on what is usable*. We use our opinions every time we go shopping or show off our latest app finds; our usability opinions are strongly held and deeply embedded in our sense of self. Grimes (2008, 2009) is notable for suggesting that it is self-evident that apps are better than older technologies (such as pencil and paper drug dose calculations), a claim that is contradicted by experiments [Thimbleby & Williams, 2013].

- *What we experience as usable is not the same as whether it is usable for other people*. In particular, for most of us our dominant experience of apps is in consumer applications, where our patterns of work are very different from clinical work. We can deceive ourselves that apps are nice because they feel nice, but that is different from whether they empirically perform effectively.
- *Users are rarely properly involved until a design is finished*, and then only in the nominal sense of evaluating the finished product. Users should be involved all along to helping improve the design at every step, though this creates up-front costs and questions (just when there is a rush into development). As often happens with clinical apps, where a few people (e.g., a developer and a clinician work together), the trap is to think the single user — an expert in the design and its background — is representative of users in general. None of the papers reviewed have any evidence of initial user engagement.
- *Our review shows that research papers on health apps often use naïve usability methods*, particularly when health apps are published in clinical journals. One infers that the peer review barriers for publishing on apps are very low, and hence much of the research literature misleads developers further.

Not all the app research papers we reviewed involved mature usability studies. Those that did fell into one or more common traps:

- *They frequently confused usability for satisfaction alone*. Whether users like apps is not all there is to usability, as per ISO 9241.
- *They frequently performed summative usability studies*. Users may like the app, but this measure has not been calibrated. Under ISO 9241, you evaluate usability, then improve the app, then measure again. Thus, if, as is often the case in the literature, the single usability study has found no identifiable problems, how reliable a measure is it? None of the studies we reviewed had (or attempted to have) external validity.
- *They frequently had weak experimental methods and weak statistical analysis*. For example, the developers of the app are the same people who perform the experiment. They ignored the Hawthorne Effect, evaluator effect, blinding, none used factorial designs. They over-tested and in some cases achieved unwarranted confidence levels.[1] Cairns (2007) documents a range of

---

[1] One app paper published in a peer reviewed clinical journal achieved $p < 0.0001$, when comparable tests in the mature peer reviewed usability literature (e.g., Association for Computing Machinery Computer-Human Interaction ACM CHI conference) typically report only $p < 0.05$. Unfortunately,

problems in the general usability literature, and the app usability literature as sampled for this report seems unaware of these critical methodological issues.

- *Ad hoc usability measures*. Papers that assessed usability used *ad hoc* approaches, which were not validated. (There is also a methodological bias that an *ad hoc* usability assessment will be biased to favour the app under test.) There are many standard usability assessment tools available that provide results that can be compared against known standards.

Usability seems a self-evident concept and users have strong views about their feelings and experience of using apps. Our review of the medical app literature shows there is little mature usability research that goes beyond this level of DIY usability. The antidote is a proper understanding of usability. Indeed the international standard ISO 9241 defines the rigorous concept: it defines *user experience* (abbreviated UX) as a person's perceptions and responses that result from the use or anticipated use of a product, system or service. This is the mature version of DIY usability but it is just one factor of usability.

Generally the more important concern is whether an app is clinically effective; users would be expected to be trained and skilled in the application area, and many UX features (e.g., animations) may be distractions from effective work performance, indeed they would influence user's self-reporting of user experience. Many papers we reviewed that performed usability evaluations relied on user questionnaires alone, and therefore must be considered only initial studies.

Although UX must not be seen as all there is to usability, UX is still critical to apps. For example, UX issues determine drop out: even dedicated users will drop off using an app they do not feel they like over time, which may be counter-productive if the app is clinically effective. Unfortunately our literature review found no evidence of longitudinal studies which would reveal such usability effects. Low *et al*, 2011, is notable for mentioning (but not studying) longitudinal issues, as well as blinding, statistical power etc.

Low *et al*, 2011, again is notable for mentioning experimental effects, such as avoiding clinician embarrassment that may arise in normal use and clinicians' general lack of familiarity with cognitive aids, factors that may undermine the validity of the experimental results achieved in simulation environments.

3.2.2. Heads down and heads up user experience

Robinson *et al* (2014) is a very recent and authoritative book on app UX — it is provocative because it thoughtfully argues that most apps are dull and merely imitate PC screen applications, rather than going right in to the user experience with a mobile handheld. Robinson *et al* make the very useful distinction between "heads down" interaction — as a clinician sitting at a patient record system might work — and "heads up" — as a clinician walking around using an app must be. In heads up apps, the clinician is seen as part of a team, part of an activity, the app is supporting. Additionally, they point out that many of the usability guidelines that are in widespread use [Nielsen, 1994, being a good example; Shneiderman *et al*, 2013; Preece *et al*, 2015], and others are also well-known texts) are outdated by the new opportunities apps open up. It is thus interesting to see most usability

---

insufficient details were published to check the paper's statistics (despite, e.g., American Psychological Association (APA) standards for reporting statistics); in this case over-testing is the simplest explanation.

evaluations in the literature framed in "heads down style" Nielsen/Shneiderman/Preece terms rather than in Robinson *et al* terms.

3.2.3. User Centred Design (UCD)

If the developers follow the standards recommendations, the design process must be a multi-stage problem-solving process, during which they analyse actual user behaviour, and further test the validity of their assumptions with real users doing realistic tasks — in the case of clinical apps, real clinicians and real or simulated patients. A well-designed clinical app will not force the users to change the way they work just to accommodate its function (needing workarounds) — and it is well known that workaround can only be discovered by empirical investigation. In general, the approach is called the User Centred Design (UCD).

UCD processes consist of the *analysis phase*, where the designer studies the user, the *design phase*, where the designer create a system prototype, test its usability and generate system specification, the *implementation phase*, where the designer works closely with developers to bring the concept into reality, iteration, followed by the *deployment phase*, where the designer studies how the system is used in the real world. To achieve effectiveness and efficiency, a cohort of clinicians with domain specific knowledge and experience must be actively involved in the iterative design process for considerable amount of time. Recruiting the right participants to contribute to the design, and maintaining this relationship throughout the process can be challenging. These resource challenges can be addressed by adopting *participatory design strategies*, which support the combination of both local and remote collaboration (which takes advantage of the internet to recruit participants in evaluation).

However, to app developers, complying with these standards is admittedly time consuming and costly — and may seem counter-productive when the clinical need and the app usability is "obvious."

3.2.4. Cost effective UCD: Discount usability versus RCTs

Usability is expensive to evaluate thoroughly. Usability is therefore often measured using "discount usability" techniques [Nielsen, 1994], where a small sample of participants (eg., n=5) are used to assess UX on the basis that it is more cost-effective to study small samples. Indeed, if iterative design is used (as ISO 9241 requires) initial results from small n would be used to improve the app, as larger studies are wasting effort evaluating the obsolete design (much as clinical trials need protocols for adapting to patient outcomes). However, small n is inappropriate for summative evaluation (other than pilot studies), which is what is generally presented in the literature.

Although UX is important for user acceptance, it is not all there is to usability. Even if people like them (good UX), clinical apps cannot survive on UX alone – UX does not emphasise the rare but critical problems that can arise in real use. Apps must also be effective over the long term, efficient and 'bug-free' — both in their clinical recommendations or calculations or in their user interface design — in the first instance. Are they safe? Are they clinically reliable? Do they reduce or otherwise manager user error? Does their use induce new sorts of error or incident that are not immediately apparent in a quick UX trial? It is beyond the scope of this report to expand on these issues, except to note that the ISO standard 62508:2010 provides good guidance on human aspects of dependability and should be referred to for developing safe and effective clinical systems. (There are of course many more standards also highly relevant to health apps, particularly ISO 14971, but

beyond the scope of the review undertaken here. Note that ISO standards typically contain useful bibliographies and other background material — they are not just standards, but useful learning material.)

There is a widely-held view that accepted techniques used in pharmaceuticals, such as randomized controlled trials, are inappropriate for health apps. Generally, RCTs take a long time, and app technology is advancing very fast: so performing RCTs (or similarly rigorous trials) could delay the adoption of important technologies. However, this argument is flawed [Liu & Wyatt 2012]. First, large RCTs do not just evaluate drug effectiveness, they also help detect low incidence side-effects. Do apps have side effects? No study to date has been large enough to find out (nor, to our knowledge, did any seek to detect *problems* with app usability). Secondly, RCTs are not the only way to perform rigorous evaluation, and as a methodology in fact they may not be the best way to evaluate most health apps, so they are just a straw man in the common argument. Thirdly, several apps *have* been evaluated in RCTs [eg., Carter, *et al*, 2013; Irvine, *et al*, 2015; Low, *et al*, 2011].

We naturally expect pharmaceuticals to be manufactured in sterile conditions, so we at least can be sure we do not get infections as a direct side effect of using them. There are analogous techniques available for app development, comparable to asepsis, which guarantee apps are free from bugs — indeed, even the word *bug* is the same! These methods include human factors engineering, software engineering, and formal methods. We note that these methods, let alone using state of the art methods, are notably absent from all apps reviewed here. Until rigorous development methods are used (and regulated as such) then there will be no guarantees apps will be dependable for their intended use.

Only in exceptional cases do we need to test pharmaceuticals for microbial contamination, and it would certainly be impractical to test everything routinely. Instead, pharmaceutical companies use state of the art techniques to avoid contamination, and they avoid obviously sloppy production methods. In app development, unfortunately, the world has a strong culture of "anyone can program" — as indeed they can — just as "anyone can make a pill" yet while we might encourage more clinicians to program we would not expect them to distribute home-made pills made in their kitchens. Why then do we think app development is different or exempt from the professionalism, rigor and quality control expected of any other medical intervention?

We may look forward to when app developers use state of the art techniques to avoid software bugs. This has yet to happen, and will remain unachievable while software warranties and business models (and regulatory frameworks) provide few incentives to improve quality. For example, the high-profile Mersey Burns app says in its warranty that the user indemnifies the NHS. This is surely the reverse of the business incentives we need for safer apps?

The preceding discussion sounds harsh, but is consistent with leading research (e.g., http://www.chi-med.ac.uk). On the contrary: there are standard techniques to produce dependable apps; we need to see wider uptake of them.

For apps to become more usable, in the full sense of that word, including more dependable for clinical use, we would like to see:

- *More informed usability evaluation*. Fortunately there are plenty of good textbooks on usability; we recommend these: [MacKenzie, 2013; Cairns & Cox, 2008]; and relevant ISO standards, such as 9241, which also provide substantial bibliographies). It should be noted that many textbooks

on usability are undergraduate overviews (of a vast subject) and are inadequate on their own for informing professional app development.

- *More informed software development*. Uability techniques assume the software conforms to requirements, and a specification, and that it can be modified reliably, but in our review of apps (for more details see below) we found numerous software bugs that would have been avoided by competent software development. It is too easy to write programs; but it is very hard to write *good* ones. Fortunately there are plenty of textbooks on software engineering; we recommend these: (Knight, 2012; Sommerville, 2015; Thimbleby, 2008).

None of the papers we reviewed here appear to have adopted either approach in their app development or published papers.

3.2.5. Usability case study 1: Peek Eye app

The Peek eye testing app is designed for low-cost eye tests, and in testing on 300 participants had comparable performance to much more costly and far less mobile eye testing equipment [Bastawrous, *et al*, 2015]. The user interface to Peek does not resemble conventional equipment, and its success seems to be a testament to applying new thinking in design appropriate to handheld apps. Note, however, that Bastawrous did not evaluate usability, and (from that point of view) their study has some minor methodological problems: for instance, the same experimenters performed app and conventional eye tests on the same participants. The apparent high level of expertise with the app might reflect good training rather than usability for users in the field, and the repeat tests would be expected to give a higher correlation than, say, a blinded approach. As a Class I device, Peek is currently applying for CE registration.

Innovative health apps are ideal for achieving widespread clinical benefits.

The picture shows the Peek eye testing app in use in the Masai Mara — achieving comparable results to the usual expensive and heavy equipment, which would be impractical to take to the patient.

Copyright: Peek, 2015. Used with permission.

3.2.6. Usability case study 2: Children's Interactive Art in Therapy

The Children's Interactive Art in Therapy (CIAT) project aims to improve parents' involvement in children's therapy. CIAT has produced an app that supports the therapist, the parent and the disabled child work together to improve the child's fine motor skills in handwriting and drawing. CIAT adopted the recommended iterative design process fully, and with very little cost. The work was led by Dr Karen Li, http://www.karenyli.me

The app developers conducted three focus group meetings with a small group of therapist representatives working in a children's centre. Between each meeting, there was a 6 week exploration phase, where the developers sent probing artefacts to all therapists to assess and post back. The artefacts were redesigned several times during the process, iteratively developed from the findings from the previous meetings and data collected during the exploration phases. Thematic analysis revealed several possible design questions, which were then evaluated with the therapists in further meetings.

Once the design questions were finalized, the developers went through several rounds of iteratively refining online interactive prototypes. The prototypes had ability to gather the user comments on the design in high fidelity. Since this process was online, the iterative cycle was repeated rapidly and with low cost, adhering to ISO 9241.

The outcomes of this iterative process are three integrated apps for the child, the parent, and the therapist respectively. The child uses their app to practice and assess their handwriting and drawing skills by completing a drawing. The parent uses their app to select a task for their child from a task library, based on the therapist's recommendation, and they can adjust the difficulty level of that task within the therapist's recommended range. The therapist can then review the parent's report alongside the child's performance data over time. The therapist can then prescribe suitable exercise and assessment remotely, which the parent and child can then pursue on the app.

3.3. Quality of apps designed for patients and the public

This section identifies evidence relating to the quality of apps. It uses two measures of quality; the quality of the evidence underpinning the apps and the accuracy of the data collected by the apps.

Quality of the evidence underpinning apps.

An assessment of asthma apps [Huckvale et al 2015] showed that between 2011 and 2013 the number of asthma apps more than doubled from 93 to 191, despite the withdrawal of 25% (n = 23/93) of existing apps. Newer apps were no more likely than those available in 2011 to include high quality information, such as the use of action plans, or to offer guidance consistent with evidence; 13% (n = 19/147) of all apps, and 39% (n = 9/23) of those intended to manage acute asthma, recommended self-care procedures unsupported by evidence. Despite increases in the

numbers of apps targeting specific skills such as acute asthma management (n = 12 to 23) or inhaler technique (from n = 2 to 12), the proportion consistent with guidelines (17%, n = 4/23) and inhaler instructions (25%, n = 3/12), respectively, was low, and most apps provided only either basic information about asthma (50%, n = 75/147) or simple diary functions (24%, n = 36/147). In Abroms 2013 study of 47 iPhone apps for smoking cessation [Abroms 2013], the median score for adherence to US Preventive Service Task Force evidence-based guidelines was only 13.5 (21%) on a scale from 0 to 64; for the dozen apps in which Abroms had also scored adherence to evidence 3 years earlier, the median score dropped slightly from 14.5 to 14. This suggests poor use of underpinning evidence or reference to national guidelines during app development.

Accuracy of data collected by apps to identify exacerbations of long term conditions:

One 2013 study of 50 participants [Johnston et al 2013] successfully showed a smartphone based collection system for exacerbations of COPD enabled near complete identification of exacerbations at inception. This could potentially be used to identify and initiate treatment earlier with a view to reducing hospital length of stay. Participants transmitted 99.9% of 28,514 possible daily diaries. All 191 COPD exacerbations meeting Anthonisen criteria (2.5 exacerbations/participant-year) were detected.

Mood monitoring: one 2014 validation study from Denmark [Faurholt-Jepsen 2014] compared smartphone self-monitoring in 17 patients with bipolar disorder over a period of 3 months. This showed that objective phone-captured data such as physical and social activity corresponded closely to clinically rated depressive symptoms. Self-rated depressive symptoms using the app also correlated well with Hamilton Depression Rating Scale item scores.

3.4 Benefits of apps used by patients and the public

Considerations when reviewing studies of benefit and impact of apps

RCTs are rightly the "gold standard" for assessing the impact of health apps on quality (eg. for patient-centred care, costs of care delivery, impact on resource use, cost effectiveness) and patient outcomes, as they address unknown variation across patients. This is important as for example, when assessing the impact of apps to support self-management of blood glucose levels, diabetic patients may be expected to have variation in their vision that will affect the usability of apps designed for them. As the results of more high quality RCTs of mobile apps become available, we will be able to conduct SRs and meta-analyses to identify the overall impact of apps for patients and the public.

The aim is to review the best evidence relating to the beneficial or harmful impact of the technologies on patient or clinician decisions, actions or behaviours and health systems (eg. cost, drug or service usage) for both mobile apps and app-like web services.

Outcomes of interest when evaluating impact and benefits of apps

When considering the impact of a digital health intervention such as a mobile app it is important to consider two broad categories of impact measure (Friedman & Wyatt 2005). These are:

H. Direct impacts on system users (clinicians or patients), or on the patients of clinician users, eg.:

- Clinical decisions, eg. diagnosis, test interpretation, prognosis
- Clinical actions, prescribing, referral, test ordering, hospital admission, discharge, counselling, outpatient review
- Patient decisions and actions: self-management (eg., knowledge or confidence about the disease and how to manage it; self-efficacy); drug dose adjustment (eg. in diabetes, asthma); drug adherence; picking up prescribed drugs; self-referral; keeping appointments; usage of health services
- Behavioural change, eg. smoking cessation, exercise (preferably objectively measured by pedometer, not self-report), weight reduction, alcohol intake, dietary improvement (eg. 5 a day), hazardous behaviours (eg. drunk driving, risky sexual encounters)
- Patient outcomes or surrogate outcomes: risk score (eg. Qrisk2 for cardiovascular disease); clinical outcomes and surrogates (eg. blood pressure (BP) for stroke, HbA1C for diabetes), disease progression or complication rates; drug/surgery side effects; quality of life; length of (disease free) survival; disability rates; mortality rate.

2. Impacts on health services and health systems, eg.

- Service utilisation eg. A&E/emergency room, outpatients, inpatient admission, length of stay; DNA rates for appointments/investigations
- Costs of running the service per patient/encounter
- Drug or investigation costs; cost effectiveness of services or health technologies (eg. PHR might improve communication & thus make it more cost effective to run a home dialysis service)
- Rates of accidents, HIV/STD transmission, adverse events or near misses.

Hierarchy of relevant study designs:

In assessing the evidence on impact, we make use of Sackett's well-known hierarchy of study designs [Friedman & Wyatt 2014]:

1. Systematic review of randomised trials
2. Large well designed trials
3. Smaller trials or trials with defects (eg. less than 80% follow up; allocation concealment unclear)
4. Other study designs: well conducted interrupted time series or controlled before and after studies (see Cochrane EPOC guidance on critical appraisal of these)
5. Uncontrolled studies, expert opinion and reasoning from first principles

3.4.1. Apps for Diabetes

There are a number of SRs of diabetes mobile apps available from 2010 to 2015 across multiple online vendors [Chomutare 2011, Arnhold 2014, Eng 2013, Brooke 2013, Demidowich 2012, Rao 2010,El-gayer 2013].

A 2011 meta-analysis of 1657 pooled patients from 22 RCTs [Liang et al 2011] showed that mobile phone interventions for diabetes self-management reduced HbA1c values by a mean of 0.5% [6 mmol/mol; 95% confidence interval, 0.3-0.7% (4-8 mmol/mol)] over a median of 6 months follow-up duration. In subgroup analysis, 11 studies among Type 2 diabetes patients reported significantly greater reduction in HbA(1c) than studies among Type 1 diabetes patients [0.8% (9 mmol/mol) vs. 0.3% (3 mmol/mol); p=0.02]. The effect of mobile phone intervention did not significantly differ by other participant characteristics or intervention strategies

A 2015 SR [Hunt 2015] focusing on diabetes self-management identified 14 studies focusing on apps and internet based services. Fourteen studies including qualitative, quasi-experimental, and RCT designs were identified and included in the review. The review found that technological interventions had positive impacts on diabetes outcomes including improvements in haemoglobin A1C levels, diabetes self-management behaviours, and diabetes self-efficacy. Of the six mobile phone intervention studies, three were qualitative and three were RCTs. Participants in the qualitative studies generally reported positive outcomes from using the mobile phone intervention. Participants appreciated the personalized feedback and education received from the intervention. Participants in RCTs using a mobile phone intervention noted improvements in HbA1C levels. Overall, mobile phone interventions had small sample sizes making generalization of study findings difficult.

Internet interventions identified in the Hunt SR include education, goal-setting, tracking of behaviours, patient feedback and support. Of the eight internet studies reviewed, seven were RCTs and the remaining study had a quasi-experimental design. All studies that measured changes in HbA1C levels noted improvements and all improvements were significant with one exception. In two of the studies, short-term improvements were noted in HbA1C, but not at the second, long-term follow-up. Several studies noted improvements in outcomes in both intervention and control groups

A 2012 SR [Holtz 2012] identified 21 articles published between 2000 and 2010. Seventy-one percent of the studies used a study-specific application, which had supplemental features in addition to text messaging. The outcomes assessed varied considerably across studies, but some positive trends were noted, such as improved self-efficacy (a psychological construct about a person's confidence that they can improve their health, believed to be a reliable predictor of behaviour change), haemoglobin A1c, and self-management behaviours. The studies evaluated showed promise in using mobile phones to help people with diabetes manage their condition effectively. However, many of these studies lacked sufficient sample sizes or intervention lengths to determine whether the results might be clinically or statistically significant.

One RCT published in 2011 [Quinn 2011] evaluated the effectiveness of the WellDoc system, a patient-coaching and provider clinical decision support system. The multimodal tool enables patients to wirelessly upload blood glucose readings and other diabetes-related information, and receive real-time feedback either via the health care provider (HCP), caregiver or WellDoc research team. In a 1-year cluster-randomized clinical trial, the intervention group's A1c decreased by 1.9% compared to the usual care group that decreased by 0.7%. This is one of the longest RCTs identified investigating a mobile app.

A 2014 SR [Cotter 2014] identified 9 papers focusing on internet interventions to support lifestyle modification for diabetes management. Two studies demonstrated improvements in diet and/or physical activity and two studies demonstrated improvements in glycemic control comparing web-based intervention with control. At 6 month follow up, 1 paper reported significant improvement in A1C in the web based treatment arm compared to the usual-care arm (p<0.05), significance decreased slightly in intent-to-treat analyses (p<0.06). The other study reporting a significant result showed that the web-based intervention group showed a 2.18 point decrease in HbA1c versus 0.9 in the usual care control group (p<0.05). However the study was small (n=74) with limited retention (63%) and a high baseline A1C of 8.9. Of the 4 studies that measured BMI, only one reported a significant change in weight. Four of the 9 studies measured blood pressure but none demonstrated a decrease when comparing web-based intervention arm with the control. Similarly, of the 5 studies that measured participants' lipid levels none reported a significant improvement compared with control. Successful studies were theory-based, included interactive components with tracking and personalized feedback, and provided opportunities for peer support. Website utilization declined over time in all studies that reported on it. Two studies focused on high risk, underserved populations.

One SR from 2012 [Yu et al 2012] aimed to evaluate any association between effectiveness and "clinical usefulness", "sustainability" or "usability" of web-based diabetes self-management tools. The authors defined a clinically useful tool as one that provides clinically useful answers and is easy to use, access, and read. Sustainability was defined as the degree to which an innovation continues to be used after initial effort to secure adoption is completed, which is a critical component in addressing the gap between research and practice, yet is often not addressed or assessed. Usability of the tool was defined as the extent to which a product can be used by specified users to complete tasks successfully, in time, and with satisfaction in a specified context. This may be underemphasized in research studies, where participants are routinely oriented to and trained on the use of the tool. In this SR, 57 studies were reviewed (25 RCT, 1 CCT, 14 before-after studies, 17 observational studies) and the authors found moderate but inconsistent improvements in a variety of psychological and clinical outcomes including HbA1c and weight. Surprisingly, meta-regression of adequately reported studies (12 studies, 2731 participants) demonstrated that, although the interventions studied resulted in positive outcomes, effectiveness was not associated with clinical usefulness nor usability. This may be due to limited power of the meta regression to identify small but useful association, or to the fact that most of web based tools included were clinically useful, sustainable and usable, using the author's definitions.

One 2014 SR focused on the use of mobile phones for weight loss in the general public; (included here because of the relevance of weight gain to the risk of developing diabetes) [Aguilar-Martinez 2014]. A total of 10 studies were identified on obese or overweight adults. The mean body mass index (BMI) of the subjects varied from 22 to 36 kg/m^2. Two studies used text messaging or multimedia messaging. All the other studies used mobile-phone apps or web-based programmes that could be accessed from mobile phones as a part of a weight-loss intervention or for evaluating their potential for use and their acceptance. Most studies lasted 2-4 months and the maximum duration was 1 year. All but two studies showed reductions in the participants' bodyweight, BMI, waist circumference and body fat in the various interventions. There appeared to be a proportional relationship between weight loss and programme use. The programmes most benefited those who took a pro-active approach to everyday problems. Frequent self-recording of weight seemed to be important, as was the personalisation of the intervention (counselling and individualized feedback). Finally, a social support system acted as a motivational tool.

A number of other SRs identified in the literature that showed benefits of technology for the effective management of diabetes were excluded as they were not deemed to be sufficiently recent or sufficiently focus on apps/app-like web services [eg. Russell-Minda 2009; Connelly 2013].

3.4.2. Apps for Smoking Cessation

A number of studies have shown that digital interventions can be effective in promoting smoking cessation. One SR for smoking cessation [Chen et al 2012] is detailed and has very sound methodology including 60 RCTs/quasi-RCTs reported in 77 publications. Pooled estimate for prolonged abstinence (relative risk (RR) = 1.32, 95% confidence interval (CI) 1.21 to 1.45) and point prevalence abstinence (RR = 1.14, 95% CI 1.07 to 1.22) suggested that computer and other electronic aids increase the likelihood of cessation by 32% compared with no intervention or generic self-help materials. There was no significant difference in effect sizes between aid to cessation studies (which provide support to smokers who are ready to quit) and cessation induction studies (which attempt to encourage a cessation attempt in smokers who are not yet ready to quit).

The vast majority of studies focusing on mobile interventions and smoking cessation evaluated SMS messaging systems [Koel et al 2014]. This SR identified 15 RCTs with 13094 participants which showed that mobile based interventions are effective for smoking cessation. We were unable to identify any SR/RCTs for smoking cessation with mobile apps, however it was possible to identify 3 protocols for RCTs to compare app-based intervention for smoking cessation (Baskerville 2015, Bindhim 2014 and Valdivieso-Lopez 2013).

Although not strictly apps, a 2012 Cochrane review [Whitaker 2012] of 5 RCTs with at least six month cessation outcomes showed positive benefits of mobile phones for smoking cessation. Three studies involved a purely text messaging intervention adapted for different populations and contexts. One study was a multi-arm study of a text messaging intervention and an internet QuitCoach separately and in combination. The final study involved a video messaging intervention delivered via the mobile

phone. When all five studies were pooled, mobile phone interventions were shown to increase the long term quit rates compared with control programmes (RR 1.71, 95% CI 1.47 to 1.99, over 9000 participants), using a definition of abstinence of no smoking at six months since quit day but allowing up to three lapses or up to five cigarettes.

One 2012 SR investigating use of app-like web services for smoking cessation identified eight studies [Munoz 2009, Etter 2009, Munoz 2006, Pike 2007, Prokhorov 2008, Rabius 2008, Graham 2011, Norman 2008] measuring the effect of five online smoking prevention and cessation websites on clinical outcomes: (smoking initiation, cigarette use,1-day smoking abstinence, 7-day smoking abstinence, 30-day point prevalence). One study of an interactive, multimedia smoking prevention and cessation curriculum demonstrated lower rate of smoking initiation compared with use of a self-help booklet. With respect to smoking cessation, one study showed no change in cigarette use at 6 months, three studies demonstrated no difference in quit rate (as measured by 1-day reported abstinence, 7-day reported abstinence, 30-day point prevalence), and one study showed an improvement in quit rate. Participants who visited a site more than five times were twice as likely to quit than participants who visited a site less than five times (20.0% vs 9.8%, p<0.001). In addition, higher quit rates were found with more interactive, tailored sites compared with the static control site (13% vs 10%, p=0.04).

3.4.3. Apps for Dementia

We were unable to identify SRs or RCTs to assess the impact of mobile apps for patients with dementia to improve care or overall patient outcome.

A literature review from 2007 [Lauriks 2007] reviewed ICT-based services for identified unmet needs in patients with dementia. This concluded that informational websites offer helpful information for carers but seem less attuned to the person with dementia and do not offer personalized information. ICT solutions aimed at compensating for disabilities, such as memory problems and daily activities demonstrate that people with mild to moderate dementia are capable of handling simple electronic equipment and can benefit from it in terms of more confidence and enhanced positive effect

One SR from 2014 [McKechnie 2014] evaluated 14 empirical studies evaluating a range of complex, multifaceted computer-mediated interventions (none were mobile apps) for carers of people with dementia. Most studies found that positive aspects of caring were increased through these interventions, as was carer self-efficacy. There were mixed results in relation to social support, and physical aspects of caring did not seem to be affected .

A survey from 2014 of family caregivers (RR 560/881) provided with iPads equipped with a suite of mobile health apps designed to support family care givers [Frisbee et al 2014] did not find a statistically significant difference between treatment and control groups (no iPad).

A controlled parallel group trial (n=84) indicated that a carer held record [Simpson 2006] provides some benefit for carers in the domains of carer strain and locus of control however these are not specific to mobile apps/devices.

3.4.4. Apps for Mood Monitoring

In most cases, the comparison for the studies reviewed here is no intervention versus a computer based intervention. Many of the digital interventions are simply standard proven face to face or text based interventions adapted for digital/online use. There were no studies which compared digital vs face to face delivery of the same materials, however.

The utility of computer-aided psychotherapy has been shown in a SR to be as effective as face to face psychotherapy [Cuijpers et al 2009]. A considerable subsection of the literature utilising psychotherapy on mobile phones involves conversion of a validated psychotherapy tool and adapting it for mobile phone use.

A 2014 SR [Ye et al 2014] examining the effectiveness of internet-based interventions for children, youth, and young adults with anxiety and/or depression found 7 RCTs. Meta-analysis suggested that, compared to waitlist control (a group of participants included in the study that is assigned to a waiting list and receives intervention only after the end of the study), internet-based interventions were able to reduce anxiety symptom severity (standardized mean difference and 95% CI = -0.52 [-0.90, -0.14]) and increase remission rate (pooled remission rate ratio and 95% CI =3.63 [1.59, 8.27]). The effect in reducing depression symptom severity was not statistically significant (standardized mean difference and 95% CI = -0.16 [-0.44, 0.12]). There was no statistical difference in anxiety or depression symptoms between internet-based intervention and face-to-face intervention (or usual care).

A 2011 SR examining use of mobile phones for psychiatric interventions [Ehrenreich 2011] found 8 RCTs (5 focusing on smoking cessation and 3 on anxiety disorders). Of the three studies focusing on anxiety disorders, Two studies examined augmentation of Cognitive Behaviour Therapy (CBT) with a handheld computer compared to standard CBT of six weeks and 12 weeks. Treatment with CBT augmented by handheld computers led to significant reductions in panic symptoms compared to the wait-list control group but standard 12-week CBT interventions had the largest and most lasting treatment effects. The third showed no difference in the self-report measures of social phobia between wait-list controls and a handheld computer intervention group.

An RCT from 2014 with 52 participants [Daggo 2014] concluded that a guided Internet-based cognitive behaviour therapy for social anxiety disorder (SAD) adapted for mobile phone administration (mCBT) had significantly better patient reported outcomes when compared with a guided self-help treatment based on interpersonal psychotherapy (mIPT). Measures were collected

at pre-treatment, during the treatment, post-treatment and 3-month follow-up. On the primary outcome measure, the Liebowitz Social Anxiety Scale – self-rated, both groups showed statistically significant improvements

A 2013 SR [Donker 2013] identified 8 studies describing 5 apps targeting depression, anxiety, and substance abuse met the inclusion criteria. Four apps provided support from a mental health professional. Results showed significant reductions in depression, stress, and substance use.

Anxiety: a meta analysis of 40 RCTs (n=2648) from 2014 [Adelman 2014] showed that computerised CBT (cCBT) was significantly more effective than wait-list control in the treatment of anxiety disorders with longitudinal studies showing that individuals undergoing cCBT tended to improve after completion of treatment with longer follow up periods associated with greater symptom reduction.

A Cochrane review of 4 RCTs (607 pooled patients) from 2013 [Laugharne 2013] identified that currently there is no evidence that user-held clinical information has an effect on hospital or outpatient use for patients with psychotic disorders and deem that further evidence is required.

3.4.5. Apps etc. for other long term conditions

A 2013 Cochrane review of 2 RCTs with 408 pooled participants [Marcano Belisario 2013] failed to find any statistically significant improvement of asthma symptoms/asthma-related complications when comparing a mobile app based intervention versus a traditional method.

A 2013 SR focusing on web-based interventions for patient empowerment and physical activity in chronic diseases [Kuipers et al 2013] found 19 papers including patients with diabetes (n=11), heart failure (n=3), COPD (n=1), cardiovascular disease (n=1), cancer (n=1), and mixed patient groups (heart disease, lung disease, type 2 diabetes; n=1). Interventions varied greatly between studies although 6 key themes were identified. These were:

1) Education
2) Self-monitoring
3) Feedback/individual tailored information
4) Self-management training
5) Personalised exercise programme
6) Communication with healthcare providers or fellow patients

The RCTs varied greatly in content, duration, and frequency. Significant, positive effects on patient empowerment were found in 4 studies and 2 studies reported positive effects on physical activity. The remaining studies reported mixed results or no significant differences between intervention and comparison groups (ie, either both groups or neither group improved) on these outcomes.

3.5. Risks and challenges of apps used by patients and the public

3.5.1. Accuracy of apps for patients and the public

A study of the accuracy of 4 image analysis apps for patients to use to assess the risk of a pigmented lesion (Wolf et al, 2013) found that three of them were very inaccurate at diagnosing melanoma; the remaining accurate app simply took a photograph and sent it to a Board certified dermatologist.

Our work in Leeds on the accuracy of apps for calculating risk of cardiovascular disease has demonstrated that:

- None of the 19 apps communicated risk in terms that members of the public can understand (ie. X per thousand rather than percentage probability)

- Only one suggested that the user talk to a clinician – and it suggested this whatever the risk

- Several apps ignored key data items or prevented the user from entering an age greater than 74, for example

- The misclassification rate per app (with 20% threshold for significant risk of CVD) varied from 7% to 33%

- The median error rate for free apps (13%) was significantly less than the error rate for paid apps (26%), p = 0.026

3.5.2. Linking apps for patients to the electronic health record

We found no rigorous empirical studies describing the impacts of linking apps to the electronic record. Some likely theoretical benefits and dis-benefits we hypothesise are shown in the table below.

Theoretical Benefits	Potential Dis-benefits
Allows user access to more complete, more accurate patient data	Difficulty establishing user identity (no NHS smart card reader)
Possibility of entering data into HER from mobile device at bedside etc.	Wide range of EHRs to which app must connect
Aligns with Bring Your Own Device policy for health professionals	Concern about privacy and information governance, eg. if mobile device lost
Allows patient generated data / PROMs to be incorporated in the NHS record	Technical difficulties of establishing and maintaining interoperability at both semantic and syntactic levels
	App functionality curtailed when user moves out of connectivity by wifi or 3G
	Possibility of two different values existing for same data item if app temporarily disconnected

4. Review results on apps used by clinicians

4.1.Clinical acceptability and adoption rates of apps

We have identified some published surveys on this topic but they suffer from low response rates while others are focused on a particular niche or speciality. This makes it difficult to draw generalizable, valid conclusions. We are also aware of two current surveys which will report soon:

- Royal College of Physicians Research Panel survey. Data was collected via a Research by Design (RbD) survey to RCP members and fellows.

- a survey of usage of apps by all mental health professionals around Nottingham and Leeds – results due in July).

Methods for the RCP app survey:

- Research findings are based on a survey made available to the Royal College of Physicians member/fellows research panel [self selected] between 14 April and 6 May 2015.

- All 2,658 panel members were invited to take part in an online survey designed and hosted by Research by Design.

- Two reminder emails were sent to non-respondents.

Selected interim results of the RCP survey are as follows:

- 1,104 responses were received after 2 reminders, a response rate of 42%. 62% of respondents were Fellows, 62% were consultants and 27% were female. 55% were born from 1960-1979 with 7% since 1980 and 38% pre 1960.

- There is good uptake of health apps by respondents: more than half of respondents (586, 54%) use Apps to support their clinical work while 42% do not. Usage is significantly higher amongst international members compared with UK members

- Unsurprisingly, usage of Apps decreases with age, with greater usage seen amongst younger physicians.

- Over three quarters of respondents who use apps (77%) use just 1 to 5 Apps to support their clinical work.

- Two fifths of app users believe that these apps are either "very important" (29%) or "essential" (13%) to their clinical work.

- The most common clinical tasks for which app users use apps are to access guidelines (73% of users) and calculate risk (61% of users). Over half (55%) of the 586 app-using respondents also use apps to inform prescribing decisions and 39% use them in diagnosis.

- There was a wide range in the frequency of app usage: app users used the apps monthly (12%), weekly (44%), daily (27%) or several times a day (12%)

- The biggest concerns about clinical Apps are about the quality of the underlying evidence base on which the app is based (43%) and the accuracy of information generated (43%); information governance is also a concern (25%); apprehension is highest amongst UK physicians.

- The majority of respondents (73%) feel that RCP should be responsible for recommending Apps that are of high quality and safe to use; many respondents also expect specialist societies (69%) or the NHS (50%) to play a role in app recommendation.

- 58% of respondents are supportive of RCP produced / commissioned Apps; though proportionally more members than fellows

- The kind of apps favoured by the full set of 1054 respondents include apps to help follow clinical guidelines (78%), apps to support training and professional development (57%) and apps to help navigate RCP member services (31%)

We located no published studies of the usage by clinicians of apps specifically for diabetes, dementia, mood monitoring or smoking cessation.

4.2.Usability of apps designed for clinicians

We now present some simple examples of user interface design problems found with a few health apps, selected because they were CE marked and ran on the Apple iPad (the iPad can emulate iPhone apps). The problems have been selected for their simplicity to depict; unfortunately more complex design problems (which may have more impact in clinical practice) are too complex to explain briefly in this report. All the apps reviewed here have CE marks as Class I devices. A wider selection of medical examples — these sorts of user interface design problems do not just affect health apps — can be found in Thimbleby, *et al* 2015. Note that evaluation of the impact of the described usability defects or an estimate of the representativeness of the sample is beyond the scope of this report. In any case, usability is not the sole factor that should be applied in app evaluation, and other criteria must be applied as well: see Appendix 5: Example mHealth app checklist from Royal College of Physicians.

We have notified the app developers of our findings. An important advantage of apps is that they are easy to update, and well-designed apps will notify their users that upgrades are available. Hence the specific findings we show below are no doubt obsolete, but as they stand they illustrate well some of the types of typical design issues that are widespread in medical apps more generally.

Mersey Burns app

Mersey Burns is an innovative app that allows the user to use gestures to draw the patient burn area, and it then uses the Parkland formula to calculate a fluid prescription. To perform the calculation, the app needs the age and weight of the patient and the time of the burn. Mersey Burns is available for iPad/iPhone, Android, Blackberry PlayBook and on the web. This is a significant advantage, since a user might get stuck (for any reason, say a dead battery) and need to use their colleague's device — which may not be the same on as they own. Available on the web, Mersey

Burns can be used on any device with a web browser, including a desktop PC. The Mersey Burns user interface is identical across all platforms ("platform independence"), which is an important usability feature.

Mersey Burns is implemented in HTML5, which gives it these advantages. Unfortunately, HTML5 is not a dependable implementation environment and creates usability problems of its own. (For example, it makes user input validation very hard, because HTML5 was designed for ease of development, not for dependability.)

The following list of sample usability issues for Mersey Burns is longer than for other apps in this review. Because of the platform independence it is easy to view the source code of the app, and hence allow us to find design bugs just by inspecting it. Ironically, then, because Mersey Burns is designed prioritizing platform independence, it is easier for us to identify bugs in it!

- When the app is launched, the patient data is from the previous patient. The app does not check new patient data is entered.
- The app can send summary emails to any recipient. Apart from the basic data, there is no attempt at patient identification so it might be very easy to mix up emails sent on several patients. (The bulk of the email is the text of a disclaimer — see below.)
- The app does not provide data validation. It is possible to accidentally enter a patient weight higher than the weight of the universe, and because of input field overflow (digits in the data are not shown to the user) the user may believe the number entered is reasonable.
- Burn times can be miscalculated: a burn time entered as 23:56 hours:minutes ago is taken to be now. (The problem arises because the app's clock or the patient's idea of the correct time may be approximate, so the app rounds times; unfortunately, 23:56 ago is treated as 4 minutes ago, which is under 5 minutes, so it is rounded to zero, i.e., now.)
- The prescription is given to a high degree of precision (e.g., 3144ml) which will be harder to read than recommended notations 3,200 mL to enhance readability (ISMP, 2007).
- Burn surface area can be entered either by haptic gesture (drawing) or by a number. A user can draw on the anterior body, turn it over (tapping a button to do so) then enter a number. Unfortunately, the drawn burns are deleted. The design assumes a user enters either a drawing or a number; the danger is that the drawn side of the body, which is deleted, is invisible to the user, and the deletion happens without warning or consent — the user may have taken a few minutes working on it.
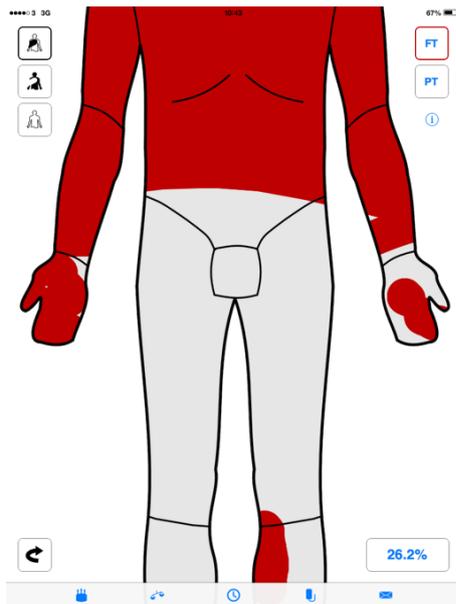
Screenshot of Mersey Burns, showing drawn burn locations. As the user draws the burn area, the numeric body surface area (shown here as 26.2%) adjusts automatically. If the user taps on the numeric burn area and adjusts it, the drawn area is cleared without warning (and possibly without the user seeing the change — as might happen here with any burns on the head or feet). This feature is thus *misleadingly* not quite equal opportunity (Thimbleby & Runciman, 1986).

The icons at the bottom of the picture are obscure. What might be thought to look like a "3 pin plug" is a birthday cake for date of birth; the "mouse" is an infusion bag to display the prescription. No tooltips appear to explain them.

- There are minor version control problems. It is likely that the numeric burn area was added after the inventive idea of haptic drawing; some of the help and warnings do not mention it. For example, if an attempt is made to send an email before specifying a burn area, the app asks for the burn area to be drawn (reinforcing the user problem described above) — not entered in either drawing or percentage area.
- Text for the prescription on an iPad is only 2 mm high. This is too small to be easily readable. There are no accessibility options to change font or size to improve legibility for the user.
- Like most health apps (including all of those reviewed here), the warranty (copied into any emails sent from the app as a disclaimer) is "No warranties." The warranty also says "You agree to indemnify and hold … NHS harmless from any claim … as a result of your use or misuse of the app" and "[the NHS] may, at its sole discretion, modify this disclaimer … at any time … without giving notice to you" which probably means the warranty contract is unenforceable.
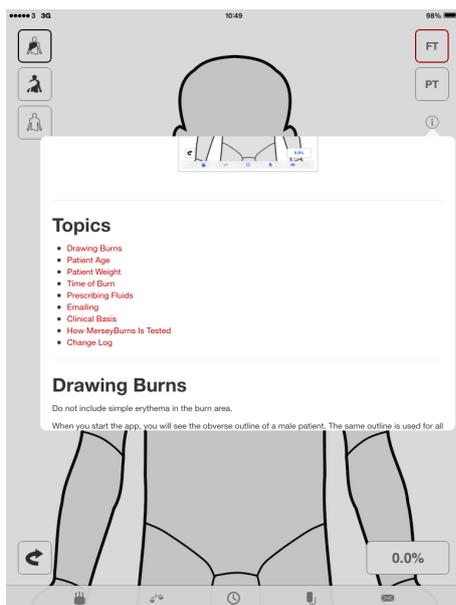
Unusually for a health app, Mersey Burns has a substantial user manual, which is clearly structured. The manual is built into the app and can be used when there is no internet connection.

Like many apps, tooltips are not available, so if a user wonders what FT means, say (highlighted top right of image) or what any of the icons mean there is no way to find them in the manual, except by reading the whole manual. There is no search in the manual (a keyboard search would help, but would generally not help search for icons, since there is no way to type them).

Overall, however, Mersey Burns is one of the most carefully developed and tested apps for professional use that we have seen, and has doubtlessly saved many lives and much unnecessary disfigurement by bringing complex calculations to the bedside in a format that makes them accessible to clinicians and easy to follow. Minor changes, such as those suggested above, could make it even better.

Sepsis 6 app
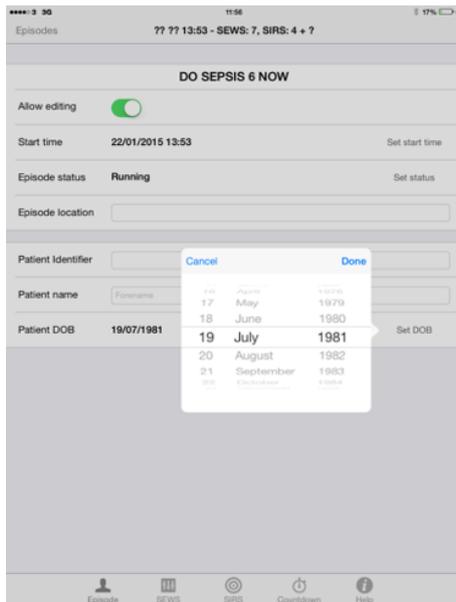
The Sepsis 6 app is another app intended to bring a clinical score to the bedside, in this case to help identify and manage people with sepsis. Usability issues again arise, for 38ompute:

Screenshot. What are the "??" marks at the top of the screen?

A date dialogue box is shown, apparently selecting 19 July 1981. If the user taps "Done" the date is entered as the patient date of birth. If the user *slightly* misses "Done" the dialogue is cancelled and the date is not changed: when the dialogue box is displayed only Cancel and Done should work or there should be a warning that the user has dismissed the dialogue (by tapping anywhere else) and the data has not been changed.
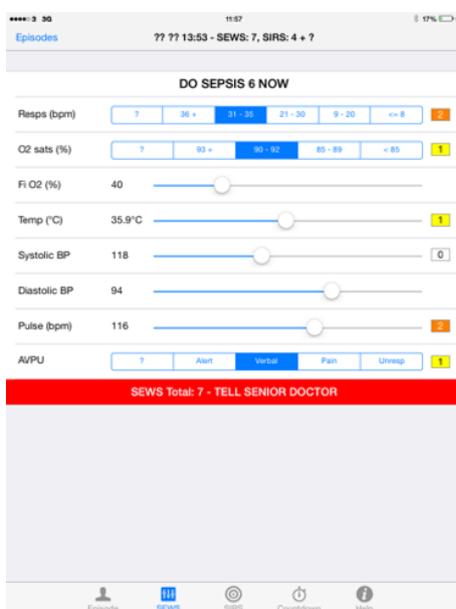
Sepsis 6. Like Mersey Burns there is no check that the calculations are based on the current patient. Shown here, the SEWS score is an amalgam of two patients.
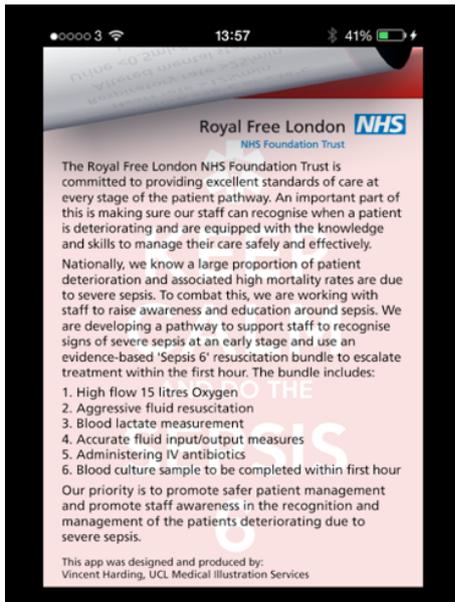
Sepsis 6 help screen. The help screen may appear to have frozen, since the only way of leaving is to click on the upturned page graphic, which has no affordance to click or go back: it is just a visual skeuomorph and not a button or finger target consistent with others used in the app.

The app does not provide any effective contact information, problem reporting, or support. The named contact is a "digital designer" — who presumably cannot provide clinical help — even if the user has the internet to help figure out how to contact them.

A different Sepsis 6 app
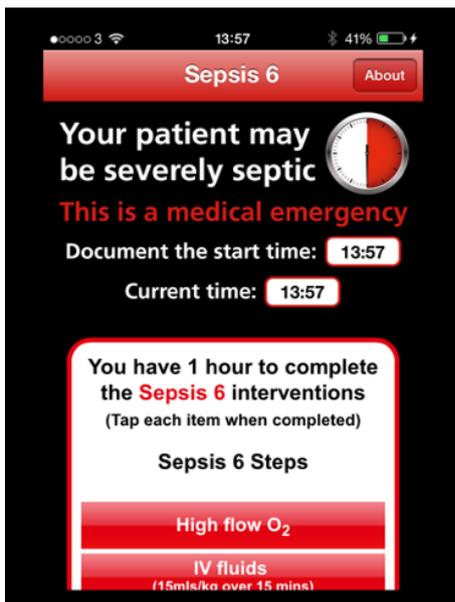
Confusingly, another app for sepsis management by a different developer is also called Sepsis 6. This app shown here on an iPhone. Also notice how the screen contents are clipped and it is not possible to see the full menu of choices presented to the user.

The "clock" (top right) is animated, and seems to serve no purpose except as a visual distraction — which is probably not helpful for a sepsis app.
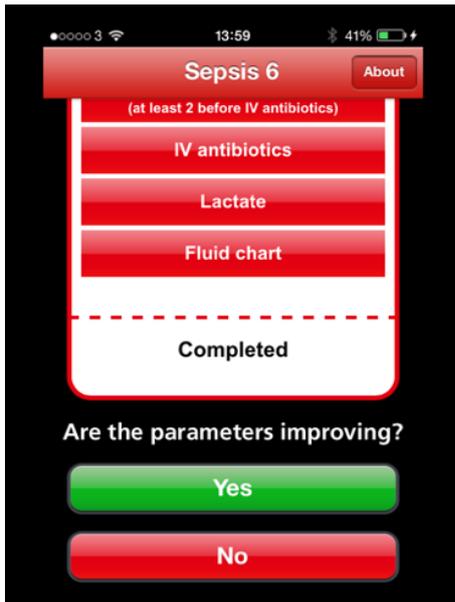
Sepsis 6, shows "completed" when nothing has been completed.

The user can say parameters are improving when no parameters have been selected — another form of failure to validate user input. Standard user interface guidelines would suggest dimming or otherwise disabling inappropriate options.

ONCOassist app
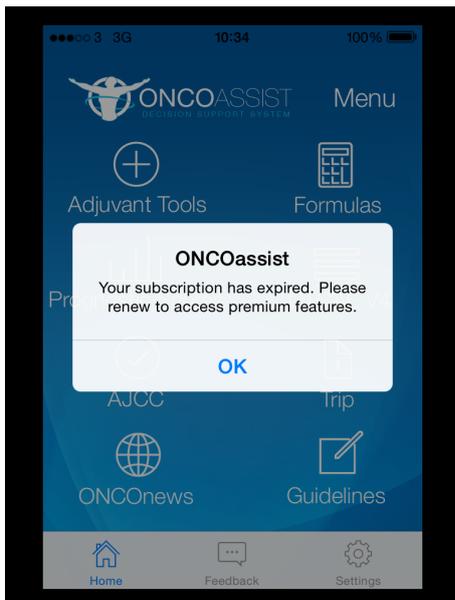
The ONCOasssist app is provides clinicians looking after cancer patients with up to date knowledge about chemotherapy regimes, practice guidelines, calculation tools etc.. Like many apps, ONCOassist has a subscription model, which will tend to be disruptive to clinical use. It would be easy to be using it just as the subscription expires. In general, to continue, one would need an account already set up (or credit card details to hand) as well as a wifi connection.

Of course apps need a business model, but they need to balance that against clinical use, or hospitals should have policies about subscription apps.

Most of the time, an app will look like it is working well, then without warning it will freeze while it makes internet connections. Shown here is ONCOassist; note that it gives the user no idea of how long they will need to wait before they can use the app properly (or even how they can go back and perhaps try some other feature of the app).
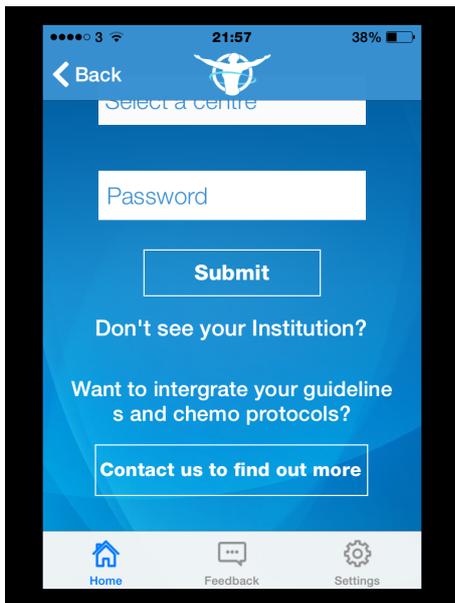
Simple spelling mistakes shown in ONCOassist, but nevertheless a clear example of poor quality control.

Note that the screen is also clipped, so the user cannot see all information on it.
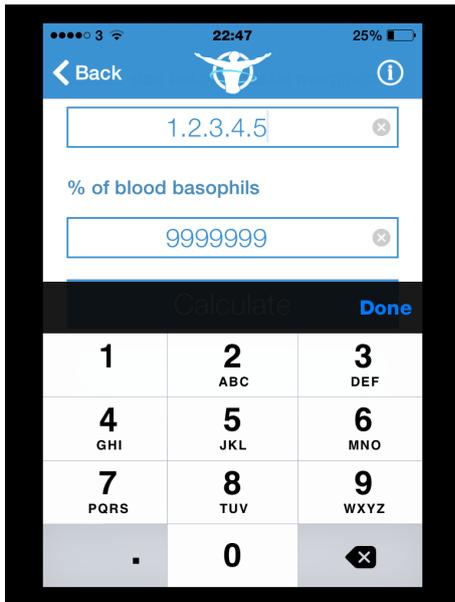
ONCOassist does not validate user input: this extreme example shows how a user might enter a syntax error (top) and an out of range number (bottom). Less egregious errors may be less obvious and contribute to patient harm.
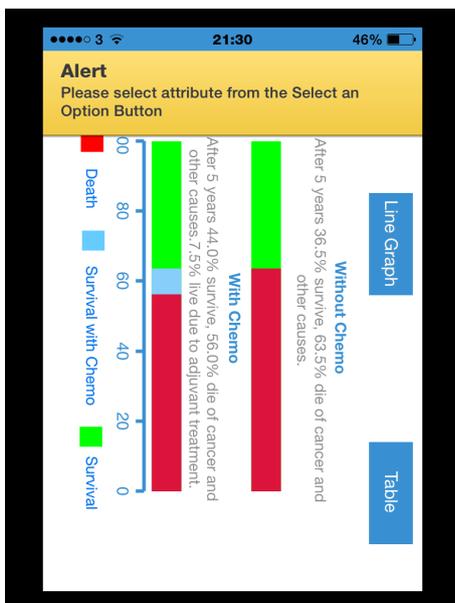
The yellow Alert at the top of the ONCOassist screen asks the user to select an attribute and press a button. Neither actions are possible. The alert is a bug.
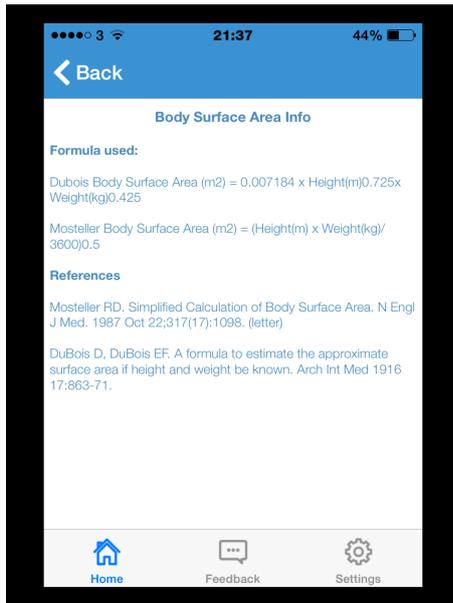
Some of the formulas ONCOassist uses are incorrect and incorrectly formatted (e.g., exponents are confusingly written inline).

4.3. Benefits of apps used by clinicians

No studies have yet been identified that evaluate the impact of apps used by clinicians on patient-centred care, costs of care delivery, resource use, or cost effectiveness, for the key focus areas of diabetes, smoking cessation, dementia or mood monitoring, nor for other long term conditions (LTCs).

4.4. Risks and challenges of apps used by clinicians

No studies yet located evaluate the impact of apps used by clinicians on potential risks eg. clinical safety, information governance or privacy concerns.

4.4.1. Accuracy of apps designed for clinicians

Studies have identified a number of health apps that can compromise patient safety and are potentially dangerous in clinical use. For example, some apps designed for opioid dosage conversion demonstrate dangerously poor accuracy. A study of 23 opioid drug dose calculators [Haffey et al, 2013] converting between fentanyl, methadone doses, etc., found dangerously large variations in calculated dose equivalence. Conversion of a 1mg dose of oral morphine to methadone resulted in a methadone dose of 0.05 to 0.67mg (a 13:1 range) with only 4 (44%) of the methadone conversion apps commenting that the conversion formula depends on the actual dose as well as the drugs. Only 11 (48%) of the 23 apps provided evidence sources and 12 (52%) had no stated involvement of a medical professional.

Another study of generic medical calculator apps tells a slightly more positive story [Bierbrier et al 2014]. Six (43%) of the 14 medical calculating apps studied provided completely accurate and reliable results when tested across 13 functions. Three free apps were 100% accurate and contained

the most functions desired by general medicine consultants (CliniCalc, Calculate by QxMD, and Medscape). The functions that were least likely to be accurate were the Child Pugh score (inaccurate in 8 apps, 57%) and Model for End stage Liver Disease (MELD, 8 apps). The authors commented that, while many apps were reliable, it is important to be careful when calculating MELD scores or Child-Pugh scores on some apps, and greater scrutiny is warranted to ensure full accuracy of smartphone medical calculator apps.

The problem that many developers of apps designed for clinical use have little or no formal medical training and do not involve clinicians in the development process (and may therefore be unaware of patient safety issues raised by inappropriate app content or functioning) is supported by other studies [Hamilton et al 2012, Rodrigues et al 2013].

5. Review results on personal health records used by patients and the public

To assist the reader in understanding this area, we have developed a simple typology showing the different types of PHRs, as shown in the table:

Type of PHR	Functions	Example
Paper PHRs	A paper record to support shared care between primary & secondary care	Shared maternity care card; red book for child's immunisations
Type 1 PHRs	Patient access to read official records	Current GP systems open to patients to view
Type 2a PHRs	Patient's own private records	Microsoft HealthVault
Type 2b PHRs	Patient's own private records that patient may link to official record, or authorise clinicians to view	Original HealthSpace model; EMIS Personal Health Record
Type 3 PHRs	Official records that patient is allowed to edit/add to	PatientView; US Veteran's Agency records

Patient controlled electronic PHRs allow a number of information collection, sharing, exchange and self-management functions to be supported, as shown in the diagram below [Kaelber, 2008].

Figure: Idealized hub and spoke concept for an electronic PHR
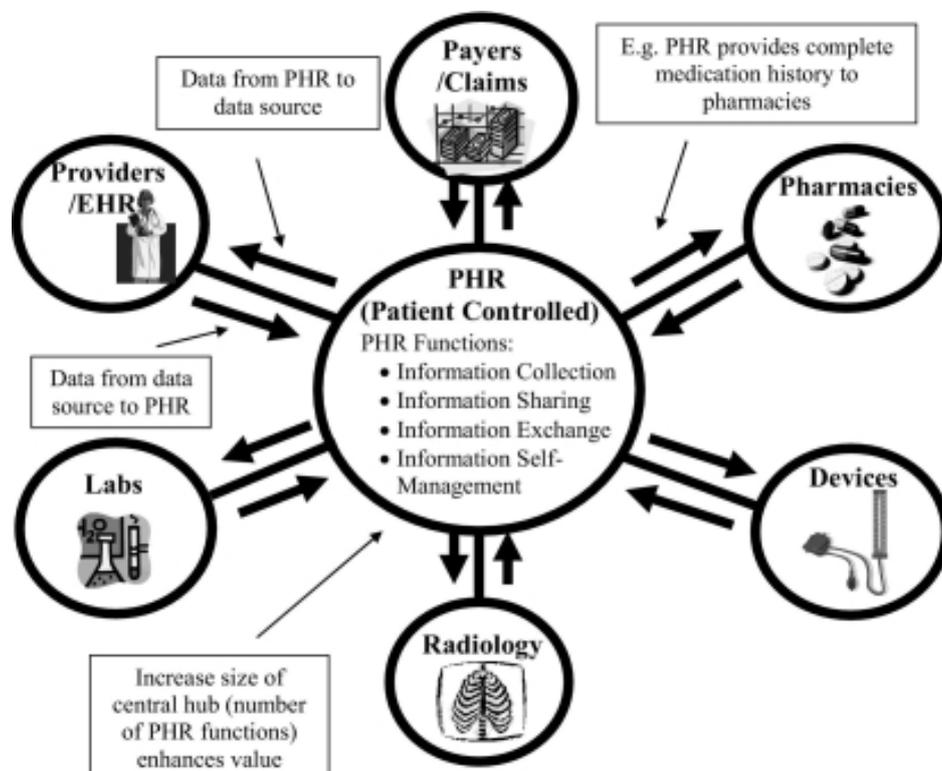
5.1.Patient acceptability and adoption of PHRs

Acceptability and factors influencing this

Acceptability is only the first stage of technology adoption; if the idea of the technology is not acceptable, people will not sign up, let alone use the technology. This section reviews studies of PHR acceptability, the next section discusses studies of adoption, which consists of enrolment (signing up) and usage.

Patient portals (personal electronic health records) should provide patients with tools to better understand and manage their illness and risk factors. However, there are several studies showing resistance to these tools from both patients and providers, for a number of reasons. To establish preferences for new functions to be included in one of the largest PHR projects, the US Veterans' Administration "My HeatheVet" PHR, a random sample of 4% of the users visiting four or more pages from Oct 2007 – Oct 2008 were asked to complete a survey (Nazi, 2010). Of the 585,039 eligible site visitors, just over 100,000 (100,617) surveys were completed (17.2% response rate). The results are shown in the following table.

Figure: Preferences of My HealtheVet respondents (from Nazi 2010)

Preferences of American Customer Satisfaction Index My HealtheVet Respondents (N=100 617)

	N[*]	Percentage
Additional services desired[‡]		
View my upcoming appointments	79892	87
Schedule or change my appointments	68395	74
Look at information in my VA medical record	67714	73
Online, secure communication with my doctor	58878	64
Checking that different medications I take are safe when used together	45986	50
Reminders of preventive care I need (eg, shots, cancer screening)	34707	38
Notification of new content/features on the site	32418	35
Advance check-in for my VA clinic visits	31863	35
Monthly email newsletter	24186	26
Share information that I have stored in My HealtheVet with other people	23088	25
Advanced directive (eg, living will, durable power of attorney for healthcare)	20418	22
Educational programs	18800	20
Information about the quality of VA healthcare	11231	12
Other	8791	10
Total	92160	
Medical records wanted in My HealtheVet Personal Health Record[‡]		
My VA medical records	18122	90
My military service medical records	12973	64
My medical records from other non-VA medical providers	6665	33
Not sure	1108	5
I don't want my records to be available on My HealtheVet	634	3
Other	451	2
Total	20244	

One recent SR [Kruse et al 2015] reviews good quality studies that evaluated the acceptability of portals. Specifically, they examined the characteristics of patient portals to support long term condition self-management that received positive remarks from patients or healthcare providers. The SR was of good quality (the authors searched both the CINAHL (Cumulative Index to Nursing and Allied Health Literature) and PubMed databases; two reviewers analysed the articles) and the authors analysed the results of 27 articles that passed explicit quality criteria. The results showed that:

- Patient portals show significant improvements in patient self-management of chronic disease and improve the quality of care provided by providers.

- The most prevalent positive attribute noted by both patients and healthcare providers was patient-provider communication, which appeared in 11 of 27 articles (41%).

- The most prevalent negative perceptions are security concerns and user-friendliness, both of which occurred in 11 of 27 articles (41%).

- User-friendliness was a common concern for those patients and healthcare providers unfamiliar with the internet, who therefore find it difficult to navigate the patient portal.

- The high cost of installation and maintenance of a portal system deters some healthcare providers from implementing such technology into their practice, but this was only mentioned in 3 of the 27 articles (11%). It is possible that US incentives for "meaningful use" will attenuate this cost barrier.

Patient and provider acceptability factors identified in a SR of 27 studies on acceptability of patient portals for chronic disease management		
Acceptability factor	No. of studies in which factor mentioned	Was factor positive or negative? No. of studies (percentage)
Patient-provider communication	17	11 (41%) Positive 3 (11%) Negative
Security	11	11 (41%) Negative
User-friendliness	16	7 (26%) Positive 11 (41%) Negative
Secure messaging	15	10 (37%) Positive 5 (19%) Negative
Quality of care	10	10 (37%) Positive
Disease outcomes	10	10 (37%) Positive 2 (7%) Negative

Educational resources	12	9 (33%) Positive 3 (11%) Negative
Time	5	5 (19%) Positive
Cost	3	3 (11%) Negative

Table adapted from: Kruse CS, Argueta DA, Lopez L, Nair A. Patient and Provider Attitudes toward the Use of Patient Portals for the Management of Chronic Disease: A Systematic Review. Brixey J, ed. *Journal of Medical Internet Research*. 2015;17(2):e40. Doi:10.2196/jmir.3703.

This recent comprehensive SR of positive and negative attitudes to portals / PHR for use in long term conditions revealed mixed attitudes from patients and healthcare providers about using patient portals to manage chronic disease. The authors suggest that a standard patient portal design that provides patients with the resources to understand and manage their chronic conditions will promote the adoption of patient portals in health care organizations. However, it is currently unclear what these key elements of an effective portal/PHR are (see section below on portal benefits and outcomes).

Factors determining adoption rates

When considering empirical studies of adoption, we need to distinguish enrolment (signing up for a PHR) from utilisation (actually using the PHR on a continuing basis, perhaps for 3-6 months or more). The first merely represents a single decision, while the second reflects a series of decisions taken over a period and establishment of a new pattern of user behaviour. Such a behaviour change is a much more difficult task to achieve.

A 2014 SRSR [Amante et al 2014] examined factors associated with enrolment to and utilization of portals among patients with diabetes, to identify the barriers and facilitators of portal use. This SR was of good quality (the authors searched PubMed and CINAHL for papers reporting original research using quantitative or qualitative methods on characteristics, barriers, and facilitators associated with portal enrolment and utilization) but was limited to studies on patients with diabetes in the United States since 2005. Of the sixteen articles identified which passed critical appraisal criteria, nine were quantitative, three were qualitative, and four used mixed-methods. The results were:

- A number of demographic characteristics, having better-controlled diabetes, and healthcare providers who engaged in and encouraged portal use were all associated with increased portal enrolment and utilization

- Barriers to portal enrolment included a lack of patient capacity (eg. poor memory or eyesight), poor motivation to use the portal and low awareness of the portal, or of specific portal functions.

- Barriers to portal utilization included lack of patient capacity, lack of healthcare provider or patient buy-in to apparent portal benefits, and negative patient experiences when using a portal

- Facilitators of both portal enrolment and utilization included healthcare providers and family members recommending, endorsing or engaging in portal use, good usability, good internet access, patients were educated about how to use and benefit from portals.

They concluded that as more healthcare organisations offer portals, addressing barriers and exploiting facilitators may help patients with diabetes achieve potential benefits. While this SR focused on patients with diabetes in the USA, it seems likely that the findings will generalise to people with other LTCs and to the UK setting. The author's focus on studies published in the last 10 years helps improve the generalizability of these findings to current portal/PHR technologies.

5.2. Clinical acceptance of PHRs for patients

Wynia et al (2011) conducted a survey in 2008/9 of the views of US physicians about the use of patient PHRs with 856 eligible physicians (63 % response rate). They found that 14% of physicians were frequent (daily or more often) users of patient PHRs, and analysed the views of this group, those who were willing in theory and those who were unwilling users separately. The results are shown in the following table.

Figure – Physician opinions on PHRs from Wynia et al 2011

Physicians' Beliefs And Concerns About Patient Electronic Personal Health Records, According To Reported Willingness To Use The Records In Clinical Practice And Experience Using The Records, 2008–09

Belief or concern	Weighted percent	Weighted percent among		
		Those willing to use PHRs	Those unwilling or neutral	Frequent PHR users
Agree or strongly agree that				
PHRs empower patients to participate in their care	48	68	34	56
If patients had access to their electronic records, they would provide a valuable check on record accuracy	38	48	30	27
Using PHRs will improve quality of care	30	53	13	56
Using PHRs will improve my relationship with patients	22	46	5	53
PHRs will save me time in the long run	28	48	13	49
Somewhat or very concerned that				
Patients might omit important information in their PHR	85	79	89	77
PHRs might contain incorrect information	79	71	84	67
Physicians might be held liable for knowing all the information in a PHR	76	65	83	70
PHRs might not have adequate privacy protections	70	60	77	65
Physicians won't be reimbursed for time spent reviewing PHRs	63	52	70	51
If patients had access to their electronic records, including my notes, I would censor what I write	36	26	44	20

5.3.Benefits and outcomes of PHRs used by patients

Davis et al conducted a SR in 2014 to determine the impact of providing patients with access to their medical records (electronic or paper-based) on healthcare quality, specifically examining measures of safety, effectiveness, patient-centeredness, timeliness, efficiency, and equity as suggested by the Institute for Healthcare Improvement in Boston. This SR was of good quality (the authors searched for articles in English indexed in PubMed from January 1970 to January 2012). After critical appraisal, 20 RCTs and 7 uncontrolled case series were included. Studies were categorized by measured outcome: effectiveness (n=19), patient-centeredness (n=16), and efficiency (n=2). Surprisingly, no study addressed patient safety, timeliness, or equity of access. The results showed that:

- Access to health records appeared to enhance patients' perceptions of control of their illness and reduced or had no effect on patient anxiety.

- Outcomes were equivocal with respect to several aspects of effectiveness and patient-centeredness.

- Efficiency outcomes were mixed; some studies showed a reduction in the frequency of face to face and telephone encounters while others showed an **increase** following introduction of the PHR. The largest study [Pallen 2012] showed significantly more visits to after hours, emergency room and hospitals following access to a PHR.

- Only four studies looked at the impact of PHRs on health related behaviours; of these, two showed a significant improvement, and two did not. In the single good quality study looking at health behaviours, an RCT showed no impact on use of prescription drugs, alcohol, tobacco or awareness of healthy living.

- There was no measurable improvement in quality of life due to the PHR (1 study).

Turning to the impact of PHRs on specific disease outcomes in this SR:

- Four studies covered diabetes, but only one showed a significant drop in HbA1C.

- Two studies covered hypertension, but both showed no objective benefit of patient access to data in the PHR.

- An RCT of access to an electronic PHR for pregnant women in addition to a pregnancy information website showed no change in perceived usefulness of the information.

They concluded that, although a few positive findings generally favoured patient access, the literature is unclear on whether providing patients access to their medical records actually improves quality.

Please see the following table summarising major study results, and Appendix 6 for details of all the PHR impact studies located.

Sample electronic Patient health records (PHR) and outcome of studies			
Electronic PHR	**Study**	**Results**	**Outcome**
"My HealtheVet", Veterans Health Administration, USA	Nazi, 2010	Satisfaction with My HealtheVet is high (8.3/10.0), and users are highly likely to return to the site (8.6/10.0) and recommend the site to other veterans (9.1/10.0).	Positive – Patient satisfaction
Two urban general practices in Manchester, UK	Fitton, 2014	Patient record access likely to save time for patients and practices. If 30% of patients accessed their electronic general practice record online at least twice a year, a 10,000-patient practice is likely to save 4,747 appointments and 8,020 telephone calls each year – about 11% of appointments.	Positive – Decreased appointments
"MyHealthManager", Kaiser Permanente, USA	Palen, 2012	Significant increase in office visits , telephone encounters, emergency department encounters and hospitalizations for PEHR users.	Negative – Increased appointments and hospitalisations
11 primary care practices in the Partners HealthCare system, USA	Wright, 2012	Patients receiving reminders via a secure PHR "eJournal" more likely to receive mammography (48.6% vs 29.5%, p = 0.006) and influenza vaccinations (22.0% vs 14.0%, p = 0.018).	Positive – Better health promotion
University of Pittsburgh Medical Center (UPMC) "HealthTrak", USA	Hess, 2007	No change in number of patient encounters or telephone calls, but messages increased. Participants favoured enhanced communication and access to laboratory tests but felt messages not being answered.	Neutral

"MyChildren's" portal, Children's Hospital Boston, USA	Bourgeois, 2009	Most accessed features were the lab results (82%) and summary form (21%). Patient edited their medication list (5.4%), allergy list (2%), problem list (1%) and immunization list (1%).	N/A
PHR tethered to the Military Health System, that leverages MicrosoftHER HealthVault and GoogleHER Health, USA	Do, 2011	169 (67.6%) selected Microsoft® HealthVault, and 81 (32.4%) selected Google® Health as their PHR of preference. Sample evaluation of users reflected 100% (n = 60) satisfied with convenience of record access and 91.7% (n = 55) satisfied with overall functionality of PHR.	Positive – Patient satisfaction
"My dispensed medications", Sweden	Montelius, 2008	The usefulness was rated high (total median grade 5; Inter Quartile Range [IQR] 3, on a scale 1-6). Positive about the design (total median grade 5; IQR 1, on a scale 1-6). Elderly patients felt they were 'getting control' (P < .001).	Positive – Patient satisfaction
Free "Personal Health Information Management System" (PHIMS), in a housing facility for low-income and elderly residents, USA	Kim, 2009	Of the eligible residents, only 13% (44/330) used the system; 77% of user activities occurred with assistance. Residents' ability to use the PHR system was limited by poor computer and internet skills, technophobia, low health literacy, and limited physical/cognitive abilities.	Negative – those who can benefit the most from a PHR system may be the least able to use it
"eCleveland Clinic MyChart", USA	Miller, 2007	Main predictors of use are the number of diagnoses and number of clinical encounters. Hence PHR is most useful to sicker patients who are greater consumers of healthcare.	N/A

5.3.1.Impact of PHRs on patient decisions and actions

No. / Type of studies included: 1 SR (14 RCTs, 21 observational and hypothesis testing studies, 5 quantitative and descriptive studies and 6 qualitative studies included), 1 SR (26 publications including 8 RCTs), 2 RCTs and 2 case studies.

Diseases included in analysis: Chronic health conditions including diabetes, heart failure, hypertension, depression and preventive services, rheumatoid arthritis, major depression and COPD.

A qualitative study of 305 adults in Taiwan reported 69.8% patient reported improvement in quality of healthcare with implementation of a chronic disease management portal [Tang, 2003].

One study on heart failure demonstrated no significant difference in the "self-efficacy" part of the Kansas City Cardiomyopathy questionnaire. Intervention group showed more adherences to medical advice but no difference in adherence to medications [Ross 2004]. In one SR acceptance of portals was higher in younger, computer literate and more enthusiastic patients. Patient satisfaction was generally high among the portal users [Goldzweig 2013].

A before and after study of use of web portals in rheumatoid arthritis patients found that lack of internet facility was the most common reason why patients had not logged on to the portal (Van der Vaart 2014). Of the 194 patients, 115 (55%) had used the web portal at least once, and 27 patients (13%) had used the portal over 3 times. Non users were more likely to be older, single, lower educated and unemployed. During the login 15/86 patients had a single problem with the portal. 33% of all logged in patients felt that the quality of care was higher as a result of the portal. Satisfaction with the rheumatologist/nurse or perceived self-efficacy in patient-provider communication did not show any significant change with the implementation of the portal. Similarly, patients did not perceive any significant change in personal control, illness coherence, treatment control and medical adherence.

In a case study of diabetic patients using a web portal in the US, patients felt that the system would enhance communication, allow checking of lab results, remind patients about appointments and allow them to contact physicians (Hess 2007). After the implementation, patients felt more empowered and able to communicate easily. The blood glucose tracking tools were regarded as the most beneficial.

5.3.2.Impact of PHRs on behavioural change

No. / Type of studies included: 1 SR (60 RCTs/quasi-RCTs reported in 77 publications), 1 SR (9 RCTs), 1 SR (18 RCTs) and 3 RCTs.

Diseases included in analysis: Smoking cessation, diabetes, chronic diseases, cancer, behavioural change, preventive care and cancer.

A smoking cessation SR including 60 RCTs found that in comparison with no intervention or generic self-help material, interventions using electronic aids significantly increased the likelihood of achieving prolonged abstinence or point prevalence abstinence from smoking, measured at the longest follow-up (Chen 2012). Pooled relative risks were 1.32 (95% CI 1.21 to 1.45) for prolonged abstinence and 1.14 (95% CI 1.07 to 1.22) for point prevalence abstinence at follow-up. There was no substantial heterogeneity in these analyses. There were no substantial differences in effect size between aid to cessation and cessation induction studies. The mixed-treatment comparison showed a small but statistically significant positive intervention effect on time to relapse (mean HR 0.87, 95% CI 0.83-0.92). Six studies compared different electronic interventions with a single tailored component against each other. As the settings and contents of the electronic interventions being assessed were different, it was not possible to include them in a

meta-analysis and compare the electronic interventions with each other. Also, there were no comparisons made between the use of web based PHR versus mobile technology.

A 10 month website based intervention with professional coaching help; a USA Study (320 patients), found no significant increase in physical activity [Glasgow, 2003]. Participant rate was 82%. There was greater use of the website in the first 3 months, then a progressive drop in rates over time.

A 16 week study (324 patients) with the study group having access to an interactive website vs. the control group only having access to part of the website found a significant increase of 42.6% in steps taken in a day [Richardson, 2010].

In 2012, Glasgow conducted a 12 month study with 463 patients on website intervention or website intervention plus follow up telephone call and found a significant increase of 10.6% improvement in calories expended through physical activity at 4 months. However the effect significantly decreased when the patients were followed up to 12 months duration (decrease of 18.7% at 12 months)..

Significant positive effects on patient empowerment reported in 4 studies and positive physical activity reported in 2 studies. The interventions were often used in different combinations and adapted to specific patient populations; hence, the individual contribution of the effects of the intervention was not feasible. Based on the evidence the authors identified 7 elements of web based interventions that could benefit cancer patients, including the provision of a cancer survivorship care plan [Kuijpers 2012].

An RCT on interactive health communication application for cancer follow up showed significant within-group improvements in depression in the experimental group only. In the control group, self-efficacy deteriorated significantly over time [Rutland 2013]. Another 6 month RCT concluded that behaviour change can be assisted using interactive websites, however recruitment can be difficult and the authors suggested that this can be improved if primary care information was integrated to the PHR [Dickinson 2013].

An RCT on the use of a web portal for preventive and cancer care found that at 4 months, delivery of colorectal, breast, and cervical cancer screening increased by 19%, 15%, and 13%, respectively, among users [Kirst 2012].

5.3.3.Impact of PHRs on clinical outcomes

No. / Type of studies included: 1 SR (14 RCTs, 21 observational and hypothesis testing studies, 5 quantitative and descriptive studies and 6 qualitative studies included), 1 SR (23 RCTs), 1 SR (9 RCTs), 1SR (26 publications including 8 RCTs) and 3 RCTs.

Diseases included in analysis: Chronic health conditions including diabetes, heart failure, hypertension, depression and preventive services. Anxiety disorders (phobia, panic disorder, obsessive compulsive disorder and PTSD), cancer and COPD.

In one RCT evaluating diabetes care, patients receiving a web based and nurse care management had significantly lower HbA1c compared to the control group [Ralston 2009]. There was no difference in outpatient visits or primary care / specialty visits or inpatient days.

Patients receiving face to face visits and secure messaging in addition to a hypertension web portal (included input of BP readings and HER data), showed significant improvements in BP compared to usual care and only portal access [Green 2008]. Another study found no significant difference in BP.

Another US based study exhibited no difference in HbA1c, BP or low-density lipoprotein (LDL) cholesterol levels between the groups after 12 months of access to an electronic personal health record (Grant 2008), while a different RCT showed lower HbA1c at 6 months but not at 12 months' time [Tang 2013].

A SR on anxiety disorders found that computer-aided psychotherapy had a larger effect on anxiety related conditions compared to contrast conditions (d=1.08, 95%CI 0.84-1.32; 21 studies) [Cuijpers 2009]. Two RCTs were deemed as 'outliers' and removal of these two studies decreased the heterogeneity from (I^2=65.6% to 36.2%). This also reduced the effect size (d=0.94, 95%CI 0.80-1.08; 19 studies). There was no significant difference in anxiety at 3 months follow up. Computer-aided psychotherapy had a small to moderate effect on quality of life (d=0.46, 95% CI 0.30-0.62; 12 studies) and depression (d=0.56, 95%CI 0.41-0.71) compared to contrast conditions. There was no significant difference between computer-aided and face-to-face psychotherapies at 3 or 6 months follow up.

Another 3 month study in South Korea (73 patients) found a significant decrease in HbA1c by 0.6% (7mmol/mol). Only 27% of participants had computer access. 83% satisfaction with the use of the website was shown [Kim and Kang, 2006]. An RCT including 11 primary care practices with 244 diabetic patients showed no significant decline in HbA1c levels between intervention and control groups, with both groups showing three-quarters of all patients at their HbA1c goal at the end of the study [Grant 2008]. A similar pattern was seen for BP and LDL cholesterol control.

A Canadian study of 511 patients who received 6 month intervention to track quality of diabetes care indicated an increase in exercise by 125% (extra 67.5%) and decrease in HbA1c by 0.2% (2 mmol/mol). A similar study in the USA (761 patients) showed no significant increase in physical activity, but a decrease in HbA1c by 0.6% (2 mmol/mol) [Holbrook, 2009].

An RCT(12 month duration), including 104 patients in the USA with chronic disease showed significant improvement in cognitive status (+0.8% in control group vs. -1.0% in intervention group) and functional level (19.4 vs. 20.0) in the intervention group receiving a home electronic portal compared to usual home health care [Noel 2004].

An RCT in the USA including both Type 1 and 2 diabetes patients (n = 104) showed a significant reduction in HbA1c levels between the control group and an intervention group receiving web based care management (-1.2 vs. -1.6%) in 12 months' time. In addition, there was significant difference in HbA1c reduction between high users and low users (-1.2 vs. -1.6%). There was also a significant better reduction in the systolic BP (-7 vs. -10). There was also significant better reduction in the Triglyceride levels and increase in high-density lipoprotein (HDL) levels in the intervention group (McMahon, 2005).

In another RCT, HbA1c levels did not show any significant reduction when patients were followed up for 3 months duration. They also reported no significant change in BP and exercise. The portal also had poor usability due to technical complications [Faridi 2008]. An RCT on interactive health communication applications for cancer follow up found significant within-group improvements in depression in the experimental group only. In the control group health-related quality of life deteriorated significantly over time [Rutland 2013].

5.3.4.Impact of PHRs on clinical decisions and actions

No. / Type of studies included: 1 RCT, one 12 month quasi-experimental study on diabetes conducted in the USA and Puerto Rico showed a large reduction in prescriptions of 59% in the PHR group, while control patients experienced a +37% increase over baseline rates. [Kobb, 2003].

Diseases included in analysis: Diabetes.

5.3.5.Impact of PHRs on health service utilisation
No. / Type of studies included: 1 SR (14 RCTs, 21 observational and hypothesis testing studies, 5 quantitative and descriptive studies and 6 qualitative studies included), 1 SR (26 publications including 8 RCTs), 1 SR (18 RCTs), 1 RCT, 3 case studies and 3 surveys.

Diseases included in analysis: Chronic health conditions including diabetes, heart failure, hypertension, depression and preventive services, chronic diseases, paediatrics and COPD.

Two observational studies from the Kaiser healthcare system showed different results, with one cohort study showing significantly higher rates of office visits, telephone call, clinic visits, emergency visits and hospitalisations (Palen 2012), while another found a decrease in visits and less increase in telephone contacts (Zhou 2007). The figure below shows the mean number of office visits per Health Maintenance Organisation (HMO) (Kaiser Permanente, USA) member per month for the 12 months before and after the PHR ("MyHealthManager", MHM in figure) was made available, distinguishing users from non-users [Palen, 2012]. There is a statistically significant increase in the number of office visits for users totalling 0.7 extra visits per year compared to non-users.
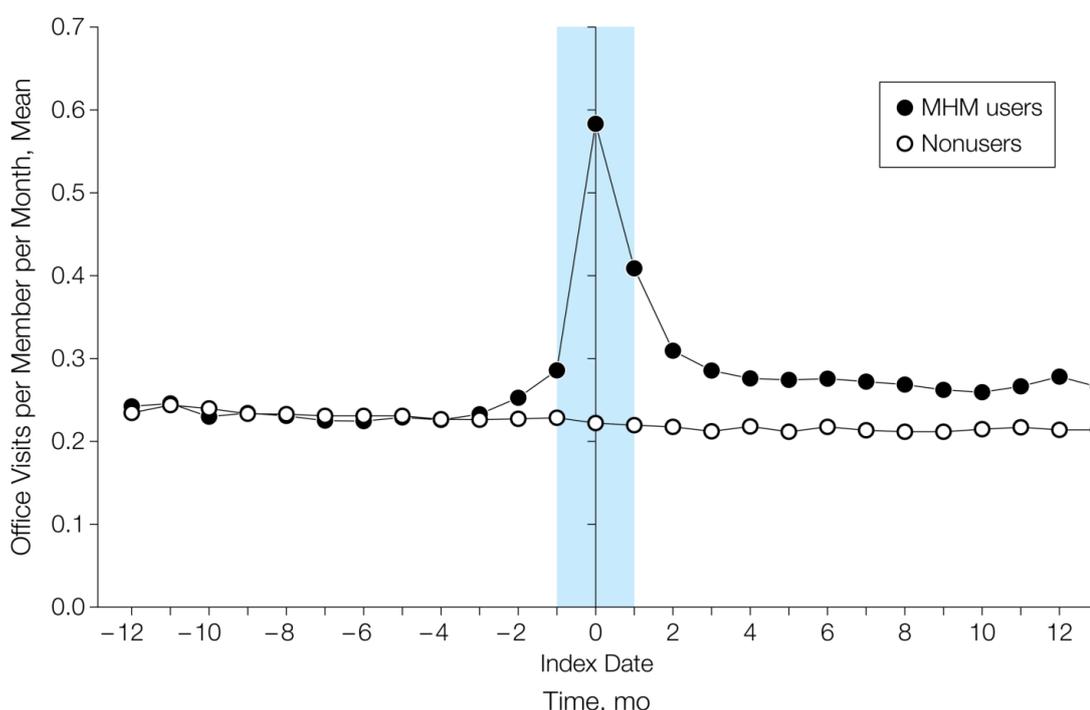
Note – Each data point represents mean office visits from the preceding 30 days. The tinted area indicates a 30-day period on either side of the index date.

One 12 month quasi-experimental study on diabetes conducted in the USA and Puerto Rico showed reduction in hospital admissions (+27% in control group vs. -60% in intervention group), bed days of care (+32% vs. -68%), emergency room visits (+22% vs. -66%). However it is not clear whether these findings were significant. Patient satisfaction was 97% and provider satisfaction was 100% [Kobb 2003].

A RCT (12 month duration), including 104 patients in the USA with chronic disease indicated significant reduction in total number of urgent visits (+5 vs. – 83), however there was no difference in the total number of nurse visits during the study period (Noel 2004).
A retrospective audit of 10,746 adults with diabetes showed that better diabetic profile in PHR users compared to non-users was likely secondary to their engagement with their health rather than the PHR itself [Tenforde 2011].

A case study of diabetic patients using a web portal in the US indicated over three years there was no change in the number of patients seen per month or in the number of phone calls received (Hess 2007).

The number of electronic messages increased through the years (from less than 50 per month to over 400 messages per month).

Another case study of paediatric and adolescent PHR use showed user login an average of 6.3 times in 3 months of a pilot study [Bourgeois 2009]. The most frequently used function was lab results (82%). Users had also accessed/edited medication list (5.4%), problem/allergy/immunisation (1 – 2%).

A survey of 12,793 diabetic patients in the Netherlands on the use of a web portal found a lack of interest or time (18.1%), questions too difficult (6.6%), too personal (3%) and other reasons (22%) as reasons why patients were not willing to participate in the study (Ronda 2013). Of the respondents 45.5% had a login and they were significantly younger (59.7 years vs. 67.4 years). Type 1 diabetics had significantly more access than Type 2 patients (89.8% vs. 41%). In both groups, patients were more likely to have a login if there were younger and more frequently treated by an internist.

A cross sectional web based survey on the willingness to share paediatric patient information found that of the 261 respondents (56% response rate), more reported they would share all information with the state/local public health authority (63.3%) than with an out-of-hospital provider (54.1%) [Weitzman 2012]. However a few would not share any information with these parties (respectively, 7.9% and 5.2%).

5.3.6. Impact of PHRs on costs of running the service / cost effectiveness

No. / Type of studies included: 1 SR (60 RCTs/quasi-RCTs reported in 77 publications).

Diseases included in analysis: Smoking cessation.

Cost-threshold analyses in a smoking cessation SR indicated some form of electronic intervention is likely to be cost-effective when added to non-electronic behavioural support, but there is substantial uncertainty with regards to what the most effective (thus most cost-effective) type of electronic intervention is, which warrants further research [Chen 2012]. Expected value of perfect information (EVPI) calculations suggested the upper limit for the benefit of this research was around £2000–3000 per person.

6. Review results on other aspects

H.4. Quality assurance and assessment methods for Apps & PHRs

No empirical studies have yet been identified relevant to this section. However, drawing on our experience, the following table lists the eight main methods available to improve the quality of apps or related eHealth tools, the likely resource implications of each for the health and care system, potential advantages and disadvantages and an example of each approach.

Method	Resource implication	Other advantages	Disadvantages	Examples
Informal user rating ("Wisdom of the crowd")	Low	Simple user ranking	Hard for users to assess quality; likelihood of click factory bias	Current app stores, MyHealthApps
Users apply explicit quality criteria	Low	Explicit approach	Requires widespread dissemination; uncertainty about who can apply them	RCP checklist
Classic peer reviewed article	Low	Rigorous (?)	Slow, resource intensive, doesn't fit App model	47 PubMed articles on app quality so far
Physician peer review	Low	Timely Dynamic	Not as rigorous Scalable?	iMedicalApps, MedicalAppJournal RCP checklist (Appendix 5)
Developer self-certification	Low	Dynamic	Requires developers to understand & comply; checklist must fit apps	Based on Health on the Net Foundation Code of Conduct (HON Code)? RCP checklist
Developer support	Medium	Low resources required	Technical knowledge needed Multitude of developers	British Standards Institute (BSI) Publically Accessible Standard (PAS) 277
CE marking, external regulation	Low	Credible	Slow, expensive, apps don't fit national model	US Food and Drug Administration (FDA), MHRA
Curated app store	High	Credible	Is it scalable? Perceived as useful by all users?	NHS App Store
Centrally managed app distribution	High	In use in some enterprise settings	Not scalable – may work on local level. Requires investment in mobile devices	Beth Israel Hospital

Health apps are regulated to a certain extent, and for specified types of clinical use they must be CE marked. Use of generic apps in a clinical context, for instance for calculations that could be done on an

office calculator, are not regulated. The CE marking regulation is a weak hurdle that is easily overcome, with many health apps being particularly weak on detecting user error. Other problems include app use being interrupted by the tablet or phone's other apps for other purposes, such as getting a text message while using a health app. Or Wi-Fi can drop, the battery can go flat, and some apps may restrict use until fees are paid — all of these are likely to be inappropriate interruptions during clinical use. The RCP has recently published guidance on health apps that covers these practical issues (https://www.rcplondon.ac.uk/resources/using-apps-clinical-practice-guidance).

Clear identification of high quality apps by assessment using appropriate criteria should enable clinicians, patients and other healthcare professionals to make an informed decision of which apps to choose for a particular situation.

A recent stakeholder meeting of about 70 people, ranging from public authorities, ICT industry, and academia to healthcare professionals was held to discuss the challenges facing the app market under the auspices of EU eHealth week in Riga (http://ec.europa.eu/digital-agenda/en/news/mhealth-green-paper-next-steps). Delegates were asked to vote on which model for improving the safety and quality of lifestyle and well-being apps was most acceptable. The results from the 58 people who voted are shown in the following graph.
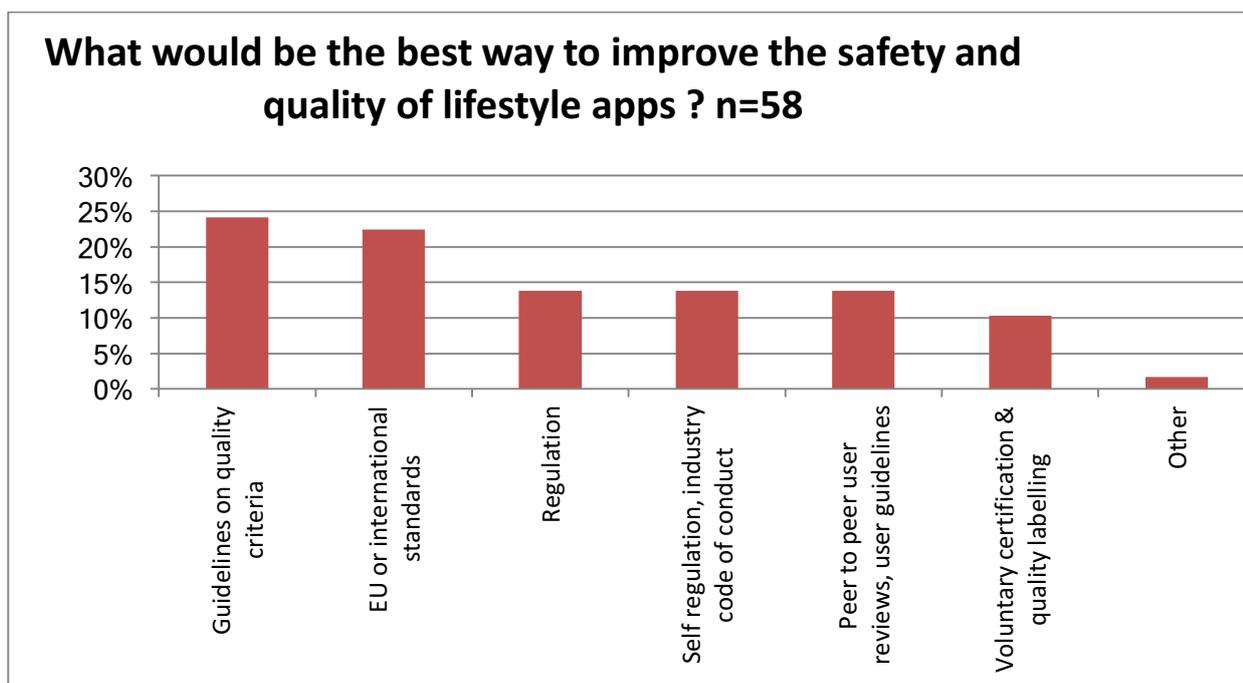
Unfortunately, the options presented are not mutually exclusive (eg. guidelines on quality criteria and European or international standards are closely related, while self-regulation/ industry code of conduct and voluntary certification and quality labelling are both actions that app developers can take). However, it does appear that, to the 58 members from this mixed stakeholder group who voted, "guidelines on quality criteria" (24% of votes) and "EU or international standards" (22%) were the most acceptable strategies and together accounted for 46% of the votes, while regulation, self-regulation (industry code of conduct) and peer to peer reviews or guidelines for users (at 14% each) also gained significant support. However, if one added together self-regulation (14%) and voluntary certification and labelling (10%) this

adds up to 24% of the votes, so becomes the joint leader with guidelines on quality criteria. These two actions are also complementary.

6.2. The variety of health apps, app taxonomy and features relevant to assessment

We have not yet identified a useful paper setting out a taxonomy of health apps likely to be relevant to the NIB 1.2 Core Group. However, from our experience, it may be useful to describe several broad classes of health apps for each broad user group, including:

For the public:

Risk calculators allow members of the public to assess their risk of eg. Heart attacks to motivate them to change their behaviours.

Pregnancy calculators and advisor apps are a large and commonly used category.

Behaviour change apps eg. For smoking cessation help people to monitor and improve health-related behaviours and manage risk factors.

Fitness devices, such as wristband sensors, can monitor physical activity and sleep patterns and have been used successfully in rehabilitation.

Crowdsourcing apps tap into a network of people suffering similar conditions and who may provide advice and remedies. These may be useful for exploring options but may mislead users.

For patients:

Self-monitoring/self-management apps may remind patients to take medications, record measurements such as blood glucose or blood pressure, or collect information on mood changes.

Medical portal apps allow patients to log into patient record databases and view lab and test results, request prescriptions and sometimes communicate with professionals.

Disease monitoring apps are typically used for chronic conditions such as diabetes and hypertension. If used successfully, these apps should reduce visits to GPs and hospitals.

Remote consultation apps such as for stroke provide rapid preliminary diagnosis by facilitating consultation with an expert, either for the patient or from GP or junior doctor to a consultant. These are related to **virtual doctor** apps which provide quick 24/7 (paid) access to a nurse or doctor. These apps are familiar in the US, but there are concerns that the provider may over-treat.

Virtual counselling apps are similar to remote consultations apps by facilitating consultation with a therapist used in mental illness and conditions such as post-traumatic stress disorder and anorexia.

For health professionals:

Information apps such as the British National Formulary (BNF) or National Institute for Health & Care Excellence (NICE) guidelines app act as reference sources.

Medical calculators eg. for drug dosage, body surface area, BMI etc.

Risk calculators allow clinicians to assess the risk posed by a patient before deciding on a course of action such as a prescription.

Decision support apps allow the clinician to enter patient data (or link to that patient's record) and deliver advice or interpretation based on these data. They may be based on an encoded clinical practice guideline or other source of computer-readable knowledge.

General purpose apps, include PDF viewers can be used for reading medical textbooks, email for sending patient pictures and notes, calculators, word processors, cameras (eg., photographing rashes for help in diagnosis), providing information to users and colleagues and getting advice in *ad hoc* ways. General purpose apps are not regulated as health apps *per se*, but may be used in ways (e.g, managing patient data) that are medical in the sense of the European regulations.

While the type and complexity of the app is one indicator of clinical and other risks associated with app use, and thus the need for assessment, the context of use also needs to be taken into account [Lewis & Wyatt 2014]. Thinking about app attributes that may be important to the assessment process leads to a list of attributes that may be relevant to either the intensity of assessment required, or the user's need for an assessment. See table below which are inferences based on evidence and practices seen throughout the mobile app ecosystem.

Fundamental app features relevant to app assessors or users of assessments (Source: Lewis & Wyatt 2014 and authors of this report)

App feature or attribute	Why important to assessor	Why important to user or assessment
Clinical risk associated with use of the app	Requires that more effort is taken with assessment, more certainty needed	Users will place more emphasis on assessment for high risk apps
Complexity of the app	Adds cost to the assessment process	Increases uncertainty about the app, thus reliance on assessment
Transactions eg requesting prescriptions, booking appointment	Makes it more complex to assess	Makes it more useful to the user
Linkage with NHS records / spine services	Makes it much more complex to assess	Makes it more useful to user
Does app capture / store identifiable patient data on device or Cloud	Need to check data protection	Makes assessment more necessary
App is (or should be) classified as a medical device (ie. Supports diagnosis, treatment etc.)	MHRA approval/CE mark necessary	If CE marked, assessment not necessary
Free or paid	Paid apps generate revenue directly that can fund assessment process	Paid makes assessment more necessary
NHS funder / publisher	If NHS, no assessment needed – but reputational risk	If NHS, assessment not necessary
Rarity of the condition the app supports	If rare, fewer experts available to advise on assessment process	
Source of the information provided by the app	If all sources are already assessed (eg. NICE, Patient Information Forum), little further work may be needed	If not well known, makes assessment more necessary
Whether app provides generic information (like a textbook) or personalised output / advice (like an advisor)	Assessing an app that provides personalised advice carries a higher risk and legal liability for the assessor than a textbook-like app	If personalised output, makes assessment more necessary
Whether there is a published study of the accuracy of the advice / output generated or the impact of the app	Published studies make the accreditation process easier	

App feature or attribute	Why important to assessor	Why important to user or assessment
Whether there is a robust quality assurance process In the development of the app	Use of robust quality assurance method eg BSI (PAS 277) make the accreditation process easier	

This in turn leads to a list of some sample functions that are included in health and clinical apps, in ascending order of risk, with 11 being highest risk (source: authors of this report):

1. Display of local static information (maintained within an organisation eg. hospital)

2. Display of static information from a recognised external source eg. NICE guideline, NHS Choices

3. A game to help users learn more about a condition – using recognised content

4. Display of simple measurement, eg. heart rate

5. Capture and logging of a behaviour eg. step count

6. Helping the user access a condition-specific forum and share their experiences

7. Calculation of a simple result with minimal risk of miscalculation from 2 or more easily obtained variables, eg. BMI from data that user enters

8. Sharing of anonymised user results with others as a league table to encourage competition

9. Calculation of a complex result with potentially serious consequences, eg. the personal risk of a cardiac event over next 10 years

10. Giving the user advice about what to do to reduce their risk, suggesting a revised insulin dose etc.

11. Allowing the user to control a complex medical device requiring significant expertise and with potentially fatal consequences eg. ventilator, radiotherapy machine, surgical robot.

6.3. Suggested evaluation cascade for apps and PHRs

The evaluation methods used need to be appropriate to the risk and stage of development of the technology. This is illustrated in the table below (source: based on Wyatt & Spiegelhalter1990; Wyatt & Altman 1995, Wyatt 1997; Wyatt 2000), which proposes appropriate evaluation methods for low, medium and high risk clinical apps. Note that the criteria are cumulative, ie. For a high risk app, all of the criteria for low and medium risk apps need to be fulfilled before checking for impact and software engineering.

Risk	Area	Topics	Methods
Low	Use case	• Purpose, sponsor • User, cost	Inspection (end user appraisal of a particular app for use in a particular setting for a specific purpose)
	Content	• Based on sound evidence? • Proven behaviour change methods?	Inspection
	Safety	• Data protection • Usability	Inspection
Medium	Safety	• Usability	HCI lab/user tests

		• Interoperability	Adherence to standards
	Accuracy	• Calculations • Advice	Scenario-based studies
High	Impact	• On knowledge, attitudes, self-efficacy • On health behaviours, outcomes	Within-subject experiments on decisions, self-efficacy Field trials of impact
	Software engineering	• Were appropriate tools used? • Is the code well written?	Inspect team credentials etc. Inspect source code

6.4. General issues and digital healthcare myths

6.4.1. The challenges of evaluating evolving technology

While the increment in benefit resulting from using each new version of an app or website may be small (Wyatt 1998), there is a perception that evaluation studies on technologies that were common even five years ago may have limited applicability today (Liu & Wyatt 2012). Several consequences flow from this:

Rapid cycle evaluation may be preferable to our current slow model for carrying out studies.

Technology developers and researchers should focus evaluation studies on uncovering generic principles that govern the acceptability, usability and effectiveness of these technologies [Wyatt 2009]. Examples include how to display risk to ensure that users act appropriately; how to provide feedback to users on their behaviour (eg. as the daily or total weekly step count; as an absolute figure or percentage of target; for that individual or for a team?) to maximise behaviour change. These "design principles" will not change over time, as they concern inhere the psychology of users rather than fleeting aspects of the technology.

Internet-based randomised trials ["A-B testing"] can be carried out extremely fast: eg. [Nind et al 2011] – we overshot our recruitment target of 850 in one study by failing to close recruitment soon enough, and recruited a total of 900 participants in 5 days

All the major e-commerce and social media providers use A-B testing techniques to refine and improve their websites, often in hours they can randomise tens of thousands of users to two or more versions of their site and find out which leads to more clicks or income. The health and care system should probably invest in tools to enable similar studies to be carried out using eg. the NHS Choices website.

6.4.2. Could digital health worsen the cyber divide ?

Worsening of health inequalities is certainly a risk if common sense is ignored in the design and delivery of digital health systems and services. However it can be argued that only digital channels will provide the kind of tailored, use anywhere, immersive experience needed to draw some hard-to-reach people into the health system. See table below (source: the authors):

Ways to ensure that digital health services lessen health inequalities	Ways that digital health could worsen health inequalities
Adopt a multi-channel strategy including delivery via low tech channels eg. SMS, teletext, email, textphone. (NB. High market penetration of portable/tablet devices and interactive television)	Tendency to design for high tech high cost channels and media, eg. high bandwidth broadband, apps on smart phones

Use accessible interfaces and media (eg. video) for people with learning or sensory difficulties or limited reading skills	Poorly designed user interfaces with complex structure or language
Bring health services to people in familiar environments at home / at work / at leisure	Failure to take advantage of the potential geographical separation of health facilities and services (eg. kiosks in GP practices)
Integrate digital health services into online games, social media, other online activities	Failure to integrate with other online services / activities
Provide immediate user gratification for health related behaviour change via points, league tables, social recognition etc. ("Gamification")	Failure to recognise the need of many users for immediate gratification to reward small achievements
Seek opportunities to co-design, co-produce and co-deliver health services with specific user groups	Assume that older people and those with less education will shun digital media; failure to design with their interests in mind

This implies that developers and those who assess digital health services need to consider:

- Cost of devices to access the service

- Inclusive interface design

- Accessibility guidelines such as those from the World Wide Web consortium, W3C.

The author suggests carrying out a "health equality impact assessment" for every significant new digital health service, analogous to a privacy impact assessment.

6.4.3. How can digital health improve the last year of life?

Superficially, it appears that people entering the last years of life – many of whom are likely to be elderly or have a serious long term condition such as diabetes, with reduced dexterity and eyesight or memory problems – could avoid use of digital health services. However, for the reasons argued in the earlier section about cyber divide, we would dispute that. Some specific arguments are:

1. A proportion of those in the last year of life are quite fit and die of a sudden illness such as a stroke or heart attack. The arrival of this terminal event may actually be delayed for months or years as a result of using lifestyle or self-management apps.

2. Even if the person themselves may be unable to use a digital service, many people have close family or friends who can use these services on their behalf. There are also charities such as Net Neighbours [www.ageuk.org.uk/york/our-services/net-neighbours, [Blythe et al 2012] who arrange for a volunteer to access internet services for those who are unable to do this for themselves. In the authors opinion an implication for digital service designers is to anticipate that a proportion of users will be acting on behalf of another person.

6.4.4. How much is information governance a real barrier ?

There is concern that current information governance is "acting as a brake" on innovation in digital healthcare. However, others observe that there is currently a free for all in data sharing from apps etc. with only partial/presumed consent: "*Even as you're reading this, your smart phone can reveal your location… data brokers are going to know more about us than we know ourselves*". [Madhumita Venkataraman, Wired, November 2014]. The recently announced EU Code of Conduct on mobile health

apps [http://ec.europa.eu/digital-agenda/en/news/mhealth-green-paper-next-steps] covers privacy and security as well as some other aspects of quality, so may help address current fears about lack of effective regulation in this area. Implementation in 2017 of the EU Good Data Protection Regulation with its focus on opt-in consent, privacy by design and fines of up to 100M euros should improve the position.

6.4.5.How might the entry of major corporates impact the market?

In the last year we have seen increasing interest in the mHealth market from major corporate players such as IBM, Apple and Google labs. This could bring welcome professionalism to the apps market. For example, Apple has tightened the criteria for health apps to forbid apps that carry out a number of behaviours it considers unacceptable [https://developer.apple.com/app-store/review/guidelines/#healthkit]. These prohibited behaviours include:

- Failure to comply with relevant applicable law for each Territory in which the App is made available,
- Writing false or inaccurate data into Healthkit
- Strong users' health data in iCloud
- Sharing user data from Healthkit to third parties for advertising or data mining purposes (except for health research)
- Sharing user data with third parties without explicit user consent
- Marketing apps that use the Healthkit framework without notifying users of this fact
- Providing an app that uses the Healthkit framework without a privacy policy
- Providing diagnoses, treatment advice, or controlling hardware designed to diagnose or treat medical conditions that do not provide written regulatory approval
- Conducting health-related human subject research without approval from an independent ethics review board and obtaining consent from participants or, in the case of minors, their parent or guardian

The USA is a dominant market, with a very different regulatory and incentive structure, as well as different medical practices (eg. including units of measurement). The way in which US apps are used and deployed in the UK may induce clinical errors, and inappropriate advice or suggestions about treatments that are not licensed in the UK. The US FDA has been a pro-active regulator and may tighten health app regulation, but it remains to be seen whether this has any impact on the diverse market, let alone has any impact outside of the USA.

6.4.6. The influence of technology and health trends

Future trends will include much wider uptake of health apps by patients as well as increasing worries about security and privacy issues, which, because of high profile media interest, have been more prominent than safety worries.

There is considerable enthusiasm for health apps which Wyatt has called "Apptimism" (TEDx talk http://tedxtalks.ted.com/video/Avoiding-apptimism-in-digital-h;search%3Aleeds), and some apps certainly show considerable promise and may even disrupt medical practice in beneficial ways. Yet very little of the discussion around apps is critical or informed, and reliable and objective research is noticeable by its absence. It is to be hoped that more research will be done on appropriate methodologies and that research standards will improve, facilitating more reliable comparisons on apps and their relative benefits.

6.4.7 App development – a multi-disciplinary approach

The RCP held a national workshop on health apps (11 May 2015; funded by EPSRC), deliberately including a wide variety of participants: including app developers, regulators, clinicians, and researchers. Some participants had won prizes for their work.

A key insight emerging from the workshop was the importance of multidisciplinary approaches, in all phases of production — design, development, evaluation, regulation, procurement, post-market surveillance.

The RCP workshop noted that everyone had previously under-estimated the importance of multidisciplinary work before participating in the workshop: being exposed at the workshop to complimentary skills and knowledge highlighted everyone's blind spots. This was a surprise: everybody at the workshop previously thought they were experts!

For example, software engineers can make critical contributions to apps, but one may not know this until one has worked with competent software engineers. Other areas of concern were the frequent absence of participatory design, user centred design (UCD), and rigorous methodologies applied in evaluation.

The workshop also noted that the common app development model ("a couple of clinicians and a couple of developers"), though it may be fine for innovation, is inadequate in the long run.

The workshop participants are currently collaborating on writing a brief "app manifesto" explaining the necessity and value of multidisciplinary perspectives.

(The workshop took the Mersey Burns app and its peer reviewed publications as a "worked example" and this resulted in some of the observations on Mersey Burns described in the section on usability, above.)

7. Conclusions

7.1.Summary of the main results

Summary of evidence on acceptability, usability and impact of the use of apps by patients/the public

There is some evidence from surveys about the factors that make apps more acceptable to patients. There is also some evidence about usage rates for apps in specific areas such as diabetes, but there is no study that allows us to predict the likely influence of the proposed national app assessment programme on patient/public app download or usage rates., let alone effectiveness.

The usability of apps for patients remains very variable, with many examples of poor usability. In addition, there are several studies showing poor accuracy of apps that attempt to make a diagnosis (eg. of melanoma from images of pigmented skin lesions) or predict risks from clinical features (eg. of cardiac risk prediction for patients).

There is some evidence from randomised trials (RCTs) about the positive impact of apps on patient decisions and actions, including behaviour change. For example, one RCT showed significantly increased weight loss over 6 months in users of the MyMealMate app. There is also some evidence of the modest benefit of patients using apps on clinical outcomes, eg. on HbA1C in diabetics. However, there is no rigorous evidence about the impact of patients using apps on their utilisation of health service resources.

Summary of evidence on acceptability, usability and impact of the use of apps by clinicians

We have included early results of a recent internet survey of app use by 1104 physician respondents to the RCP Panel survey (panel size 2658, response rate 42%). Overall, 54% of respondents use apps to support their clinical work (especially to look up guidelines for risk assessment and to inform prescribing); and rated these as either "Very important" (29%) or Essential" (13%) to their clinical work. The biggest concerns about clinical apps are about the quality of the underlying evidence base on which the app is based (43%) and the accuracy of information generated (43%); information governance is also a concern (25%). The majority of respondents (73%) felt that the RCP should be responsible for recommending apps that are of high quality and safe to use; many respondents also expected specialist societies (69%) or the NHS (50%) to play a role in app recommendation.

There is some concern about the accuracy of apps intended to be used by clinicians, for example for drug dose conversion or calculation of risk scores. This may be explained by the low rates of engaging clinicians, software engineers and usability experts, in the design and testing of apps. The recent move by the Royal College of Physicians to remind clinicians that they should only use CE marked apps for such purposes and the proposed national assessment process, emphasizing testing of the accuracy of such apps before use by clinicians, is likely to cause app developers to engage more clinicians in the development and testing process.

There is clear evidence that the design and usability of apps for clinicians – even those that have passed CE marking – fails to take account of well-known human factors issues. We gave several examples for well-

known CE marked apps, with brief critiques of screen shots. We were unable to locate any studies of the impact of app use by clinicians on the quality and safety of care nor on efficiency / resource utilisation.

Summary of evidence on acceptability and impact personal health records – PHRs

The evidence on patient decisions and actions indicates better adherence to medical advice, enhanced empowerment and better communication with the use of PHRs. This evidence is largely based on single studies of acceptable quality. Problems with web portal, technology availability and poor socio economic and educational status are seen as barriers. The evidence on behavioural change is more robust and includes large SRs across different chronic conditions. Significant positive behavioural changes, e.g. smoking cessation, increased physical activity, mood improvement in cancer patients and better health promotion are seen with the use of PHRs. The effects are largely seen in the shorter duration with studies reporting significantly decreasing effects with time.

The largest volume of evidence on PHRs is on patient clinical outcomes. Various positive outcomes have been outlined – including significant reduction in HbA1c levels in Diabetic patients, improvement in daily functional effect and moderate improvements in quality of life. Although occasional studies report no effect, the vast majority are positive, especially for HbA1c reduction. The reports on other outcomes such as cholesterol improvement and blood pressure control have not shown any significant benefits.

Evidence on the cost of running PHRs / cost effectiveness as well as the decisions and actions taken by clinicians is scarce and we are unable to provide with conclusive statements on these measures.

Younger patients were more likely to access PHRs and many studies have reported the access of laboratory test results as the most useful feature of the PHRs. Patient satisfaction was generally high across the majority of the studies and individual studies have identified a list of barriers to usage. These frequently included poor technology access, elderly and low educational status. The utilisation of healthcare services has often been reported positive in small individual studies showing evidence of reduction in admission rates and emergency visits. However a large US based study has found that there could be significant increase in patient visits, telephone calls and hospitalisations.

Summary of other findings and insights

To support the health and care system in developing robust methods for assuring quality and assessing apps and related products, we provide:

- A table listing nine potential methods for assuring the quality of apps and web delivered services, the likely implications for the health and care system of adopting each of these, and the pros & cons of each (our work, no published evidence)

- The results of recent survey of EU stakeholders on their preferences about app assessment and quality improvement methods

- A taxonomy or ontology of apps etc. that we developed to identify those factors most useful to the organisation running an assessment process, and to the users of assessed products

- Our analysis of a number of "digital healthcare myths"

- Our proposals for a risk-related evaluation strategy.

7.2. Key review findings relevant to the proposed NIB 1.2 assessment model

The table below brings together material already presented in the literature review and some new material, where relevant.

Feature of the proposed NIB 1.2 assessment model	Relevant review finding, or new evidence identified
Need for an assessment process	Many articles demonstrate the poor quality of the underlying evidence base, lack of health professional involvement and some evidence of inaccuracy in health related apps. The survey responses of 58 stakeholders attending the recent 2015 EU eHealth week in Riga confirms these concerns and the need for some kind of assessment process, with 24% voting for guidelines on quality, 22% for EU or international standards, and 14% each for regulation, self-regulation and peer to peer user reviews. 10% voted for voluntary certification. [http://ec.europa.eu/digital-agenda/en/news/mhealth-green-paper-next-steps]
Designing the process to be scalable	In June 2015 there were 1.5M apps on Google Play, 1.4M on apple, and 830k apps on other app stores [http://www.statista.com/statistics/270291/popular-categories-in-the-app-store/]. Apple's app store alone sees about 2 billion downloads per month (100 billion cumulative since 2008) and is growing by about 40,000 new apps per month, or more than 1000 per day. Lifestyle apps accounted for 8.6%, health and fitness apps for 2.8% and health apps for 2.1% of all apps, making 13.5% of all apps or about 503,000 health-related. This demonstrates the need for a scalable assessment process.
Making the process risk-related	Conforms with recommendations of Better Regulation Task Force 1997 [https://en.wikipedia.org/wiki/Better_Regulation_Commission]. A risk framework [Lewis & Wyatt 2014] for app usage and an "App Usage Factor" to quantify population app impact [Lewis & Wyatt JMIR, in press] have been proposed.
Developer self assessment	Self-assessment is widely used in the regulation of low risk Class 1 medical devices; no evidence was found about the truth or otherwise of self-declared product features, but random checks by authorities are often advocated.
Crowd sourcing, community feedback	Some evidence found for the validity of health-related crowd sourcing for research purposes, eg. use of SMS messages to declare infant feeding methods [Whitford 2012]. Some evidence of biased crowd sourced data in other areas eg. ecology. Risk of manufacturers recruiting click factories or 'Black hat' web optimisation methods to boost product reputation [David Segal. The dirty little secrets of search. New York Times 12/2/11, http://www.nytimes.com/2011/02/13/business/13search.html?_r=0]
Independent impact evaluation	Garg demonstrated in their SR of 100 RCTs of decision support systems [Garg 2005] that RCTs conducted by independent evaluators are about one third as likely to show improved clinical processes or outcomes compared to RCTs conducted by system developers – 28% versus 74% success, respectively, p = 0.001
Need for evaluation support	Many published studies of apps etc. suffer from serious biases or restricted generality, so external expert evaluation support is likely to help improve study quality and rigour. It may also help bring a wider perspective, eg. health economic, patient and professional concerns, leading to studies that address a wider range of questions.
Involvement of patients / public at all stages	Involving patients in planning and conducting research improves research quality [Goodare 1999]
What to assess	The interim results of the RCP survey of 1100 physicians supports the idea of assessing the quality of the evidence underlying apps as well as the accuracy of app outputs / advice

Who to assess it	The RCP survey of 1100 physicians supports the idea of professional organisations quality assessing apps and related products for physicians, with rates of trust in assessment by the College (73%) or specialty societies (69%) being higher than trust in the NHS (50%). There is no evidence about whom the public and patients would trust as a source of assessments.
How to assess – new models needed	While the need for new evaluation methods is often repeated by technologists, there are powerful arguments (ranging from the philosophical to practical) marshalled by [Liu & Wyatt 2012] to retain standard methods such as randomised trials, even when the intervention of interest is new technology. AB testing is merely a rapid randomised study conducted using the internet to complete the study in hours not months; while regulation and ethics are light touch, the study design and data analysis methods are the same as for conventional RCTs.

7.3. Review limitations and comments on the quality of studies and on the literature

The evidence base regarding these services is extremely limited. This is due to the relatively recent development of these services in healthcare provision and inconsistent nomenclature, which makes the identification of relevant papers challenging. It was also difficult to compare and contrast findings from many RCTs, as there was significant variation in the intervention components, the outcome studied or the population targeted. Consequently there were few useful meta-analyses.

We noted a number of other significant limitations during the literature review, including:

1. Out of date technology or out of date searches in SRs. A common issue encountered throughout the review was the limited lifetime of the evidence base. Many SRs (particularly of web based interventions) included studies that are 5-10 years old. The rapid rate of development of web technology, mobile apps and mHealth interventions has previously been identified as a major issue for those engaged in this research. While the increment in benefit resulting from each of new version of an app or website may be small, there is a perception that evaluation studies of technologies that were common even 5 years ago may have limited applicability today.

2. Inconsistent naming of interventions. A similar issue is the inconsistent naming by researchers and indexing by librarians of these technologies; this has previously been identified as an inhibitor to developing a clear overview of the evidence [Lewis et al 2014]. This risk was reduced by repeating searches using a range of common nomenclature terms

3. Developer involvement in studies: Many of the RCTs and other studies cited identified were carried out by developers of the application/PHR etc. This is a potentially significant sources of bias, which could lead to over estimates of benefit by a factor of three [Garg et al 2005]. This highlights the need for independent evaluation of mHealth interventions, and clear declaration of competing interests in journal articles.

4. Recruitment and drop-out rates. A number of studies were limited by poor recruitment rates in the target population, high drop-out rates and limited follow up durations. These are known causes of bias in RCTs and limit the usefulness of a study. We also know that short follow up durations are unlikely to be helpful as there is a large fall off in adherence to eg. smoking cessation interventions with time.

5. Clinical significance versus statistical significance. Many of the larger studies generated statistically significant outcomes, however a small but statistically significant difference may not outweigh the logistical and other challenges of implementing the intervention, ie. The result may not be clinically significant. There is often a lack of data regarding how big a change in each outcome measure would

be clinically useful. It is therefore often difficult to assess the likely practical impact of a statistically significant finding on a target population.

6. Use of non-validated outcome measures. A number of studies used non-validated instruments to measure outcomes, eg patient self-report for smoking cessation rather than the recognised gold standard of breath carbon monoxide or saliva cotinine. This introduces a level of uncertainty into that study's results.

7. Duplicate inclusion of studies across SRs: It is important to note that due to the methodology of SRs, a number of studies are cited by more than one SR so appear more than once in our summary of these.

7.4. Identified research questions and suggestions for future work

To provide a more evidence-based response to many of the questions arising in this report, the following research topics need to be addressed:

1. The likely influence of the proposed national app assessment programme on patient/public app download or usage rates; specifically, who would the public and patients trust as a source of assessments ?

2. How to improve the usability, reliability and accuracy of apps intended for public use in potentially safety critical scenarios

3. The impact of patients using apps on their utilisation of health service resources.

4. The factors that influence app use by clinicians

5. The accuracy of apps intended for use by clinicians

6. The impact of app usage by clinicians on patient outcomes and healthcare resource utilisation

7. The impact of PHR use on patient outcomes and resource utilisation in the UK

8. Which PHR functions are most useful and contribute to clinical and efficiency impacts

9. How to encourage more clinicians to engage in patient centred PHRs, eg. to respond to patient messages (in a US study, only 14% of health professionals used the patient PHR at least once a day).

We suggest that NHS England discuss setting up a programme of R&D to address this list of questions with the NIHR, as answers are needed before the health and care system can move confidently towards the intended digital healthcare vision. Such evidence will also be a major driver of patient and clinical adoption of these technologies and the confidence of service commissioners, as well as informing NICE guidance production processes.

References

(In alphabetical order by surname of first author)

Abroms LC, Padmanabhan N, Thaweethai L, Phillips T. iPhone apps for smoking cessation: a content analysis. *American Journal of Preventative Medicine* 2011; 40(3):279-85.

Abroms LC, Lee Westmaas J, Bontemps-Jones J, Ramani R, Mellerson J. A content analysis of popular smartphone apps for smoking cessation. *American Journal of Preventive Medicine* 2013; 45(6):732-6.

Adelman CB, Panza KE, Bartley CA, Bontempo A, Bloch MH. A meta-analysis of computerized cognitive-behavioral therapy for the treatment of DSM-5 anxiety disorders. J Clin Psychiatry. 2014 Jul;75(7):e695-704. Doi: 10.4088/JCP.13r08894.

Aguilar-Martínez A, Solé-Sedeño JM, Mancebo-Moreno G, Medina FX, Carreras-Collado R, Saigí-Rubió F. Use of mobile phones as a tool for weight loss: a systematic review. J Telemed Telecare. 2014 Sep;20(6):339-49. Doi: 10.1177/1357633X14537777. Epub 2014 May 29.

Alzheimer's Association. *Alzheimer's Caregiving – There's An App For That!* http://www.alzheimersblog.org/2013/08/19/alzheimers-caregiving-theres-app-that/ (accessed 27 April 2015).

Amante DJ[1], Hogan TP, Pagoto SL, English TM. A systematic review of electronic portal usage among patients with diabetes. Diabetes Technol Ther. 2014 Nov;16(11):784-93. Doi: 10.1089/dia.2014.0078. Epub 2014 Jul 7.

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. *Journal of Medical Internet Research* 2014; 16(4):e104.

E Årsand, D H Frøisland, S O Skrøvseth, T Chomutare, N Tatara, G Hartvigsen, J T Tufano, "Mobile Health Applications to Assist Patients with Diabetes: Lessons Learned and Design Implications," *Journal of Diabetes Science and Technology*, 6(5), 2012.

J Barnes, A Duffy, N Hamnett, J McPhail, C Seaton, K Shokrollahi, M I James, P McArthur, R Pritchard Jones, "The Mersey Burns App: evolving a model of validation," *Emergency Medicine Journal*, 2014, doi: 10.1136/emermed-2013-203416

Baskerville NB, Struik LL, Hammond D, Guindon GE, Norman CD, Whittaker R, Burns CM, Grindrod KA, Brown KS. Effect of a mobile phone intervention on quitting smoking in a young adult population of smokers: randomized controlled trial study protocol. *JMIR Res Protoc* 2015; 4(1):e10.

H. Bastawrous, H. K. Rono, I. A. T. Livingstone, H. A. Weiss, S. Jordan, H. Kuper, M. J. Burton, P, "Development and Validation of a Smartphone-Based Visual Acuity Test (Peek Acuity) for Clinical Practice and Community-Based Fieldwork," JAMA Ophthalmology, published online May 28, 2015. Doi 10.1001/jamaophthalmol.2015.1468

Bierbrier R, Lo V, Wu RC. Evaluation of the accuracy of smartphone medical calculation apps. J Med Internet Res. 2014 Feb 3;16(2):e32. Doi: 10.2196/jmir.3062.

BinDhim NF, McGeechan K, Trevena L. Who Uses Smoking Cessation Apps? A Feasibility Study Across Three Countries via Smartphones. *JMIR mHealth uHealth* 2014; 2(1):e4.

Blythe M, Monk A. Net Neighbours: Adapting HCI Methods to Cross the Digital Divide.
http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.128.9059&rep=rep1&type=pdf

Bourgeois FC, Mandl KD, Shaw D, Flemming D, Nigrin DJ. Mychildren's: integration of a personally controlled health record with a tethered patient portal for a pediatric and adolescent population. AMIA Annu Symp Proc. 2009;2009:65–9.

Bourgeois FC, Taylor PL, Emans JS, Nigrin DJ, Mandl KD. Whose personal control? Creating private, personally controlled health records for pediatric and adolescent patients. Journal of the American Medical Informatics Association. 2008;15:737–743.

Brooke MJ, Thompson BM. Food and Drug Administration regulation of diabetes-related mHealth technologies. *Journal of Diabetes Science & Technology* 2013; 7(2):296-301.

P. Cairns, "HCI... Not As It Should Be: Inferential Statistics in HCI Research," *Proceedings of the 21st British HCI Group Annual Conference on People and Computers*, volume 1, pp195-201, 2007.

P. Cairns, A. L. Cox, *Research Methods for Human-Computer Interaction*, Cambridge University Press, 2008.

M. C. Carter, J. V. Burley, C. Nykjaer, E. J. Cade, "Adherence to a Smartphone Application for Weight Loss Compared to Website and Paper Diary: Pilot Randomized Controlled Trial," *J Med Internet Research*, 15(4):e32, 2013.

Chen YF, Madan J, Welton N, Yahaya I, Aveyard P, Bauld L, Wang D, Fry-Smith A, Munafò MR. Effectiveness and cost-effectiveness of computer and other electronic aids for smoking cessation: a systematic review and network meta-analysis. Health Technol Assess. 2012;16(38):1-205, iii-v. doi: 10.3310/hta16380.

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. *Journal of Medical Internet Research* 2011; 13(3):e65.

Connelly J, Kirk A, Masthoff J, MacRury S. The use of technology to promote physical activity in Type 2 diabetes management: a systematic review. Diabet Med. 2013 Dec;30(12):1420-32. Doi: 10.1111/dme.12289.

Cotter AP, Durant N, Agne AA, Cherrington AL.Internet interventions to support lifestyle modification for diabetes management: a systematic review of the evidence. J Diabetes Complications. 2014 Mar-Apr ;28(2) :243-51. Doi : 10.1016/j.jdiacomp.2013.07.003. Epub 2013 Dec 12.

Cuijpers P, Marks IM, van Straten A, Cavanagh K, Gega L, Andersson G. Computer-aided psychotherapy for anxiety disorders: a meta-analytic review. Cogn Behav Ther. 2009;38(2):66-82. Doi: 10.1080/16506070802694776.

Jesper Dagöö, Robert Persson Asplund, Helene Andersson Bsenko, Sofia Hjerling, Anna Holmberg,Susanne Westh, Louise Öberg, Brjánn Ljótsson, Per Carlbring, Tomas Furmark, Gerhard Andersson. Cognitive 73omputeri therapy versus interpersonal psychotherapy for social anxiety disorder delivered via smartphone and computer: A randomized controlled trial. Journal of Anxiety Disorders 2014**.** Volume 28, Issue 4, , 410–417

Davis Giardina T, Menon S, Parrish DE, Sittig DF, Singh H. Patient access to medical records and healthcare outcomes: a systematic review. J Am Med Inform Assoc. 2014 Jul-Aug;21(4):737-41. Doi: 10.1136/amiajnl-2013-002239. Epub 2013 Oct 23.

Dayer L, Heldenbrand S, Anderson P, Gubbins PO, Martin BC. Smartphone medication adherence apps: Potential benefits to patients and providers. *Journal of the American Pharmacists Association* 2013; 53(2), 172–181.

Demidowich AP, Lu K, Tamler R, Bloomgarden Z. An evaluation of diabetes self-management applications for Android smartphones. *Journal of Telemedicine & Telecare* 2012; 18(4):235-8.

DiabetesMine. http://www.slideshare.net/amytenderich/diabetesmine-patient-survey-results-2013-summit-final (accessed 27 April 2015).

Dickinson WP, Glasgow RE, Fisher L, Dickinson LM, Christensen SM, Estabrooks PA, Miller BF. Use of a website to accomplish health 74omputeri change: if you build it, will they come? And will it work if they do? J Am Board Fam Med. 2013 Mar-Apr; 26(2):168-76.

Donker, T., Petrie, K., Proudfoot, J., Clarke, J., Birch, M.-R., & Christensen, H. (2013). Smartphones for Smarter Delivery of Mental Health Programs: A Systematic Review. *Journal of Medical Internet Research*, *15*(11), e247. Doi:10.2196/jmir.2791

B Ehrenreich, B Righter, DA Rocke, L Dixon, and S Himelhoch. Are mobile phones and handheld computers being used to enhance delivery of psychiatric treatment? A systematic review. J. Nerv Ment Dis 2011 Nov;199(11):886-91

El-Gayar O, Timsina P, Nawar N, Eid W. Mobile applications for diabetes self-management: Status and potential. *Journal of Diabetes Science and Technology* 2013; 7: 247–62.

Eng DS, Lee JM. The promise and peril of mobile health applications for diabetes and endocrinology. *Pediatric Diabetes* 2013; 14(4):231-8.

Etter J. Comparing computer-tailored, internet-based smoking cessation 74omputerize reports with generic, untailored reports: a randomized trial. J Health Commun 2009;14:646e57

European Commission. *Public consultation on the Green Paper on mobile Health.* https://ec.europa.eu/digital-agenda/en/public-consultation-green-paper-mobile-health (accessed 27 April 2015).

Faridi Z, Liberti L, Shuval K, et al. Evaluating the impact of mobile telephone technology on type 2 diabetic patients' self-management: the NICHE pilot study. J Eval Clin Pract. 2008;14:465–469

Farrelly S, Brown GE, Flach C, Barley E, Laugharne R, Henderson C. User-held personalised information for routine care of people with severe mental illness. *Cochrane Database of Systematic Reviews* 2013; 10:CD001711.

Faurholt-Jepsen M, Frost M, Vinberg M, Christensen EM, Bardram JE, Kessing LV. Smartphone data as objective measures of bipolar disorder symptoms. Psychiatry Res. 2014 Jun 30;217(1-2):124-7. Doi: 10.1016/j.psychres.2014.03.009. Epub 2014 Mar 13.

Fonda Stephanie J., Richard J. Kedziora, Robert A. Vigersky, Sven-Erik Bursell, ''Evolution of a web-based, prototype Personal Health Application for diabetes self-management'', Journal of Biomedical Informatics, Volume 43, Issue 5, Supplement 1, Project HealthDesign, October 2010, Pages S17-S21.

Friedman CP, Wyatt JC. Evaluation methods in biomedical informatics (2nd edition). New York: Springer-Publishing, October 2005. 386 pages, ISBN 0-387-25889-2.

Friedman C, Wyatt J. Evaluation & technology assessment. Chapter 11 in Biomedical Informatics: Computer Applications in Health Care and Biomedicine, 4th Ed. (EH Shortliffe, Cimino JJ eds). New York: Springer-Verlag, 2014 (book chapter)

Kathleen Frisbee, Neil Evans, Thomas Houston, Kevin Todd, Thomas English, Erin Borglund. Assessing the Impact of Mobile Health Apps on Family Caregiver Burden Levels and the Factors Predicting Mobile Health App Use
http://www.medicine20congress.com/ocs/index.php/med/med2014/paper/view/2660

Garg AX, Adhikari NKJ, McDonald H et al. Effects of 75omputerized clinical decision support systems on practitioner performance and patient outcomes, A systematic review. JAMA 2005; 293: 1223-38

Ghorai K, Akter S, Khatun F, Ray P. mHealth for Smoking Cessation Programs: A Systematic Review. *Journal of Personalized Medicine* 2014; 18;4(3):412-23.

Glasgow RE, Boles SM, McKay HG, Feil EG, Barrera M Jr. The D-Net diabetes self-management program: long-term implementation, outcomes, and generalization results. Prev Med 2003; 36: 410–419.

Glasgow RE, Kurz D, King D, Dickman JM, Faber AJ, Halterman E et al. Twelve-month outcomes of an internet-based diabetes self-management support program. Patient Educ Couns 2012; 87: 81–92

Goldzweig CL, Orshansky G, Paige NM, Towfigh AA, Haggstrom DA, Miake-Lye I, et al. Electronic patient portals: Evidence on health outcomes, satisfaction, efficiency, and attitudes: A systematic review. Annals of Internal Medicine. 2013;159(10):677-687.

Goldzweig CL, Towfigh AA, Paige NM, Orshansky G, Haggstrom DA, Beroes JM, Miake-Lye IM, Shekelle PG. Systematic Review: Secure Messaging between Providers and Patients, and Patients' Access to Their Own Medical Record. Evidence on Health Outcomes, Satisfaction, Efficiency and Attitudes. VA-ESP Project #05-226, 2012.

Goodare H, Lockwood S. Involving patients in clinical research improves the quality of research. BMJ 1999 Sep 18; 319(7212): 724–725. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1116584/

Graham AL, Cobb NK, Papandonatos GD, et al. A randomized trial of Internet and telephone treatment for smoking cessation. Arch Intern Med 2011;171:46e53.

Grant RW, Wald JS, Schnipper JL, Gandhi TK, Poon EG, Orav EJ, et al. Practice-linked online personal health records for type 2 diabetes mellitus: a randomized controlled trial. Arch Intern Med. 2008;168(16):1776–82. Doi: 10.1001/archinte.168.16.1776.

Green BB, Cook AJ, Ralston JD, Fishman PA, Catz SL, Carlson J, et al. Effectiveness of home blood pressure monitoring, Web communication, and pharmacist care on hypertension control: a randomized controlled trial. JAMA. 2008;299:2857-67.

D A Grimes, "The nomogram epidemic: Resurgence of a medical relic," *Annals of Internal Medicine*, 149(4):273–275, 2008.

D A Grimes, "Response: Where is the evidence that the resurgence of nomograms is harmful?" *Annals of Internal Medicine*, 150(4): 282, 2009.

Gysels M, Richardson A, Higginson IJ. Does the patient-held record improve continuity and related outcomes in cancer care: a systematic review. Health Expect. 2007;10:75–91. Doi: 10.1111/j.1369-7625.2006.00415.x.

Haffey F, Brady RR, Maxwell S. A comparison of the reliability of smartphone apps for opioid conversion. Drug Saf 2013 Feb;36(2):111-117. [doi: 10.1007/s40264-013-0015-0] [Medline: 23322549]

Hamilton AD, Brady RR. Medical professional involvement in smartphone 'apps' in dermatology. Br J Dermatol 2012 Jul;167(1):220-221. [doi: 10.1111/j.1365-2133.2012.10844.x] [Medline: 22283748]

Henderson C, Laugharne R. Patient held clinical information for people with psychotic illnesses. Cochrane Database Syst Rev. 2000 CD001711. Doi:10.1002/14651858.cd001711. PubMed PMID: 10796656

H. Hern, "App fail on iPad grounds 'a few dozen' American Airlines flights," *The Guardian*, 29 April 2015. http://www.theguardian.com/technology/2015/apr/29/apple-ipad-fail-grounds-few-dozen-american-airline-flights

R Hess, C L Bryce, S Paone, G Fischer, K M McTigue, E Olshansky, S Zickmund, K Fitzgerald and L Siminerio, "Exploring Challenges and Potentials of Personal Health Records in Diabetes Self-Management: Implementation and Initial Assessment," *Telemedicine and e-Health*, 13(5), 2007. Doi: 10.1089/tmj.2006.0089

Hess R., Bryce C.L., Paone S., Fischer G., McTigue K.M., Olshansky E., and al. Exploring challenges and potentials of personal health records in diabetes self-management: implementation and initial assessment Telemed J E Health 2007 ; 13 : 509-517

Holbrook A, Thabane L, Keshavjee K, Dolovich L, Bernstein B, Chan D et al. Individualized electronic decision support and reminders to improve diabetes care in the community: COMPETE II randomized trial. CMAJ 2009; 181: 37–44.

Holtz B, Lauckner C. Diabetes management via mobile phones: a systematic review. Telemed J E Health. 2012 Apr;18(3):175-84. Doi: 10.1089/tmj.2011.0119. Epub 2012 Feb 22.

Huckvale K, Morrison C, Ouyang J, Ghaghda A, Car J. The evolution of mobile apps for asthma: an updated systematic assessment of content and tools. BMC Med. 2015 Mar 23;13:58. Doi: 10.1186/s12916-015-0303-x.

Caralise W Hunt. Technology and diabetes self-management: An integrative review World J Diabetes. 2015 Mar 15; 6(2): 225–233. Published online 2015 Mar 15. Doi: 10.4239/wjd.v6.i2.225

Institute for Safe Medication Practices, *List of Error-Prone Abbreviations, Symbols, and Dose Designations*, 2007.

International Standards Organisation, ISO 9241-210: 2010(E), 2010.

H. B. Irvine, H. Russell, M. Manocchia, D. E. Mino, T Cox Glassen, R. Morang, J. M. Gau, A. J. Birney, D. V Ary, "Mobile-Web app to self-manage low back pain: randomized controlled trial," *J Med Internet Research*, **17**(1):e1, 2015 doi: 10.2196/jmir.3130.

Johnston NW, Lambert K, Hussack P, Gerhardsson de Verdier M, Higenbottam T, Lewis J, Newbold P, Jenkins M, Norman GR, Coyle PV, McIvor RA. Detection of COPD Exacerbations and compliance with patient-reported daily symptom diaries using a smart phone-based information system [corrected]. Chest. 2013 Aug;144(2):507-14. Doi: 10.1378/chest.12-2308.

Kim CJ, Kang DH. Utility of a web-based intervention for individuals with type 2 diabetes: the impact on physical activity levels and glycemic control. Comput Inform Nurs 2006; 24: 337–345

Kobb R, Hoffman N, Lodge R, Kline S. Enhancing elder chronic care through technology and care coordination: report from a pilot. Telemed J E Health. 2003;9:189–195

Koel Ghorai, Shahriar Akter, Fatema Khatun, and Pradeep Ray¨ mHealth for Smoking Cessation Programs: A Systematic Review. J Pers Med. 2014 Sep; 4(3): 412–423. Published online 2014 Jul 18. Doi: 10.3390/jpm4030412

J. Knight, *Fundamentals of Dependable Computing for Software Engineers*, Chapman & Hall/CRC Innovations in Software Engineering and Software Development Series, 2012.

Krist A H, S. H. Woolf, S. F. Rothemich, R. E. Johnson, J. E. Peele, T. D. Cunningham, D. R. Longo, G. A. Bello, G. R. Matzke. Interactive Preventive Health Record to Enhance Delivery of Recommended Care: A Randomized Trial. The Annals of Family Medicine, 2012; 10 (4): 312 DOI: 10.1370/afm.1383

Krist AH, Peele E, Woolf SH, et al. Designing a patient-centered personal health record to promote preventive care. BMC Med Inform Decis Mak 2011 Nov 24;11:73.

Kruse CS[1], Argueta DA, Lopez L, Nair A. Patient and provider attitudes toward the use of patient portals for the management of chronic disease: a systematic review. J Med Internet Res. 2015 Feb 20;17(2):e40. Doi: 10.2196/jmir.3703.

Wilma Kuijpers, MSc,1 Wim G Groen, PhD,1 Neil K Aaronson, PhD,1 and Wim H van Harten, MD, PhD. A Systematic Review of Web-Based Interventions for Patient Empowerment and Physical Activity in Chronic Diseases: Relevance for Cancer Survivors. J Med Internet Res. 2013 Feb; 15(2): e37. Published online 2013 Feb 20. Doi: 10.2196/jmir.2281

Lauriks S1, Reinersmann A, Van der Roest HG, Meiland FJ, Davies RJ, Moelaert F, Mulvenna MD, Nugent CD, Dröes RM. Review of ICT-based services for identified unmet needs in people with dementia. Ageing Res Rev. 2007 Oct;6(3):223-46. Epub 2007 Aug 2.

Lewis TL[1], Boissaud-Cooke MA, Aungst TD, Eysenbach G. Consensus on use of the term "App" versus "Application" for reporting of mHealth research. J Med Internet Res. 2014 Jul 17;16(7):e174; discussion e174. Doi: 10.2196/jmir.3460.

Lewis TL, Wyatt JC. A new metric to assess the population impact of mobile medical Apps: app usage factor. Submitted to **J Med Internet Res** January 2015.

Lewis TL, Wyatt JC. mHealth and mobile medical Apps: a framework to assess risk and promote safer use. **J Med Internet Res.** 2014 Sep 15;16(9):e210. Doi: 10.2196/jmir.3133.

Liang X, Wang Q, Yang X, Cao J, Chen J, Mo X, Huang J, Wang L, Gu D. Effect of mobile phone intervention for diabetes on glycaemic control: a meta-analysis. *Diabetic Medicine* 2011; 28(4):455-63.

Liu JLY, Wyatt JC. The case for randomized controlled trials to assess the impact of clinical information systems. JAMIA 2011 http://jamia.bmj.com/content/early/2011/01/26/jamia.2010.010306

D. Low, N. Clark, J. Soar, A. Padkin, A. Stoneham, G. D. Perkins, J. Nolan, "A randomised control trial to determine use of the iResus© application on a smart phone improves the performance of advanced life support provider in a simulated medical emergency," *Anaesthesia*, **66**:255–262, 2011. Doi 10.1111/j.1365-2044.2011.06649.x

I. S. Mackenzie, *Human-Computer Interaction: An Empirical Research Perspective*, Morgan Kaufmann, 2013.

Marcano Belisario JS[1], Huckvale K, Greenfield G, Car J, Gunn LH. Smartphone and tablet self-management apps for asthma. Cochrane Database Syst Rev. 2013 Nov 27;11:CD010013. Doi: 10.1002/14651858.CD010013.pub2.

Martinez CH, St Jean BL, Plauschinat CA, Rogers B, Beresford J, Martinez FJ, Richardson CR, Han MK. Internet access and use by COPD patients in the National Emphysema/COPD Association Survey. BMC Pulm Med, 2014;14(1):66.

B Martínez-Pérez, I de la Torre-Díez, M López-Coronado, B Sainz-de-Abajo, M Robles & J M García-Gómez, "Mobile Clinical Decision Support Systems and Applications: A Literature and Commercial Review," *J Med Syst*, 38(4), 2014. Doi: 10.1007/s10916-013-0004-y

McKechnie V, Barker C, Stott J. Effectiveness of computer-mediated interventions for informal carers of people with dementia-a systematic review. Int Psychogeriatr. 2014 Oct;26(10):1619-37. Doi: 10.1017/S1041610214001045. Epub 2014 Jul 3.

McMahon GT, Gomes HE, Hickson Hohne S, et al. Web-based care management in patients with poorly controlled diabetes. Diabetes Care. 2005;28:1624–1629

Mold F, de Lusignan S, Sheikh A, Majeed A, Wyatt JC, Quinn T, Cavill M, Franco C, Chauhan U, Blakey H, Kataria N, Arvanitis TN, Ellis B. Patients' online access to their electronic health records and linked online services: a systematic review in primary care. Br J Gen Pract. 2015 Mar;65(632):e141-51. Doi: 10.3399/bjgp15X683941.

Munoz RF, Barrera AZ, Delucchi K, et al. International Spanish/English Internet smoking cessation trial yields 20% abstinence rates at 1 year. Nicotine Tob Res 2009 ;11 :1025^{e}34.

Munoz RF, Lenert LL, Delucchi K, et al. Toward evidence-based internet interventions: A Spanish/English web site for international smoking cessation trials. Nicotine Tob Res 2006;8:77e87.

Nguyen HQ, Donesky D, Reinke LF, Wolpin S, Chyall L, Benditt JO, Paul SM, Carrieri-Kohlman V. Internet-based dyspnea self-management support for patients with chronic obstructive pulmonary disease. J Pain Symptom Manage. 2012 Oct 13; doi: 10.1016/j.jpainsymman.2012.06.015.

J. Nielsen, *Usability Engineering*, Academic Press, 1994.

Nind, T., Wyatt, J., Ricketts, I., Mcpate, P., & Liu, J. (2009). Website credibility and intervention effectiveness. Proceedings of Persuasive Technology and Digital Behaviour Intervention, Symposium A at the AISB 2009 Convention (pp. 36–39). Heriot-Watt University, Edinburgh: SSAISB: The Society for the Study of Artificial Intelligence and the Simulation of Behaviour. http://www.aisb.org.uk/convention/aisb09/Proceedings/PERSUASIVE/FILES/Proceedings.pdf

Noel HC, Vogel DC, Erdos JJ, et al. Home telehealth reduces healthcare costs. Telemed J E Health. 2004;10:170–183.

Norman CD, Maley O, Li X. Using the internet to assist smoking prevention and cessation in schools: a randomized, controlled trial. Health Psychol 2008;27:799e810.

Osborn CY, Mayberry LS, Mulvaney SA, Hess R. Patient web portals to improve diabetes outcomes: a systematic review. Curr Diab Rep. 2010 Dec;10(6):422–35. Doi: 10.1007/s11892-010-0151-1.

Palen TE, Ross C, Powers JD, Xu S. Association of online patient access to clinicians and medical records with use of clinical services. JAMA. 2012;308: 2012-9.

PatientView. https://alexwyke.wordpress.com/2014/11/23/what-do-people-want-from-health-apps/ (accessed 27 April 2015).

Pike KJ, Rabius V, McAlister A, et al. American cancer society's QuitLink: randomized trial of internet assistance. Nicotine Tob Res 2007;9:415e20.

J. Preece, H. Sharp, Y. Rogers, *Interaction Design*; 4th edition, John Wiley & Sons, 2015.

Prokhorov AV, Kelder SH, Shegog R, et al. Impact of a smoking prevention interactive experience (ASPIRE), an interactive, multimedia smoking prevention and cessation curriculum for culturally diverse high school students. Nicotine Tob Res 2008;10:1477e85.

Proudfoot J, Nicholas J. Monitoring and evaluation in low intensity CBT interventions 2010. In J. Bennett-Levy J, Richards D, Farrand P, Christensen H, Griffiths K, Kavanagh D, Klein B, Lau M, Proudfoot J, Ritterband L, White J, Williams C (Eds.), *Oxford Guide to Low Intensity CBT Interventions*. (pp. 97-104). Oxford: Oxford University Press.

Quinn CC, Shardell MD, Terrin ML, Barr EA, Ballew SH, Gruber AL. Cluster-randomized trial of a mobile phone personalized behavioural intervention for blood glucose control. Diabetes Care. 2011;34

Rao A, Hou P, Golnik T, Flaherty J, Vu S. Evolution of data management tools for managing self-monitoring of blood glucose results: a survey of iPhone applications. *Journal of Diabetes Science & Technology* 2010; 4(4):949-57.

Rabius V, Pike KJ, Wiatrek D, et al. Comparing internet assistance for smoking cessation: 13-month follow-up of a six-arm randomized controlled trial. J Med Internet Res 2008;10:e45.

Ralston JD, Hirsch IB, Hoath J, Mullen M, Cheadle A, Goldberg HI. Web-based collaborative care for type 2 diabetes: a pilot randomized trial. Diabetes Care. 2009;32:234-9.

Research2guidance 2013: http://research2guidance.com/currently-only-12-of-diabetics-that-have-a-smartphone-use-a-diabetes-app-until-2018-the-share-will-rise-to-78-globally/ (accessed 27 April 2015)

Research2Guidance. http://research2guidance.com/the-market-for-mhealth-app-services-will-reach-26-billion-by-2017/ (accessed 27 April 2015).

Richardson CR, Buis LR, Janney AW, Goodrich DE, Sen A, Hess ML et al. An online community improves adherence in an internet-mediated walking program. Part 1: results of a randomized controlled trial. J Med Internet Res 2010; 12: 138–153.

S. Robinson, G. Marsden, M. Jones, *There's Not an App for That: Mobile User Experience Design for Life*, Morgan Kaufmann, 2014, ISBN-13: 978-0124166912

Rodrigues MA, Visvanathan A, Murchison JT, Brady RR. Radiology smartphone applications; current provision and cautions. Insights Imaging 2013 Oct;4(5):555-562 [FREE Full text] [doi: 10.1007/s13244-013-0274-4] [Medline: 23912880]

Ronda MC, Dijkhorst-Oei LT, Gorter KJ, Beulens JW, Rutten GE. Differences between diabetes patients who are interested or not in the use of a patient Web portal. Diabetes Technol Ther 2013 Jul;15(7):556-563

Ross SE, Haverhals LM, Main DS, et al. Adoption and use of an online patient portal for diabetes (Diabetes-STAR) AMIA Annu Symp Proc. 2006:1080.

Ross SE, Moore LA, Earnest MA, Wittevrongel L, Lin CT. Providing a web-based online medical record with electronic communication capabilities to patients with congestive heart failure: randomized trial. J Med Internet Res. 2004; 6:e12.

Ruland CM, Andersen T, Jeneson A, Moore S, Grimsbø GH, Børøsund E, Ellison MC. Effects of an internet support system to assist cancer patients in reducing symptom distress: a randomized controlled trial. Cancer Nurs. 2013;36(1):6–17. Doi: 10.1097/NCC.0b013e31824d90d4.

Russell-Minda E, Jutai J, Speechley M, Bradley K, Chudyk A, Petrella R. Health technologies for monitoring and managing diabetes: a systematic review. J Diabetes Sci Technol. 2009 Nov 1;3(6):1460-71.

M Schmettow, W Vos, J M Schraagen, "With how many users should you test a medical infusion pump? Sampling strategies for usability tests on high-risk systems," Journal of Biomedical Informatics, 46:626–641, 2013.

B. Shneiderman, S. M. Jacobs, M. S. Cohen, C. Plaisant, *Designing the User Interface*, 5th edition, Pearson, 2013.

K Shaw, J Fudin, "Evaluation and Comparison of Online Equianalgesic Opioid Dose Conversion Calculators," *Practical Pain Management*, 61–66, 2013.

Shen N, Levitan MJ, Johnson A, Bender JL, Hamilton-Page M, Jadad AR, Wiljer D. Finding a Depression App: A Review and Content Analysis of the Depression App Marketplace. *JMIR mHealth uHealth* 2015; 3(1):e16.

Simpson R, Wakefield P, Spiers N, Jagger C, Lindesay J. Carer-held records for dementia: a controlled trial. Int Psychogeriatr. 2006 Jun;18(2):259-68.

H. Sommerville, *Software Engineering*, Addison Wesley; 10th edition, 2015.

Sox CM, Gribbons WM, Loring BA, Mandl KD, Batista R, Porter SC. Patient-centered design of an information management module for a personally controlled health record. Journal of Medical Internet Research 2010; 12 (3): e36-e.

Sposaro F, Danielson J, Tyson G. iWander: An Android application for dementia patients. *Conf Proc IEEE Eng Med Biol Soc*. 2010; 2010:3875-8.

Tang CH, Li CC, Chang GH, Chang P. Implementing a personalized portal combined with workflow management tools used in diabetes care. AMIA Annu Symp Proc. 2003:1026.

Tang PC, Overhage JM, Chan AS, Brown NL, Aghighi B, Entwistle MP, et al. Online disease management of diabetes: engaging and motivating patients online with enhanced resources-diabetes (EMPOWER-D), a randomized controlled trial. J Am Med Inform Assoc. 2013;20:526-34.

Tenforde M, Nowacki A, Jain A, Hickner J. The association between personal health record use and diabetes quality measures. J Gen Intern Med. 2012 Apr;27(4):420-4. Doi: 10.1007/s11606-011-1889-0. Epub 2011 Oct 18.

H. Thimbleby, C. Runciman, "Equal Opportunity Interactive Systems," *International Journal of Man-Machine Studies*, **25**(4):439–451, 1986. Doi 10.1016/S0020-7373(86)80070-0

H. Thimbleby, A. Lewis, J. G. Williams, "Making Healthcare Safer by Understanding, Designing and Buying Better IT," *Clinical Medicine*, 2015.

H Thimbleby, "Ignorance of Interaction Programming is Killing People," *ACM Interactions*, 52–57, September+October, 2008. Doi: 10.1145/1390085.1390098

H Thimbleby, D Williams, "Using nomograms to reduce harm from clinical calculations," Proceedings IEEE International Conference on Healthcare Informatics, 461-470, 2013.

Torous J, Chan SR, Yee-Marie Tan S, Behrens J, Mathew I, Conrad EJ, Hinton L, Yellowlees P, Keshavan M. Patient Smartphone Ownership and Interest in Mobile Apps to Monitor Symptoms of Mental Health Conditions: A Survey in Four Geographically Distinct Psychiatric Clinics. *JMIR Mental Health* 2014; 1(1):e5.

Valdivieso-López E, Flores-Mateo G, Molina-Gómez JD, Rey-Reñones C, Barrera Uriarte ML, Duch J, Valverde A. Efficacy of a mobile application for smoking cessation in young people: study protocol for a clustered, randomized trial. *BMC Public Health* 2013; 1;13:704.

Van der Vaart, Drossaert CH, Taal E, Drossaers-Bakker KW, Vonkeman HE, van de Laar MA. Impact of patient-accessible electronic medical records in rheumatology: use, satisfaction and effects on empowerment among patients. BMC Musculoskelet Disord. 2014 Mar 26;15:102. Doi: 10.1186/1471-2474-15-102.

Van Velsen L, Beaujean DJ, van Gemert-Pijnen JE. Why mobile health app overload drives us crazy, and how to restore the sanity. BMC Med Inform Decis Mak 2013;13:23 [FREE Full text] [doi: 10.1186/1472-6947-13-23] [Medline: 23399513]

Wai Cheung K. The role of emergency physicians in tobacco cessation and British Columbia's Smoking Cessation Program. *Canadian Journal of Emergency Medicine* 2013; 15(0):1-2.

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.

Weitzman, E.R, Kelemen, S, Kaci, L, Mandl K.D. 2012. Willingness to share personal health record data for care improvement and public health: a survey of experienced personal health record users. BMC Med Inform Decis Mak. May 22;12(1):39.58

Heather M Whitford, Peter T Donnan, Andrew Symon, Gillian Kellett, Ewa Monteith-Hodge, Petra Rauchhaus, Jeremy C Wyatt. Evaluating the reliability, validity, and practicality of SMS messaging to collect research data, JAMIA April 2012. http://jamia.bmj.com/cgi/rapidpdf/amiajnl-2011-000785?ijkey=etOqtEpwhQMS5mm&keytype=ref

Whittaker R, McRobbie H, Bullen C, Borland R, Rodgers A, Gu Y. Mobile phone-based interventions for smoking cessation. Cochrane Database Syst Rev. 2012 Nov 14;11:CD006611. Doi: 10.1002/14651858.CD006611.pub3.

J A Wolf, J F Moreau, O Akilov, T Patton, J C English III, J Ho, L K Ferris, "Diagnostic Inaccuracy of Smartphone Applications for Melanoma Detection," *JAMA Dermatology*, 149(4):422-426, 2013. Doi: 10.1001/jamadermatol.2013.2382

Wyatt J, Spiegelhalter D. Evaluating medical expert systems: what to test and how ? Medical Informatics 1990; 15: 205-217

Wyatt JC. Measuring quality and impact of the World Wide Web. BMJ 1997; 314: 1879-81 (commentary)

Wyatt JC, Altman DG. Prognostic models: clinically useful, or quickly forgotten ? BMJ 1995; 311: 1539-41 (commentary)

Wyatt J. Evaluating electronic consumer health material. BMJ 2000; 320: 159-60 (commentary)

Wyatt J. Research in Health Informatics: how do we assess it and how to improve it ?. Chapter in: Health Informatics: An Overview, 2nd Edition. Eds. Evelyn Hovenga, Michael Kidd, Carola Hullin, Sebastian Garde. Elsevier US, 2009

Wyatt JC, Sullivan F. ABC of Health Informatics 12: eHealth and the future: promise or peril ? BMJ 2005; 331: 1391-3

Wyatt JC, Wright P. Medical Records 1: Design should help use of patient data. Lancet 1998; 352: 1375-8

Ye X, Bapuji SB, Winters SE, Struthers A, Raynard M, Metge C, Kreindler SA, Charette CJ, Lemaire JA, Synyshyn M, Sutherland K. Effectiveness of internet-based interventions for children, youth, and young adults with anxiety and/or depression: a systematic review and meta-analysis. BMC Health Serv Res. 2014 Jul 18;14:313. Doi: 10.1186/1472-6963-14-313.

Yu CH1, Bahniwal R, Laupacis A, Leung E, Orr MS, Straus SE. Systematic review and evaluation of web-accessible tools for management of diabetes and related cardiovascular risk factors by patients and healthcare providers. J Am Med Inform Assoc. 2012 Jul-Aug;19(4):514-22. Doi: 10.1136/amiajnl-2011-000307. Epub 2012 Jan 3.

Zhou YY, Garrido T, Chin HL, Wiesenthal AM, Liang LL. Patient access to an electronic health record with secure messaging: impact on primary care utilization. Am J Manag Care. 2007;13:418-24.

Appendices

Appendix 1: List of abbreviations used

AB testing	Testing of alternative screen layouts etc. in a randomised fashion
A&E	Accident & Emergency dept
APA	American Psychological Association
App	Application for a smart phone, tablet device etc.
BMI	Body Mass Index
BNF	British National Formulary
BP	Blood pressure
BSI	British Standards Institute
CBT	Cognitive behaviour therapy
CCG	Clinical Commissioning group
CE mark	Authorisation mark showing the device / app has passed relevant standards specified by the proper authority – the MHRA for medical devices & software
CI	Confidence interval
CINAHL	Bibliographic database for nursing and allied health professionals
COPD	Chronic obstructive pulmonary disease
DM	Diabetes mellitus
DNA	Did not arrive [at a clinic / appointment]
HER	Electronic health record
EPOC	Cochrane Effective Practice & Organisation of Care review group
EU	European Union
EVPI	Expected value of perfect information, a technique to calculate the value of a piece of research
FDA	US food & Drugs Administration, the equivalent of MHRA in UK
HbA1C	Glycosylated haemoglobin, an indicator of diabetes control and risk of developing complications
HCI	Human computer interaction
HCP	Health care provider
HDL	High density lipoprotein
HIV	Human immune deficiency virus
HMO	Health Maintenance Organisation
HON	Health on the Net Foundation (Code of Conduct)
IT	Information technology
ICT	Information and communications technology
ISO	International Organisation for Standardisation
LDL	Low density lipoprotein
LTC	Long term condition
mHealth	Mobile health ie. Use of phones etc. to promote / manage health
MHRA	Medicines and healthcare products regulatory agency
NIB	National Information Board
NICE	National Institute for Health & Care Excellence
PAS	Publically Accessible Standard (BSI)
PDA	Portable data assistant
PDF	Portable document format (Adobe)
PHR	Personal health record
PEHR	Personal electronic health record – a kind of PHR delivered electronically, cf. on paper
PROMs	Patient recorded outcome measures
PubMed	The largest medical bibliographic database
RCP	Royal College of Physicians
RCT	Randomised controlled trial
RR	Risk ratio
SAD	Social anxiety disorder
SMS	Short message service
SR	Systematic Review
STD	Sexually transmitted diseases

T1 / T2 diabetes	Type 1 / 2 diabetes
UCD	User-centred design
UX	User experience
W3C	World Wide Web Consortium

Appendix 2: Glossary of main terms used

AB testing	Testing of alternative screen layouts etc. in a randomised fashion to determine which ones produce positive results.
App	An application, typically a small, specialised program downloaded onto mobile devices, such as a smart phone, tablet device etc.
Big Data	Extremely large and often heterogeneous data sets that may be analysed computationally to reveal patterns, trends, and associations, especially relating to human behaviour and interactions.
Body Mass Index(BMI)	A simple index of weight-for-height that is commonly used to classify underweight, overweight and obesity in adults.
British National Formulary (BNF)	An information source for prescribing, dispensing and administering medicines, intended primarily for use by physicians, pharmacists, nurses and other health-care professionals in the UK.
Blood Pressure (BP)	The pressure of the blood in the circulatory system, often measured for diagnosis since it is closely related to the force and rate of the heartbeat and the diameter and elasticity of the arterial walls.
British Standards Institution (BSI)	A service organization that produces standards across a wide variety of industry sectors. Its codes of practice and specifications cover management and technical subjects ranging from business continuity management to quality requirements.
Cognitive behaviour therapy (CBT)	A talking therapy that can help you manage your problems by changing the way you think and behave. It is most commonly used to treat anxiety and depression, but can be useful for other mental and physical health problems.
Clinical Commissioning Group (CCG)	NHS organisations set up by the Health and Social Care Act 2012 to organise the delivery of NHS services in England.
CE mark	A symbol applied to products to indicate that they conform to relevant EU directives regarding health and safety or environmental protection. The MHRA is the authority with responsibility for this process in respect of medical devices and software in the UK.
Confidence interval (CI)	The confidence interval represents a range of values for a population parameter (a statistic such as the mean or an F value etc) for which the difference between the parameter and the observed estimate is not statistically significant at the stated level, typically 95%. Normally written <statistic> 95% CI [5.62, 8.31] etc.
Cumulative Index to Nursing and Allied Health Literature (CINAHL)	An index of English-language and selected other-language journal articles about nursing, allied health, biomedicine and healthcare.
Chronic obstructive pulmonary disease (COPD)	A lung disease characterized by chronic obstruction of lung airflow that interferes with normal breathing and is not fully reversible
Diabetes mellitus (DM)	The commonest form of diabetes, caused by a deficiency of or resistance to the pancreatic hormone insulin, which results in a failure to metabolize sugars and starch.
Electronic health record (HER)	A digital version of a patient's paper chart. EHRs are real-time, patient-centred records that make information available instantly and securely to authorized users.
Cochrane Effective Practice & Organisation of Care (EPOC)	A review group which undertakes systematic reviews of educational, behavioural, financial, regulatory and organisational interventions designed to improve health professional practice and the organisation of health care services.
Digital divide	The gulf between those who have ready access to computers, smart phones or and the Internet, and those who do not.
Expected value of perfect information (EVPI)	A technique to calculate the value of a piece of research.
US Food and Drug	An agency within the US Public Health Service with responsibility for monitoring the safety of food, drugs and medical devices.

Administration (FDA)	
Glycosylated haemoglobin test (HbA1C test)	An indicator of diabetes control and risk of developing complications.
Human computer interaction (HCI)	The study of how people interact with computers and to what extent computers are or are not developed for successful interaction with human beings.
Information and communications technology (ICT)	Technologies that provide access to information through telecommunications. It is similar to Information Technology (IT), but extends the focus to communication technologies. This includes the internet, wireless networks, cell phones, and other communication mediums
Long term condition (LTC)	A condition that cannot, at present be cured; but can be controlled by medication and other therapies. Examples of Long Term Conditions are diabetes, heart disease and chronic obstructive pulmonary disease.
mHealth	Mobile health ie. Use of phones, tablet computers etc. to promote or manage health
Medicines and healthcare products regulatory agency (MHRA)	A government body responsible for regulation of medicines and medical devices and equipment used in healthcare and the investigation of harmful incidents in the UK.
National Information Board (NIB)	The Board is responsible to the Department of Health for setting the strategy and direction for the health and care system on information technology and information.
National Institute for Health & Care Excellence (NICE)	NICE is the organisation responsible for providing national guidance and advice for improving health and social care.
Patient empowerment	The philosophy that patients are active participants in, not passive recipients of, the caring process, and thus should be well informed about all aspects of their health, 'wellness status' and disease state, to gain maximum health benefit within the context of their social demands
Patient recorded outcome measures (PROMs)	An umbrella term that covers a range of potential types of measurement but is used specifically to refer to self-reports by the patient. Data may be collected via self-administered questionnaires completed by the patient or via interviews.
Personal electronic health record (PEHR)	A kind of Personal Health Record delivered electronically, rather than on paper
Personal health record (PHR)	A health record, maintained in electronic or other format by an individual, that can be shared with anyone of the patients choosing.
Portable data assistant (PDA)	A palmtop computer that functions as a personal organizer but also provides email and Internet access.
Publically Available Specification (PAS)	A service provided by the BSI, which provides a route to standardisation and can provide product specifications, codes of practice, guidelines and vocabularies or be used as an assessment benchmark
Public Health England (PHE)	PHE is the organisation responsible for protecting and improving the nation's health and wellbeing, and reducing health inequalities.
PubMed	A free search engine accessing primarily the MEDLINE (Medical Literature Analysis and Retrieval System Online) database of life sciences and biomedical information. It includes bibliographic information for articles from academic journals covering medicine, nursing, pharmacy, dentistry, veterinary medicine, and health care.
Randomised controlled trial (RCT)	A study in which people or other "allocation units" are allocated at random to receive one of several interventions. One of these interventions is the standard of comparison or control.
Risk ratio (RR)	A measure of the risk of a certain event happening in one group compared to the risk of the same event happening in another group. Also known as relative risk.
Self-efficacy	Self-efficacy is the extent to which people believe that they are capable of performing specific behaviours in order to attain certain goals.
Social anxiety	An anxiety disorder in which a person has an excessive and unreasonable fear of social

disorder (SAD)	situations. Also called social phobia.
Short message service (SMS)	Text messaging service component of phone, web, or mobile communication systems.
Systematic Review (SR)	A critical assessment and evaluation of all identifiable research studies that address a particular clinical issue. The researchers use an explicit method for locating, assembling, and evaluating a body of literature on a particular topic using a set of stated criteria
T1 / T2 diabetes	Type 1 diabetes is less common, usually occurs under the age of 40 and is treated using daily insulin doses. Type 2 diabetes is more common, usually occurs over the age of 40 and is mainly treated using diet and exercise.
User centred design (UCD)	A process (not restricted to interfaces or technologies) in which the needs, wants, and limitations of end users of a product, service or process are given extensive attention at each stage of the design process.
Usability	The extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use.
User experience (UX)	A person's behaviours, attitudes, and emotions about using a particular product, system or service.
Waitlist control	A group of participants included in the study that is assigned to a waiting list and receives intervention after the active treatment group.
Web portal	Portal is a term, generally synonymous with gateway, for a World Wide Web site that is or proposes to be a major starting site for users when they get connected.
World Wide Web Consortium (W3C)	An industry consortium which seeks to promote standards for the evolution of the Web and interoperability between WWW products by producing specifications and reference software.

Appendix 3: Details of the review scope and four focus areas

The agreed scope [based on Jan Hoogewerf email to HSCIC project manager and others of 9-4-15] is

1. General literature about adoption and potential benefits of health and well being apps and personal health records. Also studies about methods used to define and realise benefits of health and well being apps. This is not part of the literature search, but the team will identify any good single papers.

2. Apps and PHRs in the four NIB focus areas (mood monitoring, dementia, smoking cessation and diabetes).

3. The literature review will be extended beyond the four focus areas to:

   - 'Behavioural change', as a search topic, identifying research papers which cover this specifically in relation to either PHR or apps.

   - Long term conditions will be extended to include COPD and hypertension.

4. Topics of interest to the project: adoption, benefits and dis-benefits, usability, patient perceptions, channels for interaction, challenges and information governance. We will also identify benefits and dis-benefits of linkage of apps to electronic health records.

5. Personal electronic health records: in addition to the four focus areas, we will also search literature for end of life (eg recording end of life preferences and advanced directives) and early years (eg electronic PHR or red book).

6. Information governance issues, including data privacy concerns, will be covered in the retrieved articles and in our reports. However, our expertise does not extend to technical aspects of data protection, so that will be excluded.

Appendix 4: Details of bibliographic databases, search & critical appraisal methods used

Which databases were searched ?

We focused on PubMed initially, especially via the Clinical Queries interface Raw PubMed was only used if CQs did not reveal useful results. Further searches were then conducted on Embase, Medline, Cochrane and Google Scholar.

What studies were excluded?

We limited the search to publications from year 2000 on, in English, on human not animal participants, and to exclude poor quality or duplicate studies. We excluded studies that related solely to fitness, diet and exercise where they were not addressing one of the four focus areas or other long term condition. We also excluded mobile SMS text messaging as a sole intervention.

What types of study were looked for ?

We looked for both UK and international studies. We started with SRs of relevant studies (from Clinical Queries), only moving to individual studies if necessary (ie. there is no SR or the SR is 3-4 years out of date based on the *search dates* in the SR, not the publication date of the SR. We chose the type of study according to which attribute of the app, PHR etc. the study needed to measure:

Attribute measured by the study	Preferred study type	Critical appraisal checklist from EQUATOR website
Need or requirements for the technology	Surveys, qualitative studies (ie. interviews, focus groups etc.); formal systems analysis / business process modelling work	Surveys: Kelley K et al. Good practice in conduct & reporting of surveys. Int J Qual Health Care. 2003;15:261-6 Qualitative: SRQR
Attitudes to or perceptions of the technology	Surveys, qualitative studies	Surveys: Kelley et al. Qualitative: SRQR
Usage rates for the technology	Analysis of log file data; *possibly* surveys of eligible users	?
Usability	Formal usability studies; user centred design workshops; task analysis, eye tracking studies etc.; checklist-based assessment of the system using a reputable checklist of desirable system attributes	
Inclusiveness of the technology (Cyberdivide)	Assessment of technology usability / usage rates by different age or sensory limitation groups in the four focus areas	

Quality of the data captured / shared	Analysis of the accuracy & completeness of data against a reliable gold standard source	Accuracy studies: STARD
Benefits or impact of the technology on clinical outcomes, knowledge about disease, self efficacy (empowerment), drug adherence, health related behaviours, NHS resource usage, etc.	Randomised trial; *possibly* a controlled before-after or interrupted time series study if no RCTs. Exclude simple before after and other study types.	CONSORT
Safety or risks of the technology	Analysis of adverse incidents or near misses using root cause analysis; analysis of the accuracy of system output or advice against a reliable gold standard	Accuracy studies: STARD
Privacy risks	Checklist based assessment of privacy risks / threats and controls	
Value for money, cost effectiveness	Formal cost effectiveness, cost utility or cost consequence analysis	CHEERS

We referred to the suggested checklist to appraise the selected studies, where possible – see the EQUATOR website for checklists http://www.equator-network.org/

Example general search terms

In general, we structured the search terms using PICO (population, intervention, control, outcome) elements:

P = patients (eg. with Type 1 / 2 diabetes, dementia, low mood or anxiety / depression), members of public / citizens (eg. smokers), informal carers (for dementia), parents (for T1 diabetes), health professionals

I = health app, self-monitoring via smartphone / cell phone, web service, personal electronic health record…

C = usual care

O = acceptability; concerns; usability; usage rates; impact on health outcomes (see below) or usage of healthcare resources / cost saving, etc.

We focussed on studies of technologies relevant to the focus areas below:

Focus area - intervention type	Suggested impacts to search for
Mood monitoring tool	Impact on mood (measured by HADS scale ?) or self reported mental state
Diabetes self-management tools	Diabetes control (blood sugar, HbA1C levels), hypos, hospital admissions
Smoking cessation support tools	Behaviour change - smoking rates / smoking cessation (not self report - measured using exhaled carbon monoxide or serum / saliva cotinine levels) Change in calculated cardiovascular risk

Dementia brain training tools	Mental abilities (memory, arithmetic ?); impact on carers
Dementia memory prosthesis tools	Rates of completing activities of daily living (eg. eating meals, making bed, contacting friends and family); weight loss / gain; impact on carers
Tools to support people with dementia keep in touch with friends and family	Quality of life; number of weekly contacts with friends and family; amount of support needed from social / voluntary care; NHS resources used.

To cover self-monitoring, we also ran some more general searches without mentioning the focus areas.

Other topics searched for

1. Proposed / actual App quality metrics or accreditation methods – and how well they perform, what resources it takes to apply them, how acceptable they are to stakeholders.

2. Professional implications of shift to apps / PHR etc. – opinion pieces

3. Privacy implications of mHealth and how data will be managed in future

4. End of life care personal health records

5. Early years personal health records

6. Apps and web services for other long term conditions (eg COPD and hypertension).

Details of PubMed search terms used

For the four focus areas, we searched using combinations of items from columns 1 and 3 of the table below.. We also reverse engineered the MeSH terms assigned by librarians from eligible studies and used the PubMed "Show related" function to identify other studies indexed using the same terms.

Intervention terms		Focus area terms	No of studies found[2]	No of systematic reviews
person* health record*	AND	diabet*	47	8
Apps	AND	diabet*	8	6
health records, personal[MeSH Terms] OR personal health record[Text Word]	AND	diabet*	21	4
social media[MeSH Terms] OR social media[Text Word]	AND	diabet*	7	6
Apps	AND	anxiety	5	1
Apps	AND	dementia	14	2
Smartphones	AND	dementia	7	5
social media[MeSH Terms] OR social media[Text Word]	AND	dementia	2	0
mobile app	AND	dementia	1	2
Smartphone	AND	diabet*	112	23

---

[2] The numbers given for studies and SRs are not yet checked for eligibility nor for duplicates

Apps	AND	smoking cessation	7	1
mobile app	AND	smoking cessation	12	6
Smartphone	AND	smoking cessation	64	20
Smartphone	AND	anxiety	64	9
Smartphone	AND	mood disorders	15	2
mobile app	AND	"Anxiety Disorders"[Mesh]		5
mobile app	AND	"Depressive Disorder"[Mesh]		1
Smartphone	AND	"Depressive Disorder"[Mesh]		13
Smartphone	AND	"Anxiety Disorders"[Mesh]	11	2
smartphone	AND	Depression[Mesh]	50	5
health records, personal[MeSH Terms] OR personal health record[Text Word]	AND	"Anxiety Disorders"[Mesh]	52	10
social media[MeSH Terms] OR social media[Text Word]	AND	"Anxiety Disorders"[Mesh]	230	128
Apps	AND	Blood Glucose Self-Monitoring[Mesh]		5
Self Care[Mesh]	AND	"Anxiety Disorders"[Mesh]	189	34
Self Care[Mesh]	AND	Depression[Mesh]	555	65
Self Care[Mesh]	AND	Smoking Cessation"[Mesh]	183	22

Notes on specific search terms

Expanded "mood monitoring" to anxiety and depression

Used MeSH term "Health records, personal" for PHR

Expanded dementia to include delirium

Expanded apps to include online forums and social media (eg. Big White Wall)

Reran the searches about 2-3 weeks before final report to capture any new material, and existing material that was not previously indexed for Medline

.

Appendix 5: Example mHealth app checklist from Royal College of Physicians

App name: _____ For iPhone / Android / Windows / other: _____

Date of filling out this checklist: _____   Start time:_____   End time: _____

Your name and email : _____

1. Who developed the App, and what's inside it ?

a) Is it clear who this App is for, and how people should use it ? Yes / No / Don't know

b) Is it clear what problem it's designed alleviate, or which outcome it helps promote ? Yes / No / Don't know

c) Does the App developer / funder seem well informed about this problem or outcome, and likely to be unbiased in their approach to it ? Yes / No / Don't know

d) Have they located sound, relevant, up-to-date evidence, images etc. to put into their App ? Yes / No / Don't know

e) If the App is designed to help people change their behaviour, does the design appear to be based on a proven behaviour change method ? Not applicable / Yes / No / Don't know

f) Does the App keep user data secure and private ? Yes / No / Don't know

g) Can the App be used by people with cognitive, sensory or other disabilities ? Yes / No / Don't know

2. How well does the App work ?

a) Is the App easy and fun to use ? Yes / No / Don't know

b) Is it clear what information the App needs from the user, and when ? Not applicable / Yes / No / Don't know

c) Does it give the user sensible answers or advice, quickly ? Not applicable / Yes / No / Don't know

d) Is there a way to feedback user comments to the App developer ? Yes / No / Don't know

3. Is there any evidence that the App does actually alleviate the problem ?

a) Have any studies been carried out measuring the impact of the App on people's health knowledge, behavioural intentions, health-related actions or (preferably) outcomes? Yes / No / Don't know

b) Did the study also examine any harms caused by the App or quantify costs ? Not applicable / Yes / No / Don't know

c) Were these studies independent well designed, large enough, and applicable to the user ? Not applicable / Yes / No / Don't know

d) Overall, do the benefits of using this App seem likely to outweigh inconvenience and costs to the user, and help them in the long term ? Yes / No / Don't know

Appendix 6: Detailed results on PHR study data extraction

Outcomes

1 .Direct impacts on system users

1a. Clinical decisions

1b. Clinical actions, prescribing, referral, test ordering, hospital admission, discharge, counselling, outpatient review

1c. Patient decisions and actions: self-management drug dose adjustment drug adherence; self-referral; appointments; usage of health services

1d. Behavioural change, e.g. smoking cessation, exercise weight reduction, alcohol intake, dietary improvement hazardous behaviours

1e. Patient outcomes or surrogate outcomes

2. Impacts on health services and systems

2a. Service utilisation

2b. Cost of running the service per patient / encounter

2c. Drug or investigation costs; cost effectiveness of services or health technologies

2d. Rates of accidents, HIV / STD transmission, adverse events or near misses

Study ID	Type	Year / Country	Participants	Setting	Outcomes measured	Results
Goldzweig 2013	Systematic review Not a rigorous SR – authors state that primary importance was given to studies based in USA, followed by other English speaking	Publications in Pubmed and Web of Sciences between 1990 to 2013.	**Chronic health conditions** included Diabetes, Heart Failure, Hypertension, Depression and preventive services.	Electronic patient portals 14 RCTs, 21 observational and hypothesis testing studies, 5 quantitative and descriptive studies and 6 qualitative studies included.	1c 1e 2a	Acceptance of portals was higher in younger, computer literate and more enthusiastic patients. Patient satisfaction was generally high among the portal users. Utilisation of services varied, most studies not having any significant differences in admission rates or home visits. Where differences have been found, the results have been contradictory. In one RCT evaluating Diabetes care, patients receiving a web based and nurse care management had significantly lower

	countries and Western Europe.					HbA1c compared to the control group (Ralston 2009). There was no difference in outpatient visits or primary care / specialty visits or inpatient days.

countries and Western Europe.

HbA1c compared to the control group (Ralston 2009). There was no difference in outpatient visits or primary care / specialty visits or inpatient days.

Another US based study showed no difference in HbA1c, Blood pressure or LDL levels between the groups after 12 months of access to an electronic personal health record (Grant 2008), while a different RCT showed lower HbA1c at 6 months but not at 12 months' time (Tang 2013).

One study on Heart failure demonstrated no significant difference in the "self – efficacy" part of the Kansas City Cardiomyopathy questionnaire. Intervention group showed more adherences to medical advice but no difference in adherence to medications (Ross 2004).
The study also found a significantly high emergency department visits in the intervention group (20 vs. 8 visits), with no difference in hospitalisations.

Patients receiving face to face visits and secure messaging in addition to a Hypertension web portal (included input of BP readings and EHR data), showed significant improvements in blood pressure compared to usual care and only portal access (Green 2008). Another study

						showed no significant difference in blood pressure.

Two observational studies from the Kaiser healthcare system showed different results with one cohort study showing significantly higher rates of office visits, telephone call, clinic visits, emergency visits and hospitalisations (Palen 2012), while another found a decrease in visits and less increase in telephone contacts (Zhou 2007).

Contradictory results from RCTs even among similar disease conditions. Many studies included web based interventions but also included nurse visits to patients home. |
| Chen 2012 | Systematic review and Meta-analysis

Quality assessment was assessed by method of randomisation, allocation concealment, baseline similarity between | Publications till December 2009 | Adult **smokers** | 60 RCTs/quasi-RCTs reported in 77 publications were included in the review; Trials that used conventional mass media interventions were excluded.

The vast majority of included studies evaluated interventions with a single tailored component, | 2b 1d | Well conducted review, however, the meta-analysis included all types of technologies for smoking cessation and it was not possible within the scope of this review to carry out separate meta-analysis including only the web based RCTs.

Cost-threshold analyses indicated some form of electronic intervention is likely to be cost-effective when added to non-electronic behavioural support, but there is substantial uncertainty with regard to what the most effective (thus most cost-effective) type of electronic intervention is, |

groups, similarity in co-interventions between groups, biochemical validation, extent of drop-out, differential drop-out between groups and use of intention-to-treat analysis.				with or without additional generic component. Randomised controlled trials (RCTs) and quasi-RCTs evaluating smoking cessation programmes that utilise computer, internet, mobile telephone or other electronic aids in adult smokers were included in the effectiveness review. Included trials compared a wide range of electronic aids versus no intervention, self-help materials or another electronic or non-electronic intervention. Trials recruited smokers interested in quitting (aid to cessation studies), smokers who were not ready to quit (cessation induction studies) or a mixed population.		which warrants further research. EVPI calculations suggested the upper limit for the benefit of this research is ,around £2000–3000 per person. Compared with no intervention or generic self-help material, interventions using electronic aids significantly increased the likelihood of achieving prolonged abstinence or point prevalence abstinence from smoking, measured at the longest follow-up. Pooled relative risks were 1.32 (95% CI 1.21 to 1.45) for prolonged abstinence and 1.14 (95% CI 1.07 to 1.22) for point prevalence abstinence at follow-up. There was no substantial heterogeneity in these analyses. There were no substantial differences in effect size between aid to cessation and cessation induction studies. The mixed-treatment comparison showed a small but statistically significant positive intervention effect on time to relapse (mean HR 0.87, 95% CI 0.83-0.92). Six studies compared different electronic interventions with a single tailored component against each other. As the settings and contents of the electronic interventions being assessed were different, it was not possible to include

						them in a meta-analysis and compare the electronic interventions with each other. Also, there were no comparisons made between the use of web based PHR versus mobile technology. Authors conclude that computer and other electronic aids increased the likelihood of smoking cessation compared with no intervention or generic self-help materials, but the effect was small. Decision-analytic modelling indicated that adding an electronic intervention to non-electronic behavioural support was likely to be cost-effective but there was substantial uncertainty as to the most cost-effective type of intervention.
Cuijpers 2009	Systematic review and Meta-analysis Quality criteria was assessed on 3 of 4 Cochrane Handbook criteria (including randomisation, blinding of assessors and completeness	Publications till 2008	Patients with **anxiety disorders** including phobia, panic disorder, obsessive compulsive disorder and PTSD.	23 RCTS were included in the systematic review and meta-analysis. The review included computer aided psychotherapy delivered by internet linked computers, standalone computers, palmtops phone interactive voice response, DVDs and mobile phones. There were 10 RCTs on	1e	The meta-analysis included all types of technologies for psychotherapy. It was not possible within the scope of this review to carry out separate meta-analysis including only the web based RCTs (12). However, the authors have performed Subgroup analyses which showed no significant difference in the effect size between the web based and other modes of technology. Computer- Aided psychotherapy had a larger effect on anxiety related conditions compared to contrast conditions (d=1.08, 95%CI 0.84-1.32; 21 studies). Two RCTs were deemed as 'outliers' and removal of these two studies decreased the

	of follow-up data. Validity (adequacy of random allocation concealment) was not possible in the trials included.			phobia, 9 on panic disorder, 3 on PTSD and 1 on obsessive compulsive disorder. Of the studies included, relevant to this review were 12 RCTs - these were studies which looked at web based interactive interventions compared to DVDs and mobile phones. These included 8 studies on Panic disorder, 3 on PTSD and 1 on social phobia. There were 6 Swedish, 3 Australian, 2 Dutch and 1 US based trial.		heterogeneity from (I2=65.6% to 36.2%). This also reduced the effect size (d=0.94, 95%CI 0.80-1.08; 19 studies). There was no significant difference in anxiety at 3 months follow up. Computer-aided psychotherapy had a small to moderate effect on quality of life (d=0.46, 95% CI 0.30-0.62; 12 studies) and depression (d=0.56, 95%CI 0.41-0.71) compared to contrast conditions. There was no significant difference between computer-aided and face-to-face psychotherapies at 3 or 6 months follow up. Authors concluded that Computer-aided psychotherapy was an effective treatment for anxiety compared to contrast conditions and was as effective as face-to-face psychotherapy. The results should be treated with caution due to a number of methodological limitations and study heterogeneity.
Connelly 2013	Systematic Review Systematic review critical appraisal based on Cochrane Collaboration Back Review.	2001 – 2013 publications	**Type 2 Diabetes patients**	Total of 15 RCTs on use of technology to promote physical activity in Type 2 DM patients – 9 web-based, 3 mobile phone, 2 CD ROM and 1 computer based. **(only web based records analysed)**	1d 1e	A 10 month website based intervention with professional coaching help; a USA study (320 patients), found no significant increase in physical activity. Participant rate of 82%. Greater use of the website in the first 3 months, then progressive drop in rates over time (Glasgow, 2003). Another 3 month study in South Korea (73 patients) found a significant decrease in HbA1c by 0.6% (7mmol/mol). Only 26.9 %

| | Good quality | | | 6 studies did not mention randomisation methodology.

6 studies were deemed 'high methodological quality'. | | participants had computer access. 82.9% satisfaction with the use of the website (Kim and Kang, 2006).

Canadian study of 511 patients who received 6 month intervention to track quality of diabetes care showed increase in exercise by 125% (extra 67.5%) and decrease in HbA1c by 0.2% (2 mmol/mol). Similar study in the USA (761 patients) showed no significant increase in physical activity, but a decrease in HbA1c by 0.6% (2 mmol/mol) (Holbrook, 2009)

A 16 week study (324 patients) with study group having access to interactive website vs. only access to part of website in control group found significant increase in 42.6% in steps (Richardson, 2010)

A 463 patients 12 month study on website intervention or website intervention plus follow up telephone call found significant increase of 10.6% improvement in calories expended through physical activity at 4 months. However the effect significantly decreased when the study was follow up to 12 months duration (decrease of 18.7% at 12 months) (Glasgow, 2012).

Summary of the SR: All except one web based intervention found an increase in physical activity with a range of 3% to |

						125%. Six of the studies showed a significant increase compared to the control group. A website to use intention to treat found an increase of 10.6% in calories expended through physical activity at 4 months but there was a decrease of 18.7% at 12 months' time. Two web based interventions found decrease in HbA1c levels with an average of 1 mmol/mol and 7mmol/mol. Participation rate was over 80% in four web based studies and two interventions showed a steep decline in use over time.
Osborn 2010	Systematic Review

Methods explained, limitations included – no unpublished searches.

Acceptable quality | 2000 to 2010 publications | **Type 1 and Type 2 Diabetes** patients (17 studies focused only on Diabetic patients; 1 focused only on Type 1 DM, 8 only on Type 2 DM and 8 studies included both group of patients.

26 studies included 2436 patients including 271 patients for usability studies.

Study times were between 3 months to 12 months | PHRs integrated with EHRs – termed as "patient web portals"

26 publications included.
RCTs (8), quasi-experimental studies(4), pre-post evaluations, Portal system design & function and qualitative studies of usability were included in the SR. | 1b
1c
1e
2a | One 12 month Quasi-experimental study conducted in USA and Puerto Rico showed reduction in hospital admissions (+27% in control group vs. -60% in intervention group), bed days of care (+32% vs. -68%), emergency room visits (+22% vs. -66%) and prescriptions (+37% vs. -59%). However it is not clear whether these findings were significant. Patient satisfaction was 97% and provider satisfaction was 100% (Kobb, 2003).

A qualitative study of 305 adults in Taiwan reported 69.8% patient reported improvement in quality of healthcare with implementation of a chronic disease management portal (Tang, 2003).

A Randomised control trial, 12 month duration, including 104 patients in the USA |

			durations.			with chronic disease showed significant improvement in cognitive status (+0.8% in control group vs. -1.0% in intervention group) and functional level (19.4 vs. 20.0) in the intervention group receiving home electronic portal compared to usual home health care.. However patient satisfaction and self-related health were not significant. The total number of urgent visits were significantly reduced (+5 vs. – 83), however there was no difference in the total number of nurse visits during the study period (Noel 2004).

An RCT in the USA including both Type 1 and 2 DM patients (n = 104) showed a significant reduction in HbA1c Levels between control group and an intervention group receiving web based care management (-1.2 vs. -1.6%) in 12 months' time. In addition, there was significant difference in HbA1c reduction between high users and low users (-1.2 vs. -1.6%). There was also a significant better reduction in the systolic blood pressure (-7 vs. -10). There was also significant better reduction in the Triglyceride levels and increase in HDL levels in the intervention group (McMahon, 2005).

Usage of a Type 2 diabetes portal in an RCT was higher when the information was personalised to individual patient (319 days |

						vs. 772 days) (Ross 2006). HbA1c levels did not show any significant reduction when patients were followed up for 3 months duration. They also reported no significant change in blood pressure and exercise. The portal also had poor usability due to technical complications (Faridi 2008).
Kuijpers 2012	Systematic Review Good quality	RCTs published between 1990 and 2012, web based and interactive interventions for chronic diseases. All studies year 2000 or later, with most after 2005. 12 US based studies, 2 Canadian and 2 Norwegian and 1 from Australia and Korea. 2 studies of high quality, 13 studies of moderate quality and 3 low quality as	18 unique studies of **chronic diseases** included in the review. Overall percentage of dropout between 0.0% and 52.3% with a median of 17.5% (intervention group 19.7% and control group 14.0%) Types of diseases include – Diabetes, Heart failure, COPD, Cardio vascular, cancer and mixed groups	Interventions included education, self-monitoring, feedback, training, personal exercise program and communication using web based programs.	2a 1d	Significant positive effects on patient empowerment reported in 4 studies and positive physical activity reported in 2 studies. The interventions were often used in different combinations and adapted to specific patient populations; hence, the individual contribution of the effects of the intervention was not feasible. Based on the evidence the authors identified 7 elements of web based interventions that could benefit Cancer patients, including the provision of a cancer survivorship care plan.

				per the Authors classification based on a 13 point rating system. Only 7 described method of randomisation.		
Rutland 2013

Included in Kuijpers 2012 SR | 1 year repeated-measures randomized controlled trial

Acceptable | Norway | 325 breast and prostate **cancer** patients
Overall dropout rate was 24.6%

Control group received URLs of publicly available cancer Web sites and treatment group received access to the internet based application. | Internet-based, interactive health communication application that allows cancer patients to monitor their symptoms and problems and provides individually tailored information and self-management support, e-communication with expert cancer nurses, and an e-forum for group discussion with other patients. | 1d
1e | Group differences on symptom distress were significant only for the global symptom distress index on the Memorial Symptom Assessment Scale There were no significant group differences on secondary outcomes. Additional analyses showed significant within-group improvements in depression in the experimental group only. In the control group, self-efficacy and health-related quality of life deteriorated significantly over time. |
| Van der Vaart 2014 | Before – after study | Netherlands | 360 patients diagnosed with **Rheumatoid Arthritis**. Patients excluded: those reviewed over 1 year back or with severe co morbidities. | Intervention included web portal with medication, blood results, disease activity, QOL instrument data, all accompanied by graphs. Questionnaires were sent to the patients 1 | 1c
2a | Response rate (for both pre and post questionnaires) was 54% (194 patients). Lack of internet facility was the most common reason why patients had not logged on to the portal. Of the 194 patients, 115 (55%) had used the web portal at least once, with 27 patients (13%) had used the portal over 3 times. Eighty six |

			Analysis done using total number of respondents who had completed the questionnaire 5 months after the portal went online (n = 214 patients)	month before the portal went online and 5 months after.		(40%) patients reported to have logged in to their personal space in the portal during the 5 month period. Non users were more likely to be older, single, lower educated and unemployed. During the login 15/86 patients had a single problem with the portal. 33% of all logged in patients felt that the quality of care was higher as a result of the portal. Satisfaction with the rheumatologist / nurse or perceived self – efficacy in patient – provider communication did not show any significant change with the implementation of the portal. Similarly patients did not perceive any significant change in personal control, illness coherence, treatment control and medical adherence.
Fonda 2010	Case study 2010	USA	Patients with **Diabetes Mellitus**. Prototype development included 3 focus group sessions of about 7 participants each.	Participants in the focus groups were met twice, once to assess the needs and once to assess their reaction. The prototype web portal included various applications which captured nutrition, physical activity, glucose levels, medications, mood and education tips.	2a	Based on the focus group sessions, the prototype portal was reorganised to grouping of certain gadgets and addition of various functionality. Some of the concerns included the ability of the PHR to transmit large amounts of data rapidly and the need for local data repository to handle this volume of data. Also, the portal was modified as per patient choices with both manual data entry and automated device entry choices.
Tenforde 2011	Retrospective audit	USA 2008 - 09	10,746 adults of age 18 – 75 years with	Patients engaged in a variety of activities with	2a	Better diabetic profile in PHR users compared to non – users likely secondary

			Diabetes Mellitus. 4,036 patients (37.6%) of the 10,746 patients were enrolled in the PHR.	the PHR, including secure messaging, diagnoses, co-morbidities , lab tests and enter glucometer readings		to their engagement with their health rather than the PHR itself. 4,036 patients (37.6%) of the 10,746 patients were enrolled in the PHR. Median number of login days in 12 month was 9 96% of PHR users reviewed lab tests ordered by the doctor (mean 7.8 times in 12 months). 94% of users read messages (mean 8.3 times) 91% reviewed lab results (mean 5.7 times) Baseline characteristics showed users significantly younger (59 years vs. 62 years), had commercial insurance, Caucasian ethnicity, more household income and better educated compared to non – users. Even after adjustment of demographic and provider characteristics, users had more favourable diabetes measures including documented eye examination, pneumococcal vaccination, prescribed ACE inhibitors / ARB medications and HbA1c recorded in the study period.
Hess 2007	Case Study Focus groups	2004 to 2007 USA	39 **diabetic** patients involved in focus group sessions. Patients were recruited from three primary care practices using two different EHRs.	Web portal included patient provider messaging, ability for patients to enter blood glucose values and generation of alerts to physicians.	1c 2a	Before the implementation of the web portal, patients felt that the system would enhance communication, checking lab results, remind appointments and contact physicians. After the implementation, patients felt more empowered and able to easily communicate. They blood glucose tracking tools were regarded as most

				Five pre implementation (21 patients) and five post implementation focus groups (18 patients). Each lasted 90 minutes.		beneficial. Over the three years there was no change in the number of patients seen per month or in the number of phone calls received. The number of electronic messages increased through the years (from less than 50 per month to over 400 messages per month).
Bourgeois 2009	Case study of a Paediatric / Adolescent PHR	2009 USA	**Paediatric / Adolescent PHR** 929 active users representing 403 parent accounts and 526 patient accounts. Mean age was 9.4 years. Three types of groups included – guardians, adult patients and selected minor patients.	Web based application, allowing patient's access, messaging, billing and edit, annotate and hide data fields. Separate access for each user and linked according to guardian – user relationship.	2a	User login was an average 6.3 times in 3 months of pilot study. Most frequently used function was lab results (82%). Users had also accessed / edited medication list (5.4%), problem / allergy/ immunisation (1 – 2%).
Bourgeois 2008	Opinion on sharing of electronic health data in adolescent and paediatric population	2008, USA	**Paediatric / Adolescent PHR** Patients are classified in 3 groups: <13 years, 13 – 18 years and over 18 years.	Regulations are based on the US State laws governing consent		Authors describe the design of a Paediatric / Adolescent population PHR. Access control is governed by patients age: < 13 years – Parent / guardian having all access and no access for the patient. 13 – 18 years – Parent / Guardian having most access except sensitive data, patients having all access of self related data >18 years, only patient have all access. Authors conclude that information sharing

						in this population requires review of access policies conforming to local laws, professional guidelines, privacy expectations and clinical judgements.
Sox 2010	Case study Usability testing Focus groups	2007 USA	**Paediatric / Adolescent PHR** Total of 15 parents / guardians of shool age children aged 5 – 18 years with a diagnosis of Attention deficit hyperactivity disorder (ADHD) or hyperactivity or impulsivity.	3 focus group sessions: One involving 4 English speaking parents with lower literacy / education achievement, other involving 7 English speaking parents with higher literacy and finally 4 Spanish speaking parents with diverse health literacy. Prototype ADHD portal was tested in the study.	Usability	Very small and un-representative sample size. Overall participant reaction was positive. Most found the homepage easy to use. Of 7 subjects, 5 completed all tasks successfully with one subject failing all 3 tasks. Following concerns were raised during the usability testing: 2 participants felt the web page visually unpleasant, 8 out of 10 failed medication data entry task. Subject with high literacy did not understand how exactly to describe dosage and two subjects entered incorrect drug name and strength.
Grant 2008	Randomised control trial Methodology detailed and good quality	USA 2005 - 2007	11 primary care practices including 244 patients with **Diabetes Mellitus** and who had visited the practice at least once in the prior year	PHR included patient demographics, reviewing their medication list, edit inaccuracies and answer questions on barriers and adverse effects. There was also a function to generate a Diabetes care plan based on the patient response.	1e	An average of 7% to 14% of each practice population accessed the web portal. Users were significantly younger (mean age, 56.1 years vs 60.3 years), Caucasian (89% vs 67%), commercially insured (72% vs 47%), and at or below their HbA1c goal (54% vs 47%) compared with non-users. However both groups had similar glycaemic, blood pressure, and LDL-C control at baseline measurement. There was no significant decline in HbA1c levels with both groups showing three-

						quarters of all patients at their HbA1c goal at the end of the study. A similar pattern seen for blood pressure and LDL-C control. The study included a non-representative sample as only a small proportion of the population enrolled and patients with poor metabolic control were less likely to enrol. Users were likely to be younger, less likely to be from an ethnic minorities and lived in higher income neighbourhoods.
Ronda 2013	Questionnaire Survey Use of Diabetes portal	Netherlands 2011 - 2012	Random selection from 12,793 **Type1 and 2 Diabetic** patients aged 18 – 85 years old. Diabetic patients of whom 1,500 had a web portal login and 3,000 patents without a login, stratified according to the type of Diabetes	62 general practices and one outpatient clinic that use diabetes web portal. Initial survey, followed by two reminders at 3 week intervals. The portal grants patient access to their personal electronic health records which includes information from the medical records. Patients can upload their glucose levels and also communicate with the health care provider.	2a	Of the 4,500 questionnaires resent – 101 were undeliverable, 33 patients had died and 68 did not have correct addressed. Therefore of the valid 4,399 questionnaires there was a response rate of 66.6%. Only 31.6% of the completed questionnaires with 35% of respondents not wanting to participate and 33.4% never responded. Mean age of participants was significantly lower than the non-respondents. Reasons were given for not willing to participate in the study, these included; lack of interest or time (18.1%), questions too difficult (6.6%), too personal (3%) and other reasons (22%). Of the respondents 45.5% had login and they were significantly younger (59.7 years vs. 67.4 years). Type 1 DM had significantly

						more access than Type 2 DM patients (89.8% vs. 41%). In both groups, patients were more likely to have a login if there were younger and more frequently treated by an internist.
Dickinson 2013	Randomised trial 6 months	USA	**Behaviour change** in patients aged 18 to 65 years 7646 patients were mailed, 169 patients enrolled of which 88 patients received basic website access and 81 patients received enhanced website access. Significantly higher number of smokers and of poorer physical health status in the enhanced website group.	A total of 6 practices were recruited (4 urban and 2 rural practices). Basic website included education materials whereas the enhanced sire included individualised action plans which patients can modify, secure messaging and forum section.	1d	Of the 169 patients, 48 patients (21 in basic website and 27 in enhanced website group) completed the six month follow up. No difference in demographics between the groups. There was overall healthful eating scores, physical activity levels, decline in risk factors and 8 item PHQ scores. However there was no difference between the two groups. Authors concluded that behaviour change can be assisted using interactive websites. However they found recruitment difficult and suggested that this can be improved if primary care information was integrated to the PHR.
Kirst 2011	Case study 6 months	USA 2010 - 2011	Patients of age 18 to 75 years with at least one practice visit 14 primary care practices	**Preventive Care** electronic health records which allowed data collection from the patient; interpret and provide personalised recommendations.		Percentage of patients who used the web records ranged from 1.5% to 28.3%, with people who self-selected being slightly older and male. Of those who had an access, 49% had at least one return visit in 3 months and 10~% at least one return visit in 3 – 6 month's time. Average time spent

						was 7 minutes and 21 seconds. Patients generally expressed positive comments to the portal.
Kirst 2012	Randomised controlled trial 16 months Non blinded study	USA	4,500 patients randomly selected to receive a mailed Invitation to use a PHR or usual care.	**Preventive Care and Cancer** 8 primary care practices. Three mailed invitations between 2008 to 2009	1d	Users were significantly older than nonusers (mean age 55.9 years vs 49.5 years) and more likely to be male (55.8% vs 48.9%), non-Hispanic (97.3% vs 93.1%,), and college educated (69.9% vs 62.8%,). Users also had significantly more comorbidities and were more likely to be daily Internet users. At 4 and 16 months, 229 (10.2%) and 378 (16.8%) of invited patients used the IPHR. At 4 months, delivery of colorectal, breast, and cervical cancer screening increased by 19%, 15%, and 13%, respectively, among users.
Weitzman 2012	Cross sectional web survey	USA 2009	Patients over 18 years old or parents of younger patients. 68% of patients were aged 0 – 12 years. Patients should have logged in at least on 5 separate occasions.	**Willingness to share Paediatric patient information**	2a	Of 261 respondents (56% response rate), more reported they would share all information with the state/local public health authority (63.3%) than with an out-of-hospital provider (54.1%). However a few would not share any information with these parties (respectively, 7.9% and 5.2%).

Watts 2013	Randomised controlled trial 3months follow up Described as CONSORT 2010 compliant – not detailed randomisation.	Australia 2012	35 participants with **Major depression,** aged over 18, with access to technology. 15 patients in mobile group and 20 patients in computer group had completed the pre – treatment questionnaires.	Patients were randomly allocated to mobile or computer program. The web based program allowed evaluation of the patient and encouraged activities and also included secure messaging.	1c	68% completed 6 lessons and 65.7% completed the 6 months follow up. Both group of patients showed significant benefits in the Patient Health Questionnaire. Patient satisfaction was 54% very satisfied in mobile group and 64% in computer group, with the rest being somewhat satisfied.
Martinez 2014	Survey Internet use No specific PHR	USA, 2007 to 2008	914 **COPD** patients Response rate: 7.2% (1,077 out of 15,000)	Source: National sample of hou seholds with one COPD patients / Patient support groups / COPD patients receiving Oxygen / COPD related website	2a	Participants without internet access were older (75.1 +/- 7.3 years), less educated and lower income. No difference between gender and ethnicity. Frequent users had at least one COPD exacerbation (58.3% vs 39.1%) and poor QOL scores. Patients with Anxiety, depression and obesity were more likely to use internet frequently than hypertension, arthritis and diabetes. Patients who frequently used the internet were more dissatisfies with their physicians / care – they felt they were treated poorly by the healthcare system (OR 2.46, 95% CI 1.15, 5.24) and that the doctors did not listen to their concerns (OR 3.14, 95% CI 1.42, 6.95)
Nguyen 2012	Randomised trial Methodology:	USA 2007 to 2010	125 **COPD** patients, 43 in internet based Dyspnoea self-management	12 months Dyspnoea with activities measured at 3, 6 and 12 months' time.	1c 1e 2a	No significant differences in dyspnoea with activities between the groups over 12 months, however, all groups showed significant improvement in 6Minute Walk

Randomisation procedure, sample selection not detailed. Dyspnoea self-management program (DSMP)		program (eDSMP), 41 patients in face to face program (fDSMP) and 41 patients in general health education (GHE). (Alerts to nurses sent only for the internet group) 1 lost in follow up in the GHE and eDSMP groups, 2 in the fDSMP group	Consultations, exercise program, self-monitoring and education sessions were delivered. The internet group received web based goal setting tools, diaries, and live chat sessions vs. face to face involving meetings, paper diaries and education modules. The GHE group received home visits and telephone calls / paper copy of education material.	Test (6MWT), duration and frequency of endurance exercise and frequency of strengthening exercise. The DSMP groups performed significantly more arm lifts at 6 and 12 months compared to the GHE group. The eDSMP group logged in a median of 148 times over 12 months 75% using the services at least once in 12 months. The face to face group attended significantly more educational sessions. There were no differences between the number of logged entries between paper and electronic form.

Sample of excluded studies

Study ID	Type	Year / Country	Participants	Setting	Reason for exclusion	Results
Yu 2012	Systematic review	2012	2731 Diabetes patients in Meta-analysis	Usefulness of web-compatible Diabetes related tools	Use of static websites, games and CD ROMS which do not involve patient input as an electronic health record	Moderate but inconsistent effects on clinical and psychological outcomes. Significant heterogeneity on meta-analysis of 12 studies.
Walton	Cross sectional Survey 2000 and 2002	UK 2002	18, 503 mothers of children born between 2000 and 2002, living in the UK at 9 months of age	Personal Child Health record – booklet given to new parents in the UK to record child's growth, development, health service utilisation.	Not Electronic personal health record	Less uptake by women living in disadvantaged community, young maternal age, large family size and lone parent status.
Gysels 2007	Systematic Review	2007	Patients with cancer – either newly diagnosed or at any stage in illness	Thirteen studies involving 1,975 patients There were 7 RCTs and 6 non-randomised studies	Not Electronic Health Record	Six of the 7 RCTs did not find any significant beneficial effect of PHRs One RCT reported that both patients as well as professionals were better informed. Three out of 4 RCTs found no difference in patient satisfaction; one RCT found usual acre patients were more satisfied. (86% versus 58%)
Russell – Minda 2009	Systematic review 1985 to 2008	2009	Diabetes patients	18 RCTs selected across 4 different types of interventions – self monitoring of glucose, pedometer, cell phones	Not a PHR specific SR	Self-monitoring of blood glucose can be an effective tool. Wireless technologies can improve self-care and pedometers are effective in lifestyle modifications

				or wireless technologies		
Smith 2005	Literature review	2005	Study on development and validation of a psychometrically rigorous measure of health-related quality of life (HRQoL) for people with dementia: DEMQOL	Measurement of health-related quality of life for people with dementia: development of a new instrument (DEMQOL) and an evaluation of current methodology	Not a PHR – quality of life tool, not electronic	Not relevant
Ito 2015	Literature review	2015, Japan	Review of integrating psychiatric services into comprehensive dementia care in the community	Patient and family held records of patients with dementia in the community.	Not electronic records	There is potential for family / patient held records in the community.
Varroud-Vial 2011	Authors review and opinion	-	Use of Electronic medical records	Diabetes management	Narrates France government initiative toward PHRs	
McElligott 2010	Survey analysis	-	2004–2006 National Immunisation Survey, a national, validated survey of households with children 19 to 35 months of age	Immunisation	Not Electronic records	Children with vaccination records were more likely to be up to date (83.9% vs 78.6%). The largest effects associated with multiple providers, low maternal education, and those with ≥4 children.
Rada 2005	Case study	2005	Analysis of a retracted Systematic review on interactive health communication applications.	Coding incorrectly performed and meta-analysis of heterogeneous studies	No evidence based information available	
Farrelly 2013	Systematic review	2011	Review of user held personalised information for patients with severe mental		Includes non-electronic information	

			illness			
Mandl 2014	Case study	2011 - 2012	Volunteers from a group within a Diabetes social network.	Surveillance of members of a non-profit online social network	Does not involve individual PHR	
Gega 2007	Case study	UK	Patients with anxiety and depression. 3 case studies described in detail. Most of the users had chronic problems (mean duration of 8 years).	Use of internet based CBT	Not clear if patients had any role in adding / editing data in the intervention.	Of 210 suitable patients, 20% refused the computer based CBT and 29% dropped out early. Patients were 'fairly' satisfied with the system and live support. Authors conclude that computer aided CBT enabled more patients to be treated per hour. Anxiety and depression suffered improved significantly and were satisfied with the online technology. Authors state that roughly the cost advantage has potential to rise from about 15% per patient for 350 patients a year to 41% per patient for 1,350 patients per year. Major barriers identified include funding and training issues.
Rozbroj 2014	Review	-	Mood disorders in lesbians and gay men	24 web and mobile phone based therapies	Includes both computer based and mobile / app based interventions. Analysis not done separately.	
Healey 2014	Case study	New Zealand	Smoking cessation	Analysis of online smoking cessation support network	Not a PHR, public portal (blog) which involved	

					participants to port questions and comments.	