Evaluation of Digital Health Interventions and Medical Devices

A brief report of a seminar held on the 16th September 2014

Dr Sonali Wayal, Dr Rosie Webster, and Dr Chris Vincent

Organizers:
eHealth Unit, UCL and UCL Interaction Centre
in collaboration with Science, Medicine, and Society Network of UCL

Table of Contents

Abstract

Digital interventions for health (eHealth interventions) offer tremendous potential for health promotion, care, and treatment. Smoking cessation apps and medicine dosage calculators are examples of such interventions that health care professionals and patients are increasingly using in their daily lives. eHealth interventions are an evolving science, and ensuring the effectiveness, safety, and security of such interventions is vital. This seminar was organised with the aim of discussing the methods, challenges, and opportunities for evaluating eHealth interventions, and to examine the standards and guidelines relating to such interventions. Several speakers highlighted the feasibility of conducting online randomised controlled trials for the evaluation of eHealth interventions, in particular the potential for recruiting a large sample size in a short time-span. Generic methodological challenges for evaluating eHealth interventions were highlighted such as ensuring and maintaining engagement with the intervention, defining appropriate dosage/exposure and measuring exposure to the intervention, and retention in trials. Challenges specific to the evaluation of eHealth interventions were also highlighted such as lengthy evaluation processes, risk of contamination in trials through potential exposure to other eHealth interventions. Process evaluation was thought to be important; using mixed methods to understand contextual and psychological factors that may impact on trial procedures. It was suggested that the use of robust theoretical principles to inform the development and evaluation of eHealth interventions may enable effective adaption of intervention content to changes in technological developments. Developing a theory of generalisability and using study designs alternative to traditional randomised controlled trials, were proposed as the future direction for evaluation of eHealth interventions.

Introduction

Digital health interventions offer opportunities for health promotion, care, and treatment. These may include activity trackers to monitor health and fitness, digital healthcare systems to monitor patients remotely, and mobile apps to help with smoking cessation. Healthcare professionals and patients are increasingly using them in their daily lives. In an era of increasing demand for healthcare services and resource constraints, these interventions offer an alternative to or complement clinic-based care and health promotion. eHealth interventions are an evolving science, and ensuring the effectiveness, safety, and security of such interventions is vital. This seminar was organised with the aim of discussing the methods for, challenges with, and alternatives to evaluating eHealth interventions, and the universality of standards and guidelines for such interventions. The seminar was organised by the eHealth unit at University College London (UCL) in collaboration with University College London Interaction Centre (UCLIC) and with sponsorship from the Science, Medicine, and Society Network at UCL. eHealth research experts within and outside UCL participated in the seminar. The seminar included a series of presentations on evidence based evaluation methods for eHealth technologies, which were defined as "….health services and information delivered or enhanced through the Internet and related technologies." [1] The seminar concluded with a panel discussion.

Randomised control trials for evaluating eHealth interventions

The opening presentation was '***An overview of evaluation methods in eHealth',*** delivered by Dr Sonali Wayal from eHealth unit, UCL. Based on findings of a literature review, predominantly of randomised controlled trials (RCTs) of sexual health interventions from high-income countries, the feasibility and benefits of conducting online trials of eHealth interventions were discussed. Benefits included widespread reach of eHealth interventions in short timespans, including among hard-to-reach groups like men who have sex with men. Online automatic randomization enables blinding, and can reduce selection bias. However, there are various challenges associated with online trials, such as multiple enrolment, high dropout rate, legal constraints for recruiting participants below the age of 16, lack of internet access, and concerns about data security. Retention may be improved by collecting multiple contact details of participants and offering incentives for participation in a trial/follow-up. Digital data collection can enhance internal validity and internal consistency, and reduce missing data. Furthermore eHealth interventions can be administered in a standardized format, which can enhance fidelity. Ensuring engagement with digital interventions can be a challenge, but it can be addressed by involving users in the development of the intervention, usability testing, sending email reminders and offering incentives to participants to engage with the intervention. It was also highlighted that it is equally important to carefully define 'engagement' in eHealth interventions; tracking the number of log-ins and number of intervention web-pages visited can be used as proxy measures of exposure to the intervention. In conclusion, online trials for the evaluation of eHealth interventions are feasible; however, robust measures to enhance retention in trials are required, alongside facilitation of exposure and engagement with the intervention.

Dr Julia Bailey from the UCL eHealth unit, presented insights from an online RCT conducted to evaluate the effectiveness of an online web-based intervention: 'Sexunzipped' for sexual health promotion. During her talk ***'Optimizing the Design of Online Randomized Controlled Trials: The Sexunzipped website'*** she highlighted that recruitment via Facebook and automatic online randomization, consent, and data collection are feasible and acceptable to young people. A factorial trial design tested the effect of requesting a Chlamydia urine

sample by post or not, and the effect of offering two different levels of incentives (£10 or £20). Offering £20 vouchers led to greater response to an online survey at three month follow-up (77%). Offering £20 led to a 47% response rate for postal urine Chlamydia sampling kits. Although biological sampling is considered 'gold standard', postal Chlamydia sampling is not a good outcome measure because the point prevalence of positive results was low (2.4%), and the overall response rate was low. To address the challenge of multiple enrolment in online studies, details including date of birth and gender were collected from trial participants at both baseline and follow-up (3 months). Participants with discrepant details were identified and excluded from the analysis. Sending multiple reminders via email and post had a positive impact on retention at 3 months: 33% participants responded on first contact, 11% to second reminder, and 5% responded to third contact. A further 10% responded to a postal follow up questionnaire. Participants appreciated the online format for the questionnaire measuring sexual health outcomes. Measurement reactivity is likely to be a potential challenge. In conclusion, the Sexunzipped online trial shows that repeated collection of personal identifiers can identify potential multiple registrations, and incentives and multiple reminders (via email and post) can enhance retention at follow-up.
Link to paper: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3868980/

Following this, Prof Robert West from UCL Epidemiology and Public Health department gave a presentation focused '***Evaluating digital behaviour change interventions: challenges and opportunities'***. He shared his experience and concerns regarding the evaluation of eHealth interventions, focusing on apps used for smoking cessation. One of the key challenges when evaluating eHealth interventions is the potential for contamination. In particular, in trials of smoking cessation apps there is a likelihood that participants of control arm may be motivated to search for and use existing smoking cessation apps (some of which may be effective). Therefore, using effect sizes derived from such studies to inform cost-effectiveness analysis can be a challenge, and thus cannot be used in the same way as those in surgical or drug intervention trials. The focus, therefore, should not only be on understanding if an intervention works but also how we can make it work better, i.e., optimisation rather than validation. Another challenge is that technologies rapidly become obsolete; for example, due to changes in devices, operating systems, and browsers. This affects the look and feel of interventions post-evaluation and may also impact upon effectiveness. Behaviour change theories used to design the apps/health interventions, and how they interact with the delivery platform, should therefore be clearly delineated. An appropriate working relationship between behavioural scientists and the developers of the digital technology should be established. Tendering and contracting for developers may not be the best approach, as they should be an integral part of the research team from outset, and even post-development. Furthermore, there is often an incompatibility between research governance requirements and developers' priorities like speed of development. Nevertheless, it is easy to recruit participants to trials of eHealth interventions like smoking cessation apps and the cost of data collection via the intervention itself is low. However, this raises the issue of whether the data collected within the intervention itself is valid. In these cases, construct validity can be examined by scrutinising the correlation between outcome measures of interest and factors known to be associated with them, such as age, social class, etc. Collecting objective outcome data can be a challenge; however, there is potential in the use of wearable technologies to collect objective data. In conclusion, this talk highlighted the opportunities and caveats that researchers involved with evaluation of health interventions should address.

Non-randomized study designs for development and evaluation

Next, a number of presentations addressed how other study designs, such as qualitative studies or reviews can enable assessment of effectiveness of eHealth interventions, and process or contextual evaluation.

In her presentation titled ***'Evaluating Health-Related Smartphone Apps',*** Katarzyna Stawarz from UCLIC presented data from two reviews. The reviews focused on understanding the functionality of medication reminder apps and how they support adherence, and on investigating whether 'habit formation apps' truly support habit formation. The findings of this review showed that 229 medication reminder/birth control pill reminder apps focused on simply reminding the patient. In addition, most of them did not include a 'snooze' option, nor did they offer functions supporting medication-taking routines. 115 apps claiming to support habit formation were identified; however, the quality of this support varied. Contextual cues, implementation intentions and positive reinforcement can support habit formation, but only a few apps provided such features. Most of them focused on self-tracking and reminders: features that do not support habit formation. She concluded that there is a need to focus on developing medication reminder apps that use contextual cues and creating apps that help people change their behaviour using habit instead of helping them develop a habit of using the app.

Aisling O'Kane from UCLIC presented findings from her study ***'The Influence of Situated User Experience on the Use of Type 1 Diabetes Technologies',*** which used qualitative methods to understand situated experiences of the users of Type 1 diabetes (T1D) technologies. Autoethnography was used to explore the experience of regular use of a mobile medical device (a digital wrist blood pressure monitor). This occurred in 3 different settings (Toronto, Austin, London). Contextual interviews were conducted with 19 people with T1D to understand their use of glucose meters, insulin pumps and continuous glucose meters. A sub-set of users also kept a diary. Furthermore, observations were collected at a participatory session at a tech meet-up of London based T1D insulin pump users. The findings highlighted that the use of medical devices for management of T1D is influenced not only by individual differences and personal preferences, but also by the social context. For example, uncertainty in social situations (in a new job, or when on holiday) can lead to hiding of such devices. On the other hand, some people show off their devices, to demonstrate that T1D self-care is normal, or to get perks like getting an upgrade on a plane. In conclusion, qualitative methods help understand how 'social context' influences the adoption, carrying, and use of digital medical devices. It is important to understand situated experiences of using mobile medical devices in everyday life because it can influence adherence to self-management plans.

Aneesha Singh from UCLIC presented results from a study about ***'Motivating People with Chronic Pain to do Physical Activity: Opportunities for Technology Design'***. She highlighted that physical rehabilitation of people with chronic pain can enhance confidence in physical capacity and inhibit spread of pain; however, emotional barriers can affect this process. Technology can help in augmenting pleasurable sensations, regulate emotions, and build positive self-awareness. Sound has been shown to facilitate introspection and reduce anxiety. Sound feedback technology can help self-define exercise space and can be personalized. In-depth interviews were conducted with patients with chronic pain to evaluate a smartphone app-based sound technology to increase awareness of movement based on self-rated perceived performance. Findings showed that participants found the auditory feedback useful and motivating. It enabled them to gain confidence, set challenges, visualize efforts and focus on pleasurable sensations. In conclusion, this study highlights that

sound feedback app can help persons with chronic pain address the psychological barriers to exercise.

eHealth interventions: From craft to science

Prof Jeremy Wyatt from the University of Leeds talked about '***The urgent need for research designed to uncover eHealth theories'***. He highlighted the urgent need for identifying eHealth theories, including identifying predictive theories and testing them. He posed the question "What kind of theories are important for eHealth?", and highlighted various theories of information retrieval, communication, decision-making, and personal and organizational behaviour change. However, he emphasized the need to determine the most appropriate theories to inform eHealth interventions. In order to do so, it is important to identify a digital health problem, and then understand the problem and key challenges and possible solutions by doing a literature review, ethnography, or using various other methods, to inform the choice of an appropriate and promising theory to address the problem. Choosing an appropriate theory is important because theory is enduring and enables an understanding of behaviour change, unlike technology that is fleeting and transient. The focus should be on understanding the interaction between the theory and the information system. Various study designs can be used to test theories underlying eHealth interventions; for example, a version of the information system that doesn't use the theory can be compared with a version that applies it to develop new knowledge about the problem and the theory (also known as A/B testing or split testing).

However, designing evaluation procedures for eHealth interventions is complex; for example, the mere use of app may lead to improvement in health due to the increased attention that patients may subsequently give to their disease. If the intervention app is also used for data collection in the intervention arm, then there is risk of data collection bias. Therefore, clearly describing the intervention is incredibly important. This can be done using the template for intervention description and replication (TIDieR) checklist and guide [2]. In addition, intervention modelling experiments can be conducted to optimize the intervention itself prior to conducting a RCT. Multiphase optimization for complex interventions (MOST) can involve screening intervention components for effectiveness; for example, using lab experiments on simulated decisions, RCTs, and factorial designs. This can be followed by fine-tuning the combination of intervention components using sequential multiple assignment randomized trials (SMART) for time varying interventions based on appropriate theory.

Conducting rigorous, theory-based research will enable development of generic, reliable, and actionable knowledge for designers and technology developers which is also generalizable. Thus evaluation informed by theory, rather than a technology, will enable us to understand the underlying mechanisms of effectiveness of an interventions. It would therefore not be essential to evaluate every version of every technology. In conclusion, Prof Wyatt emphasized that eHealth research should test theories from computer sciences, information systems, as well as organizational science, and results from such studies should be shared with practitioners and students.

Regulations for eHealth interventions

The last presentation was given by **Dr Chris Vincent** from UCLIC about the ***'Regulations surrounding eHealth',*** focusing on the challenges of evaluating safety and usability for novel eHealth technologies. Despite the tremendous opportunities offered by advances in

technology, there are instances where new approaches may not be compatible with existing regulatory systems. There is confusion over which regulations apply, or a lack of applicable regulation. The differences between how technology is meant to be used and how it is actually used in practice need to be monitored. Above all, exhaustively testing software that is continuously being adapted remains a challenge (for example, can results from initial testing be generalized when software functionality or the operating system changes?). In addition, there are various standards, evaluation criteria and guidelines that could potentially apply to the development of eHealth technology (e.g. apps). For technology that falls under the medical device directives, and is regulated by the MHRA[1], the current framework involves the adoption of voluntary harmonized standards. For safety and usability: 62366; 14971 and 62304 would likely apply (amongst others). These standards adopt a 'process-based' approach to define who can use a product (and how), what risks exist, what functionality is required, and how the design has been tested with users (e.g. how it minimizes the risk). The process is applied prior to placing a product on the market and is different from a clinical evaluation. There is equivalent USA guidance for healthcare apps[2]. For technology that does not fall under the medical device directive but could compromise safe practice, there is the option to adopt the process used to manage the risk associated with Health IT Systems (ISB 0129 and ISB 0160). There will soon be the option to adopt a new guidance (currently in draft), entitled 'PAS 277:2015: Health and wellness apps – Quality criteria across the life cycle – Code of practice)'. The PAS is targeted towards mobile applications, but could apply to content delivered through a browser (for example a web based app). Although all of these approaches (medical devices framework, ISBs 0129 / 0160 and PAS 277) apply to the lifecycle of the technology, it could be argued that efforts are focused predominately on pre-market/deployment phases, with a need to develop post market investigation. In addition, given the need to collect data to demonstrate the benefit of a solution prior to releasing it, the devices framework allows for a "clinical investigation"[3]. Alternatively, PAS 277 allows for field-testing; it also encourages monitoring in use, but preserves the need to maintain privacy. The process applied to Health IT (ISB 0129 and 0160) can accommodate development driven by demonstration and "user stories" (e.g. agile development).

Panel Discussion

Panel Members: Prof Robert West, Prof Jeremy Wyatt and Dr Julia Bailey

The seminar concluded with a panel discussion. The following key issues were discussed:

1. CE marking for apps/ehealth interventions

Currently it is unclear whether apps used for clinical purposes would need to be CE marked. It is also unclear whether websites that offer cognitive behaviour therapy may also be considered as a 'medical device' that need CE marking. Therefore, there is a need to understand the regulatory frameworks and its implications on developing and evaluating eHealth interventions. Standard approaches are needed, to help decide whether an eHealth

[1] https://www.gov.uk/government/publications/medical-devices-software-applications-apps

[2] http://archpsyc.jamanetwork.com/article.aspx?articleid=1847578
http://www.fda.gov/MedicalDevices/ProductsandMedicalProcedures/ConnectedHealth/MobileMedicalApplications/ucm255978.htm

[3] https://www.gov.uk/notify-mhra-about-a-clinical-investigation-for-a-medical-device

intervention can be ratified via a checklist approach or whether further evaluation is needed. Some standards can apply to process of development rather than the outcome.

2. Effect size of intervention, contamination and trial designs

Often RCTs of behaviour change interventions show a null effect or small effect. This is mostly due to contamination which influences the true effect size of the intervention, and leads to a negative Hawthorne effect/"evaluation paradox". Measurement reactivity is also a challenge, whereby participants improve their behaviour as a result of completing a questionnaire. One approach to address this is not to test the intervention per se but to test the introduction of an intervention in a milieu. Post-randomisation consent designs (i.e. Zelen's design) can help address some of these challenges. As consent is obtained only from the participants randomised to intervention arm, the potential of Hawthorne effect is minimised. Equivalence trials can also be used to test the impact of changes in technology/software platform of an intervention. In equivalence trials, however, it can be difficult to interpret results due to power issues.

3. Engagement, retention and effectiveness of interventions

It was highlighted that there is a need to distinguish between engagement with an intervention or retention in trials. It is also conceptually important to separate data collection from active intervention, but often for ease these are integrated. When the experimental integrity is undermined due to lack of engagement or attrition, one moves away from the advantage of experimental study compared to observational study designs where there is confounding by indication. Various methods like instrumental variable analysis, regression discontinuity design, or propensity score matching can be used to address these challenges. However, in instrumental variable design, variables not correlated to the outcome should be chosen.

The broader issue is that no one study design is without flaws. Therefore the study hypothesis should be addressed in several different ways, accepting that all designs have their own problems. Therefore it is extremely important to examine the explanatory theoretical mechanisms underlying the intervention. If we triangulate evidence from different sources with different weaknesses but they all point in the same direction, then there is stronger evidence to support the findings, compared to findings from a single trial or observational study. Mixed methods should be used to also understand the processes and effects of interventions.

4. Testing theories or individual versions of technologies:

There can be subsequent unforeseen effects of interventions as technologies evolve. These may be harmful or beneficial. In such situations, A/B testing can be used to test the effect of change rapidly. Similarly, SMART trials using proximal measures can be conducted in the short-term. However, it is important to address the fundamental question regarding how interventions work. Technology develops rapidly, but people don't; therefore, if we understand how and why an intervention works, it will enable us to adapt as the technology grows. We need to focus on developing theory of generalisability. We need to understand impact of different contexts to draw out our theoretical understanding across different settings, across populations, types of behaviours and across delivery systems. It is important to understand how the frequency and degree of change in app/website/technology features is acceptable/feasible without changing effectiveness.

Fractional factorial study designs can be reliably used, if there are no higher order interactions, to do several experiments within one trial, and reduce issues with (for

example) sample size. The outcome will provide an understanding of which components of the intervention work. This information will enable us to develop effective interventions and the process of optimization of an intervention. However, the chosen components of interventions that are tested must be theoretically or empirically justified as factors enabling the outcome of interest.

Conclusion

This seminar highlighted the feasibility of evaluating eHealth interventions using online RCTs and qualitative methods. The significance of robust theoretical principles to inform the development and evaluation of eHealth interventions to enable effective adaption to changes in technological developments was acknowledged. Developing a theory of generalisability and using alternatives to traditional RCTs were proposed as the future direction for evaluation of eHealth interventions.

References

1. Eysenbach G. What is e-health? *J Med Internet Res* 2001;**3**(2):e20 doi:10.2196/jmir.3.2.e20.
2. Hoffman T, Glasziou P, Boutron I et al. *BMJ* 2014; 348 :g1687 doi: 10.1136/bmj.g1687.

Acknowledgements

We thank all the speakers and seminar participants. We want to particularly thank Prof Elizabeth Murray, Dr Julia Bailey, Prof Susan Michie, Prof Robert West and Dr Henry Potts, for their contributions to the discussion during the seminar. We also want to thank the Science, Medicine, and Society Network of UCL for financial support for this seminar and Mr Guilio Gobbetti for filming this event.