Drawing on human factors engineering to reconsider paradigms for evaluating the effectiveness of health information technology

Kathrin M. Cresswell, Chief Scientist Office Postdoctoral Fellow, Centre for Medical Informatics, Usher Institute of Population Health Sciences and Informatics, The University of Edinburgh, UK

Ann Blandford, Professor of Human–Computer Interaction and Director of UCL Institute of Digital Health, University College London, UK

Aziz Sheikh, Professor of Primary Care Research & Development and Co-Director, Centre for Medical Informatics, Usher Institute of Population Health Sciences and Informatics, The University of Edinburgh, UK

Correspondence to: K Cresswell Kathrin.Beyer@ed.ac.uk

Keywords: Health information technology, evaluation, human factors, effectiveness

Word count: 1431

Introduction

There are substantial global investments being made in health information technology (HIT) as governments and providers strive to improve the safety, quality and efficiency of healthcare.(1) Accompanying this trend is a growing HIT industry projected to be worth around $230 billion by 2020, with a range of new applications becoming available daily.(2) These vary significantly in complexity, ranging from foundational technologies that impact on a range of care providers (e.g. electronic health records (EHRs)) to relatively discrete functionality affecting a limited range of settings and users (e.g. specialist apps for patients with chronic conditions) (see Box 1).

Randomized controlled trials (RCTs) of complex HIT interventions are important and appropriate to evaluate effectiveness in many contexts, particularly if they are complemented by process and in-depth qualitative evaluations which yield insights into why the intervention was effective/ineffective and its likely generalizability.(3) But HIT represents some challenges for this traditional biomedical evaluation approach. It may have relatively small diffuse effects that are often difficult to trace and attribute as it involves re-designing existing organizational processes and ways of working (e.g. when implementing EHRs).(4) The effect of the intervention can also be heavily shaped by the context and the way it is implemented;(5) this may include, for example, varying organizational strategies, shifting responsibilities of healthcare professionals (e.g. towards a greater emphasis on data entry), increased managerial control (e.g. through review of data held within systems), changing power structures and whether or not computerized decision support (CDS) is switched on. HIT is also often rapidly evolving and a certain user interface, for example, may no longer exist by the time an RCT evaluative cycle (which typically takes years) has been completed.(6)

New design and evaluation approaches are thus needed alongside RCTs to provide a broader array of evaluative approaches to investigating the effectiveness of different forms of HIT. Building on our previous work highlighting the importance of continuous systemic HIT evaluation,(7) we consider how evaluation approaches based on human factors engineering (HFE) may help to address some of these needs.

In which contexts are RCTs appropriate?

RCTs are of considerable importance to evaluating the effectiveness of interventions – including HIT – because they have the potential to minimize the risk of bias, allocate confounders randomly between intervention arms, and produce generalizable results.(8) This design is appropriate and indeed essential for some HIT contexts where clear effects on a limited number of outcomes can be anticipated. For instance, if aiming to investigate the effectiveness of a safety/business-critical medical prototype device with discrete effects on health outcomes or if interventions are particularly expensive, then RCT designs are likely to be an appropriate summative evaluative tool. Indeed, RCTs and, alternatively quasi-RCTs, have been very useful in relation to evaluating HIT, particularly when combined with embedded process and qualitative evaluations.(5)

Why are RCTs problematic for some types of technology/context?

Our experiences have however indicated that there are a number of contexts where RCTs and quasi-RCTs are not feasible and/or appropriate. Firstly, some HIT, such as EHRs, is foundational and therefore has multiple small effects that are often hard to measure. Furthermore, with these types of technology, it can be difficult to establish appropriate controls, as HIT implementations often involve large organizational transformations across care settings that are not directly comparable. Secondly, implementation context matters and should not be treated as a confounder. This may include the hospital environment into which

systems are implemented, the leadership style and implementation strategy pursued, and user attributes such as personality and competencies. Thirdly, many technologies rapidly evolve (e.g. by being refined, upgraded, customized) or they cease to exist altogether (e.g. apps), which means that the intervention may vary over time. As a result, many complex HIT systems that are used across a range of contexts lack usability and fail to achieve their true potential as they are often used in ways other than those intended.(9)

To address some of these challenges, a HFE approach may provide an alternative more agile approach to evaluating HIT, as this takes into account the changing nature of technology whilst paying close attention to the context of system use and the variety of settings in which technologies may be implemented.

What is HFE and when is it appropriate?

User-centered approaches to evaluation tend to be practically oriented, focusing on the iterative evaluation of the technology at hand; they are typically less concerned with producing generalizable results.(10) Although some traditional experimental evaluation approaches exist in these settings (e.g. performance and effectiveness analysis), there is an explicit focus on evaluating technology throughout the lifecycle, often involving iterative development and evaluation.(11) An example of such an iterative approach to system development is rapid application development (RAD). This employs iterative methods that allow technological prototypes to be developed quickly and then tested and refined in real-world settings, often based on user feedback. It is well-suited for complex environments where effects of technologies and user requirements are hard to predict in advance (e.g. when technologies require a high degree of interactivity). This may draw on cognitive task analysis (CTA), where user needs and workflows are assessed before new technology is introduced, followed by detailed analysis surrounding how new systems affect existing practices.(12) Although somewhat overlapping (e.g. action research can be used as a methodology in HFE), HFA differs from action research in that it has a more explicit focus on technological design whereas action research is a research methodology.

HFE is thus a usability engineering-based approach to “designing systems for human use”, helping to develop systems over time to bring maximum benefits to users.(12) This includes prospective evaluation approaches that allow systems to be refined in ways that promote their effective use over time. The underlying assumption is that HIT only works if it is usable and fits with users’ practices, and HFE helps to design usable and useful systems.(13) Methods focus on obtaining a clear understanding of what task needs to be undertaken (the aim) and helping to design systems that assist and motivate users to accomplish these tasks (the tool). From this perspective, evaluation of technology should therefore focus on exploring its suitability for accomplishing a task from the perspective of users, the degree to which technologies support and enhance human effectiveness, and the degree to which they fit within their context of use.

Although attention to HFE in design and development of systems is important, there is also growing recognition that ongoing design iterations are an inherent feature of the technology lifecycle.(14) These iterations may involve co-evolution of systems and users, where technologies and ways of using them are changed over time to suit contexts of use. These iterations are often difficult to anticipate ahead of time and can create opportunities for system improvement as well as expose design limitations that may be important to address to improve workflows and mitigate potential safety risks.(15) Steps involved in this prospective evaluation process include understanding current practices and needs, identifying possible design solutions, checking proposed solutions against needs, testing of implemented systems with users, and continuous co-evolution between design and use (see Table 1).(16) Techniques and tools used in HFE are outlined in Table 2; see also texts such as (17,18).

Towards re-conceptualizing HIT evaluation approaches

Both RCT-based and HFE-based approaches to evaluation are essential for evaluating the effectiveness of HIT, but it is important to recognize that the appropriateness of methods depends on the needs emerging from different contexts/technologies. The choice of methods should be determined by the type of HIT to be evaluated, its stage of development, and the key evaluation questions in that situation. For example, HFE approaches work well for technologies that have distributed effects and can be tested and refined in collaboration with users in real-world settings. In comparison, RCTs are necessary for pre-market testing of safety-critical medical devices and for determining the impact of technologies on health outcomes.(19) We propose a decision tree to guide evaluators in relation to choice of methods in Figure 1.

A key future activity should also include establishing consensus amongst evaluators on which approaches are best suited to different types of technologies and evaluation questions, and establishing key influence diagrams to consider appropriate surrogate markers.(20) For instance, safety/business-critical technologies (e.g. infusion devices, hemodialysis machines, clinical decision support systems) are likely to require extensive up-front development (ideally in close collaboration with users) and extensive effectiveness testing, whilst prototypes of less critical devices (e.g. activity trackers) may be tested and refined in close collaboration with users in real-world settings without significant safety implications.

Conclusions

Traditional RCT-based evaluation paradigms, although suitable for determining effectiveness surrounding health outcomes, are not appropriate for evaluating the effectiveness of different types of HIT at different stages of development. HFE-based evaluation approaches can help to address these shortcomings, particularly in relation to designing systems that fulfill user needs and involve users throughout the development process.

Funding: This work has drawn on data funded by the NHS Connecting for Health Evaluation Programme (NHS CFHEP 001, NHS CFHEP 005, NHS CFHEP 009, NHS CFHEP 010), the National Institute for Health Research under its Programme Grants for Applied Research scheme (RP-PG-1209-10099) and The Commonwealth Fund. KC is supported by a Chief Scientist Office (CSO) of the Scottish Government Post-doctoral Fellowship and AS is supported by the Farr Institute. AB has been funded by EPSRC (EP/G059063/01). The views expressed are those of the author(s) and not necessarily those of the NHS, the CSO, EPSRC, the NIHR or the Department of Health.

Competing interest declaration: The authors declare that they have no competing interests.

Guarantor: AS is the guarantor for this work.

Contributorship: AS and AB conceived this work. AS is currently leading a National Institute for Health Research (NIHR) funded national evaluation of electronic prescribing and medicines administration systems. KC led on the write-up and drafting the initial version of the paper, with AS and AB commenting on various drafts.

References

1. Black AD, Car J, Pagliari C, et al. The impact of eHealth on the quality and safety of health care: a systematic overview. PLoS Med 2011;8(1):e1000387.

2. Healthcare IT Market worth 228.8 Billion USD by 2020. Available from: http://www.marketsandmarkets.com/PressReleases/healthcare-it-market.asp (last accessed: 15/05/16).

3. Avery AJ, Rodgers S, Cantrill JA, Armstrong S, Cresswell K, Eden M, Elliott RA, Howard R, Kendrick D, Morris CJ, Prescott RJ. A pharmacist-led information technology intervention for medication errors (PINCER): a multicentre, cluster randomised, controlled trial and cost-effectiveness analysis. The Lancet 2012;379(9823):1310-9.

4. Brown C, Lilford R. Evaluating service delivery interventions to enhance patient safety. BMJ 2008;337:a2764.

5. Takian A, Petrakaki D, Cornford T, et al. Building a house on shifting sand: methodological considerations when evaluating the implementation and adoption of national electronic health record systems. BMC Health Services Research 2012;12(1):105.

6. Campbell M, Fitzpatrick R, Haines A, et al. Framework for design and evaluation of complex interventions to improve health. British Medical Journal 2000;321(7262):694.

7. Catwell L, Sheikh A. Evaluating eHealth interventions: the need for continuous systemic evaluation. PLoS Med 2009;6(8):e1000126.

8. Sheikh A, Smeeth L, Ashcroft R. Randomised controlled trials in primary care: scope and application. Br J Gen Pract 2002;52(482):746-51.

9. Cresswell KM, Bates DW, Williams R, et al. Evaluation of medium-term consequences of implementing commercial computerized physician order entry and clinical decision support prescribing systems in two 'early adopter' hospitals. Journal of the American Medical Informatics Association 2014;21(e2):e194-202.

10. Kitchenham BA, Dyba T, Jorgensen M. Evidence-based software engineering. In Proceedings of the 26th international conference on software engineering 2004 May 23 (pp. 273-281). IEEE Computer Society.

11. Kushniruk A. Evaluation in the design of health information systems: application of approaches emerging from usability engineering. Computers in biology and medicine. 2002 May 31;32(3):141-9.

12. Kushniruk AW, Patel VL. Cognitive and usability engineering methods for the evaluation of clinical information systems. Journal of biomedical informatics. 2004;37(1):56-76.

13. Sanders MS, McCormick EJ. Human factors in engineering and design. McGRAW-HILL book company. 1987.

14. Rosson MB, Carroll JM. Usability engineering: scenario-based development of human-computer interaction. Morgan Kaufmann; 2002.

15. Cresswell K, Mozaffar H, Lee L, et al. Workarounds to hospital electronic prescribing systems: a qualitative study in English hospitals. BMJ Quality & Safety 2016;bmjqs-2015.

16. Cresswell K, Coleman J, Slee A, et al. A toolkit to support the implementation of electronic prescribing systems into UK hospitals: preliminary recommendations. Journal of the Royal Society of Medicine 2014;107(1):8-13.

17. Cassano-Piche A, Trbovich P, Griffin M, et al. Human Factors for Health Technology Safety: Evaluating and Improving the use of Health Technology in the Real World. Available from https://www.dropbox.com/s/t9vpg8kkvoxmvp7/CED%20HF%20Health%20Technology%20Safety.pdf?dl=0 (last accessed: 27/06/16).

18. Preece J, Rogers Y, Sharp H. Interaction Design. Wiley. 2015.

19. Hulstaert F, Neyt M, Vinck I, et al. Pre-market clinical evaluations of innovative high-risk medical devices in Europe. International Journal of Technology Assessment in Health Care. 2012;28(03):278-84.
20. Woolf S, Schünemann HJ, Eccles MP, et al. Developing clinical practice guidelines: types of evidence and outcomes; values and economics, synthesis, grading, and presentation and deriving recommendations. Implementation Science 2012;7(1):61.

Figures and tables

Box 1: Examples of existing HITs and evaluation questions

EHRs (also called electronic medical records (EMRs), electronic patient records (EPRs), depending on the focus): Issues including usability by non-clinical staff, changing relationships between clinician and patient, and new models of care.

Clinical decision support systems, computerized physician order entry, Electronic Medication Administration Records, patient barcode scanning, and related technologies for patient identification and medication management: Safety, interoperability and usability are all important for such technologies.

Technologies for medication administration and other therapies, including infusion pumps and syringe drivers (sometimes with dose error reduction software) for medication administration, and hemodialysis technology: Safety and usability are paramount across all settings, but particularly challenging for home use.

Monitoring technologies: Ease of use and of interpretation of the information displayed are important.

In vitro diagnostics: As well as ease of use (particularly for delivering reliable results), it may be important to consider the care pathway to ensure people receive appropriate support depending on the diagnostic results.

Digital behavior change interventions: Effective engagement is essential if the intervention is to be effective. Individual differences (knowledge, motivations, self-efficacy, etc.) influence engagement.

Table 1: Lifecycle perspective of implementing HIT and HFE-informed lines of inquiry

Lifecycle stage	HFE questions
Conceptualization	What are the organization's or individual's needs?
Project initiation	What kinds of solutions (not necessarily technological) are there to address those needs?
Functional specification	If the solution is technological, then what functionalities are likely to address the specified needs? Who will the users of the technology be?
Drafting a business case	[For each group of users]: how will this fit in their workflow? How will their workflow need to be adapted to exploit the new technology? What are the costs and benefits to them (in their role) in using the new system, and how can we make the benefits outweigh the costs?
Procurement/tendering	What technologies exist and which ones are likely to address the organizational/user needs? Does a new technology need to be designed and implemented?
System choice	How does the new technology support the long-term agenda of the organization (e.g. for patient engagement / shared care, interoperating with other new systems)? How easy is the new system to use? How safe is it (e.g. limiting effects of human error)? (Particularly important if it's only used occasionally by that particular user group)
Contracting	How can organizations and suppliers work together to implement a technology that is usable and brings benefits to users?
Pre-implementation	How exactly do workflows need to be changed across different professions and settings?
Implementation	How have planned changes to workflows played out in reality? Are there adverse consequences for some groups of users?
System optimization	How can users and organizations help to shape systems to better fit with their contexts of use? What are the opportunities for improvement and what are design limitations?

Table 2: Methodologies commonly used in human factors engineering

Tool	Features	Useful for	Considerations
Observation (sometimes called 'ethnography')	Observing people working and using HIT	Understanding what people do in practice and what they need	Usefully complemented by interviews or Contextual Inquiry, to better understand what is observed
Interviews (usually semi-structured)	Interviewing people about their work, their experiences of HIT, their requirements for future HIT, etc.	Understanding people's perceptions and experiences	People have difficulty reporting accurately on what they do
Contextual Inquiry	Combining observations and interviews to understand work and the use of HIT	Gaining insights for design based on a better understanding of people's work and activities	Takes place within the context where people use HIT
Diary studies	Participants maintain a diary of relevant thoughts and experiences when using the HIT	Gathering information about situated use (e.g. of mobile HIT)	Quality of data is dependent on commitment of participants.
Task analysis	Systematically decomposing tasks (that the HIT supports) into sub-tasks to analyse the sequence and performance criteria	Supports systematic thinking about user tasks and how they are achieved with the HIT	Should be based on empirical data of real user tasks
Cognitive Task Analysis	Systematic analysis of the cognitive tasks (user's thought processes) when using the HIT	Analysis of issues such as likely errors, mental workload and compatibility between tasks as defined by HIT and the way people think about their tasks	Requires expertise in cognitive science
Rapid application development (RAD)	Employs iterative methods that allow technological prototypes to be developed quickly and then tested and refined in real-world settings, often based on user feedback	Well-suited for complex environments where effects of technologies and user requirements are hard to predict in advance (e.g. when technologies require a high degree of interactivity)	Close collaboration between users, implementers and developers is essential

Heuristic evaluation	A checklist approach to checking the device interface for usability and safety based on 'rules of thumb'	Checking for obvious problems at early stages of development	Needs expertise in understanding and interpreting the heuristics. Dependent on the expertise (and biases) of the evaluators
Cognitive Walkthrough	An expert review approach that involves 'walking through' the steps of an interaction between user and device, reasoning about possible user errors. One common form of cognitive task analysis.	Early review, focusing on user cognition, requiring access to a prototype HIT	Should be conducted by experts in cognitive science. Assumes that the HIT is 'walk up and use'.
User testing (often with think-aloud)	Testing the HIT with representative users in a simulated use environment. Users are often invited to articulate thoughts while interacting with HIT	Identifying which HIT features people find easy to use, and which cause problems	Participants should be representative of the intended user population(s), and tasks used in should provide good coverage of real-world use. Yields insights into HIT design but not how it fits in the broader work context.
RCT	Testing the HIT with representative users under controlled conditions to test a hypothesis about health outcomes	Testing the effectiveness of HIT for achieving health outcomes	Need a valid comparator, and conditions need to be as realistic as possible while remaining controlled.

Figure 1: Flow diagram guiding the choice of evaluation approaches