Evaluating Digital Health Interventions

Key Questions and Approaches

Elizabeth Murray, FRCGP, FRCP(Edin), PhD,[1] Eric B. Hekler, PhD,[2] Gerhard Andersson, PhD,[3,4] Linda M. Collins,[5] Aiden Doherty,[6] Chris Hollis, PhD, FRCPsych,[7] Daniel E. Rivera, PhD,[8] Robert West, PhD,[9] Jeremy C. Wyatt, DM, FRCP[10]

Digital health interventions have enormous potential as scalable tools to improve health and healthcare delivery by improving effectiveness, efficiency, accessibility, safety, and personalization. Achieving these improvements requires a cumulative knowledge base to inform development and deployment of digital health interventions. However, evaluations of digital health interventions present special challenges. This paper aims to examine these challenges and outline an evaluation strategy in terms of the research questions needed to appraise such interventions. As they are at the intersection of biomedical, behavioral, computing, and engineering research, methods drawn from all of these disciplines are required. Relevant research questions include defining the problem and the likely benefit of the digital health intervention, 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. Although RCTs are important for evaluation of effectiveness and cost effectiveness, they are best undertaken only when: (1) the intervention and its delivery package are stable; (2) these can be implemented with high fidelity; and (3) there is a reasonable likelihood that the overall benefits will be clinically meaningful (improved outcomes or equivalent outcomes at lower cost). Broadening the portfolio of research questions and evaluation methods will help with developing the necessary knowledge base to inform decisions on policy, practice, and research.

(Am J Prev Med 2016;51(5):843–851)

From the [1]Research Department of Primary Care and Population Health, University College London, London, United Kingdom; [2]Designing Health Lab, School of Nutrition and Health Promotion, Arizona State University, Phoenix, Arizona; [3]Department of Behavioural Sciences and Learning, Linköping University, Linköping, Sweden; [4]Department of Clinical Neuroscience, Karolinska Institute, Stockholm, Sweden; [5]The Methodology Center and Department of Human Development and Family Studies, The Pennsylvania State University, State College, Pennsylvania; [6]MRC Clinical Trial Service Unit Hub, Nuffield Department of Population Health, University of Oxford, Oxford, United Kingdom; [7]NIHR MindTech HTC, University of Nottingham, Nottingham, United Kingdom; [8]School for the Engineering of Matter, Transport, and Energy, Ira A. Fulton Schools of Engineering, Arizona State University, Phoenix, Arizona; [9]Research Department of Epidemiology and Public Health, University College London, London, United Kingdom; and [10]Wessex Institute, University of Southampton, Southampton, United Kingdom

Address correspondence to: 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. E-mail: elizabeth.murray@ucl.ac.uk.

This article is part of a theme section titled Digital Health: Leveraging New Technologies to Develop, Deploy, and Evaluate Behavior Change Interventions.

0749-3797/$36.00

http://dx.doi.org/10.1016/j.amepre.2016.06.008

Introduction

There is enormous potential for digital health interventions (DHIs; i.e., interventions delivered via digital technologies such as smartphones, website, or text messaging) to provide effective, cost effective, safe, and scalable interventions to improve health and healthcare. DHIs can be used to promote healthy behaviors (e.g., smoking cessation,[1] healthy eating,[2] physical activity,[3] safer sex,[4] or alcohol consumption[5]); improve outcomes in people with long-term conditions[6] such as cardiovascular disease,[7] diabetes,[8] and mental health conditions[9]; and provide remote access to effective treatments (e.g., computerized cognitive behavioral therapy for mental health and somatic problems).[10–13] 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 behaviors, titrate medication, clarify health priorities and reach treatment decisions congruent with these, and improve communication between patients and healthcare professionals (HCPs). Active components may include information, psycho-education, personal stories, formal decision aids, behavior change support, interactions with HCPs 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 behavioral therapy or mindfulness training.

To date, the potential of DHIs has scarcely been realized, partly because of difficulties in generating an accumulating knowledge base for guiding decisions about DHIs. These include the rapid change of the wider technology landscape,[14] which requires DHIs 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 an RCT may not be published for 5–6 years, by which time the iPhone operating system has 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, policymakers, and healthcare commissioners 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 DHIs, 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 DHIs 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 United Kingdom National Institute for Health and Care Excellence.

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]

Structure

The paper starts by defining the research questions (RQs) that, in the authors' opinion, should form the basis for an appraisal of a DHI (Table 1). It then considers appropriate research methods for each of these RQs. 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 that arise, or are particularly salient, in evaluation of DHIs, the main areas of consideration for each issue are outlined. Throughout, the paper emphasizes that the RQs 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 DHIs becomes standard clinical practice,[17] ad hoc referral from a clinician,[18] supported access (e.g., face to face,[19] by telephone,[20] or by e-mail[21]), hosting on a trusted portal (e.g., National Health Service Choices), marketing via public health campaigns, or embedding in a social network.

Defining the Problem

1. Is There a Clear Health Need That This Digital Health Intervention Is Intended to Address?

2. Is There a Defined Population That Could Benefit From This Digital Health Intervention?

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

Defining the Likely Benefit of the Digital Health Intervention

3. Is the Digital Health Intervention 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 DHIs, as impact and cost effectiveness

Table 1. Key RQs for an Appraisal of a DHI

RQ
Defining the problem
1. Is there a clear health need that this DHI is intended to address?
2. Is there a defined population that 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 key components are needed for the DHI? Which components 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 that 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 that is meaningful for the stage of the research process?
9. Has the possibility of harms been adequately considered? And the likelihood of risks or adverse outcomes assessed?
10. Has DHI cost and its cost impact on users and health systems been adequately considered and measured?
11. What is the overall assessment of the utility of this intervention? 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?

DHI, digital health intervention; RQ, research question.

are highly dependent on the total number of users,[25] 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 users' 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. Hence, an early component of any evaluation of a DHI should be a determination and optimization 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 or acceptability) and to the delivery package around the DHI. For many DHIs, "users" will include HCPs who "prescribe" the DHI and monitor outcomes. Thus, RQs 3–6 require work with HCPs as well as patients or the public.

Establishing and optimizing potential reach and uptake require methods used in engineering and computer science, collectively referred to as "human-centered design."[26–28] These include concept sketching,[29] co-design strategies,[26] low-fidelity or "Wizard of Oz" prototyping,[30,31] and user experience testing.[28] In the business world, there is increasing interest in "lean" principles that specify methods for early-stage testing of features related to feasibility, including[32]:

- 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 Digital Health Intervention 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?[33,34]

5. What Key Components Are Needed for the Digital Health Intervention? Which Components 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 DHIs are highly complex interventions containing multiple components, so the development process needs to include a period of optimization. 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 Optimization Strategy,[35,36] 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[37]; factors affecting compliance with, adherence to, fidelity of, or scalability of the intervention[38]; variables and decision rules used to tailor intervention strategy, content, or intensity to individuals[39]; 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 the most effective intervention that requires no more than 1 hour per week of participant time.

The experimental approaches used for optimization include full or fractional factorial experiments,[40,41] the Sequential Multiple-Assignment Randomized Trial (SMART),[42] and system identification techniques.[43,44] 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. This is discussed further in Collins et al.[40,45]

6. What Strategies Should Be Used to Support Tailoring the Digital Health Intervention to Participants Over Time?

Where the RQ focuses on tailoring the DHI to participants over time (e.g., non-responders, or daily adjustments reflecting changing needs or context) a SMART design,[46] 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.[35,47]

System identification approaches are used in engineering to obtain dynamic systems models; these in turn are the basis for the design of control systems that achieve optimization.[48] 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 before dictating the usefulness of the model for controller design. Examples can be found in Timms and colleagues[49] and Deshpande et al.;[43] experimental procedures involving pseudo-random multisine signals are currently being evaluated in a physical activity intervention based on Social Cognitive Theory.[50]

7. What Is the Likely Direction and Magnitude of the Effect of the Digital Health Intervention or Its Components Compared to a Comparator That Is Meaningful for the Stage of the Research Process?
8. How Confident Are We About the Magnitude of the Effect of the Digital Health Intervention or Its Components Compared to a Comparator That Is Meaningful for the Stage of the Research Process?

Once RQs 3–6 have been addressed, the research team is 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 plus delivery package plus context of use) has 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 RCT to establish the magnitude of the effect (effect size) of the DHI compared to a meaningful comparator. "Relatively stable" is a matter for investigator judgment, guided by the causal explanation and optimization data.[51] The wider technologic landscape is likely to continue to evolve, and investigators must judge what impact this will have on the generalizability of their findings. The importance of undertaking an RCT and not relying solely on formative studies is evidenced

by the fact that RCTs have repeatedly overturned assumptions drawn from observational or non-randomized studies.[52,53] 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 DHIs that 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; and
- establishing separate data collection methods from the DHI itself.

The importance of context has been described in RQs 3 and 5. Understanding, defining, and describing the context in which an RCT is undertaken is necessary to inform judgments around the generalizability 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,[54] but is particularly salient for trials of DHIs. External validity refers to the extent to which the results apply to "a definable group of patients in a particular setting," whereas internal validity is based on how the design and conduct of the trial minimizes potential for bias.[55] The emphasis in trials of pharmaceutical products is on internal validity and reducing bias, and extensive work has confirmed the importance of this.[56] 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[54] and to digital interventions in particular, including concerns about the degree to which design features that enhance internal validity jeopardize 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 generalizability.

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. Here, specification means having an agreed framework for classifying the intervention components, including the degree of human input and components that are individually tailored. Such specification is required for replication of trial results, comparison between DHIs, synthesizing data across trials in systematic reviews and meta-analyses,[57] and may help with determining the criteria for "substantial equivalence" of DHIs. The concept of substantial equivalence is used for medical device and pharmaceutical regulation by the U.S. Food and Drug Administration 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 DHI for depression, then each new mindfulness app for depression would not be required to undergo RCT testing, but instead to demonstrate substantial equivalence to existing "predicate" interventions.[58] The relevant data to collect would then focus on usage, adherence, demographic access parameters, and user preferences.

The selection of a suitable comparator is determined by the RQ that is addressed, which will vary with the stage of the research. In pragmatic trials that aim to determine the effectiveness of a new treatment compared to current best practice, the comparator is typically "treatment as usual." However, in trials of DHIs, the participants in the treatment-as-usual group may have access to a myriad of other DHIs. People accustomed to using DHIs are often also accustomed to searching online for resources. Someone who has sought help for a particular problem, entered a trial, been randomized to the comparator arm, and who finds the comparator intervention unhelpful, may well search online until they find a better resource.[59] 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 here.[57]

There is a temptation in RCTs of DHIs to embed data collection into the intervention, but this may introduce systematic bias or confound the intervention with the measurement method. This bias may favor 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 Harms Been Adequately Considered? And the Likelihood of Risks or Adverse Outcomes Assessed?

DHIs are not harm free, although to date, the data on actual harms are relatively sparse. There are various mechanisms by which DHIs could result in harm. First, they could be designed to achieve an outcome that is widely viewed as harmful; for example, websites that promote suicide. Second, DHIs 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 to enable users to determine whether they are safe to drive, but do not, in fact, have any capacity to estimate blood alcohol concentration.[60] Alternatively, a DHI could contain inaccurate information or advice. Third, a DHI could provide accurate information and advice, but this could be misinterpreted or wrongly applied, leading to decisions that harm health. Alternatively, this accurate information could lead to increased anxiety or depression. Fourth, ineffective DHIs 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. Fifth, 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 DHIs are useless and no further effort should be invested. Finally, DHIs may "leak" personal data because of inadequate security and encryption functions.[47]

All developers of DHIs 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 optimization work, whereas others may require long-term observational studies during widespread implementation.

10. Has Digital Health Intervention Cost and its Cost Impact on Users and Health Systems 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 are addressed in detail by McNamee and colleagues.[25]

11. What Is the Overall Assessment of the Utility of This Intervention? How Confident Are We in This Overall Assessment?

12. Should We Change Research Priorities?

13. Should We Change Clinical Practice?

Answers to RQs 1–10 should enable an assessment of the overall utility of the DHI (e.g., balancing its effects, usage, scalability, costs, safety), along with an estimate of 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 realizing 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 an RQ-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 multidisciplinary collaborations that draw on insights and experience from multiple fields, including clinical medicine, health services research, behavioral science, education, engineering, and computer science. Researchers from an engineering or computer science background may be surprised by the reliance on RCTs, whereas those from a biomedical or behavioral 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, in which interventions undergo multiple cycles of development and optimization. A definitive trial should be undertaken only once: the intervention together with the delivery package around it have reached a degree of stability such that future developments can be considered relatively minor; there is reasonable confidence that the intervention plus delivery package can be implemented with high fidelity; and there is a reasonable likelihood that the overall benefits will be clinically meaningful and lead to either improved outcomes or equivalent outcomes at lower cost (Table 2).

How best to combine rigor with efficiency in evaluating DHIs requires a great deal of methodologic research

Table 2. 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-centered 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 Optimization Strategy.
6. 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.
Priority topics for future research
The key priority is to improve the efficiency of evaluations without jeopardizing 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 jeopardizing external validity; improved methods for enhancing DHI uptake and minimizing 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.

DHI, digital health intervention.

(Table 2). Areas to explore in future methodologic research include:

Enabling Individual Studies to Generate More-Useful Data

- 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;
- 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;
- better understanding of how to improve the internal validity of RCTs of DHIs in terms of retention and follow-up without jeopardizing external validity in terms of the recruited population or impact on the intervention;
- improved methods for reducing the large amounts of missing data that may occur, and addressing the inevitable biases this raises; and
- 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.

Enabling More-Useful Synthesis and Comparison of Data Generated by Different Studies

- identification, specification, and classification of important contextual factors;
- specification and classification of target populations;
- specification and classification of DHIs, to gain an understanding of the important active components and mechanism of action, to replicate and synthesize evidence across DHI evaluations, and begin to address the issue of determining substantial equivalence between DHIs;
- specification and determination of appropriate comparators, according to the stage of the research process; and
- improved reporting of studies of DHIs, building on initiatives such as the Template for Intervention Description and Replication reporting guideline[57] and the CONSORT–EHEALTH statement.[61]

This 2016 theme issue of the *American Journal of Preventive Medicine* is supported by funding from the NIH Office of Behavioral and Social Sciences Research (OBSSR) to support

the dissemination of research on digital health interventions, methods, and implications for preventive medicine.

This paper is one of the outputs of two workshops, one supported by the Medical Research Council (MRC)/National Institute for Health Research (NIHR) Methodology Research Program (PI Susan Michie), the OBSSR (William Riley, Director) and the Robert Wood Johnson Foundation (PI Kevin Patrick); and the other by the National Science Foundation (PI Donna Spruitj-Metz, proposal # 1539846).

Author Aiden Doherty is supported by the British Heart Foundation Centre of Research Excellence at Oxford [grant number RE/13/1/30181].

Elizabeth Murray is the Managing Director of a not-for-profit Community Interest Company that disseminates a digital health intervention to support self-management of people with type 2 diabetes (HeLP-Diabetes) to the NHS in England. She receives no financial remuneration from this. No other financial disclosures were reported by the authors of this paper.

References

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*. 2011;378(9785):49–55. http://dx.doi.org/10.1016/S0140-6736(11)60701-0.
2. 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. http://dx.doi.org/10.3310/hta15370.
3. Vandelanotte C, Spathonis KM, Eakin EG, Owen N. Website-delivered physical activity interventions a review of the literature. *Am J Prev Med*. 2007;33(1):54–64. http://dx.doi.org/10.1016/j.amepre.2007.02.041.
4. Bailey JV, Murray E, Rait G, et al. Interactive computer-based interventions for sexual health promotion. *Cochrane Database Syst Rev*. 2010;9:CD006483. http://dx.doi.org/10.1002/14651858.cd006483.pub2.
5. 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. http://dx.doi.org/10.1111/j.1360-0443.2010.03214.x.
6. Murray E, Burns J, See Tai S, Lai R, Nazareth I. Interactive Health Communication Applications for people with chronic disease. *Cochrane Database Syst Rev*. 2005;4:CD004274. http://dx.doi.org/10.1002/14651858.cd004274.pub4.
7. McLean G, Band R, Saunderson K, et al. Digital interventions to promote self-management in adults with hypertension systematic review and meta-analysis. *J Hypertens*. 2016;34(4):600–612. http://dx.doi.org/10.1097/HJH.0000000000000859.
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. http://dx.doi.org/10.1111/jebm.12042.
9. Hollis C, Morriss R, Martin J, et al. Technological innovations in mental healthcare: harnessing the digital revolution. *Br J Psychiatry*. 2015;206(4):263–265. http://dx.doi.org/10.1192/bjp.bp.113.142612.
10. Arnberg FK, Linton SJ, Hultcrantz M, Heintz E, Jonsson U. Internet-delivered psychological treatments for mood and anxiety disorders: a systematic review of their efficacy, safety, and cost-effectiveness. *PLoS One*. 2014;9(5):e98118. http://dx.doi.org/10.1371/journal.pone.0098118.
11. Olthuis JV, Watt MC, Bailey K, Hayden JA, Stewart SH. Therapist-supported Internet cognitive behavioural therapy for anxiety disorders in adults. *Cochrane Database Syst Rev*. 2015;3:Cd011565. http://dx.doi.org/10.1002/14651858.cd011565.
12. McCombie A, Gearry R, Andrews J, Mikocka-Walus A, Mulder R. Computerised cognitive behavioural therapy for psychological distress in patients with physical illnesses: a systematic review. *J Clin Psychol Med Settings*. 2015;22(1):20–44. http://dx.doi.org/10.1007/s10880-015-9420-0.
13. van Beugen S, Ferwerda M, Hoeve D, et al. Internet-based cognitive behavioral therapy for patients with chronic somatic conditions: a meta-analytic review. *J Med Internet Res*. 2014;16(3):e88. http://dx.doi.org/10.2196/jmir.2777.
14. Patrick K, Hekler EB, Estrin D, et al. Rapid rate of technological development and its implications for research on digital behavior interventions. *Am J Prev Med*. 2016 In press.
15. Murray E, Khadjesari Z, White IR, et al. Methodological challenges in online trials. *J Med Internet Res*. 2009;11(2):e9. http://dx.doi.org/10.2196/jmir.1052.
16. Michie S, West R. Guide to Development and Evaluation of Digital Behaviour Change Interventions in Healthcare. London: UCL Centre for Behaviour Change; In press.
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*. 2015;66(10):1043–1050. http://dx.doi.org/10.1176/appi.ps.201400477.
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. http://dx.doi.org/10.1136/bmj.f540.
19. Hamilton FL, Hornby J, Sheringham J, et al. DIgital Alcohol Management ON Demand (DIAMOND) feasibility randomised controlled trial of a web-based intervention to reduce alcohol consumption in people with hazardous and harmful use versus a face-to-face intervention: protocol. *Pilot Feasibility Stud*. 2015;1(1):1–8. http://dx.doi.org/10.1186/s40814-015-0023-1.
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. http://dx.doi.org/10.2196/jmir.3199.
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. http://dx.doi.org/10.1371/journal.pone.0010939.
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. http://dx.doi.org/10.1136/bmj.a1655.
23. Campbell NC, Murray E, Darbyshire J, et al. Designing and evaluating complex interventions to improve health care. *BMJ*. 2007;334(7591):455–459. http://dx.doi.org/10.1136/bmj.39108.379965.BE.
24. Murray E, Treweek S, Pope C, et al. Normalisation process theory: a framework for developing, evaluating and implementing complex interventions. *BMC Med*. 2010;8:63. http://dx.doi.org/10.1186/1741-7015-8-63.
25. McNamee P, Murray E, Kelly M, et al. Designing and undertaking a health economics study of digital health interventions. *Am J Prev Med*. 2016. In press.
26. Agarwal R, Anderson C, Crowley K, Kannan PK. *Understanding Development Methods From Other Industries to Improve the Design of Consumer Health IT: Background Report*. (Prepared by Westat, under Contract No. HHSA290200900023I.) AHRQ Publication No. 11-0065-EF. Rockville, MD: Agency for Healthcare Research and Quality; 2011.
27. Maguire M. Methods to support human-centred design. *Int J Hum Comput Stud*. 2001;55(4):587–634. http://dx.doi.org/10.1006/ijhc.2001.0503.

28. Preece J, Sharp H, Rogers Y. *Interaction Design-Beyond Human-Computer Interaction*. New York: John Wiley & Sons; 2015.
29. Buxton B. *Sketching User Experiences: Getting the Design Right and the Right Design*. Burlington, MA: Morgan Kaufmann; 2010.
30. Detmer WM, Shiffman S, Wyatt JC, Friedman CP, Lane CD, Fagan LM. A continuous-speech interface to a decision support system: II. An evaluation using a Wizard-of-Oz experimental paradigm. *J Am Med Inform Assoc*. 1995;2(1):46–57. http://dx.doi.org/10.1136/jamia.1995.95202548.
31. Li Y, Hong J, Landay J. Design challenges and principles for Wizard of Oz testing of location-enhanced applications. *IEEE Pervasive Comput*. 2007;6(2):70–75. http://dx.doi.org/10.1109/MPRV.2007.28.
32. Bowen DJ, Kreuter M, Spring B, et al. How we design feasibility studies. *Am J Prev Med*. 2009;36(5):452–457. http://dx.doi.org/10.1016/j.amepre.2009.02.002.
33. Hekler EB, Michie S, Rivera DE, et al. Developing and refining models and theories suitable for digital health interventions. *Am J Prev Med*. In press.
34. Yardley L, Spring B, Riper H, et al. Understanding and promoting engagement with digital behavior change interventions. *Am J Prev Med*. In press.
35. Collins LM, Nahum-Shani I, Almirall D. Optimization of behavioral dynamic treatment regimens based on the sequential, multiple assignment, randomized trial (SMART). *Clin Trials*. 2014;11(4):426–434. http://dx.doi.org/10.1177/1740774514536795.
36. Collins LM, Kugler KC, Gwadz MV. Optimization of multicomponent behavioral and biobehavioral interventions for the prevention and treatment of HIV/AIDS. *AIDS Behav*. 2016;20(Suppl 1):S197–S214. http://dx.doi.org/10.1007/s10461-015-1145-4.
37. Wyrick DL, Rulison KL, Fearnow-Kenney M, Milroy JJ, Collins LM. Moving beyond the treatment package approach to developing behavioral interventions: addressing questions that arose during an application of the Multiphase Optimization Strategy (MOST). *Transl Behav Med*. 2014;4(3):252–259. http://dx.doi.org/10.1007/s13142-013-0247-7.
38. Caldwell LL, Smith EA, Collins LM, et al. Translational research in South Africa: evaluating implementation quality using a factorial design. *Child Youth Care Forum*. 2012;41(2):119–136. http://dx.doi.org/10.1007/s10566-011-9164-4.
39. 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. http://dx.doi.org/10.1016/j.amepre.2007.12.024.
40. Collins LM, Dziak JJ, Li R. Design of experiments with multiple independent variables: a resource management perspective on complete and reduced factorial designs. *Psychol Methods*. 2009;14(3):202–224. http://dx.doi.org/10.1037/a0015826.
41. Dziak JJ, Nahum-Shani I, Collins LM. Multilevel factorial experiments for developing behavioral interventions: power, sample size, and resource considerations. *Psychol Methods*. 2012;17(2):153–175. http://dx.doi.org/10.1037/a0026972.
42. Almirall D, Nahum-Shani I, Sherwood NE, Murphy SA. Introduction to SMART designs for the development of adaptive interventions: with application to weight loss research. *Transl Behav Med*. 2014;4(3):260–274. http://dx.doi.org/10.1007/s13142-014-0265-0.
43. 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*. 2014;4(3):275–289. http://dx.doi.org/10.1007/s13142-014-0282-z.
44. Rivera DE, Pew MD, Collins LM. Using engineering control principles to inform the design of adaptive interventions: a conceptual introduction. *Drug Alcohol Depend*. 2007;88(Suppl 2):S31–S40. http://dx.doi.org/10.1016/j.drugalcdep.2006.10.020.
45. Collins LM, Dziak JJ, Kugler KC, Trail JB. Factorial experiments: efficient tools for evaluation of intervention components. *Am J Prev Med*. 2014;47(4):498–504. http://dx.doi.org/10.1016/j.amepre.2014.06.021.
46. Lei H, Nahum-Shani I, Lynch K, Oslin D, Murphy SAA. "SMART" design for building individualized treatment sequences. *Annu Rev Clin Psychol*. 2012;8:21–48. http://dx.doi.org/10.1146/annurev-clinpsy-032511-143152.
47. Huckvale K, Prieto JT, Tilney M, Benghozi PJ, Car J. Unaddressed privacy risks in accredited health and wellness apps: a cross-sectional systematic assessment. *BMC Med*. 2015;13:214. http://dx.doi.org/10.1186/s12916-015-0444-y.
48. Ljung L. *System Identification—Theory for the User*. 2nd ed, Upper Saddle River, NJ: PTR Prentice Hall; 1999.
49. Timms KP, Rivera DE, Collins LM, Piper ME. A dynamical systems approach to understanding self-regulation in smoking cessation behavior change. *Nicotine Tob Res*. 2014;16(Suppl 2):S159–S168. http://dx.doi.org/10.1093/ntr/ntt149.
50. Martin CA, Deshpande S, Hekler EB, Rivera DE. A system identification approach for improving behavioral interventions based on Social Cognitive Theory. Paper presented at: American Control Conference (ACC), July 1-3, 2015; Chicago, IL.
51. 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. http://dx.doi.org/10.2196/jmir.4391.
52. Baron JA, Barry EL, Mott LA, et al. A trial of calcium and vitamin D for the prevention of colorectal adenomas. *N Engl J Med*. 2015;373(16):1519–1530. http://dx.doi.org/10.1056/NEJMoa1500409.
53. Fenton A, Panay N. Hormone therapy and cardiovascular disease—are we back to the beginning? *Climacteric*. 2015;18(4):437–438. http://dx.doi.org/10.3109/13697137.2015.1057958.
54. Kennedy-Martin T, Curtis S, Faries D, Robinson S, Johnston J. A literature review on the representativeness of randomized controlled trial samples and implications for the external validity of trial results. *Trials*. 2015;16(1):495. http://dx.doi.org/10.1186/s13063-015-1023-4.
55. Rothwell PM. External validity of randomised controlled trials: "to whom do the results of this trial apply?". *Lancet*. 2005;365(9453):82–93. http://dx.doi.org/10.1016/S0140-6736(04)17670-8.
56. Savovic J, Jones H, Altman D, et al. Influence of reported study design characteristics on intervention effect estimates from randomised controlled trials: combined analysis of meta-epidemiological studies. *Health Technol Assess*. 2012;16(35):1–82. http://dx.doi.org/10.3310/hta16350.
57. 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. http://dx.doi.org/10.1136/bmj.g1687.
58. *Benefit-Risk Factors to Consider When Determining Substantial Equivalence in Premarket Notifications [510(k)] with Different Technological Characteristics Draft Guidance for Industry and Food and Drug Administration Staff*. Rockville, MD: U.S. DHHS; 2014.
59. 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*. 2015;18(6):2011–2020. http://dx.doi.org/10.1111/hex.12351.
60. Weaver ER, Horyniak DR, Jenkinson R, Dietze P, Lim MS. "Let's get wasted!" and other apps: characteristics, acceptability, and use of alcohol-related smartphone applications. *JMIR Mhealth Uhealth*. 2013;1(1):e9. http://dx.doi.org/10.2196/mhealth.2709.
61. Eysenbach G, CONSORT−EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of web-based and mobile health interventions. *J Med Internet Res*. 2011;13(4):e126. http://dx.doi.org/10.2196/jmir.1923.