HHS Public Access

Evaluating digital health interventions: key questions and approaches

Elizabeth Murray, FRCGP, FRCP(Edin), PhD,
Director, eHealth Unit and Head of Department, Research Department of Primary Care and Population Health, University College London, London, UK

Eric B. Hekler, PhD,
Assistant Professor and Director, Designing Health Lab, School of Nutrition and Health Promotion, Arizona State University, Phoenix, Arizona, US

Gerhard Andersson, Ph.D,
Department of Behavioural Sciences and Learning, Linköping University, Linköping, Sweden. Department of Clinical Neuroscience, Karolinska Institute, Stockholm, Sweden

Linda M. Collins,
The Methodology Center and Department of Human Development and Family Studies. The Pennsylvania State University, State College, USA

Aiden Doherty,
MRC Clinical Trial Service Unit Hub, Nuffield Department of Population Health, University of Oxford, Oxford, UK

Chris Hollis, PhD, FRCPsych,
Director, NIHR MindTech HTC, University of Nottingham, Nottingham, UK

Daniel E. Rivera, PhD,
Professor, School for the Engineering of Matter, Transport, and Energy, Ira A. Fulton Schools of Engineering, Arizona State University, Phoenix, Arizona, US

Robert West, PhD, and
Professor of Health Psychology and Director of Tobacco Studies, Research Department of Epidemiology and Public Health, University College London, London, UK

Jeremy C. Wyatt, DM, FRCP
Director, Wessex Institute, University of Southampton, Southampton, UK

Abstract

Corresponding author information: Elizabeth Murray FRCGP FRCP(Edin) PhD, Research Department of Primary Care and Population Health, University College London, Upper Floor 3, Royal Free Hospital, Rowland Hill Street, London NW3 2PF, United Kingdom, Elizabeth.murray@ucl.ac.uk, Tel: +44 (0) 207 794 0500 ext 36747, Fax: +44 (0)20 7472 6871.

Digital health interventions (DHI) have enormous potential as scalable tools to improve health and healthcare delivery by improving effectiveness, efficiency, accessibility, safety and personalisation. Achieving these improvements requires a cumulative knowledge base to inform development and deployment of DHI. However, evaluations of DHI present special challenges. This paper aims to examine these challenges and outline an evaluation strategy in terms of the Research Questions (RQs) needed to appraise DHIs. As DHI are at the intersection of biomedical, behavioural, computing and engineering research, methods drawn from all these disciplines are required. Relevant RQs include defining the problem and the likely benefit of the DHI, which in turn requires establishing the likely reach and uptake of the intervention, the causal model describing how the intervention will achieve its intended benefit, key components and how they interact with one another, and estimating overall benefit in terms of effectiveness, cost-effectiveness and harms. While Randomised Controlled Trials (RCTs) are important for evaluation of effectiveness and cost-effectiveness, they are best undertaken only when: a) the intervention and its delivery package are stable; b) these can be implemented with high fidelity and c) there is a reasonable likelihood that the overall benefits will be clinically meaningful (improved outcomes or equivalent outcomes at less cost). Broadening the portfolio of RQs and evaluation methods will help with developing the necessary knowledge base to inform decisions on policy, practice and research.

Background & Aims

There is enormous potential for digital health interventions (i.e. interventions delivered via digital technologies such as smartphones, website, text messaging) to provide effective, cost-effective, safe, and scalable interventions to improve health and healthcare. Digital health interventions (DHI) can be used to promote healthy behaviours (e.g. smoking cessation, [1, 2] healthy eating, [3] physical activity, [4] safer sex [5] or alcohol consumption [6]), improve outcomes in people with long term conditions [7] such as cardiovascular disease (McClean in press), diabetes [8] and mental health conditions [9] and provide remote access to effective treatments, for example computerised cognitive behavioural therapy for mental health and somatic problems. [10–14] They are typically complex interventions with multiple components, and many have multiple aims including enabling users to be better informed about their health, share experiences with others in similar positions, change perceptions and cognitions around health, assess and monitor specified health states or health behaviours, titrate medication, clarify health priorities and reach treatment decisions congruent with these, and improve communication between patients and health care professionals (HCP). Active components may include information, psycho-education, personal stories, formal decision aids, behaviour change support, interactions with HCP and other patients, self-assessment or monitoring tools (questionnaires, wearables, monitors, and effective theory-based psychological interventions developed for face-to-face delivery such as cognitive behavioural therapy or mindfulness training.

To date, the potential of DHIs has scarcely been realized, partly because of difficulties generating an accumulating knowledge-base for guiding decisions about DHI's. Difficulties include the rapid change of the wider technology landscape (Patrick 2016), which requires DHI's to constantly evolve and be updated just to remain useful, let alone improve. For example, imagine an iPhone app promoting physical activity, with development and

evaluation starting in 2008. Results from a randomized controlled trial may not be published till 5–6 years later, by which time the iPhone operating system (iOS) had undergone substantial changes to functionality, design, and overall use. These operating system changes would result in the evaluated app feeling out-of-date at best and non-functional at worst. As such, the knowledge gained from that efficacy trial would be minimally useful for supporting current decisions about using that app. Other difficulties include the idiosyncratic wants and needs of users and the influence of context on effectiveness.

However, the public, patients, clinicians, policy-makers and healthcare commissioners all have to make decisions on DHI now, and researchers need to support such decision-making by creating an actionable knowledge base to identify the most effective, cost-effective, safe, and scalable interventions (and components) for improving individual and population health. These decisions are particularly important in resource-constrained contexts.

This paper explores issues that arise in developing an accumulating knowledge base around DHI, and how this knowledge can be generated in a timely manner, using scarce resources efficiently. The approach is pragmatic, with a focus on decision-making and moving the science forward, generating cumulative knowledge around identifying important components and working out how to test them with a view to improving the quality and effectiveness of DHI and the efficiency of the research process. This paper is written from the perspective of a body charged with appraising evidence for using specific DHI within a publically-funded, resource-limited health system, such as the UK National Institute for Health and Care Excellence (NICE).

This paper does not seek to provide detailed analysis of appropriate design features of evaluation studies such as choice of comparators, outcome measures, mediator and moderator variables, study samples, or the occasions when particular study designs are a better fit with the evaluation context. These are important issues for which a literature is beginning to emerge. [15, 16]

Paper structure

The paper starts by defining the Research Questions (RQ) which, in our opinion, should form the basis for an appraisal of a DHI (Table 1). It then considers appropriate research methods for each of these RQ. Where the appropriate methods are largely similar to those used in research of other (non-digital) complex interventions, readers are referred to the appropriate references. Where there are novel or specific issues which arise, or are particularly salient, in evaluation of digital health interventions, the main areas of consideration for each issue are outlined. Throughout, the paper emphasises that the RQ apply not just to the digital components of the DHI, but also the surrounding "delivery package". This package will vary according to the nature and functions of the DHI, but often requires as much thought and study as the DHI itself. Example components of delivery packages could include system redesign where use of DHI becomes standard clinical practice, [17] ad hoc referral from a clinician, [18] supported access (eg. face-to-face, [19] by telephone, [20] or email [21]), hosting on a trusted portal (e.g. NHS Choices), marketing via public health campaigns, or embedding in a social network.

Research Questions

Defining the problem

1 *Is there a clear health need which this DHI is intended to address?*

2 *Is there a defined population who could benefit from this DHI?*

As with any complex intervention, consideration of the likely benefits of a digital health intervention starts with a detailed and often theory-based characterisation of the nature of the problem and the context in which the intervention will be used. [22–24]

Defining the likely benefit of the DHI

3 *Is the DHI likely to reach this population, and if so, is the population likely to use it?*

The concepts of reach, uptake and context are particularly salient for DHI, as impact and cost-effectiveness are highly dependent on the total number of users (McNamee 2016), and effectiveness may be highly dependent on context. For example, effects seen when a DHI is used in a controlled environment (laboratory or clinical office) may not be replicated if used in the "wild", with many competing demands on user's attention. An important consideration is whether a DHI is accessible across a range of commonly used operating systems and devices and is interoperable with other healthcare information systems, such as electronic health records (EHRs). Hence an early component of any evaluation of a DHI should be a determination and optimisation of reach and uptake by the intended population, in the context in which the DHI will be used. This will often require iterative adaptations both to the DHI itself (e.g. to improve usability, acceptability) and to the "delivery package" around the DHI. For many DHI, 'users' will include healthcare professionals (HCPs) who 'prescribe' the DHI and monitor outcomes. Hence RQs 3 - 6 require work with HCPs as well as patients or the public.

Establishing and optimising potential reach and uptake requires methods used in engineering and computer science, collectively referred to as human-centred design [25–27]. These include concept sketching, [28] co-design strategies [25] and low-fidelity or "Wizard-of-Oz" prototyping, [2930] and user experience testing. [27] In the business world there is increasing interest in "lean" principles [31] that specify methods for early-stage testing of features related to feasibility [32] including:

- *Acceptability* and usability (will the target audience (e.g. patients, HCPs) incorporate and sustain the intervention into their lives/clinical practice?);

- *demand*, (will relevant stakeholders use it?);

- *implementation*, (will it have high fidelity within real-world use?);

- *practicability*, (can it be delivered with minimal burden?);

- *adaptation*, (can it be adapted to novel contexts without compromising fidelity and integrity?); and

- *integration*, (can it be integrated successfully into existing healthcare systems?).

4 **_Is there a credible causal explanation for the DHI to achieve the desired impact?_**

Establishing a credible causal explanation for the DHI is essential and must address not only the DHI, but also the "delivery package". For example, if there is a human support element, is that element aimed entirely at improving engagement with the DHI, or will there be additional therapeutic content embedded in the human support? Are there important issues around the credibility or authority invested in those that deliver the human support? See (Hekler 2016) and (Yardley 2016) for further discussion.

5 **_What are the key components of the DHI? Which ones impact on the predicted outcome, and how do they interact with each other?_**

Understanding which components actually have the predicted impact on the outcome, and whether and how components interact, is critical. Most DHI are highly complex interventions, containing multiple components so the development process needs to include a period of optimisation. This entails evaluating the performance of individual components of the intervention, and how the presence, absence, or setting of one component impacts the performance of another. One efficient method is the Multiphase Optimisation Strategy (MOST), [33, 34] which involves establishing a set of components that are candidates for inclusion, specifying an optimization criterion for the entire intervention, and then collecting experimental data to identify the subset of components that meet the criterion. Here the term "component" is broadly defined, and may refer to aspects of the content of the intervention, including any human input; [35] factors affecting compliance with, adherence to, fidelity of, or scalability of the intervention; [36] variables and decision rules used to tailor intervention strategy, content, or intensity to individuals; [37] or any aspect of an intervention that can profitably be separated out for examination. Two example optimization criteria are: the most effective intervention that can be delivered for < £100 per participant; or most effective intervention that requires no more than one hour per week of participant time.

The experimental approaches used for optimization include full or fractional factorial experiments [38, 39], the sequential multiple-assignment randomized trial (SMART), [40] and system identification techniques [41, 42]. The factorial experimental design can be a useful and economical approach for examining the effects of individual intervention components, and is the only experimental design that enables full examination of all interactions. [38] For further discussion see Collins et al. [38, 43]

6 **_What strategies should be used to support tailoring the DHI to participants over time?_**

Where the research question focuses on tailoring the DHI to participants over time (e.g., non-responders, or daily adjustments reflecting changing needs or context) a SMART design, micro-randomized trial, or system identification experiment may be appropriate. A SMART is a special case of the factorial experiment involving randomization at several stages, where each stage corresponds to one of the decisions that must be made about adapting the intervention, and some or all of the randomization may be contingent on response to treatment. [34, 40, 44]

System identification approaches are used in engineering to obtain dynamic systems models; these in turn are the basis for the design of control systems which achieve optimization. [45] System identification experiments are inherently idiographic in nature, and work best when planned changes (preferably random or pseudo-random in nature) are introduced to adjustable components of an intervention (e.g. dosages). After obtaining experimental data, the system identification methodology guides decisions of model structure, parameter estimation, and model validation prior to dictating the usefulness of the model for controller design. For examples see Timms et al. [46] and Deshpande et al; [47] experimental procedures involving pseudo-random multisine signals are currently being evaluated in a physical activity intervention based on Social Cognitive Theory. [48]

7 *What is the likely direction and magnitude of the effect of the DHI or its components compared to a comparator which is meaningful for the stage of the research process?*

8 *How confident are we about the magnitude of the effect of the DHI or its components compared to a comparator which is meaningful for the stage of the research process?*

Once RQ 3–6 have been addressed, the research team are likely to be able to estimate the direction and magnitude of the effect of the DHI. If this estimate suggests that the DHI is likely to be beneficial to individuals or a population, has sufficient acceptability and feasibility to ensure adequate reach and uptake for cost-effectiveness, and when the total treatment package (i.e. DHI + delivery package + context of use) have all been iterated and adapted to the point where the treatment package is likely to remain relatively stable over the medium term, it may be appropriate to undertake a definitive randomised controlled trial to establish the magnitude of the effect (effect size) of the DHI compared to a meaningful comparator. "Relatively stable" is a matter for investigator judgement, guided by the causal explanation and optimisation data. [49] The wider technological landscape is likely to continue to evolve, and investigators must judge what impact this will have on the generalisability of their findings. The importance of undertaking a RCT, and not relying solely on formative studies is evidenced by the fact that RCTs have repeatedly overturned assumptions drawn from observational or non-randomised studies (e.g.[50, 51]). Hence the assumption of equipoise, required for a trial to be ethical, does hold. Although the general principles of designing and conducting RCTs for complex interventions [22] are applicable to DHIs, there are specific features of DHI which need consideration if a trial is to provide useful evidence that supports rational decision-making. These include:

- The context in which the trial is undertaken

- The trade-off between external and internal validity

- Specification of the intervention and delivery platform

- Choice and specification of the comparator

- Establishing separate data collection methods from the DHI itself

The importance of context has been described above (RQ 3 & 5). Understanding, defining and describing the context in which an RCT is undertaken is necessary to inform judgements

around the generalisability of the results outside the trial environment, particularly before implementing a DHI in a different context.

Deciding how to balance external and internal validity is a challenge for many trials [52], but is particularly salient for trials of DHI. External validity refers to the extent to which the results apply to "a definable group of patients in a particular setting", while internal validity is based on how the design and conduct of the trial minimises potential for bias [53]. The emphasis in trials of pharmaceutical products is on internal validity and reducing bias, and extensive work has confirmed the importance of this. [54] However, there are real questions as to how well approaches developed to reduce bias in drug trials translate to trials of complex interventions in general [52] and to digital interventions in particular, including concerns about the degree to which design features that enhance internal validity jeopardise external validity. For example, poor retention to the trial, leading to missing follow-up data, may be countered by boosting the human component of the trial by undertaking some of the trial activities face-to-face, or by recruiting highly motivated participants who may be unrepresentative of the people who would use the intervention in routine practice. Hence data from trials apparently at low risk of bias may paradoxically be less appropriate for informing policy than those with potentially greater risk of bias but better generalisability.

Detailed specification of the DHI is important, but may be hard to achieve, particularly where there is a high degree of tailoring, adaptive learning and user choice. By specification we mean having an agreed framework for classifying the intervention components, including the degree of human input and components which are individually tailored. Such specification is required for replication of trial results, comparison between DHI, synthesising data across trials in systematic reviews and meta-analyses [55] and may help with determining the criteria for 'substantial equivalence' of digital interventions. The concept of 'substantial equivalence' is used for medical device and pharmaceutical regulation by the FDA and similar regulatory bodies. Essentially, if a pivotal trial exists, interventions meeting criteria for 'substantial equivalence' would not require further RCT evidence. For example, if a pivotal RCT (or meta-analysis) demonstrated effectiveness of a mindfulness-based digital intervention for depression, then each new mindfulness app for depression would not be required to undergo RCT testing – but to demonstrate substantial equivalence to existing 'predicate' interventions. [56] The relevant data to collect would then focus on usage, adherence, demographic access parameters, user preferences etc.

The selection of a suitable comparator is determined by the research question addressed, which will vary with the stage of the research. In pragmatic trials which aim to determine the effectiveness of a new treatment compared to current best practice, the comparator is typically 'treatment as usual' (TAU). However, in trials of DHIs, the participants in the TAU group may have access to a myriad of other digital interventions. People accustomed to using digital interventions are often also accustomed to searching online for resources. Someone who has sought help for a particular problem, entered a trial, and been randomised to the comparator arm, and who finds the comparator intervention unhelpful, may well search online until they find a better resource. [57] This activity may be hard to prevent or track, but risks undermining the trial.

In head-to-head RCTs, where the effects of two (or more) DHIs are compared with each other or against a face-to-face intervention, it is important to define which components of the comparator interventions are the same and which are different. Here the specification of the comparator should follow the same principles as the specification of the intervention outlined above. [55]

There is a temptation in RCTs of DHI to embed data collection into the intervention, but this may introduce systematic bias or confound the intervention with the measurement method. This bias may favour the intervention or, by more accurately recording adverse events, it may appear to show that the intervention is causing harm.

9 *Has the possibility of harm been adequately considered? And the likelihood of risks or adverse outcomes assessed?*

Digital health interventions are not harm free, although to date, the data on actual harms are relatively sparse. There are various mechanisms by which DHI could result in harm. First, they could be designed to achieve an outcome which is widely viewed as harmful, for example websites which promote suicide. Secondly, DHI can make fraudulent claims, which if believed, can result in the user experiencing harm. Examples of this include apps that claim to promote safer consumption of alcohol, including providing estimates of blood alcohol concentration (BAC) to enable users to determine whether they are safe to drive, but which do not in fact have any capacity to estimate BAC. [58] Alternatively, a DHI could contain inaccurate information or advice. Thirdly, a DHI could provide accurate information and advice, but this could be misinterpreted or wrongly applied, leading to that decisions which harm health. Alternatively, this accurate information could lead to increased anxiety or depression. Fourthly, ineffective DHI lead to opportunity costs for users, and if paid for by a health service, opportunity costs for the system. If individuals or systems put resources (funds, time, effort) into ineffective interventions, those resources are not available for effective interventions. Fifthly, individuals (and systems) may become disillusioned and despondent if they use ineffective interventions, leading to a belief that either the individual is incapable of responding to treatment, or that all DHI are useless and no further effort should be invested. Finally, DHIs may 'leak' personal data because of inadequate security and encryption functions. [59]

All developers of DHI should actively consider the possibility of harm and include evaluations that look for potential harms including breaches of privacy and information governance. Identification and quantification of expected harms (such as increased anxiety) can be undertaken as part of an RCT, but unexpected harms will require alternative strategies for identification and quantification. Some may emerge during the development and optimisation work, while others may require long-term observational studies during widespread implementation.

10 *Has cost been adequately considered and measured?*

It is essential to consider sustainability and cost-effectiveness from the very beginning of the development of a DHI. The development phase should include consideration of the long term costs of maintenance and updating, how these costs could be met, and who will take

responsibility for them. Methods for undertaking a formal health economic analysis is addressed in detail by (McNamee et al 2016).

11 *What is the overall assessment of the utility of this intervention? And how confident are we in this overall assessment?*

12 *Should research priorities change?*

13 *Should clinical practice change?*

Answers to the previous ten questions should enable an assessment of the overall utility of the DHI (e.g., balancing its effects, usage, scalability, costs and safety), along with an estimate of how confidence in this assessment. This in turn can guide decision-making about research priorities and clinical practice. This assessment may range from considering that there is sufficient evidence of beneficial effect with sufficient confidence in the effect size, along with adequate understanding of the costs, scalability, sustainability and risks of harm for a specific DHI that it should be incorporated into routine clinical practice, to realising that a given DHI is so unlikely ever to have either sufficient clinical impact or reach that no further research resource should be invested in it.

Discussion and conclusions

This paper outlines a research question-driven approach to the evaluation of DHI, which should lead to an accumulating knowledge base around such interventions in a timely and resource efficient manner. Good research in this area requires fertile multi-disciplinary collaborations which draw on insights and experience from multiple fields including clinical medicine, health services research, behavioural science, education, engineering, and computer science. Researchers from an engineering or computer science background may be surprised by the reliance on randomised controlled trials, while those from a biomedical or behavioural sciences background may consider there is too much emphasis on methods other than RCTs. The view put forward in this paper is that definitive, well-designed RCTs remain an important part of the overall toolkit for evaluating DHI, but only one part. Researchers in this field could learn from the iterative approach adopted by engineering and computer science where interventions undergo multiple cycles of development and optimisation. A definitive trial should be undertaken only once: a) the intervention together with the delivery package around it have reached a degree of stability such that future developments can be considered relatively minor, b) there is reasonable confidence that the intervention plus delivery package can be implanted with high fidelity and c) there is a reasonable likelihood that the overall benefits will be clinically meaningful and lead to either improved outcomes or equivalent outcomes at less cost.

How best to combine rigor with efficiency in evaluating DHI requires a great deal of methodological research. Areas to explore in future methodological research include:

1. *Enabling individual studies to generate more useful data:*

a. Consideration and validation of appropriate short-term proxy outcomes, together with identification of when use of these is appropriate, and when definitive outcomes such as health status are needed;

b. Improving methods for early formative work, to make it as efficient as possible, and define if further investment in more intensive research designs and development processes is warranted.

c. Better understanding of how to improve the internal validity of RCTs of DHI in terms of retention and follow-up, without jeopardising external validity in terms of the population recruited or impact on the intervention;

d. Improved methods for reducing the large amounts of missing data that may occur, and addressing the inevitable biases this raises;

e. Better methods for determining whether and how a DHI will become scalable and sustainable, including understanding how a DHI might be supported through self-sustaining business models.

2. *Enabling more useful synthesis and comparison of data generated by different studies:*

a. Identification, specification and classification of important contextual factors;

b. Specification and classification of target populations;

c. Specification and classification of DHI, so that we can gain an understanding of the important active components, mechanism of action, replicate and synthesise evidence across DHI evaluations and begin to address the issue of determining "substantial equivalence" between DHIs;

d. Specification and determination of appropriate comparators, according to the stage of the research process;

e. Improved reporting of studies of DHI, building on initiatives such as the TIDIER reporting guideline [55] and the CONSORT–EHEALTH statement. [60]

Reference List

1. Free C, Knight R, Robertson S, et al. Smoking cessation support delivered via mobile phone text messaging (txt2stop): a single-blind, randomised trial. Lancet. Jul 2; 2011 378(9785):49–55. [PubMed: 21722952]

2. Zhuo X, Zhang P, Barker L, Albright A, Thompson TJ, Gregg E. The lifetime cost of diabetes and its implications for diabetes prevention. Diabetes Care. Sep; 2014 37(9):2557–2564. [PubMed: 25147254]

3. Harris J, Felix L, Miners A, et al. Adaptive e-learning to improve dietary behaviour: a systematic review and cost-effectiveness analysis. Health Technol Assess. 2011; 15(37):1–160.

4. Peyrot M, Rubin RR. Access to diabetes self-management education. Diabetes Educ. Jan-Feb;2008 34(1):90–97. [PubMed: 18267995]

5. Bailey JV, Murray E, Rait G, et al. Interactive computer-based interventions for sexual health promotion. Cochrane Database Syst Rev. 2010; (9):CD006483. [PubMed: 20824850]

6. Khadjesari Z, Murray E, Hewitt C, Hartley S, Godfrey C. Can stand-alone computer-based interventions reduce alcohol consumption? A systematic review. Addiction. 2011; 106(2):267–282. [PubMed: 21083832]

7. Murray, E., Burns, J., See Tai, S., Lai, R., Nazareth, I. The Cochrane Library. 2005. Interactive Health Communication Applications for people with chronic disease.

8. Pal K, Eastwood SV, Michie S, et al. Computer-based diabetes self-management interventions for adults with type 2 diabetes mellitus. Cochrane Database Syst Rev. 2013; 3:CD008776.doi: 10.1002/14651858.CD008776.pub2.:CD008776

9. Zhuo X, Zhang P, Kahn HS, Bardenheier BH, Li R, Gregg EW. Change in medical spending attributable to diabetes: national data from 1987 to 2011. Diabetes Care. Apr; 2015 38(4):581–587. [PubMed: 25592194]

10. Andersson E, Ljotsson B, Smit F, et al. Cost-effectiveness of internet-based cognitive behavior therapy for irritable bowel syndrome: results from a randomized controlled trial. BMC Public Health. 2011; 11:215. [PubMed: 21473754]

11. Kaldo V, Haak T, Buhrman M, Alfonsson S, Larsen HC, Andersson G. Internet-based cognitive behaviour therapy for tinnitus patients delivered in a regular clinical setting: outcome and analysis of treatment dropout. Cogn Behav Ther. 2013; 42(2):146–158. [PubMed: 23432207]

12. Iacobucci G. Diabetes prescribing in England consumes nearly 10% of primary care budget. Bmj. 2014; 349:g5143. [PubMed: 25121512]

13. Kerr M, Rayman G, Jeffcoate WJ. Cost of diabetic foot disease to the National Health Service in England. Diabet Med. Dec; 2014 31(12):1498–1504. [PubMed: 24984759]

14. Kaltenthaler E, Parry G, Beverley C, Ferriter M. Computerised cognitive-behavioural therapy for depression: systematic review. Br J Psychiatry. 2008; 193(3):181–184. [PubMed: 18757972]

15. Murray E, Khadjesari Z, White IR, et al. Methodological challenges in online trials. J Med Internet Res. 2009; 11(2):e9. [PubMed: 19403465]

16. RMSW. A Guide to Development and Evaluation of Digital Behaviour Change Interventions in Healthcare. Version 1. London: UCL Centre for Behaviour Change; 2016.

17. Titov N, Dear BF, Staples LG, et al. MindSpot Clinic: An Accessible, Efficient, and Effective Online Treatment Service for Anxiety and Depression. Psychiatr Serv. Jul 1.2015 appips201400477.

18. Bower P, Kontopantelis E, Sutton A, et al. Influence of initial severity of depression on effectiveness of low intensity interventions: meta-analysis of individual patient data. BMJ. 2013; 346:f540. [PubMed: 23444423]

19. Schafer I, Pawels M, Kuver C, et al. Strategies for improving participation in diabetes education. A qualitative study. PLoS One. 2014; 9(4):e95035. [PubMed: 24733428]

20. Dennison L, Morrison L, Lloyd S, et al. Does brief telephone support improve engagement with a web-based weight management intervention? Randomized controlled trial. J Med Internet Res. 2014; 16(3):e95. [PubMed: 24681761]

21. Titov N, Andrews G, Davies M, McIntyre K, Robinson E, Solley K. Internet treatment for depression: a randomized controlled trial comparing clinician vs. technician assistance. PLoS ONE. 2010; 5(6):e10939. [PubMed: 20544030]

22. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M. Developing and evaluating complex interventions: the new Medical Research Council guidance. BMJ. 2008; 337:a1655.doi: 10.1136/bmj.a1655.:a1655 [PubMed: 18824488]

23. Campbell NC, Murray E, Darbyshire J, et al. Designing and evaluating complex interventions to improve health care. BMJ. 2007; 334(7591):455–459. [PubMed: 17332585]

24. Murray E, Treweek S, Pope C, et al. Normalisation process theory: a framework for developing, evaluating and implementing complex interventions. BMC Med. 2010; %20(8):63.

25. Nouwen A, Winkley K, Twisk J, et al. Type 2 diabetes mellitus as a risk factor for the onset of depression: a systematic review and meta-analysis. Diabetologia. Dec; 2010 53(12):2480–2486. [PubMed: 20711716]

26. Nouwen A, Lloyd CE, Pouwer F. Depression and type 2 diabetes over the lifespan: a meta-analysis. Response to Mezuk et al. Diabetes Care. May.2009 32(5):e56. author reply e57. [PubMed: 19407066]

27. Nouwen A. Depression and diabetes distress. Diabet Med. Oct; 2015 32(10):1261–1263. [PubMed: 26202578]

28. Kahn R, Davidson MB. The reality of type 2 diabetes prevention. Diabetes Care. Apr; 2014 37(4): 943–949. [PubMed: 24652724]

29. Leung MY, Pollack LM, Colditz GA, Chang SH. Life years lost and lifetime health care expenditures associated with diabetes in the U.S., National Health Interview Survey, 1997–2000. Diabetes Care. Mar; 2015 38(3):460–468. [PubMed: 25552420]

30. Fisher L, Hessler D, Masharani U, Strycker L. Impact of baseline patient characteristics on interventions to reduce diabetes distress: the role of personal conscientiousness and diabetes self-efficacy. Diabet Med. Jun; 2014 31(6):739–746. [PubMed: 24494593]

31. Pouwer F, Nefs G, Nouwen A. Adverse effects of depression on glycemic control and health outcomes in people with diabetes: a review. Endocrinol Metab Clin North Am. Sep; 2013 42(3): 529–544. [PubMed: 24011885]

32. Demakakos P, Zaninotto P, Nouwen A. Is the association between depressive symptoms and glucose metabolism bidirectional? Evidence from the English Longitudinal Study of Ageing. Psychosom Med. Sep; 2014 76(7):555–561. [PubMed: 25077428]

33. Deglise C, Suggs LS, Odermatt P. SMS for disease control in developing countries: a systematic review of mobile health applications. J Telemed Telecare. Jul; 2012 18(5):273–281. [PubMed: 22826375]

34. Collins LM, Nahum-Shani I, Almirall D. Optimization of behavioral dynamic treatment regimens based on the sequential, multiple assignment, randomized trial (SMART). Clin Trials. Jun 5; 2014 11(4):426–434. [PubMed: 24902922]

35. Mukund Bahadur KC, Murray PJ. Cell phone short messaging service (SMS) for HIV/AIDS in South Africa: a literature review. Stud Health Technol Inform. 2010; 160(Pt 1):530–534. [PubMed: 20841743]

36. Cauch-Dudek K, Victor JC, Sigmond M, Shah BR. Disparities in attendance at diabetes self-management education programs after diagnosis in Ontario, Canada: a cohort study. BMC Public Health. 2013; 13:85. [PubMed: 23360373]

37. Strecher VJ, McClure JB, Alexander GL, et al. Web-based smoking-cessation programs: results of a randomized trial. Am J Prev Med. 2008; 34(5):373–381. [PubMed: 18407003]

38. Vaidya V, Gangan N, Sheehan J. Impact of cardiovascular complications among patients with Type 2 diabetes mellitus: a systematic review. Expert Rev Pharmacoecon Outcomes Res. Jun; 2015 15(3):487–497. [PubMed: 25824591]

39. Nielsen AB, Jensen P, Gannik D, Reventlow S, Hollnagel H, de Olivarius NF. Change in self-rated general health is associated with perceived illness burden: a 1-year follow up of patients newly diagnosed with type 2 diabetes. BMC Public Health. 2015; 15:439. [PubMed: 25924731]

40. Dall TM, Storm MV, Semilla AP, Wintfeld N, O'Grady M, Narayan KM. Value of lifestyle intervention to prevent diabetes and sequelae. Am J Prev Med. Mar; 2015 48(3):271–280. [PubMed: 25498548]

41. Dunkley AJ, Bodicoat DH, Greaves CJ, et al. Diabetes prevention in the real world: effectiveness of pragmatic lifestyle interventions for the prevention of type 2 diabetes and of the impact of adherence to guideline recommendations: a systematic review and meta-analysis. Diabetes Care. Apr; 2014 37(4):922–933. [PubMed: 24652723]

42. Ng CS, Lee JY, Toh MP, Ko Y. Cost-of-illness studies of diabetes mellitus: a systematic review. Diabetes Res Clin Pract. Aug; 2014 105(2):151–163. [PubMed: 24814877]

43. Roland, MBN., Howe, A., Imision, C., Rubin, G., Storey, K. The future of primray care: Creating teams for tomorrow. London: 2015.

44. Scalone L, Cesana G, Furneri G, et al. Burden of diabetes mellitus estimated with a longitudinal population-based study using administrative databases. PLoS One. 2014; 9(12):e113741. [PubMed: 25470484]

45. Roy T, Lloyd CE. Epidemiology of depression and diabetes: a systematic review. J Affect Disord. Oct; 2012 142(Suppl):S8–21. [PubMed: 23062861]

46. Pronk NP, Remington PL. Combined Diet and Physical Activity Promotion Programs for Prevention of Diabetes: Community Preventive Services Task Force Recommendation Statement. Ann Intern Med. Sep 15; 2015 163(6):465–468. [PubMed: 26168073]

47. Deshpande S, Rivera DE, Younger JW, Nandola NN. A control systems engineering approach for adaptive behavioral interventions: illustration with a fibromyalgia intervention. Transl Behav Med. Sep; 2014 4(3):275–289. [PubMed: 25264467]

48. Smith KJ, Beland M, Clyde M, et al. Association of diabetes with anxiety: a systematic review and meta-analysis. J Psychosom Res. Feb; 2013 74(2):89–99. [PubMed: 23332522]

49. Mohr DC, Schueller SM, Riley WT, et al. Trials of Intervention Principles: Evaluation Methods for Evolving Behavioral Intervention Technologies. J Med Internet Res. 2015; 17(7):e166. [PubMed: 26155878]

50. Panagioti, M., Richardson, G., Murray, E., et al. Health Services and Delivery Research. Reducing Care Utilisation through Self-management Interventions (RECURSIVE): a systematic review and meta-analysis. Southampton (UK): NIHR Journals Library Copyright (c) Queen's Printer and Controller of HMSO; 2014. This work was produced by Panagioti et al. under the terms of a commissioning contract issued by the Secretary of State for Health. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.; 2014

51. Panagioti M, Richardson G, Small N, et al. Self-management support interventions to reduce health care utilisation without compromising outcomes: a systematic review and meta-analysis. BMC Health Serv Res. 2014; 14:356. [PubMed: 25164529]

52. Betjeman TJ, Soghoian SE, Foran MP. mHealth in Sub-Saharan Africa. Int J Telemed Appl. 2013; 2013:482324. [PubMed: 24369460]

53. Rothwell PM. External validity of randomised controlled trials: "to whom do the results of this trial apply?". Lancet. 2005; 365(9453):82–93. [PubMed: 15639683]

54. Peyrot M, Rubin RR, Funnell MM, Siminerio LM. Access to diabetes self-management education: results of national surveys of patients, educators, and physicians. Diabetes Educ. Mar-Apr;2009 35(2):246–248. 246, 258–263. [PubMed: 19208816]

55. Hoffmann TC, Glasziou PP, Boutron I, et al. Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. BMJ. 2014; 348:g1687. [PubMed: 24609605]

56. Pintaudi B, Lucisano G, Gentile S, et al. Correlates of diabetes-related distress in type 2 diabetes: Findings from the benchmarking network for clinical and humanistic outcomes in diabetes (BENCH-D) study. J Psychosom Res. Nov; 2015 79(5):348–354. [PubMed: 26526307]

57. Khadjesari Z, Stevenson F, Godfrey C, Murray E. Negotiating the 'grey area between normal social drinking and being a smelly tramp': a qualitative study of people searching for help online to reduce their drinking. Health Expect. Feb 12.2015

58. Alva ML, Gray A, Mihaylova B, Leal J, Holman RR. The impact of diabetes-related complications on healthcare costs: new results from the UKPDS (UKPDS 84). Diabet Med. Apr; 2015 32(4): 459–466. [PubMed: 25439048]

59. Jones A, Vallis M, Pouwer F. If it does not significantly change HbA1c levels why should we waste time on it? A plea for the prioritization of psychological well-being in people with diabetes. Diabet Med. Feb; 2015 32(2):155–163. [PubMed: 25354315]

60. Eysenbach G, Group C-E. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res. 2011; 13(4):e126. [PubMed: 22209829]

Key guidance points and priority topics for future research

Guidance points based on existing research

1. The efficient development of safe, effective, widely accessible DHIs requires innovative research methods to generate an accumulating knowledge base that can be used to guide decision making.

2. Reach and uptake are crucial determinants of the overall impact of a DHI, and can be determined and improved using human-centred design methods.

3. Sustainability and revenue models should be considered early in the development process.

4. Defining a clear causal model that accounts for the multiple components of a DHI and the surrounding delivery package is essential.

5. Identifying the essential or active components of a DHI or its delivery package can be done using a framework derived from engineering known as Multiphase Optimisation Strategy (MOST).

6. Randomised controlled trials (RCTs) remain an important method for determining DHI impact in terms of effectiveness and cost-effectiveness, but are best undertaken once the DHI and its delivery package are stable, can be implemented with high fidelity, and are highly likely to lead to clinically meaningful benefits.

The key priority is to improve the efficiency of evaluations without jeopardising rigor. Achieving this will entail

1. *Enabling individual studies to generate more useful data through:* improving methods of early formative work; better understanding of when and how short-term proxy outcomes should be used and when definitive outcomes are needed; better methods for improving internal validity of trials without jeopardising external validity; improved methods for enhancing DHI uptake and minimising missing data; and better methods for considering whether and how DHI will become scalable and sustainable.

2. *Enabling more useful synthesis and comparison of data generated by different studies through:* improved specification and classification of context, target populations, digital health interventions and their components, using more appropriate comparators for the stage of the research process, and improved reporting of trials of DHI.

Table 1

Key Research Questions for an appraisal of a DHI.

Defining the problem

1 Is there a clear health need which this DHI is intended to address?

2 Is there a defined population which could benefit from this DHI?

Defining the likely benefit of the DHI

3 Is the DHI likely to reach this population, and if so, is the population likely to use it?

4 Is there a credible causal explanation for the DHI to achieve the desired impact?

5 What are the key components of the DHI? Which ones impact on the predicted outcome, and how do they interact with each other?

6 What strategies should be used to support tailoring the DHI to participants over time?

7 What is the likely direction and magnitude of the effect of the DHI or its components compared to a comparator which is meaningful for the stage of the research process?

8 How confident are we about the magnitude of the effect of the DHI or its components compared to a comparator which is meaningful for the stage of the research process?

9 Has the possibility of harm been adequately considered? And the likelihood of risks or adverse outcomes assessed?

10 Has cost been adequately considered and measured?

11 What is the overall assessment of the utility of this intervention? And how confident are we in this overall assessment?

Decisions to be made based on our current knowledge

12 Should we change research priorities?

13 Should we change clinical practice?