CHAPTER

10

PROGRAM EVALUATION

David H. Gustafson and Farrokh Alemi

Large-scale evaluations of health and social service programs are commonly initiated to help policymakers decide on the future direction of programs. They often examine issues such as the following:

- Approval of the program and enrollment of the patients (e.g., elapsed time from announcement of the funding to enrollment of first patient);
- Correspondence between what was approved and what was done (e.g., the percent of program objectives planned that were actually implemented);
- The demand for the proposed program (e.g., the ratio of patients to the service capacity);
- Description of the patients (e.g., the number and demographic background of patients);
- Provider satisfaction (e.g., measures of conflict among providers and satisfaction with care);
- Satisfaction with the program (e.g., patient satisfaction surveys); and
- Affect on patient outcomes (e.g., measures of patients' mortality, morbidity, or health status);

Not surprisingly, with this kind of interest, program evaluation has become a big business and an important field of study; program evaluations are requested and funded by virtually every department of health and social services, not to mention many legislatures, governors, and city administrations.

The basic concept of evaluating social and healthcare programs, as shown in Figure 10.1, is straightforward. A program is expected to meet certain performance standards (A). The program actually performs at a level (B) that may equal or exceed the standards or may fall short because of flaws or unexpected environmental influences. Actual performance is compared to expected performance (C), and decisions are made about which, if any, of the discrepancies are worrisome (D). The findings are explained and interpreted to decision makers (E); and changes are introduced in either system performance or the expectations of it (F).

This book has a companion web site that features narrated presentations, animated examples, PowerPoint slides, online tools, web links, additional readings, and examples of students' work. To access this chapter's learning tools, go to ache.org/DecisionAnalysis and select Chapter 10.

Although the basic concepts of evaluation are simple, actual implementation can be quite complex, and numerous evaluation techniques and philosophies have been introduced over the years (Alemi 1988). The major approaches are categorized as experimental, case study, and cost-benefit analysis.

Some researchers have advocated an experimental approach, with carefully designed studies using experimental and control groups, random assignment of subjects, and pre- and post tests (Reynolds 1991).

Variations on this experimental theme often remove the random assignment criterion. These "quasi-experimental" designs interject alternative explanations for findings (Campbell and Stanley 1966). Many process improvement efforts can be thought of as quasi-experimental studies (Benedetto 2003; Alemi, Haack, and Nemes 2001). An experimental evaluation is not always necessary, but when it is, random assignment must be an essential element of it.

Another school of thought advocates examining case studies (Barker and Barker 1994; Alemi 1988; Brown 2003), arguing that case studies are superior to experiments because of the difficulty of identifying criteria for experimental evaluation. Further, experiments require random subject assignment and pre- and post tests, both of which are impractical because they interfere with program operation. Case studies use unobtrusive methods to examine a broad range of objectives and procedures that are sensitive to unintentional side effects. This approach helps administrators improve programs instead of just judging them. Some case studies report on services offered and the characteristics of their use, while others are less concerned with the physical world and emphasize the values and goals of the actors. Their reports tend to contain holistic impressions that convey the mood of the program. Any evaluation must contain case studies to help people really understand and act on the conclusions. In the context of continuous quality improvement, the importance of case studies is emphasized by insisting on expressing the problem in the "customer's voice."

Cost-benefit analysis evaluates programs by measuring costs and benefits in monetary terms and calculating the ratio of costs to benefits (Neumann 2005; Weinstein et al. 1996). This ratio is an efficiency statistic

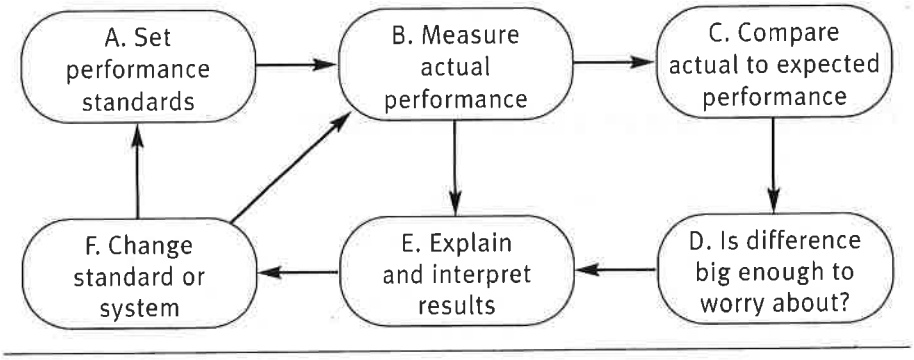

FIGURE 10.1 Schematic Representation of Program Evaluation

showing what is gained for various expenditures. There are many types of cost-benefit analysis. Some analyses assume that the market price of services fairly and accurately measures program benefits; others measure benefits on the basis of opinion surveys. Variations on the cost-benefit theme involve comparisons that do not translate everything into a dollar equivalent. The critical characteristic of those studies is an ability to compare what you get against what it costs to get it.

Many Evaluations Are Ignored

Although program evaluations take a good deal of time and money, their results are often ignored (Hoffmann and Graf von der Schulenburg 2000; Drummond 1994). Even if interesting data are collected and analyzed, evaluations have no effect if their results are not directly relevant and timely to a decision. Often, evaluation reports present a variety of unrelated findings and thus confuse rather than clarify the decision maker's choices. Despite calls for evidence-based practice, many clinicians ignore small and large evaluation studies (Zeitz and McCutcheon 2003).

Evaluation studies with little effect generally began with a poor design. An evaluation can gain influence if the evaluator understands and focuses on the options, values, and uncertainties of the decision makers. To provide the kind of evaluation that supports policy formation, relevance to the decision must be designed at the start, not tacked on at the end. Edwards, Gutentag, and Snapper (1975) wrote:

> Evaluations, we believe, exist (or perhaps only should exist) to facilitate intelligent decision making . . . an evaluation research program will often satisfy curiosity. But if it does no more, if it does not improve the basis for

decisions about the program and its competitors, then it loses its distinctive character as evaluation research and becomes simply research (p. 140).

Decision-Oriented Evaluation Design

Certain design factors can increase the relevance of an evaluation to the actual decision making. The evaluators should do the following:

- *Identify the primary users of the evaluation.* Often, evaluators say their findings are intended for policymakers, not particular individuals. The evaluators do not contact an individual policymaker because they consider such person's views irrelevant or they perceive that such individuals hold their positions temporarily, while the policy issue and the evaluation task remain more or less permanent. It is best to name the decision makers before designing evaluation studies to meet their needs. Asking decision makers is preferred even if these individuals will not be on the job when evaluation results become available because decision makers' information needs are often dictated by their positions, not by their idiosyncratic preferences. Even though decision makers change, their needs remain stable, and identifying the needs of a single decision maker is not a waste of effort because it will help meet the needs of future decision makers.
- *Identify the decision-making needs of decision makers, and provide information to meet their needs.* Gustafson and Thesen (1981) found that once decision makers' information needs were prioritized in order of importance, the top 65 percent of the priorities had nothing to do with the type of questions addressed by typical evaluation data. Decision makers tend to seek help on issues that depend more on values and expectations than on objective statistics. Although program statistics cannot be ignored, program evaluations should provide information that will actually influence decisions.
- *As part of the evaluation, suggest options that creatively address the issues.* Too often, evaluations produce evidence about the strengths and weaknesses of the existing system, touching only briefly on what improvements could be made. An effective evaluation must devote considerable effort to identifying improvements that will remove the weaknesses.
- *Identify and attempt to reduce the most important uncertainties involved in the system being evaluated.* Frequently, a system performs well under some conditions and poorly under others. Attempts to improve system performance are limited by the inability to predict

when true differentiating conditions will arise. Evaluators should identify the differentiating conditions, develop means to predict when they will arise, or propose system improvements that are less susceptible to variations in those conditions.

- *Explain how they reached their results.* Statistical analysis can satisfy detached, rational decision makers, but many policymakers find that examples underscore the justification for conclusions in a more persuasive manner than rational and statistical arguments alone. Such examples can be constructed by careful evaluation of the program. To explain the evaluation results, examples should allow analysts to observe, experience, and describe how the system actually operates. Succinct examples can make an evaluation report "come alive" and help decision makers feel and understand the consequences of their decisions at several levels. The use of examples has three benefits: (1) it permits you to describe how the system actually functions; (2) it permits you to compare actual operation to intended operation; and (3) if done by qualified personnel, it omits the need to assess the adequacy of the system's operation.
- *Examine the sensitivity of the evaluation findings to assess their practical significance.* With a sufficiently large database, almost any difference can be statistically significant. But small differences, even if statistically significant, may not matter. Analysts should conduct a sensitivity analysis to see how erroneous the assumptions can be before they cause the decision makers to act mistakenly. Evaluation studies can include sensitivity analyses but rarely do.
- *Present results more quickly and when they can be most useful.* Too often, decisions must be made before an evaluation is complete. Analysts should present reliable preliminary findings to decision makers as they go along. This shows that they are sensitive to the timing of a decision, particularly in terms of knowing the critical moments in the policy process. Reliable information that could influence policy almost always surfaces during an evaluation and not only at its end. If evaluators know the timing of a decision, they can give input when it is most useful. The decisions will be made anyway, and policymakers will act on whatever information they have.

In summary, program evaluation should be tied to the decision-making process. The remainder of this chapter presents a nine-step strategy for such a decision-oriented evaluation design. This presentation will be clarified by referring to an evaluation of the nursing home quality assurance process conducted by Gustafson and his colleagues (Gustafson, Fiss, and Fryback 1981; Gustafson et al. 1990).

Continued rises in nursing home costs in the United States have stimulated increasing debate about how regulation can improve the industry. Some critics find the government ineffective at evaluating nursing homes (Winzelberg 2003). These critics argue that current surveys of nursing home quality are too frequent, are too intensive, and have little relation to the health status and functional ability of nursing home residents. Gustafson, Fiss, and Fryback (1981) were asked to evaluate the process of surveying nursing home quality (this work was continued in Gustafson et al. 1990), and this chapter uses their experience to illustrate how a decision-oriented evaluation is done.

Step 1: Identify the Decision Makers

The first step in planning an evaluation is to examine the potential users and to invite them to devise the plan of action. In the nursing home example, three groups were expected to use the evaluation results: (1) the state government, to decide what program to implement; (2) the federal government, to decide whether to support the state's decision and whether to transfer aspects of the project to other states; and (3) several lobbying groups (nursing home associations), to choose their positions on the topic. The evaluators identified individuals from each group, asked them to collaborate with the evaluation team, and kept them informed of progress and preliminary conclusions throughout the study.

Step 2: Examine Concerns and Assumptions

Next, the evaluators talked to the chosen decision makers (the program administrator and experts on the program, for example) to determine their concerns and assumptions and to identify the potential strengths, weakness, and intended operations of the program.

In the nursing home example, a decision maker who was concerned about the paperwork burden for quality assurance deemed the effort excessive and wasteful. A second decision maker was concerned with the cost of the quality assurance process and worried that it would divert money from resident care. A third was more concerned that quality assurance funds be distributed in a way that helped not only to identify problems but also to facilitate solutions. This person preferred to find solutions that would bring nursing homes into compliance with regulations. A fourth decision maker felt that the quality assurance process should attend not only to clients'

medical needs but also to their psychological and social needs. All of these divergent objectives were important because they suggested where to look while designing a quality assurance process to address the decision makers' real needs.

The evaluators also helped identify and clarify each decision maker's assumptions. These assumptions are important because, regardless of accuracy, they can influence decisions if not challenged. One decision maker believed the state must play a policing role in quality assurance by identifying problems and penalizing offending homes. Another person believed the state should adopt the role of change agent and take any necessary steps to raise the quality of care, even if it had to pay a home to solve its problems. Arguments for and against these philosophies about the role of government were examined, and although new data on these issues were not collected, the final report reviewed others' research on the matter.

Step 3: Add Your Observations

Another important method of examining a program is to use one's own observations. The perceptions of decision makers, while very useful for examining problems in detail, do not prove that problems exist, only that they are perceived to exist. Thus, it is important to examine reports of problems to see that they are, indeed, real problems. Members of the evaluation team should watch the system from beginning to end to create a picture of its functioning. A system analyst should literally follow the quality assurance team through a nursing home and draw a flowchart of the process.

Although observational studies are not statistically valid, they can add substantial explanatory power and credibility to an evaluation and allow you to explain failure and suggest improvements. A valuable side effect of such observations is that you will gather stories describing specific successes and failures. These stories have powerful explanatory value, often more than the statistical conclusions of the evaluation. The observations not only suggest how and where to modify the program, but they also indicate areas that should be targeted for empirical data collection.

Step 4: Conduct a Mock Evaluation

The next step is performing a mock evaluation, which is a field test to refine the evaluation protocol and increase efficiency. The mock evaluation keeps the decision maker informed and involved. Too often, decision makers first

see the results of the evaluation when reading the final report. While this sequence probably allows enough time to produce a fine product, time alone guarantees neither quality nor relevance. It is preferable to inform the decision maker about the findings as the project proceeds, because information being gathered could influence the decision. Decision makers will want access to this information. The mock evaluation lets the decision maker determine which areas require more emphasis, allowing you to alter your approach while you have time.

A mock evaluation is similar to a real one except that experts' opinions replace much of the data. This "make-believe" evaluation helps estimate how much money and time are needed to complete the evaluation. It also changes the data collection procedures, sample-size requirements (because a more realistic estimate of variance in the data is gained), and analysis procedures. Finally, the mock evaluation gives a preview of likely conclusions, which allows decision makers to tell whether the projected report will address the vital issues as well as identify weaknesses in the methodology that can still be corrected.

Critics of such previews wonder about the ethics of presenting findings that may be proven wrong by careful subsequent observation. But supporters counter by questioning the ethics of withholding information that could inform policy. These questions represent two extreme positions on a difficult issue. It is true that preliminary results may receive more credibility than they deserve. Moreover, decision makers may press to alter the evaluation design to prevent reaching embarrassing conclusions. Those dangers may be outweighed by the alternatives of producing irrelevant data, missing critical questions, or failing to contribute valuable information when it could help the policy debate.

In the nursing home example, after the decision makers had read the mock report, they were asked to speculate about how its findings might affect their actions. As the evaluation team described its preliminary findings, the decision makers explained their possible courses of action and listed other information that would increase the evaluation's utility.

It is important to make sure that evaluation findings lead to action. Decision makers can react in many ways to various findings. Some consider negative findings a sufficient basis for changing their opinions and modifying the system, while others continue to adhere to existing opinions. If your findings do not motivate the decision makers to change the system, this is a signal that you could be collecting the wrong data. At this point, you can decide to collect different data or analyze it more appropriately. The goal remains to provide information that really influences the decision maker's choices.

In the nursing home study, the evaluation team observed several nursing home surveys, talked with interested parties, developed a flowchart of the process, and then asked the group to consider what they would do differently if the evaluation suggested that current efforts were indeed effective. The question was then repeated for negative findings on various dimensions. The discussion revealed that the experts, like others in the field, believed existing quality assurance efforts were inefficient and ineffective, and these people expected the evaluation to confirm their intuition. They thought evaluation findings would make a difference in the course of action they would follow. In other words, they were certain about the effectiveness of the current system but uncertain about how to improve it. This is an important distinction, because an evaluation study that only gauged the effectiveness of the current system would confirm their suspicions but not help them act. What they needed was a study that would pave the way for change, not just to criticize a system that was clearly failing.

At this point, the evaluation team and its advisory group developed an alternative method of nursing home quality assurance that helped reallocate resources by focusing on influencing the few problematic homes, not the majority of adequate ones. A brief nursing home survey was designed to identify a problem home and target it for more intensive examination. Then, an evaluation was designed to contrast this alternative approach to the existing method of evaluation. Thus, the mock evaluation led to the creation of an alternative system for improving nursing home quality; instead of just evaluating quality of the current system, the team compared and contrasted two evaluation systems.

The mock evaluation is a preview that helps the decision makers see what information the evaluation will provide and suggest improvements that could be made in the design. "Showing off" the evaluation also makes the decision makers more likely to delay their decision making until the final report is complete.

Step 5: Pick a Focus

Focus is vital. In the planning stage, discussions with decision makers usually expand the scope of the upcoming evaluation, but fixed resources force you to choose which decision makers' uncertainties to address and how to do so. For example, further examination of the potential effect of the evaluation of nursing home quality assurance revealed a sequence of decisions that affected whether evaluation findings would lead to action. The state policymakers were responsible for deciding whether to adopt the proposed

changes in quality assurance. This decision needed the approval of federal decision makers, who relied on the opinions of several experts as well as on the evaluation. Both state and federal decisions to modify the quality assurance method depended on a number of factors, including public pressure to balance the budget, demand for more nursing home services, the mood of Congress toward deregulation, and the positions of the nursing home industry and various interest groups. Each of these factors could have been included in the effort, but only a few were selected because of budget constraints.

Some factors in the decision-making process may be beyond the expertise of the evaluation team. For example, the evaluators might not be qualified to assess the mood of Congress. Although the evaluation need not provide data on all important aspects of the decision process, the choice not to provide data must be made consciously. Thus, the analyst must identify early in the process which components to include and which to exclude as a conscious and informed part of evaluation planning.

Sloppy analyses are not being advocated here. Rather, evaluations often operate on limited budgets and thus must allocate resources to produce the best product without "breaking the budget." This means that specificity in some areas must be sacrificed to gain greater detail elsewhere.

Step 6: Identify Criteria

Now the evaluation team and decision makers set the evaluation criteria, based on program objectives and proposed strengths and weakness of the program. (See Chapter 2 for a discussion of how to identify evaluation criteria, or attributes; see Chapter 6 to see how the analysis can be done using a group of decision makers.)

The nursing home evaluation focused on a number of questions, one of which was the difference between the existing method of quality assurance and the alternative method. The following criteria were used to evaluate this issue:

- *Relation to regulatory action.* The quality assurance effort should lead to consistent regulatory actions.
- *Ease of use.* Administering quality assurance should interfere with delivering nursing home care no more than necessary.
- *Reliability.* The quality assurance effort should produce consistent findings, no matter who does the reviews.

- *Validity.* The findings should correlate with adverse outcomes of nursing home care (such as an increasing rate of deterioration in residents' ability to attend to their daily activities).
- *Influence.* Quality assurance should change the way long-term care is delivered.
- *Cost.* The cost of conducting quality assurance must be measured to allow the selection of the most cost-effective method.

In the nursing home example, an evaluation design was created to divide the state into three regions. In the lower half of the state (and the most populous), nursing homes were randomly assigned to control and experimental conditions, after ensuring that an equal number of proprietary nursing homes and nonprofit nursing homes, of similar sizes and treating similar patients, would be placed in each group. The northern half of the state was divided into two regions, one receiving the new regulatory method and one not. This was done to observe how the management of the regulatory process would change the results. Such random assignment greatly increased the credibility of the evaluation.

A second aspect of design was the measures used. Previously, a nursing home's quality was judged on the basis of the number of conditions, standards, and elements found out of compliance. However, it was apparent that radical differences in the severity of violations could take place within a level (e.g., element). It was decided to convene a panel of experts to numerically rate the severity of different violations.

Step 7: Set Expectations

Once the evaluation design was completed, decision makers were asked to predict the evaluation findings and express what they expect to find. This request accomplished two things. First, it identified the decision makers' biases so the team could design an evaluation that responded to them. Second, it gave a basis for comparing evaluation findings to the decision maker's expectations, without attributing them to specific people. The effect of evaluation results is often diluted by hindsight. Reviewers might respond to the results by saying, "That is what I would have expected, but. . . ." Documenting expectations in advance prevents such dilution.

In the nursing home example, decision makers expected that the alternative method would be slightly better than the current method, but they were surprised at how much better it performed. There were substantial cost savings as well as improvements in effectiveness. Because their

expectations had been documented, their reactions were more akin to "Aha!" than to "That's what we expected." This helped create a momentum for change.

Step 8: Compare Actual and Expected Performance

In this phase, data are collected to compare actual and expected performance. The observed findings are compared to the decision maker's expected findings. (For more information on data collection and statistical comparison, consult the many books on evaluation that cover these topics in detail.) There are a variety of ways this can be done. One common way is to replace actual and expected performance comparisons with a comparison of the control (or currently operating) and experimental (new) methods.

Another way is to use statistical process control tools. First, a control chart is created, showing progression of time in the x-axis. Three lines are plotted: One line shows the actual observed performances of the program, and the other two lines show the upper and lower control limits based on the expected performance of the program. When the observed rate is outside the control limits, a statically significant change in the program is signified.

Step 9: Examine Sensitivity of Actions to Findings

Sensitivity analysis allows you to examine the practical effect of your findings. In this step, decision makers were asked to describe the various courses of action they would have taken if they had received specific evaluation findings. Then they were asked to consider what they would have done upon receiving different findings. Once the threshold above which the actions would change was identified, the probability that the findings could contain errors large enough to cause a mistake was calculated. Using the nursing home example, the evaluation might have revealed that one method of quality assurance was 5 percent more expensive than another. Decision makers might determine that savings of 20 percent would induce them to change their decision. In this case, the decision makers would be asked to identify a threshold, say 15 percent, above which they would change their decision. The evaluation team would then calculate the probability that the findings contained an error large enough to exceed the threshold. In other words, the team would then state the chance that a reported 5 percent difference in cost is indeed a 15 percent difference in cost.

Sensitivity analysis allows decision makers to modify their confidence in the evaluation findings. If the findings suggest that the reported practical differences are real and not the result of chance, then confidence increases. Otherwise, it decreases.

Summary

This chapter has provided a step-by-step approach to using decision analysis in program evaluation. The primary uses for the evaluation are identified. The decision makers' assumptions are identified, and information is provided to meet the need. A particular focus is selected, data are gathered, and the program's performance is compared to expected outcomes. The sensitivity of the conclusions are examined, and new options are proposed. A decision analytic design engages decision makers throughout the effort and actively focuses on data that are likely to lead to action. One advantage of the proposed approach is the extensive involvement of decision makers in various components of the evaluation.

Review What You Know

1. Is it necessary to identify one or more decision makers before conducting a program evaluation as described in this chapter? Why or why not?
2. Why should a program evaluation focus on decisions? Why not focus on cost or improvements in access to care or some other feature?
3. What is the point of doing mock evaluations? What are the advantages?
4. How does a decision analytic evaluation select criteria used for evaluation?
5. What are the steps in a decision analytic evaluation?
6. What is the point of doing a sensitivity analysis in a decision analytic evaluation?
7. Should an evaluator speed up the preparation of a report? Why or why not?
8. Why should an evaluator collect information about a policymaker's expectations before presenting evaluation reports?

Rapid-Analysis Exercises

This exercise is designed to help you decide whether you should evaluate a service and how you should do so. For many of the questions, there is

no right or wrong answer. Just provide an answer that fits your preferences. To proceed, you must have a specific service in mind, either a service where you work or a service where you are a customer. Assume that you are the manager of this service. The purpose of the activity is to help you become more aware of your own thoughts and reservations about conducting an evaluation of services.

This exercise is in three parts. In the first part, you will think about the need for program evaluation. In part two, you will design a way of evaluating your customers' satisfaction with your service. In the last part, you will contrast your design with what was covered in the chapter and, in this manner, gain insights about some of the ideas expressed in the chapter. Before proceeding, identify the particular service you will evaluate.

Part 1

Evaluation takes time and money. To maintain independence, an evaluation of the effect of your service is best done by independent groups of investigators. Other evaluations (e.g., market studies, studies of patient satisfaction with your service) can be done in-house. No matter who does the evaluation, it still requires much planning, data collection, analysis, and reporting. These activities could compete with your ability to focus your funds and time on organizing and improving the service. Sometimes evaluations are not done because the effects of the services are known or can be deduced from the experiences of others. The following questions will help you understand why you may be ambivalent about conducting an evaluation of your services:

1. What reservations do you have about conducting an evaluation of your service?
2. Is there sufficient evidence in the literature or in the industry to suggest that what you are doing will work well for patients of different backgrounds?

If you think about it, evaluation could have many benefits. It could tell you and your clients that the service is effective in changing lives. It can tell you how to improve, and everyone—even the best among us—needs to improve. It can help you convince third-party payers to pay for your service. In this section, you are asked a number of questions about the implications of not evaluating the service.

3. Would it help your efforts to market your service if you knew more about the people who are currently using your service and how they feel about it?

4. Describe how a survey of your clients could help you improve the service or improve the way it is marketed.
5. Would patients who use your service ask for the evaluation of the service (particularly when there is news coverage of unique events), or have your patients already evaluated your reputation by the time they come to the service?
6. Would your efforts to market the service to third-party payers be hurt if you do not have data that the service works (e.g., that it saves money, improves access, or improves quality of health services)?
7. Inside your organization, would your career be affected if you do not have data to show that what you did was reasonable? In other words, do you have a lot of support from different managers so long as the service makes money, or is the utility of the service already being questioned?
8. What might go wrong if you fail to evaluate the service? Think through the next six months to the next two years. Describe a situation where you would say "Oops, I wished I had evaluated the effect of our service."
9. Think harder. Is there some opportunity missed or negative consequences that may happen to you or to your organization as a consequence of failing to evaluate your service?

The following questions are designed to help you determine whether you should evaluate your service, what type of evaluation would be most useful, and who should conduct it:

10. Now that you have gotten this far, how do you feel about the evaluation and the type of questions that it should address?
    - There is no need for an evaluation.
    - Evaluate market, including characteristics and sizes of patient groups attracted.
    - Evaluate patients' satisfaction with service.
    - Evaluate effect of service on patients' health status and lifestyles.
    - Do other types of evaluations.
11. Do you feel that you have the sufficient training or experience to conduct the evaluation yourself, or would you like help on how to evaluate your service?

Part 2

This part is intended to help you evaluate the satisfaction with your program services. As before, you must have a specific service in mind before proceeding.

Think through the goal of the evaluation survey. Sometimes, purchasers are interested in repeated evaluation efforts that not only document problems but also show a systematic effort to resolve them. You may also engage in an evaluation to help you find and fix problems. In this case, you are not so much interested in reporting the problems you find but in fixing them and going on. There are many other reasons too.

1. What is your real reason for wanting to know if consumers are satisfied with your service?

One of the first issues you need to think through is who should do the work. Sometimes it is useful to engage a third party, which will help convince purchasers and other reviewers of the data that independent evaluations were done. If you evaluate your own efforts, there is always a suspicion that you may not report the whole story. In addition, some third-party surveyors can benchmark your service against your competitors. For example, they can report that your service is among the top 5 percent of all services. Purchasers and consumers like benchmarked data. At the same time, asking others to help you evaluate is time consuming and expensive, and it may interfere with keeping your activities secret until it is publicly released.

2. Given these issues, who do you think should evaluate your service? Why?

Next, you need to determine the type of evaluation and how often it should be conducted. This depends in part on what questions you want answered. If, for example, you want to know which target group (e.g., people of certain age or gender) is most satisfied with your service, then an occasional cross-section analysis is sufficient. In a cross-section analysis, you would survey the patients after they have been exposed to your service. Cross-section analysis can also be used to benchmark your service against other services. If you plan to regularly survey your clients over time, then you need a longitudinal study. These types of studies are best if you want to know whether exposure to your service changed the level of satisfaction patients have over time.

3. Do you think you may need to conduct a longitudinal or a cross-sectional study?
4. How often do you want to evaluate the satisfaction with your service?

Satisfaction surveys can be misleading in many ways. One possibility is that improvement in satisfaction may be related to other events and not to your service. For example, patients' lifestyle adjustments may change their satisfaction with your service. To control for this type of error, it is important to contrast the improvement against a control group

that has been exposed to another service. Another source of error could be that, over time, respondents are learning the system more and thus are more satisfied with the services they are using. Dissatisfied individuals are unlikely to use your service. Surveying only users of your services may mislead you by painting a rosy picture of clients' satisfaction. To control for these types of errors, it is important to contrast your services with others and to explicitly look for customers who are not repeat users.

5. What steps will you take to ensure that changes in patient care mix are not misleading you?

Other sources of errors are also possible. Campbell and Stanley (1966) highlight a list of common errors in survey research. Given the various sources of error, you need to choose for yourself how important it is to have an accurate picture of clients' satisfaction with your service. At the extreme, you can randomly assign people to two services: yours and an alternative placebo service. Random assignments control for most types of errors. But random assignment is expensive and, in some occasions, blind assignment may be unethical. Subjects have to volunteer to be assigned to the experimental or a placebo service, and subjects may refuse to participate. Another approach is repeated evaluation of satisfaction over time. It provides a time series of data that control for some errors. Because subjects have self-selected to be part of these studies, the result may be rosier than if all subjects were included. The least accurate analysis is to do studies without any comparison group.

6. Should your analysis consist of randomly assigned control groups, repeated evaluations over time, or no control group?

Obviously, you will survey the people who received your service, but you can also survey others to serve as a comparison group for you. If you do so, you would need to determine which type of comparison group you will include. If you need a preponderance of evidence, track a control group over time and use it as your comparison group. If you need data that are beyond reasonable doubt, then choose a control group that is randomly assigned.

7. Will you survey others who have not used your service as a comparison group?
8. If so, will you track a control group over time or use a group that is randomly assigned?

Next, you need to determine what you want to ask. Some of the items in satisfaction surveys include the following:

- Overall satisfaction with quality of the services
- Ease of use of the services
- Satisfaction with the integration of the services with other health services
- Accuracy, comprehensiveness, usefulness, and timeliness of the information and services received
- Comfort received
- Skills gained from cognitive services and support groups

You do not need to include all of the above items, nor do you need to limit your surveys to above items. There are many data banks of surveys. Keep in mind that standardized surveys allow you to benchmark your data against others. In contrast, doing your own survey helps you focus on patients' reactions to innovations in your effort. You can tailor your own surveys to fix your needs and therefore get more for the effort you are putting in.

9. What questions are you planning to ask?

It is neither necessary nor reasonable to survey all patients who use your service; instead, you can sample. *Sampling* helps reduce the data collection burden on both the patients and the analyst. The size of the sample depends on what you are trying to conclude. If there is a lot of variability in patients' satisfaction with your service, you need larger samples. If you plan to compare your service with others and the two efforts are very similar, you need larger data. More important than the size of the survey is whether it represents the population. Getting many patients to respond does not correct for the lack of a representative sample. This is one case in which more is not always better. The point of sampling is to get a representative sample of people who receive your service. Small and large samples can both be representative. The key is to examine whether there are systematic differences among people who respond and those who do not. Some examples of nonrepresentative designs include the following:

- *Surveying only those who complete your service.* Most dissatisfied patients will abandon the service before reaching the end.
- *Surveying patients in a particular month.* Patients' preferences and types of illness may be seasonally different.

You should randomly select a percentage of patients (not to be mistaken with randomly assigning patients to the service, which is a much harder task). This gives an equal chance that any particular patient may be included. In some circumstances, you may wish to over sample segments of the population. When segments of your population are small, you need to over

sample these segments so that you can obtain an accurate estimate of their satisfaction. Otherwise, too few of them will be in your sample to provide an accurate picture. Suppose few teenagers visit your service. If you want to know about their satisfaction with your service, you will need to over sample teenagers. Thus, you may sample every ten adults but every five teenagers. Over sampling helps get a more accurate picture of small subgroups of patients using your service.

10. Which sampling strategy do you wish to implement?
11. Why do you expect that satisfied or dissatisfied clients will be reached in this fashion?
12. How you plan to verify whether your sample represents the population to which you want to generalize?

Many choices are available for data collection. You can have the survey done online and automatically. In these types of surveys, a computer calls or sends an e-mail to your clients. You can also survey participants by mail, by telephone, or in person. The mode of conducting the survey may affect the results. Online, computerized telephone, and mailed surveys are self-administered. Patients are more likely to report deviant social behavior in self-administered surveys. Online surveys (if connected to an automatic reminder) have a larger response rate than offline surveys. Online surveys are less expensive than offline surveys. Among offline surveys, face-to-face interviews are most expensive but allow for longer interviews.

13. Given the trade-offs of different modes of surveys (e.g., online, face-to-face, mailed, or telephone) which is your preferred approach, and why?

How you conduct your survey will have a lot to do with its success. You should alert the respondents that you plan to survey them before you actually send them a survey. This is preemptive reminder for people who forget to respond. In this fashion, the respondents will hear about you at least four times, when you (1) invite them to participate in the survey, (2) alert them that the survey is upcoming, (3) send them the survey, and (4) remind nonrespondents to complete the survey.

The invitation to participate in a survey should highlight who is conducting the survey, what the goal of the survey is, how long it should take to complete the survey, and how nonresponse will affect the overall conclusions.

14. What will you say in your invitation to participate?

The alert of the upcoming survey should include an appreciation of respondent's willingness to participate, the day the survey will be sent, and the importance of timely responses.

15. What will you say in the alert of the upcoming survey?

A reminder to nonrespondents often includes another copy of the survey. With online surveys, it is often necessary to make sure that respondents can answer questions quickly and without much download time. In fact, if you are tracing the person through e-mail, then it is best to embed the survey in the e-mail. With mailed surveys, you should include a self-addressed, stamped envelope.

16. What will you say in your reminder to participate?
17. What method will you use to send the second copy of the survey?

No matter how you do the survey, you should provide a real benefit for the respondent in completing the survey. Altruism and voluntary requests get you far, but not far enough. Think through what questions you can add that will make the respondent feel happier and more cared for at the end of the survey. Many providers combine satisfaction surveys with surveys of patients' health status or lifestyle appraisals. Patients get the benefit of a free health appraisal and evaluation while they complete the satisfaction surveys.

18. Why should your potential respondents take time away from their busy schedules to answer your questionnaire? In other words, what will they gain from completing your survey?

The language you use for the survey is important. Keep in mind that your services are open to many people from different backgrounds, and people are more likely to respond to a questionnaire prepared in their native language.

19. What language will you use for the survey?

Before you can analyze the data, you need to code the data (i.e., assign numbers to responses). When coding the data, you should include different codes for

- no responses to any questions in the survey;
- skipped questions;
- unclear or unusable responses; and
- responses of N/A (not applicable).

Analyze the missing data codes first. If a large percent of responses is missing, then it is doubtful that you can use the survey to arrive at any conclusions. If certain patient groups tend to skip specific questions, then the survey might have a systematic bias.

20. What response rate do you think is adequate to consider the survey useful?
21. What types of patterns might indicate a systematic bias?

If you are conducting online data collection, then there is no need to spend time entering the data into the computer. If you have done mailed, telephone, or in-person surveys, you must enter the data into the computer. This is often a tedious process. To make sure that the data are correct, enter the data twice and verify the difference between the two data sets.

Another step taken in cleaning the data is to check for inconsistent or out-of-range responses. If responses 1 through 5 are expected, but response 7 is given, then the response 7 is considered out of range and erroneous. Similarly, if earlier in the survey the client indicated that he is male and later that he is pregnant, then an inconsistent response is detected. Spend time cleaning the data. It will help you when it comes to interpreting the results.

22. How would you prepare the data for analysis?

To analyze the data, begin with descriptive statistics. Check each variable for skewness, range, mode, and mean to determine if the responses seem reasonable. Next, plot the data. Satisfaction surveys usually have a number of scales. Each scale is the average of responses to a number of questions. The idea is that if there are different ways of asking the same question, then one may have a more reliable scale. If so, then data may have a normal distribution. Scale responses should look like an upside down "U" shape. Statistical theory suggests that averages of more than four numbers will tend to have a normal distribution.

23. What do you expect to be the distribution of your data?

After you have completed your descriptive data analysis and everything makes sense to you, then you should conduct other analyses. If you have a cross-sectional design, then you may use cross-tabulation to display the data and use statistical methods (e.g., a chi-square test) to examine the significance of the differences you observe in a table.

24. How do you plan to analyze the data?
25. What will the final figures and tables look like (prepare mock-ups with hypothetical data)?

Part 3

Planning for program evaluation requires you to think through your needs and activities, which you did in parts 1 and 2. Now, explain how your plans are different from the steps described in this chapter.

1. What is missing from your plans that, according to this chapter, is important in decision analytic approaches to program evaluation? List all of the steps described in this chapter and, for each step, describe if it has been addressed in part 1 and 2 of the assignment.
2. Does the decision analysis approach give you a perspective that might otherwise be missing from a program evaluation (e.g., sensitivity to the timing of decisions, the need to focus on a decision maker, the need to focus on a specific decision with various options, or the need for a sensitivity analysis)?
3. How could you do a mock evaluation as part of your efforts to survey client satisfaction?
4. How could you make sure that decision makers' expectations are assessed before your data are reported?
5. If the decision maker is under time pressure, what steps will you take to make sure that your findings are available in time or that the decision maker is aware of what you are working on so she may wait for the results.

Audio/Visual Chapter Aids

To help you understand the concepts of program evaluation, visit this book's companion web site at ache.org/DecisionAnalysis, go to Chapter 10, and view the audio/visual chapter aids.

References

Alemi, F. 1988. "Subjective and Objective Methods of Program Evaluation." *Evaluation Review* 11 (6): 765–74.

Alemi, F., M. R. Haack, and S. Nemes. 2001. "Continuous Improvement Evaluation: A Framework for Multisite Evaluation Studies." *Journal for Healthcare Quality* 23 (3): 26–33.

Barker, S. B., and R. T. Barker. 1994. "Managing Change in an Interdisciplinary Inpatient Unit: An Action Research Approach." *Journal of Mental Health Administration* 21 (1): 80–91.

Benedetto, A. R. 2003. "Six Sigma: Not for the Faint of Heart." *Radiology Management* 25 (2): 40–53.

Brown, P. 2003. "Qualitative Methods in Environmental Health Research." *Environmental Health Perspectives* 111 (14):1789–98.

Campbell, D. T., and J. C. Stanley. 1966. *Experimental and Quasi Experimental Design for Research.* Chicago: Rand McNally.

Drummond, M. 1994. "Evaluation of Health Technology: Economic Issues for Health Policy and Policy Issues for Economic Appraisal." *Social Science and Medicine* 38 (12): 1593–600.

Edwards, W., M. Gutentag, and K. Snapper. 1975. "A Decision-Theoretic Approach to Evaluation Research." In *Handbook of Evaluation Research*, edited by E. L. Streuning and W. Gutentag. London: Sage Publications.

Gustafson, D. H., C. J. Fiss, Jr., and J. C. Fryback. 1981. "Quality of Care in Nursing Homes: New Wisconsin Evaluation System." *Journal of Long Term Care Administration* 9 (2): 40–55.

Gustafson, D. H., F. C. Sainfort, R. Van Konigsveld, and D. R. Zimmerman. 1990. "The Quality Assessment Index (QAI) for Measuring Nursing Home Quality." *Health Services Research* 25 (1 Part 1): 97–127.

Gustafson, D. H., and A. Thesen. 1981. "Are Traditional Information Systems Adequate for Policy Makers?" *Health Care Management Review* 6 (1): 51–63.

Hoffmann, C., and J. M. Graf von der Schulenburg. 2000. "The Influence of Economic Evaluation Studies on Decision Making: A European Survey. The EUROMET Group." *Health Policy* 52 (3): 179–92.

Neumann, P. J. 2005. *Using Cost-Effectiveness Analysis to Improve Health Care: Opportunities and Barriers.* New York: Oxford University Press.

Reynolds, J. 1991. "A Reconsideration of Operations Research Experimental Designs." *Progress in Clinical and Biological Research* 371:377–94.

Winzelberg, G. S. 2003. "The Quest for Nursing Home Quality: Learning History's Lessons." *Archives of Internal Medicine* 163 (21): 2552–6.

Weinstein, M. C., J. E. Siegel, M. R. Gold, M. S. Kamlet, and L. B. Russell. 1996. "Recommendations of the Panel on Cost-Effectiveness in Health and Medicine." *JAMA* 276 (15): 1253–8.

Zeitz, K., and H. McCutcheon. 2003. "Evidence-Based Practice: To Be or Not to Be, This Is the Question!" *International Journal of Nursing Practice* 9 (5): 272–9.