

Regardless of how and by whom the results of the intervention will be used, the evaluation plan should be developed concurrent to the intervention. Making the evaluation an integral part of the overall intervention will avoid a common pitfall of lacking **baseline data** for comparison. It will allow feedback of information during the course of the implementation and encourage all team members to become familiar with the evaluation goals. In addition, thinking about evaluation at the same time that the intervention is being designed allows any evaluation expenses that will be incurred to be included in the proposed budget. Most funding agencies today require some kind of accountability of their funded projects and allow researchers to spend 5 to 10 percent of the overall budget on evaluation items.

PREPARING FOR AN EVALUATION

It is important to involve the advisory group and other stakeholders in the evaluation discussion and design as soon as planning the evaluation starts. This may take some time and effort because people who have never been engaged in research may not be acquainted with the rules of scientific objectivity. The rules of data collection and how an evaluation will increase the credibility of the outcomes will require some explanation to these groups. Some basic concepts of the research design and the importance of protection of human subjects will also require discussion. All this groundwork will increase the likelihood that the stakeholders will support an evaluation of the intervention.

The research team must develop an evaluation plan with a budget and timeline to enable the process to move forward in a systematic manner. Examples of two timelines for one-year interventions at the institution and the policy level appear in the next section.

DESIGNING THE EVALUATION

Four steps are required in developing an evaluation design. Each of these is described in turn in this section.

Step 1: Setting a Timeline

A **Gantt chart** is a type of timeline that details the anticipated start and stop dates for evaluation in visual form. It can be shown in units of weeks, months, or years. A sample timeline for an evaluation of a one-year intervention using dance classes as a form of exercise in a senior center to increase stamina and mobility appears in Table 10.1.

Step 2: Establishing Which Objectives and Outcomes to Evaluate

As a next step, list objectives and specific aims or outcomes of the intervention that will be evaluated. The goal of an intervention pertains to the problem that will be addressed, such as decreasing adolescent pregnancies, reducing cigarette smoking, or

improving cardiovascular health. The objectives refer to the strategies that will be used to address the problem, such as increasing opportunities for after-school activities for girls (to address the problem or overall aim of reducing adolescent pregnancy), implementing tobacco-cessation interventions in community libraries (to address the problem or overall aim of reducing cigarette smoking), or organizing a church-based walking intervention for women (to address the problem of women's declining cardiovascular health). Recall from Chapter Eight that specific goals can vary depending on the ecological focus of the intervention (whether at the group, institution, community, or policy level). In reviewing the specific objectives of the intervention, consider the results of the community assessment (Chapter Seven), feasibility of the intervention, access to the target population, and available resources. All these aspects should be considered when you are deciding which aspect of the intervention to evaluate.

Step 3: Determining the Appropriate Type of Evaluation

A process evaluation assesses the progress of the project and how well the intervention is being implemented. Process data can be collected from recruitment logs, participant records, the minutes of project staff meetings, and individual or group interviews with staff members, stakeholders, and members of the target population. Observational techniques can provide information about how the activities are being implemented and the degree of **fidelity** to the intervention protocol and design.

An impact evaluation investigates whether the intervention achieved its objectives. This type of evaluation requires careful selection of the data to be collected and the indicators to be used for data analysis. It is likely that **quantitative methods** will be used for this type of evaluation, which necessitates identifying instruments such as questionnaires that will measure the relevant constructs. Once collected, responses can be easily downloaded into a data software program (such as Stata, SPSS, or SAS) and analyzed for **statistical significance**.

Qualitative methods can be used for impact evaluation. These include strategies such as participant observation, focus groups, and **semistructured interviews**, in which text data are analyzed for recurrent themes. **Triangulation of methods**—using both quantitative and qualitative methods to examine different facets of the intervention—is also possible. The type of methodology selected depends on what is being measured and the resources and time available for measurement.

The strongest quantitative method for evaluating whether an intervention has achieved its goals is a **randomized-controlled trial**. It is considered the gold standard for experimental research because it is the least likely to produce biased results. However, few community-based interventions can use this method because of the amount of resources needed. In a randomized-controlled trial, both an experimental unit and a control unit of analysis must be identified and randomly assigned. The **experimental group** receives the intervention, and the **control group** is either left alone (except for data collection) or given some kind of **placebo** treatment, which

should have no effect. This design is expensive because it involves twice as many subjects or organizations, depending on the unit of analysis, so that the two groups can be compared to determine whether the intervention had an effect. Most of the cost of this research design depends on the required sample size, the **unit of analysis**, and the efforts necessary for follow-up.

Quasi-experimental designs are more realistic but still strong quantitative methods for evaluating community interventions (Shadish, Cook, & Campbell, 2002). The strongest form of quasi-experimental design is one in which the evaluator selects a comparison group that is similar in many respects to the intervention group. No randomization is involved, and baseline data are collected from both groups prior to the start of any intervention activities. This is called a **two-group/pre-post design**. A **one-group/pre-post design** is a weaker design that compares the intervention group to itself at baseline. Both these designs are improved by collecting data again a few weeks or months after the conclusion of the intervention, to establish two postintervention data collection points. The second data collection point demonstrates whether changes that were seen as a result of the intervention are sustained over at least the brief period of time between the end of the intervention and the third data collection. These designs are referred to as two-group/pre-post-post and one-group/pre-post-post designs. Many good books detailing research designs are available, and it is advisable to consult one while planning the evaluation (Shadish, Cook, & Campbell, 2002).

An outcome evaluation determines whether the overall goals of the intervention were met. In reality, the ultimate outcome of most community-based health interventions cannot be assessed because of the limited time and resources available or because of limited statistical power. For example, if the overall goal of an intervention implemented in one of two churches is to improve the cardiac health of women through increased levels of physical activity, the impact evaluation may assess changes in women's cardiovascular mortality rates. Detecting such a change, however, will be very difficult considering the amount of time that would be needed to examine this outcome. In addition, other extraneous factors—such as improvements in medical technology and diagnosis—can also contribute to a decrease in mortality rates, making it difficult to determine how much of the change is ultimately due to the intervention.

Step 4: Selecting Reliable Indicators

Indicators to measure the success of the intervention must be selected for each specific objective. Several factors go into the selection of indicators or variables to measure intervention success, including the availability of data, the unit of analysis, and the **reliability** of the indicator.

The availability of measures and the difficulty in collecting them should be thought through in this selection process. For example, in an intervention with the overall aim of decreasing adolescent pregnancies and the specific objective of implementing a

communitywide after-school sports intervention, it may be feasible to follow a cohort of participants using a **prospective cohort design**. If this is the case, data could be collected on the number of pregnancies and perhaps compared to a similar population not exposed to the intervention. In small-scale or pilot projects, however, such a design is seldom possible. More common is a one-year intervention where it will not be possible to use the communitywide adolescent rate as an indicator because there is a time lag in the availability of these rates. If the intervention is being implemented at the level of only one school district or in one or two individual schools, overall adolescent pregnancy rates are generally not available at that level of measurement. This is where knowledge of the literature can provide alternative measures of intervention impact, such as changes in grades or plans for the future or attitudes about pregnancy.

The evaluation plan should include an ecological unit of analysis. In other words, all analysis related to the evaluation must occur at the individual, organization, or community level (see Chapter Four). Evaluations that analyze at the organization and community level are more robust than those that use individual- or group-level measures, because an intervention that is being implemented in only one setting—say, one school or one church—can be analyzed only at the individual level. If access to several schools or churches is possible, however, analysis can occur at both the individual and the aggregate school or church level. The sample size and thus statistical power become issues here, so a statistician will need to be consulted.

The strength of the indicators in terms of predicting an outcome is an important consideration. In interventions at the ecological levels of group, organization, and system, change at the level of knowledge is a commonly used but a relatively weak indicator of effectiveness, because in general a change in knowledge has little effect on a change in behavior. Stronger indicators than knowledge that can be used to assess intervention effectiveness are changes in attitudes, including self-efficacy; skills; behaviors; and the strongest indicator: health indices (Kirkpatrick & Kirkpatrick, 2006). All health interventions should at least measure change in knowledge and attitudes. Measuring changes in skills is a stronger indicator of behavior change and the best indicator of an intervention's success. While ideally the goal of every community-based health intervention is change and improvement in communitywide indices (that is, morbidity and mortality statistics), for short-term interventions this is generally not feasible because of the time lag in the availability of such data.

An example of these levels can be seen in an intervention to improve the health of children with asthma in a specific geographic community—an intervention at the ecological level of group. The overall goal might be to decrease asthma-related mortality and the specific objectives to decrease asthma-related hospitalizations and emergency room visits, with intervention-related activities offered to children diagnosed with asthma attending all elementary and middle schools of the district. While it may be theoretically possible to actually collect baseline and follow-up data about

asthma-related emergency room visits in the community's hospitals and urgent care centers, Health Insurance Portability and Accountability Act (HIPAA) regulations generally make this form of data difficult to impossible to collect, unless the practitioner is a member of the hospital or care center staff; it would also be necessary to collect these data from each hospital or care center, since intervention activities could result in changes in the number of visits at one institution and not at the other.

If collection of such communitywide health indices is not feasible, however, data can be collected on several other indicators. In addition to knowledge and attitudes about asthma care and prevention, demonstration of correct inhaler use (skills) before and after the intervention is usually feasible. It might also be possible for the evaluation team to return three to six months after the intervention and collect data on the number of emergency room and urgent care center visits (behaviors) that occurred in the interim. Of course, baseline data on the number of such visits should also be collected in the interval before the start of the intervention. This asthma-related intervention would therefore assess change in knowledge, attitudes, and skills related to asthma care as well as behaviors, all varied and strong indicators of intervention success.

FLEXIBILITY: AN ESSENTIAL SKILL IN EVALUATION

In implementing both the intervention and the evaluation, the practitioner must be able to maintain flexibility. Many things can go wrong, and some undoubtedly will. Policy makers or public officials may object to some aspect of the evaluation design, people may object to being interviewed, an organization may drop out before the intervention is complete, or people may drop out before follow-up data are collected. Viable alternatives are almost always available, though they may not be readily apparent. For example, staff members might be available to be interviewed rather than their supervisors. The sample size can be increased to make up for dropouts, and additional subjects may be recruited at the remaining organizations. In the worst-case scenario, the planned evaluation design may need to be rethought to make the best use of the data that are available.

SUMMARY

This chapter provided the reader with the rationale for and steps of an evaluation which every intervention should contain. Distinctions between the three types of evaluation were made, as well as specifications about when to use process, impact, and outcome evaluation. The reader should now be familiar with the basics of the different types of evaluation methodologies and the data used in those methodologies. The chapter concluded with strategies for disseminating both positive and negative evaluation results.