A Framework for Health IT Evaluation

Jim Warren, Malcolm Pollock, Karen Day, and Yulong Gu
National Institute for Health Innovation, University of Auckland
PO Box 92019, Auckland, New Zealand
jim@cs.auckland.ac.nz

Sue White
Whitehouse Associates, Health Information Management Consultancy
PO Box 12323, Thorndon, Wellington, New Zealand
sue.white@whitehouse.net.nz

Abstract

This paper provides background and overview of a Health IT Evaluation Framework that has been developed to support the National Health IT Plan and New Zealand health innovation generally. The framework recommends a pragmatic approach that includes use of both quantitative data (particularly data based on the transactional logs of operational IT systems), and qualitative data systematically gathered through stakeholder interviews. An Action Research orientation is recommended where the evaluators actively seek to understand barriers and find pointers to potential solutions. The investigation protocol is recommended to be iterative and flexible, and to involve dissemination of intermediate findings for feedback and broad dissemination of final results. Moreover, evaluation should be integrated with implementation, rather than a standalone post implementation activity. No single type of measurement should dominate the evaluation, which should employ a measurement framework including work and communication patterns, organisational culture, safety, effectiveness, system integrity and usability, as well as vendor factors, project management, participant experience and governance.

1. Introduction

The National Health IT Plan proposes an improved and rationalised health IT infrastructure for New Zealand that will ultimately support a transformed and more sustainable healthcare system. A range of innovative health IT projects is required to realise the plan, and these projects require evaluation. Moreover, it is vitally important that we learn, at a national level, from the experience acquired in each project, both in terms of its successes as well as opportunities for improvement.

For this reason the National Institute for Health Innovation (NIHI) under commission of the National Health IT Board (NHITB) has developed this Framework for Health IT Evaluation. The framework provides guidelines intended to promote consistency and quality in the process of evaluation, in its reporting and in the broad dissemination of the findings. It should also promote, to some degree, the efficiency of evaluation since the effort that has gone into the framework can be leveraged by each evaluation project team.

Herein we present elements of the framework's rationale and recommendation. The framework has been tested in the context of NIHI's evaluation of electronic referral (eReferral) projects. The eReferral reports [1-4] provide exemplars of application of the framework. We are currently applying the framework to evaluation of the National Shared Care Planning pilot for long-term condition management and have recommended its use for evaluation of the national Health Identity programme.

2. Action Research (AR) approach

An AR approach to evaluation is recommended. With respect to AR, the minimal use in a project where evaluation and implementation overlap is to allow stakeholders, particularly end users of the software, to be aware of the evaluation results on an ongoing basis so that they: (a) are encouraged by the benefits observed so far and (b) explicitly react to the findings so far to provide their interpretation and feedback, as described in [5]. At the most aggressive level, one may view the entire implementation and concurrent evaluation as an undertaking of the stakeholders themselves with IT and evaluation staff purely as the facilitators of change. For instance, Participatory Action Research (PAR) methodology has

been endorsed and promoted internationally as the appropriate format for primary health care research and in particular in communities with high needs [6]. PAR is "based on reflection, data collection, and action that aims to improve health and reduce health inequities through involving the people who, in turn, take actions to improve their own health." [7] This suggests an extreme view where the *patients* are active in the implementation; a less extreme view would see just the healthcare professionals as the participants.

Even when the evaluation is clearly following the formal end of implementation activities (which, again, is not ideal; but is often the reality), an AR philosophy can still be applied. This can take the form of the evaluation team:

- Seeking to share the findings with the stakeholders in the current system implementation and taking the feedback as a further iteration of the research
- Actively looking for solutions to problems identified (e.g. adapting interview protocols to ask interviewees if they have ideas for solutions)
- Recommending refinements to the current system in the most specific terms that are supported by the findings (with the intent of instigating stakeholders to pursue these refinements)

It is likely that many of the areas for refinement will relate to software usability. It is appropriate to recognise that implementation is never really over (locally or nationally), and that software is – by its nature – amenable to modification. This fits the philosophy of Interaction Design [8], which is the dominant paradigm for development of highly usable human-computer interfaces and most notably adhered to by Apple Inc. Fundamental to Interaction Design is the continuous involvement of users to shape the product, and the willingness to shape the product in response to user feedback irrespective of the preconceptions of others (e.g. management and programmers). If possible, especially where evaluation is well integrated with implementation, Interaction Design elements should be brought to bear as part of the AR approach.

A corollary to recommending an AR approach as per above is that the evaluation process is most appropriately planned and justified along with the IT implementation itself. This leads to setting aside the appropriate resources for evaluation, and creates the expectation that this additional activity stream is integral to the overall implementation effort.

3. Criteria Pool

In a recent 'review of reviews' of health information systems (HIS) studies, Lau et al [9] offer a broad pool of HIS benefits which he bases on the Canada Health Infoway Benefits Evaluation (BE) Framework [10] (itself based on the Information Systems Success model by DeLone and McLean [11]). Lau et al further expand the Infoway BE model based on measures emerging in their review which didn't fit the existing categories. In addition, Westbrooke et al [12] take a multi-method socio-technical approach to health information systems evaluation encompassing the dimensions of work and communication patterns, organisational culture, and safety and quality. The criteria pool is also influenced by Greenhalgh and Russell's recommendation to supplement the traditional positivist perspective with a critical-interpretive one to achieve a robust evaluation of complex e-health systems that captures the range of stakeholder views [13].

From these sources we draw the following criteria pool, as listed in Table 1. Evaluators should select a mix of criteria from the major dimensions of this pool in identifying evaluation measures for a specific evaluation project. The major focus should be on criteria from the *Impact* genre. Areas that cannot be addressed in depth (which will almost always be most of them) should be addressed qualitatively within the scope of stakeholder interviews. Some areas, such as direct clinical outcomes, are likely to be beyond the scope of most evaluation studies.

Criteria from the criteria pool may be supplemented with specific functional and non-functional requirements that have been accepted as critical success factors for the particular technology in question.

Table 1 - Criteria Pool

Criteria Domain	Criteria Type	Examples / Comments
	Genre: Impact	
Work and Communication Patterns	Efficiency	Time-and-motion measurements, logging of screen access times, transactional log cycle times (e.g. received-to-actioned latency), direct expenditure (staff time or materials), self report of task time, impression of efficiency; also Safety and Quality or Clinical Effectiveness (see below) of a given resource

	Coherence	Interruptions, multi-tasking (observed or self-reported)
Organisational Culture	Positivity	Reporting feeling positive / motivated, sick leave rates, turnover
	Safety (culture of)	Reported feeling that system is safe, specific safety promoting practices (e.g. incident reporting and review) – also see Safety and Quality domain below
	Effectiveness and Quality (culture of)	Self report that efforts are effective / that quality matters, quality improvement activity
	Social networks	Levels of inter-professional communication, inter-professional trust, respect and empathy
	Patient centeredness	Patient engagement, adherence, confidence, knowledge
Safety and Quality	Safety	Incident rates, timeliness of review, potential sources of error including data inaccuracy (wrong patient details, incorrect / missing / duplicate clinical data) and illegibility; also see Clinical Effectiveness below
	Quality	See Organisational Culture above and Clinical Effectiveness below
Clinical Effectiveness	Outcome	Mortality, morbidity, readmission, length of stay, patient functional status or quality of health/life (e.g. SF-36)
	Indicator	HbA1c, blood pressure, etc.
	Process measure	Guideline adherence – also domains above
Genre: Product		
IT System Integrity	Stability	Uptime, errors (logged or self-report), disaster recovery features, maintenance effort
	Data quality	See Safety above
	Data security	IT expert opinion, standards compliance, evidence of breaches
	Standards compliance	International / national compliance, demonstrated interoperability
	Scalability	Response time, maintainability / tailorability / extensibility, IT expert opinion
Usability	Uptake / Use	Rate and extent of uptake, persistence of use of alternatives / workarounds (as measured from transactional systems, or self-report)
	Efficiency	As per *Impact* genre above
	Accuracy	Data entry / interpretation error rates – as per Safety above
	Learnability	Extent of feature use, help desk requests, rate of uptake
	Satisfaction	Overall happiness with solution (e.g. desire to continue using it)
Vendor Factors		Cost competitiveness of licensing / services, vendor support / commitment
Genre: Process		
Project Management		On time, on budget, with proposed features / benefits
Participant Experience		Disruption (self-report or using intermediate measures from the *Impact* genre), angst / anger, meeting expectations, feeling included, impact on Organisational Culture (as per *Impact* genre)
Leadership and Governance		Identification of leaders, ability to have bridged difficult transitions, role in maintaining quality of Participant Experience and meeting Project Management goals

4. Study Design Guidelines

To achieve a multi-dimensional evaluation, and best leverage the available data sources, it is recommended that an evaluation of health IT implementation include at least the following elements in the study's data collection activities:

- Analysis of documents, physical system and workflow
- Semi-structured interviewing and thematic analysis of interview content. This may take the form of one-on-one interviews or focus groups, or (ideally) a combination, and should take an iterative, reflective and interpretivist approach
- Analysis of transactional data, i.e. analysis of the records that result from the direct use of information systems in the implementation setting(s)

The findings from these data sources will support assessment with respect to criteria selected from the criteria pool listed in Table 1.

Two further elements of study design are mandatory:

- Assessment of patient safety – at least in so far as to ask stakeholders working at the point of care to explain how the implementation may be improving or threatening safety
- Benefits framework – to collect data that supports a defensibly appropriate assessment; explained further below.

Evaluation may also involve questionnaires and timed observations (automatically or manually). Defining a control group is optional, but valuable to make a more persuasive case with respect to the innovative use of IT indeed being the source of quantitative changes in system performance; a pragmatic level of control may be to draw parallel data from a health delivery unit with characteristics similar to the one involved in the implementation. It is essential to be clear about what is being evaluated and to match the study design to the evaluation objectives and available data resources.

5. Why not RCTs?

The above recommendation is made with acknowledgement of contrasting opinions in the literature about the building of evidence for the effectiveness of health information systems. One perspective is that health IT is just an extension of Evidence Based Medicine. As with other interventions – such as drug treatment or surgical procedure – the best scientific evidence is achieved by running randomised controlled trials (RCTs). The Cochrane Collaboration's library of reviews, of which there are over 4000 so far [14], and including reviews on many health IT topics (such as the effects of nursing record systems [15] and clinical decision support systems for neonatal care [16]) represent the embodiment of the philosophy that well-informed decision making is based on RCTs. As per Lau et al's recent review [9], some decision evidence about health IT can be drawn from the RCTs to date but results are not particularly consistent in many areas, and there are many gaps.

There are a number of reasons why this guideline recommends study designs other than RCTs. These include:

- RCTs on health IT *are* an important area of endeavour, but the development of this form of evidence is the precinct of the Health Research Council (HRC)
- RCTs are not designed to provide assurance that a specific health IT implementation is working well; an RCT design must centre on testing for significant outcome on one aspect of performance, whereas for the purpose of ensuring that an implementation is working well (and to identify areas for improvement) we must use other methods, even for an implementation that is the subject of an RCT.
- A well executed RCT is expensive – every innovative use of health IT should be evaluated, but New Zealand cannot afford to apply the RCT methodology in each and every case
- Health IT challenges both the internal and external validity of the RCT approach – for instance
  - Double-blinding (where study staff and participants are unaware of whether they are experiencing 'treatment' or 'control' conditions) is largely infeasible, and
  - Keeping conditions fixed for extended period of times to gather statistically significant data samples requires suspension of perfective maintenance that will generally be very active with any innovative information system.

Insight on alternatives is provided by Greenhalgh et al who have demonstrated 'realist evaluation' that is 'pragmatic and reflexive.' [17] Also, Westbrook et al's multi-method evaluation [12] demonstrates building evaluation out of a package of multiple relatively smaller study protocols as compared to a central focus on the RCT.

6. Benefits Framework

As indicated by the Criteria Pool in Table 1, there is a wide range of potential areas of benefit (or harm) for IT systems in health, constituting a spectrum of targets for quantitative and qualitative assessment. The specific criteria for a given evaluation study should not be chosen at random. Rather, the case for what to measure and report should be carefully justified.

There are several types of sources that can inform the formulation of a benefits framework for a given evaluation study:

- Necessary properties – for systems within the scope of this framework, which touch directly on delivery of patient care, it is difficult to see how patient safety can be omitted from consideration. Also health workforce issues, such as user satisfaction with the system, are difficult to ignore (at least in terms of looking out for gross negative effects).
- Standards and policies – the presence of specific functions or achievement of specific performance levels may be dictated by relevant standards or policies (or even law).
- Academic literature and reports – previous evaluations, overseas or locally, may provide specific expectations about benefits (or drawbacks to look out for).
- Project business case – most IT-enabled innovations will have started with a 'project' tied to the implementation of the IT infrastructure, or a significant upgrade in its features or extension in its use. This project will frequently include a business case that promises benefits that outweigh costs, possibly with the mapping of benefits into a financial case. The evaluation should assess the key assertions / assumptions of the business case.
- Emergent benefits – ideally the evaluation should be organised with an iterative framework that allows follow-up on leads, e.g. initial interviews might indicate user beliefs about key benefits of the system that were outside the initial benefits framework and which could then be confirmed and measured in quantitative data.

With respect to the last point above, the benefits framework may evolve over the course of evaluation, particularly if the evaluation involves multiple sites or spans multiple phases of implementation. Thus, the benefits framework may start with the business case assumptions and a few key standards and policy requirements, plus necessary attributes about patient safety and provider satisfaction; it may then evolve after initial study to include benefits that were not explicitly anticipated prior to the commencement of evaluation.

7. Why not cost/benefit analysis or cost-effectiveness analysis?

Ideally, to make the case for sustained and comprehensive roll-out of an innovation we would like to have a clear cost/benefit case that shows a net financial gain from the action. Or at least we would like a cost-effectiveness analysis (CEA) that compares the outcomes from the innovation to some 'business as usual' situation, whether or not the outcome effect is monetized.

A 'poor man's CEA' will look at the number needed to treat (NNT) for the intervention and multiply it by the cost of an adverse event avoided. For example, if one in 10 patients avoid a $10,000 procedure when operating with the new system (NNT=10), then the effect is a $1000 per patient savings (to be contrasted with whatever cost is associated with operating the new system for the 10 patients).

While superficially compelling, it has been pointed out that such use of NNT is fraught with assumptions and is, in general, misleading [18]. Major problems lie in the facts that: (a) NNT is restricted to a single outcome and (b) an NNT is based on a set time period. The concept is relatively well contained if we think of, say, patient functional status three-weeks after a surgical procedure, but becomes less applicable for long-term condition management where a range of adverse events may be avoided continuously over an indefinite time interval. As the model becomes more complex, it becomes less compelling – both due to the number of assumptions made and the difficulty in articulating the value proposition.

A formal health economic analysis is a valuable exercise. But, like the RCTs that should underpin such an analysis, it will generally be outside of the scope of an evaluation exercise associated with a given health IT implementation.

We recommend that evaluation tests the assumptions of the business case of the implementation project. In the first instance, the focus should be on the validity of the business case. A business case will be predicated on system use. Is the system widely used, or have users resisted its use? Do users believe that the assumed effects are happening? Do the users note major iatrogenic harms that offset the case for benefit? Do the data run in the direction of the predicted beneficial effect(s)?

8. A compelling case for change

The key contribution of the evaluation against the benefits framework should be to indicate whether the innovation is one that should be adopted broadly.

To warrant recommendation for emulation the innovation should be free of 'red flags' – this includes being free of evidence of net harm to patients, and having no major negative impact on the health workforce. Beyond this, the innovation must show a clear case for some benefit that is sufficiently compelling to warrant the cost and disruption of adopting the innovation.

Health workforce is a particular challenge for New Zealand, where we face a "complex demand-supply-affordability mismatch" [19]. As such, benefits that tie directly back to effective use of health workforce will be particularly compelling. If an innovation allows us to do more (at the same quality) with the same number of healthcare workers, or do better with the same number of healthcare workers, then it is compelling. An innovation that empowers and satisfies healthcare workers may also be compelling due to its ability to retain those workers. And an innovation that lets workers "practice at the top of their licence" [20] will get the most out of our limited health workforce and engender their satisfaction while doing so. In some cases this may involve changing care delivery patterns in accordance with evidence based medicine such that use of particular services or procedures is reduced (e.g. shifting service from secondary to community). Such changes should be detectable from the transactional records of health information systems.

9. Dissemination

What is learnt from an evaluation project should be made widely known. Dissemination of the findings should not be an after-thought – it is an integral part of the project plan. As such, dissemination is not something that happens *afterwards*. Rather, it is the major emphasis of the last stage of the project. Moreover, to the greatest extent feasible, it should be something that goes on as the project progresses. Feedback from dissemination of interim findings is a valuable component of the evaluation per se.

Dissemination approaches should include:

- Conventional written reporting – interim and archival records of findings
- Face-to-face reporting – meeting with key stakeholders, but at an operational level (clinical users and IT staff) and at a governance level
- Web 2.0 – utilising a space such as the Health Innovation Exchange to post documents and engender active asynchronous written feedback over the Internet
- Academic publication – to obtain academic peer-review feedback that tests the novelty of findings on the world stage, and to leverage academic publishers to archive and disseminate the findings

10. Conclusion

The IT evaluation framework is intended for projects involving innovative use of health IT. The framework recommends a pragmatic approach to evaluation, using both qualitative and quantitative data. It emphasises capture of a broad range of stakeholder perspectives and multi-dimensional evaluation on criteria related to process and culture, as well as outcome and IT system integrity. It also recommends underpinning quantitative analysis with the transactional data from the health IT systems themselves.

The recommended approach is iterative and Action Research (AR) oriented. Evaluation should be integral to implementation – it should begin, if possible, before the new technology is introduced into the health workflow and be planned for along with the planning of the implementation itself. Evaluation findings should be used to help refine the implementation and to evoke further user feedback. Dissemination of the findings is integral and should reach all stakeholders considering uptake of similar technology.

11. References

[1] Day K, Gu Y, Warren J, White S, Pollock M. National eReferral Evaluation: Findings for Northland District Health Board. Wellington: Ministry of Health; 2011.

[2] Warren J, Gu Y, White S, Day K, Pollock M. National eReferral Evaluation: Findings for the Canterbury Initiative. Wellington: Ministry of Health; 2011.

[3] Warren J, Gu Y, White S, Day K, Pollock M. Evaluating NZ Health Innovation Projects: eReferrals Evaluation. Wellington: Ministry of Health; 2011.

[4] Warren J, White S, Day K, Pollock M. National eReferral Evaluation: Findings for Hutt Valley District Health Board. Wellington: Ministry of Health; 2011.

[5] Baskerville R. Investigating information systems with action research. Communications of the Association for Information Systems. 1999;2(19).

[6] Macaulay AC, Commanda LE, Freeman WL, Gibson N, McCabe ML, Robbins CM, et al. Participatory research maximises community and lay involvement. North American Primary Care Research Group. BMJ. 1999 Sep 18;319(7212):774-8.

[7] Baum F, MacDougall C, Smith D. Participatory action research. J Epidemiol Community Health. 2006 Oct;60(10):854-7.

[8] Cooper A, Reinmann RM, Cronin D. About face 3: the essentials of interaction design (3rd ed.). Wiley; 2007.

[9] Lau F, Kuziemsky C, Price M, Gardner J. A review on systematic reviews of health information system studies. J Am Med Inform Assoc. 2010 Nov 1;17(6):637-45.

[10] Lau F, Hagens S, Muttitt S. A proposed benefits evaluation framework for health information systems in Canada. Healthc Q. 2007;10(1):112-6, 8.

[11] Delone WH, McLean ER. The DeLone and McLean model of information system success: a ten year update. J Manag Info Systems. 2003;29:9-30.

[12] Westbrook JI, Braithwaite J, Georgiou A, Ampt A, Creswick N, Coiera E, et al. Multimethod evaluation of information and communication technologies in health in the context of wicked problems and sociotechnical theory. J Am Med Inform Assoc. 2007 Nov-Dec;14(6):746-55.

[13] Greenhalgh T, Russell J. Why do evaluations of eHealth programs fail? An alternative set of guiding principles. PLoS Med. 2010;7(11):e1000360.

[14] The Conchrane Collaboration. The Cochrane Collaboration: Working together to provide the best evidence for healthcare (URL: http://www.cochrane.org/about-us, accessed 11 December, 2010). 2010 [updated 2010; cited 11 December, 2010]; Available from: http://www.cochrane.org/about-us.

[15] Urquhart C, Currell R, Grant MJ, Hardiker NR. Nursing record systems: effects on nursing practice and healthcare outcomes. Cochrane Database Syst Rev. 2009(1):CD002099.

[16] Tan K, Dear PR, Newell SJ. Clinical decision support systems for neonatal care. Cochrane Database Syst Rev. 2005(2):CD004211.

[17] Greenhalgh T, Humphrey C, Hughes J, Macfarlane F, Butler C, Pawson R. How do you modernize a health service? A realist evaluation of whole-scale transformation in london. Milbank Q. 2009 Jun;87(2):391-416.

[18] Kristiansen IS, Gyrd-Hansen D. Cost-effectiveness analysis based on the number-needed-to-treat: common sense or non-sense? Health Econ. 2004 Jan;13(1):9-19.

[19] Gorman D. The future disposition of the New Zealand medical workforce. N Z Med J. 2010 May 28;123(1315):6-8.

[20] Wagner EH. Better care for chronically ill people (presentation to the Australasian Long-Term Conditions Conference, 7 April 2011; available from URL: http://www.healthnavigator.org.nz/conference/presentations, last accessed 28 June 2011); 2011.