Section III
Organizational Impact and Evaluation

Organizational Impact and Evaluation  .............................................. 103
Cynthia S. Gadd

Chapter 10
The Implementation of Secure Messaging  ............................................. 107
Zhou Yan

Chapter 11
Who Moved My Clinic? Donnelly University Pediatric Rehabilitation: The Wheelchair Clinic  ............................................. 115
Fredrick Hilliard

Chapter 12
OncoOrders: The Early Years  ......................................................... 127
Chris Raggio and Judith W. Dexheimer

Chapter 13
Implementing a Computerized Triage System in the Emergency Department  ......................................................... 135
Scott R. Levin, Daniel J. France, and Dominik Aronsky

Chapter 14
Medication Barcode Scanning: Code “Moo”: Dead COW  .................. 155
Laurie L. Novak and Kathy S. Moss
Organizational Impact and Evaluation

CYNTHIA S. GADD

Organizational Impact and Evaluation

The cases in this section focus on system development and implementation, and may be discussed in those contexts, of course. However, they also present an opportunity to pose the question of how general evaluation principles could be applied to study their often wide-ranging impacts within health care organizations. The following discussion contrasts traditional approaches to evaluation in biomedical informatics, which typically focus on the what questions, with sociotechnical and program evaluation approaches, and which provides more opportunities to explore the how and why questions that inform our understanding of organizational impacts.

The What

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 health information technology (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), even when they are not amenable to the questions to be answered.1

The How and Why

Stoop and Berg 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. They 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.

The When

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 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 developmental stages are defined: Specification, Component development, Combination of components into a system, Integration of system into environment, Routine use; as well as five evaluation levels: Definition, Laboratory bench, Laboratory field, Remote field validity, Remote field efficacy. In their three stages of technology assessment, Fuchs and Garber 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. 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 health information exchanges: 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.

The What Revisited: Evaluation of “Messy” HIT

Berg 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.

Another Why: Evaluation as “Useful Research”

Viewing HIT 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.” (Patton, 1996, p. 23) Program evaluation developed, particularly in the U.S., in the context of the Great Society programs of the 60s and 70s, 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.

Professionalization of evaluation brought standards, foremost of which was that evaluation should be useful, i.e., evaluations should be judged by their utility 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. Cronbach and Suppes described this as the difference between conclusion-oriented and decision-oriented inquiry.

The Who

Utilization-focused evaluation takes program evaluation one step further in that it is “done for and with specific, intended users for specific, intended uses.” 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.)

Each of the cases that follow provides an opportunity to ask the questions necessary to understand the impacts of the HIT being designed and implemented:

- What impacts should be studied?
- How did these impacts happen?
- Why did these impacts happen?
- When should evaluation be done and how does timing affect the choice of methods used?
- Why and for whom do we do evaluation?
- How can we hope to evaluate impacts with the evolving, often messy nature of HIT?

References