Evaluation as a Field

The previous chapter should have succeeded in convincing the reader that evaluation in medical informatics, for all its potential benefits, is difficult in the real world. The informatics community can take some comfort in the fact that it is not alone. Evaluation is difficult in any field of endeavor. Fortunately, many good minds—representing an array of philosophical orientations, methodological perspectives, and domains of application—have explored ways to address these difficulties. Many of the resulting approaches to evaluation have met with substantial success. The resulting range of solutions, the field of evaluation itself, is the focus of this chapter.

If this chapter is successful, the reader will begin to sense some common ground across all evaluation work while simultaneously appreciating the range of tools available. This appreciation is the initial step in recognizing that evaluation, though difficult, is possible.

Evaluation Revisited

For decades, behavioral and social scientists have grappled with the knotty problem of evaluation. As it applies to medical informatics, we can begin to express this problem as the need to answer a basic set of questions. To the inexperienced, these questions might appear deceptively simple.

• An information resource is developed. Is the resource performing as intended? How can it be improved?
• Subsequently, the resource is introduced into a functioning clinical or educational environment. Again, is it performing as intended, and how can it be improved? Does it make any difference in terms of clinical or educational practice? Are the differences it makes beneficial? Are the observed effects those envisioned by the developers or different effects?

Note that we can append "why or why not?" to each of these questions. In actuality, there are many more potentially interesting questions than have been listed here.
Out of this multitude of possible questions comes the first challenge for anyone planning an evaluation: to select the best or most appropriate set of questions to explore a particular situation. This challenge was introduced in Chapter 1 and is reintroduced here. The issue of what can and should be studied is the primary focus of Chapter 3. The questions to study in any particular situation are not inscribed in stone and would probably not be miraculously handed down if one climbed a tall mountain in a thunderstorm. Many more questions can be stated than can be explored; and it is often the case that the most interesting questions reveal their identity only after a study is begun. Further complicating the situation, evaluations are inextricably political. There are legitimate differences of opinion over the relative importance of particular questions. Before any data are collected, those conducting an evaluation may find themselves in the role of referee between competing views and interests as to what should be on the table.

Even when the questions can be stated in advance, with consensus that they are the “right” questions, they can be difficult to answer persuasively. Some would be easy to answer if we possessed a unique kind of time machine which might be called an “evaluation machine.” As shown in Figure 2.1, the evaluation machine would enable us to see how our clinical environment would appear if our resource had never been introduced. By comparing real history with the fabrication created by the evaluation machine, we could potentially draw accurate conclusions about the effects of the resource. Even if we had an evaluation machine, however, it could not solve all our problems. It could not tell us why these effects occurred or how to make the resource better. To obtain this information we would have to communicate directly with many of the actors in our real history to understand how they used the resource and their views of the experience. There is usually more to evaluation than demonstrations of causes and effects.

In part because we do not possess an evaluation machine but also because we need ways to answer additional, important questions for which the machine would be of little help, there can be no single solution to the problem of evaluation. There is, instead, an interdisciplinary field of evaluation with an extensive methodological literature. This literature details many diverse approaches to evaluation, all of which are currently in use. We introduce these approaches later in the chapter. These approaches differ in the kinds of questions that are seen as primary, how specific questions get onto the agenda, and the data collection methods ultimately used to answer the questions. In informatics it is important that such a range of methods is available because the questions of interest can vary dramatically: from the focused and outcome-oriented (Does implementation of this system affect morbidity and/or mortality?) to the practical, and market-oriented questions, such as those frequently stated by Barnett.*

1. Is the system used by real people for real use with real patients?
2. Is the system being paid for with real money?

* These questions were given to the authors in a personal communication on December 8, 1995. A slightly different version of these questions is found in Blum on page 286.