Chapter 6

Pragmatic Health Information Technology Evaluation Framework

Jim Warren, Yulong Gu

6.1 Introduction

This chapter outlines a pragmatic approach to evaluation, using both qualitative and quantitative data. It emphasizes capture of a broad range of stakeholder perspectives and multidimensional 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 also integral and should reach all stakeholders considering uptake of similar technology.

This Health Information Technology (IT) Evaluation Framework was developed under the commission of the New Zealand (N.Z.) National Health IT Board to support implementation of the New Zealand National Health IT Plan (IT Health Board, 2010) and health innovation in the country in general. The framework was published in 2011 by the N.Z. Ministry of Health (Warren, Pollock, White, & Day, 2011) with a summary version of this report presented in the Health Informatics New Zealand 10th Annual Conference and Exhibition (Warren, Pollock, White, Day, & Gu, 2011). This framework provides guidelines intended to promote consistency and quality in the process of health IT evalu-

ation, in its reporting and in the broad dissemination of the findings. In the next section, we discuss key elements of the conceptual foundations of the framework. In the third section we specifically address formulation of a Benefits Evaluation Framework from a broad Criteria Pool. We conclude with the implications of applying such a framework and summary.

6.2 Conceptual Foundations and General Approach

A number of sources informed this framework's recommendations for how to design an evaluation. In a nutshell, the philosophy is:

- Evaluate many dimensions – don't look at just one or two measures, and include qualitative data; we want to hear the "voices" of those impacted by the system.

- Be adaptive as the data comes in – don't let the study protocol lock you into ignoring what's really going on; this dictates an iterative design where you reflect on collected data before all data collection is completed.

6.2.1 Multiple dimensions

Of particular inspiration toward our recommendation to evaluate many dimensions is the work of Westbrook et al. (2007) who took a multi-method sociotechnical approach to health information systems evaluation encompassing the dimensions of work and communication patterns, organizational culture, and safety and quality. They demonstrate building evaluation out of a package of multiple relatively small study protocols, as compared to a central focus on randomized controlled trials (RCTs), as the best source of evidence. Further, a "review of reviews" of health information systems (HIS) studies (Lau, Kuziemsky, Price, & Gardner, 2010) offers a broad pool of HIS benefits which the authors base on the Canada Health Infoway Benefits Evaluation (BE) Framework (Lau, Hagens, & Muttitt, 2007), itself based on the Information Systems Success model (Delone & McLean, 2003). Lau et al. (2010) further expand the Infoway BE model based on measures emerging in their review which didn't fit the existing categories. In addition, our approach is influenced by Greenhalgh and Russell's (2010) recommendation to supplement the traditional positivist perspective with a critical-interpretive one to achieve a robust evaluation of complex eHealth systems that captures the range of stakeholder views.

6.2.2 Grounded Theory (GT) and the Interpretivist view

In contrast to measurement approaches aimed at predefined objectives, GT is an inductive methodology to generate theories through a rigorous research process leading to the emergence of conceptual categories. These conceptual categories are related to each other, and mapping such relationships constitutes a

theoretical explanation of the actions emerging from the main concerns of the stakeholders (Glaser & Strauss, 1967). Perhaps the most relevant message to take from GT is the idea of a "theory" emerging from analysis and acceptance of the messages in the interview data; this contrasts with coming in with a hypothesis that the data tests.

We recommend that the evaluation team (the interviewer, and also the data analyst) allow their perspective to shift between a positivist view (that the interview is an instrument to objectively measure the reality of the situation) and an interpretivist view. Interpretivism refers to the "systematic analysis of socially meaningful action through the direct detailed observation of people in natural settings in order to arrive at understandings and interpretations of how people create and maintain their social worlds" (Neuman, 2003, p. 77). An interpretivist accepts that their presence affects the social reality. Equally importantly, the interpretivist accepts individual views as a kind of reality in their own right. The complex demands, values and interrelationships in the healthcare environment make it entirely possible for different individuals to interpret and react to the exact same health IT system in very different ways. The interpretivist takes the view that each stakeholder's perspective is equally (and potentially simultaneously) valid; the aim is to develop the understanding of why the ostensibly contradictory views are held.

Ideally, in developing themes (or GT conceptual categories) from interview data, one would conduct a complete "coding" of the interview transcripts, assigning each and every utterance to a place in the coding scheme and then allowing a theory to emerge that relates the categories. Further, since this process is obviously subjective, one should regard the emerging schema with "suspicion" and contest its validity by "triangulation" to other sources (Klein & Myers, 1999), including the international research literature. These techniques are illustrated in the context of stakeholders of genetic information management in a study by Gu, Warren, and Day (2011). Creating a complete coding from transcripts is unlikely to be practical in the context of most health IT evaluation projects. As such, themes may be developed from interview notes, directly organizing the key points emerging from each interview to form the categories to subsequently organize into a theory of the impact of the health IT system on the stakeholders. Such a theory can then be presented by describing in detail each of several relevant themes.

6.2.3 Evaluation as Action Research (AR)

An ideal eHealth evaluation has the evaluation plan integrated with the implementation plan, rather than as a separate post-implementation project. When this is the case, the principles of AR (McNiff & Whitehead, 2002; Stringer, 1999) should be applied to a greater or lesser degree. The AR philosophy can be integrated into interviews, focus groups and forums in several ways that recognize that the AR research aims to get the best outcome (while still being a faithful reporter of the situation) and will proceed iteratively in cycles of planning, re-

flection and action. With the AR paradigm — in which case the evaluation is probably concurrent with the implementation — the activities of evaluation can be unabashedly and directly integrated with efforts to improve the effectiveness of the system.

With respect to AR, at the minimum 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. 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 the 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 (Macaulay et al., 1999). 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" (Baum, MacDougall, & Smith, 2006, p. 854). 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 pursuit of these refinements by stakeholders).

It is likely that many of the areas for refinement will relate to software usability. It is appropriate to recognize 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 (Cooper, Reinmann, & Cronin, 2007) which is the dominant paradigm for development of highly usable human-computer interfaces and most notably adhered to by Apple Incorporated. 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., man-

agement 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.

6.3 Benefits Evaluation Framework

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 and assumptions of the business case.

- Emergent benefits – ideally the evaluation should be organized with an iterative framework that allows follow-up on leads; for example, 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.

From the sources cited in 6.2.1 above, and our own experience, we draw the criteria pool in Table 6.1. 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. Moreover, 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 6.1
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)
Organizational 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 centredness	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 Organizational 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., via the 36-item Short Form Health Survey, SF-36)
	Indicator	Glycated haemoglobin (HbA1c), blood pressure, etc.
	Process measure	Clinical practice guideline adherence – also domains above

Table 6.1
Criteria Pool

Criteria Domain	Criteria Type	Examples / Comments
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

Table 6.1
Criteria Pool

Criteria Domain	Criteria Type	Examples / Comments
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 Organizational 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

Note. From *A framework for health IT evaluation*, by J. Warren, M. Pollock, S. Day, and Y. Gu, 2011, pp. 3–4. Paper presented at the Health Informatics New Zealand 10th Annual Conference and Exhibition, Auckland. Copyright 2011 by Health Informatics Conference. Reprinted with permission.

6.4 Guidance on Use of the Framework and its Implications

6.4.1 Guidance

To achieve a multidimensional 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, that is, 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 6.1.

Two further elements of study design are essential:

- Assessment of patient safety – at least insofar 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, the performance expectations should be defined working from criteria as per section 6.3 above; in keeping the GT and AR, these criteria (and thus the focus of evaluation) may be allowed to adjust over the course of the project (obviously with the agreement of funders each step of the way!).

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, but it is also essential to match the study design and evaluation objectives to the available resources.

The framework has been tested in the context of evaluations of several regional electronic referral (eReferral) projects in New Zealand. The eReferral reports (Day, Gu, Warren, White, & Pollock, 2011; Gu, Day, Humphrey, Warren, & Pollock, 2012; Warren, Gu, Day, Pollock, & White, 2012; Warren, Pollock, White, & Day, 2011; Warren, Pollock, White, Day, et al., 2011; Warren, White, Day, & Pollock, 2011) provide exemplars of the application of the framework. We also applied the framework in evaluations of the N.Z. National Shared Care Planning pilot for long-term condition management (National Institute for Health Innovation, 2013; Warren, Gu, & Humphrey, 2012; Warren, Humphrey, & Gu, 2011) and the Canterbury electronic Shared Care Record View project (Gu, Humphrey, Warren, & Wilson, 2014).

6.4.2 Implications
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 many healthcare systems; certainly it is for New Zealand, where we face a "complex demand-supply-affordability mismatch" (Gorman, 2010). As such, benefits that tie directly back to effective use of health workforce will be particularly compelling. If an innovation allows more to be done (at the same quality) with the same number of health-

care workers, or allows doing 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" (Wagner, 2011) 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 hospital-based specialist care to the community). Such changes should be detectable from the transactional records of health information systems.

6.5 Summary

A pragmatic evaluation framework has been recommended for projects involving innovative use of health IT. The framework recommends using both qualitative and quantitative data. It emphasizes capture of a broad range of stakeholder perspectives and multidimensional 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.

References

Baum, F., MacDougall, C., & Smith, D. (2006). Participatory action research. *Journal of Epidemiology & Community Health, 60*(10), 854–857. doi: 10.1136/jech.2004.028662

Cooper, A., Reinmann, R. M., & Cronin, D. (2007). *About face 3: the essentials of interaction design* (3rd ed.). New York: John Wiley & Sons.

Day, K., Gu, Y., Warren, J., White, S., & Pollock, M. (2011). *National eReferral evaluation: Findings for Northland District Health Board*. Wellington: Ministry of Health.

DeLone, W. H., & McLean, E. R. (2003). The DeLone and McLean model of information system success: A ten-year update. *Journal of Management Information Systems, 19*(4), 9–30.

Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research*. Chicago: Aldine Publishing Company.

Gorman, D. (2010). The future disposition of the New Zealand medical workforce. *New Zealand Medical Journal, 123*(1315), 6–8.

Greenhalgh, T., & Russell, J. (2010). Why do evaluations of eHealth programs fail? An alternative set of guiding principles. *Public Library of Science Medicine, 7*(11), e1000360. doi: 10.1371/journal.pmed.1000360

Gu, Y., Day, K., Humphrey, G., Warren, J., & Pollock, M. (2012). *National eReferral evaluation: Findings for Waikato District Health Board*. Wellington: Ministry of Health. Retrieved from http://api.ning.com/files/FrmWaEpjYR4bwaOAyZTeTNjQ65183Caca4ZD kjOg*IDLoICJEVdtCzbAoK3XRUY27MYXWeF7536QOsGHiMfkcJG1cR Qu*StU/Waikato.eReferral.evaluation.pdf

Gu, Y., Humphrey, G., Warren, J., & Wilson, M. (2014, December). *Facilitating information access across healthcare settings — A case study of the eShared Care Record View project in Canterbury, New Zealand.* Paper presented at the 35th International Conference on Information Systems, Auckland, New Zealand.

Gu, Y., Warren, J., & Day, K. (2011). Unleashing the power of human genetic variation knowledge: New Zealand stakeholder perspectives. *Genetics in Medicine, 13*(1), 26–38. doi: 10.1097/GIM.0b013e3181f9648a

IT Health Board. (2010). *National health IT plan: Enabling an integrated healthcare model.* Wellington: National Health IT Board.

Klein, H., & Myers, M. (1999). A set of principles for conducting and evaluating interpretive field studies in information systems. *Management Information Systems Quarterly, 23*(1), 67–93.

Lau, F., Hagens, S., & Muttitt, S. (2007). A proposed benefits evaluation framework for health information systems in Canada. *Healthcare Quarterly, 10*(1), 112–118.

Lau, F., Kuziemsky, C., Price, M., & Gardner, J. (2010). A review on systematic reviews of health information system studies. *Journal of the American Medical Informatics Association, 17*(6), 637–645. doi: 10.1136/jamia.2010.004838

Macaulay, A. C., Commanda, L. E., Freeman, W. L., Gibson, N., McCabe, M. L., Robbins, C. M., & Twohig, P. L. (1999). Participatory research maximises community and lay involvement. (Report for the North American Primary Care Research Group). *British Medical Journal, 319*(7212), 774–778.

McNiff, J., & Whitehead, J. (2002). *Action research: principles and practice* (2nd ed.). London: Routledge.

National Institute for Health Innovation. (2013, May). Shared Care Planning evaluation Phase 2: Final report. Retrieved from http://www.sharedcareplan.co.nz/Portals/0/documents/News-and-Publications/Shared%20Care%20Planning%20Evaluation%20Report%20-%20Final2x.pdf

Neuman, W. L. (2003). *Social research methods: Qualitative and quantitative approach* (5th ed.). Boston: Pearson Education.

Stringer, E. T. (1999). *Action research* (2nd ed.). Thousand Oaks, CA: SAGE Publications.

Wagner, E. H. (2011). *Better care for chronically ill people* (presentation to the Australasian Long-Term Conditions Conference, April 7, 2011). Retrieved from http://www.healthnavigator.org.nz/conference/presentations

Warren, J., Gu, Y., Day, K., Pollock, M., & White, S. (2012). Approach to health innovation projects: Learnings from eReferrals. *Health Care and Informatics Review Online, 16*(2), 17–23.

Warren, J., Gu, Y., & Humphrey, G. (2012). *Usage analysis of a shared care planning system.* Paper presented at the The AMIA 2012 Annual Symposium, Chicago.

Warren, J., Humphrey, G., & Gu, Y. (2011). *National Shared Care Planning Programme (NSCPP) evaluation: Findings for Phase 0 & Phase 1.* Wellington: Ministry of Health.Retrieved from http://www.sharedcareplan.co.nz/Portals/0/documents/News-and-Publications/F.%20NSCPP_evaluation_report%2020111207%202.pdf

Warren, J., Pollock, M., White, S., & Day, K. (2011). *Health IT evaluation framework*. Wellington: Ministry of Health.

Warren, J., Pollock, M., White, S., Day, K., & Gu, Y. (2011). *A framework for health IT evaluation.* Paper presented at the Health Informatics New Zealand 10th Annual Conference and Exhibition, Auckland.

Warren, J., White, S., Day, K., & Pollock, M. (2011). *National eReferral evaluation: Findings for Hutt Valley District Health Board*. Wellington: Ministry of Health.

Westbrook, J. I., Braithwaite, J., Georgiou, A., Ampt, A., Creswick, N., Coiera, E., & Iedema, R. (2007). Multimethod evaluation of information and communication technologies in health in the context of wicked problems and sociotechnical theory. *Journal of the American Medical Informatics Association, 14*(6), 746–755. doi: 10.1197/jamia.M2462.