ELSEVIER

Playing smallball: Approaches to evaluating pilot health information exchange systems

Kevin B. Johnson *, Cynthia Gadd

Vanderbilt University Medical Center, 2209 Garland Avenue, Room 428, Nashville, TN 37232, United States

Abstract

Health information exchange (HIE) systems are large, multimillion dollar efforts that are implemented despite initial institutional apprehension, with largely unanticipated effects on the clinical workflow, and with a primary goal of establishing a reason to sustain the effort. Traditional evaluation methods that rely on viewing HIE as generalizable tools that can be used to demonstrate impact are challenged by the realities of how HIE systems are created and the numerous systems they impact on the road to improved care. In short, it appears to be unrealistic to develop the definitive "home run" evaluation of this technology. Despite the existence of traditional approaches for large-scale evaluation, a more realistic approach may be a "smallball" model based on established IT implementation phases, with appropriate evaluation dimensions linked to each phase.
© 2007 Elsevier Inc. All rights reserved.

Keywords: Health information exchange; National health information network; Electronic health records; Public health informatics; Evaluation; Biomedical informatics

1. Overview

In its report, *Crossing the Quality Chasm* [1], the Institute of Medicine (IOM) identified health information technology (HIT) as one of the most significant tools that could help improve healthcare quality. Further, the IOM, the National Committee on Vital and Health Statistics, and the President's Information Technology (IT) Advisory Committee also have recommended the development of a national health information infrastructure to help improve safety, reduce costs, and enhance the quality of healthcare. These reports have generated significant enthusiasm and significant financial support to pilot health information exchange (HIE) programs. Over the past 5 years, there have been a number of regional health information organizations put in place to develop health information exchange [2]. These programs have largely been funded by federal, state governmental, or third party groups, based on the promise of improving the quality and safety of care, improving clinician and patient satisfaction, and demonstrating an enticing return on investment.

Health information exchange technology shares many attributes of all health information technologies. It offers benefits to consumers, clinical practices, and to overall public health. However, as noted by Lorenzi [3], these benefits rarely are achieved without some level of integration of HIT into the workflow of clinical practice and recognizing that this integration is not a one-time process. Workflow integration occurs at the level of hardware, software, policies, and content.

1.1. Hardware integration

Studies suggest that from 24–30% of practices have adopted electronic health records. [4,5] The landscape, however, is far from unspoiled. Most practices have some form of electronic claims submission process, often involving a practice management system. These practices usually

* Corresponding author. Fax: +1 615 936 1427.
  E-mail address: kevin.b.johnson@vanderbilt.edu (K.B. Johnson).

have workstations in front office rooms as well as in practitioner offices in some cases. A typical exam room has limited desk space, and will be unlikely to provide sufficient room for more than one computer display. Therefore, any effort to add HIE to the typical office practice setting needs to conform to the hardware specifications already in place; otherwise, hardware must be replaced to conform to room and pre-existing needs.

1.2. Software additions and modifications

Health information technology adoption often is associated with a need to begin with software developed as a generic system that then must be customized at the back end (to support data exchange) and front end to support data aggregation and specific environments. Indeed, this level of customization is commonplace in most non-medical personal computer software today, through the use of customizable menus, button bars, colors, and screen sizes—users have grown accustomed to this level of entitlement. When HIT does not have this level of customization, invariably, some group of users will be less satisfied with the number of mouse-clicks required to perform a function they commonly perform.

In addition to these changes, HIE user interfaces are destined to evolve as new data streams are incorporated or as more codified data are available. Other changes related to views of the data – such as more efficient whiteboards to allow informed selection of records to review – may be constructed based on early feedback. Finally, as is the case with any software, HIE products are not immune from errant software that must be fixed, or software that requires enhancements to be optimally used by a team of providers.

1.3. Policy and procedure evolution

The earliest days of HIT exploration recognized the need to reconcile HIT-related policies with existing procedures. This process often requires significant changes in the content, access control, and use of HIT. In the case of HIE applications, policy regarding who may have access in the initial phases of implementation may be revised as user needs, workflows and trust evolves. Additional policies may be constructed to address patients who do not want to participate in the HIE system, or to account for providers who may wish to review printouts from the system but do not have access to HIE directly.

Furthermore, the staging of HIE implementation may radically impact adoption and use. For example, policies about the amount of retrospective data that can be loaded into these systems affect the usefulness of HIE for many months. Decisions regarding locations that will use the system or specific types of healthcare providers who will or will not have access will impact all other aspects of system development and workflow integration.

1.4. Content customization

In many EHR applications, optimal use is accomplished through customization of not just the user interface, but also the actual content. It is now commonplace for tools that support computer-based documentation to support provider-specific templates. Order entry systems have allowed for group-specific order sets, which have been shown to improve usability and safety. In the world of HIE, content customizations appear to be less frequent. However, some systems that allow EHR-like local functionality such as e-prescribing, immunization entry, or documentation face many of the challenges noted above.

1.4.1. Realizing the promise of health information exchange

As the above points demonstrate, the fluid nature of HIT and HIE applications is generally the rule rather than the exception. While these specific changes to an HIT environment are believed to be necessary for optimal adoption and maximal return on investment, the fluid nature of these systems complicates the ability to conduct studies of their effectiveness. Not only are the systems themselves in a state of flux; personal preferences coupled with personnel turnover make many analyses difficult to perform. Many of these changes, such as errant software, may appear to have minimal impact on evaluation, but, in fact, can completely disable critical functions in an HIE application. Therefore, evaluation of relatively new HIE applications needs to take into account the potential for changes that impact the internal validity of a study.

The challenge of health information exchange presents both an opportunity for health care and a danger for evaluators, who are often saddled with the burden of demonstrating the value of a tool with limited sustained use by a set of providers immersed in a healthcare system unprepared for the transformation potential of HIE, and therefore, likely to also be in a constant state of evolution during the evaluation period.

Although these realities seem daunting (and are certainly nontrivial), there is a body of the literature that addresses the intersection of traditional methods for technology evaluation and the realities of program evaluations. These two areas will be described below, after which, in the final section of this manuscript, the concepts that can be extracted from these areas and that are relevant to HIE evaluation will be defined.

1.4.1.1. Traditional HIT evaluation. Evaluation methods in biomedical informatics must address a wide range of information resources and an equally wide range of questions that can be asked about them, from technical characteristics to organizational issues. There are invariably many actors in HIT projects, including developers, users, and patients, all of whom may have different perspectives on what questions to ask and how to interpret the answers (some of which are changing over time). This complexity necessitates a wide array of empirical methods, including

subjectivist designs that emphasize qualitative approaches, and a dynamic evaluation process. In contrast, research, including the logical-positivist/objectivist (or quantitative) traditions long dominant in biomedicine, serves the focused question or problem, excluding from a study as many extraneous variables as possible. It is not surprising that the overwhelming tendency in biomedical informatics evaluation over the past 25 years has been to prefer the use of the familiar study designs that emphasize quantitative approaches and methods, such as the randomized controlled trial (RCT) [6], even when they are not amenable to the questions to be answered [7].

Stoop and Berg [8] point out that the dominance of RCTs has been questioned for years. In addition to general critiques, such as prejudice for an overly narrow definition of science and the difficulty in separating the HIT intervention from its sociological entanglements, RCTs give "hard data" on a very constrained set of variables, leaving many more valuable questions of how and why and under what circumstances unaddressed. They argue that managerially focused evaluations should emphasize designs that focus on qualitative methods *integrated* with quantitative techniques that are less rigorous (and more widely applicable) than RCTs. The further suggest that in addition to using qualitative methods as "exploratory" steps or primarily for triangulation, the outputs of quantitative research, including "modest" before-after designs, can benefit from qualitative interpretation, e.g., to understand the consequences of downtime on performance of care.

In addition to the philosophical grounds of evaluation, there is the issue of aligning evaluation questions and methods with the developmental stages through which HIT typically progresses. Stead and colleagues [9] advised investigators to subdivide applied informatics research projects into steps and tailor the evaluation to each step; the key idea is that a relationship exists between a developmental stage of a project and the level of evaluation that is appropriate. Five (5) developmental stages are defined: specification, component development, combination of components into a system, integration of system into environment, routine use; as well as five (5) evaluation levels: definition, laboratory bench, laboratory field, remote field validity, remote field efficacy. In their three stages of technology assessment, Fuchs and Garber [10] distinguish Stage 2 efficacy studies, which focus on process measures (e.g., degree of compliance with a reminder), from Stage 3 effectiveness studies, which directly evaluate health and economic outcomes (e.g., whether use of a cancer screening reminder lowered mortality).

In another approach to calibrating the type and timing of evaluation, Friedman uses an analogy from baseball to compare "powerball" evaluation, in which all evaluation resources are saved for an RCT of an extremely mature HIT project, to "smallball" evaluation, in which a succession of smaller, focused evaluation studies are conducted across the life-cycle of the project [11]. The value of smallball evaluation is seen in its potential for self-correction in the design and implementation of the project. Friedman argues that smallball evaluation studies can address needs that are of particular importance to community-based informatics interventions, such as HIE: needs assessment, prototype testing, understanding usage (or lack of), and exploring the effects of the intervention when logistical or ethical constraints operating in community settings prevent randomization and blinding.

Berg [12] offers an approach drawn from sociotechnical science for understanding how the choice of HIT evaluation methods is necessarily grounded in recognition of the "messy" nature of healthcare practice as heterogeneous networks of people, tools, routines, etc. within specific socio-political contexts. This approach casts doubt on work as "rational" – represented in workflow diagrams and clinical pathways – but rather sees it as unfolding in the doing. Additionally, qualitative methods are deemed essential to study the network of changes resulting from HIT implementation, such as tasks, roles, and responsibilities and cultural notions of privacy and quality, as well as the fluidity of structural change inherent in healthcare organizations. Taken all together, these tumultuous interactions emphasize the simultaneous transformation of tool and practice [13]. A recent evaluation of OZIS, a communication protocol that allowed independent pharmacists to share patient medication data, demonstrated the community dynamics, in which 2 groups of competitors (vendors, pharmacists) found ways to collaborate around the system's deployment – at least until the politics and competitive pressures change the dynamics again [14].

1.4.1.2. Program evaluation. Viewing HIE implementations from a sociotechnical perspective – as unfolding in heterogeneous networks of people, tools, roles, systems, processes and within specific socio-political contexts – allows us to generalize them as a type of social program (in which information technology is one component) and therefore amenable to the philosophies and techniques that are used in the field of *program evaluation* to determine if a program 'works'. Patton defines program evaluation as "the systemic collection of information about the activities, characteristics, and outcomes of programs to make judgments about the program, improve program effectiveness, and/or inform decisions about future programming. [6]" Program evaluation developed, particularly in the US, in the context of the Great Society programs of the 1960s and 1970s, including projects focused on education, health, housing, employment, urban renewal, welfare, and family programs. Extraordinary sums were invested but the means of knowing what happened and why were not available. Early expectations for evaluation were focused on guiding funding decisions, separating successful programs from unsuccessful ones, and eventually grew to include helping improve programs as they were implemented [6].

Formalization of and focus on evaluation brought standards, foremost of which was that evaluation should be useful, i.e., that evaluations should be judged by their util-

ity and actual use. From these professional standards evolved a distinction between evaluation research – undertaken to discover new knowledge, test theories, establish truth, and generalize across time and settings – and program evaluation – undertaken to inform decisions, identify improvements, and provide information about programs within contextual boundaries of time, place, values, and politics [6]. Cronbach and Suppes [15] described this as the difference between conclusion-oriented and decision-oriented inquiry.

Utilization-focused evaluation takes program evaluation one step further in that it is "done for and with specific, intended users for specific, intended uses. [6]" This approach narrows the often large field of potential stakeholders who focus the evaluation to those who will use the evaluation data – the specific people who understand and value evaluation should focus the evaluation: what questions will provide information that they care about and that will be relevant for their future action. Substantial research supports what Patton and others have identified the *personal factor* – the presence of an identifiable individual or group who personally care about the evaluation and the findings it generates (the "users") – as the single most important predictor of evaluation utilization. (Note that the "user" here is not necessarily the same as is typical in IT settings, i.e., the person who interacts with the IT system.)

2. Approaches to HIE evaluation

As the above discussion demonstrates, traditional HIT evaluation approaches to HIE systems offer credibility and have recognizable implications when conducted carefully. However, the fluid nature of these systems, coupled with the varied approaches used to develop an HIE implementation team, construct HIE software, pilot test HIE, and transform local care using HIE, mandate a stepwise approach, with carefully constructed testable hypotheses at each juncture, as has been noted by Stoop [8] and Stead [9], and summarized by Friedman's smallball approach

[11]. These iterative evaluations can and should be the first step in a larger HIE program evaluation, given the decision-oriented ("should we continue this?") focus of most initial implementation efforts.

Most health information exchange projects roughly appear to follow the process depicted in Fig. 1. In this diagram, all phases of system development may be seen (in italics). In addition there is a schematic representation of different evaluation study components. During the problem definition phase, an understanding of the environment and the capabilities of the RHIO interleave to form a proposed model for HIE at a technical and organizational/strategic level. Although qualitative studies of this phase would be valuable, they may be extremely difficult to conduct due to socio-political conflicts and drivers that often complicate RHIO formation.

Once the decision to move forward with HIE has been made, there are numerous opportunities to conduct qualitative projects. As described by Stoop, many of these early projects will focus on organization, economic, legal and ethical considerations. With each subsequent phase of the project, other types of studies become more applicable.

The phased approach, as noted by numerous authors [8,9,11,16], appears particularly relevant for consideration with nascent HIE programs, and is the approach being taken by a number of AHRQ-funded state/regional demonstration projects.

Of note, use and impact studies are among the most complex studies to design, approve from a human-subjects perspective, and implement [17–19]. Moreover, even the most carefully conducted randomized design assessing some impact variable of importance to the general public will have limited external validity, because much of what is unique to the local environment will have influenced the model specification, HIE team construction, tool design and data acquisition phases of the project. It will be important, rather, for these studies to recognize the near-term goals of the project, which will likely focus on local, rather then generalizable, models for sustainability.
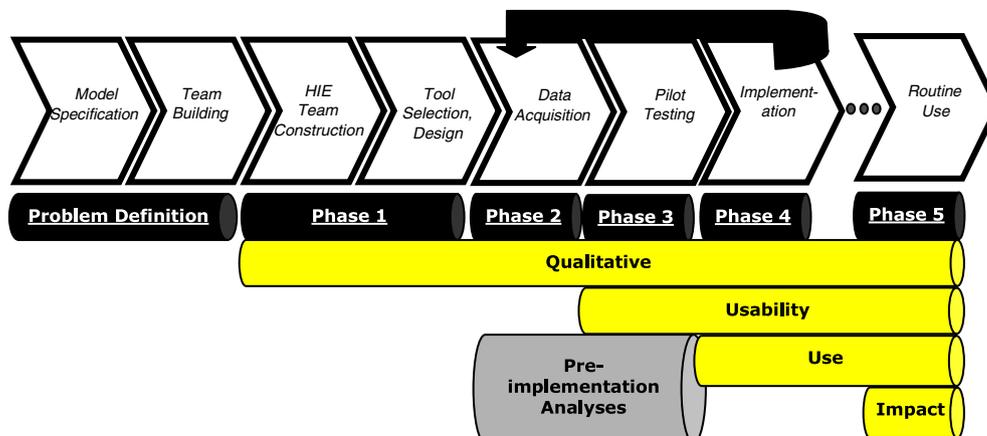

Fig. 1. Phases of system development and evaluation for new health information exchange projects.

Impact has both conclusion-oriented (determining the potential impact of HIE systems on referrals back to the medical home) and decision-oriented (determining whether the HIE program has a realizable return on investment) components. Arguably, although the conclusion-oriented metrics require more complex designs to be generalizable, the return on investment analysis needs to be based on a model that allows the conclusions to be recalculated as the HIE system evolves. This model also requires a level of stability so that the parameters (such as opt out rates) are at least discovered before it is finalized. Therefore, in almost all cases, initial evaluations of impact measures will be limited in their significance unless re-visited over time.

Table 1 summarizes the questions that might be addressed during each phase, based on work done by the MidSouth eHealth Alliance (http://www.mseha.org) evaluation team, as adopted from a recent report distributed by AHRQ [20]. Many other measures may be included.

Many phase 2 measures will require careful planning with the HIE vendor or development team. In particular, rates of exchange, latency (time between the appearance of information in a feeder system and the appearance of that information in the HIE system), and data parsing/record locating services often require logging that is not otherwise necessary, but that is critical to understanding system performance.

Design considerations will be based on time, resources, and cultural readiness. For example, although randomized, controlled trials may be designed to assess impact, ethical considerations may dictate whether and how best to construct the control condition. Furthermore, in environments with additional users being implemented each month, it may be premature to conduct many types of impact analyses, unless the effect of these changes can be accounted for or mitigated.

Because of their expense and need for often complex human subjects review, impact studies should be tailored to address specific states the system is designed to improve (such as rate of unsafe care, over-utilization, referral patterns, etc.) With these data in hand, it is often (though not always) feasible to assess how the system of care using HIE has impacted these outcomes. If specific models for sustainability are proposed, these models may be assessed for their potential feasibility, even before the system has been fully implemented. For example, if there is an agreed-upon fee structure for use of the HIE tools by a specific group (such as case managers), it may be possible to measure the number of patients impacted by the HIE tool, the change in outcome needed to justify that fee structure (such as the desired rate of referral back to a medical home), or other metrics.

Health information exchange programs are in their infancy in most settings, and promise to be both revolutionary to health care and evolutionary in their own scope and functionality. Evaluation of these systems should be aligned with the phases of their development, and with strong consideration of sociotechnical and program evaluation perspectives throughout all phases of study development. Viewing the implementation as unfolding in networks of diverse and changing people, tools, roles, systems, processes, and within specific socio-political contexts in which the HIE is but one component, reinforces the need for the evaluation to address how the tools change practice and how practice shapes the tools. By expecting the evaluation to inform decisions, identify improvements, and provide information about HIE within contextual boundaries of time, place, and values, evaluators recognize the absolute necessity of developing an evaluation that answers questions that matter to those who will use the findings it generates to make decisions about the HIE. While these

Table 1
Questions by phase

Phase	Measure	Design
1	Organizational readiness	Cross-sectional survey
	Site demographics	Site-based survey
2	Baseline clinical measures,	Primary data collection, secondary data, survey, focus group, direct observation
	Baseline workflow evaluation	Direct observation, survey
	Exchange efficiency: data latency, percent of idealized data being exchanged, percent of patients with data, validity of data parsing, record locating	Logs, data analysis
2–3	Quantitative workflow measures: admission rates, length of stay, etc.	Secondary data
3–5	Usability (measured twice a year)	Survey, administered to pre-defined cohort at each site. If necessary, survey results will target focus group meetings for clarification.
4–5	Use over time	Audit data (logins, access duration, data elements viewed, patients viewed) HELP desk use logs
	Test ordering rate over time	Internal audits of data in exchange
	Inpatient admission/re-admission rates, per month	Secondary data (if another data source is used, need to verify that definitions are identical to baseline)
	Quantitative workflow measures	Secondary data
5	Follow-up workflow evaluation	Direct observation + survey
	Follow-up clinical measures*	Primary or secondary data

* May be amenable to more elaborate design, depending on the stability of the system and the feasibility of the design.

rubrics may seem self-evident, considerable vigilance may be required to actually produce an HIE evaluation that adheres to them.

3. Conclusion

Despite pressures from all involved, the evaluation of health information exchange systems, like the evaluation of any newly adopted technology, needs to proceed in a systematic smallball fashion. HIE evaluation should begin with an assessment of the process and the overall usability at a functional level, after which there should be an analysis to assess the stability of both the HIE system and environment in which it is implemented. Only after these co-variates are assessed can a more sophisticated evaluation of HIE impact be considered to have reasonable internal validity.

Acknowledgments

References

[1] Institute of Medicine (US). Committee on Quality of Health Care in America. Crossing the quality chasm: a new health system for the 21st century. Washington, DC: National Academy Press; 2001.

[2] Bartschat W, Burrington-Brown J, Carey S, Chen J, Deming S, Durkin S, et al. Surveying the RHIO landscape. A description of current RHIO models, with a focus on patient identification. J Ahima 2006;77:64A–D.

[3] Lorenzi NM, Riley RT. Organizational issues = change. Int J Med Inform 2003;69:197–203.

[4] Jha AK, Ferris TG, Donelan K, DesRoches C, Shields A, Rosenbaum S, et al. How common are electronic health records in the United States? A summary of the evidence. Health Aff (Millwood) 2006;25:w496–507.

[5] Simon SR, Kaushal R, Cleary PD, Jenter CA, Volk LA, Orav EJ, et al. Physicians and electronic health records: a statewide survey. Arch Intern Med 2007;167:507–12.

[6] Patton MQ. Utilization-focused evaluation. 2nd ed. Beverly Hills: Sage Publications; 1986.

[7] Friedman CP, Wyatt J. Evaluation methods in medical informatics. 2nd ed. New York: Springer; 2006.

[8] Stoop AP, Berg M. Integrating quantitative and qualitative methods in patient care information system evaluation: guidance for the organizational decision maker. Methods Inf Med 2003;42:458–62.

[9] Stead WW, Haynes RB, Fuller S, Friedman CP, Travis LE, Beck JR, et al. Designing medical informatics research and library – resource projects to increase what is learned. J Am Med Inform Assoc 1994;1:28–33.

[10] Fuchs VR, Garber AM. The new technology assessment. N Engl J Med 1990;323:673–7.

[11] Friedman CP. "Smallball" evaluation: a prescription for studying community-based information interventions. J Med Libr Assoc 2005;93:S43–8.

[12] Berg M. Patient care information systems and health care work: a sociotechnical approach. Int J Med Inform 1999;55:87–101.

[13] Berg M. Rationalizing medical work: decision-support techniques and medical practices. Cambridge, Mass: MIT Press; 1997.

[14] Stoop AP, Bal R, Berg M. OZIS and the politics of safety: using ICT to create a regionally accessible patient medication record. Int J Med Inform 2007;76S1:229–35.

[15] Coleman JS, Cronbach LJ, Suppes P. National Academy of Education. Committee on Educational Research. Research for tomorrow's schools: disciplined inquiry for education; report. New York: Macmillan; 1969.

[16] Kaplan B, Shaw NT. Future directions in evaluation research: people, organizational, and social issues. Methods Inf Med 2004;43:215–31.

[17] Ammenwerth E, Aarts J, Berghold A, Beuscart-Zephir MC, Brender J, Burkle T, et al. Declaration of Innsbruck. Results from the European Science Foundation Sponsored Workshop on Systematic Evaluation of Health Information Systems (HIS-EVAL). Methods Inf Med 2006;45(Suppl 1):121–3.

[18] Ammenwerth E, Brender J, Nykanen P, Prokosch HU, Rigby M, Talmon J. Visions and strategies to improve evaluation of health information systems. Reflections and lessons based on the HIS-EVAL workshop in Innsbruck. Int J Med Inform 2004;73:479–91.

[19] Despont-Gros C, Mueller H, Lovis C. Evaluating user interactions with clinical information systems: a model based on human–computer interaction models. J Biomed Inform 2005;38:244–55.

[20] Cusack CM, Poon EG, editors. Evaluation toolkit data exchange projects. National Resource Center for Health Information Technology; 2006.