



NONPROFIT DEVELOPMENT INSTITUTE, INC.

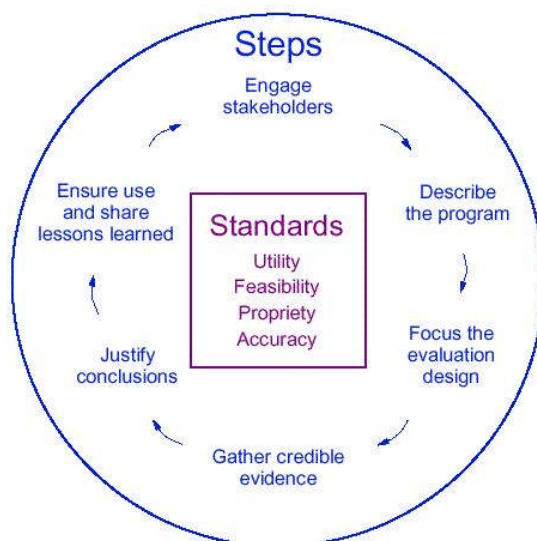
Proven Systems That Get You Results

# Program Evaluation

*A PRIMER FOR NONPROFIT ORGANIZATIONS*

## R E P O R T

**Federal funding agencies and many foundations want to see a good evaluation plan. This primer will help you create one for your next proposal.**



## TABLE OF CONTENTS

TABLE OF CONTENTS.....	1
TABLE OF CONTENTS.....	2
INTRODUCTION .....	3
DEFINING KEY CONCEPTS .....	4
INTEGRATING EVALUATION WITH ROUTINE PROGRAM PRACTICE .....	5
ASSIGNING VALUE TO PROGRAM ACTIVITIES .....	5
A FRAMEWORK FOR PROGRAM EVALUATION.....	7
STEP 1: ENGAGING STAKEHOLDERS .....	8
STEP 2: DESCRIBING THE PROGRAM.....	9
STEP 3: FOCUSING THE EVALUATION DESIGN .....	10
STEP 4: GATHERING CREDIBLE EVIDENCE .....	13
STEP 5: JUSTIFYING CONCLUSIONS .....	14
STEP 6: ENSURING USE AND SHARING LESSONS LEARNED.....	16
STANDARDS FOR EFFECTIVE EVALUATION.....	17
APPLYING THE FRAMEWORK .....	20
CONDUCTING OPTIMAL EVALUATIONS.....	20
ASSEMBLING AN EVALUATION TEAM.....	20
ADDRESSING COMMON CONCERNS.....	21
EVALUATION TRENDS.....	21
SUMMARY.....	22
REFERENCES.....	23

## INTRODUCTION

**P**rogram evaluation is an essential organizational practice in social service programs. However, it is not practiced consistently across program areas, nor is it sufficiently well-integrated into the day-to-day management of most programs. Program evaluation is also necessary for fulfilling most federal funding agency operating principles for guiding public health/social service activities, which include a) using science as a basis for decision-making and program intervention; b) expanding the quest for social equity through public health and social service actions; c) performing effectively as a service agency; d) making efforts outcome-oriented; and e) being accountable.

In 1993, Congress enacted the Government Performance and Results Act (GPRA) to ensure that the Federal government focuses programs on performance. Federal departments are now required to assess the effectiveness of their programs. In 2001, the current Administration introduced a number of management reforms, including the Budget and Performance Integration initiative, which requires Federal agencies "to identify high quality outcome measures, accurately monitor the performance of programs, and begin integrating this presentation with associated cost". As a result of these efforts, Federal departments increasingly ask for outcome-related **data** in their requirements.

This underscores the need for programs to develop clear plans, inclusive partnerships, and feedback systems that allow learning and ongoing improvement to occur. One way to ensure that new and existing programs honor these principles is for each program to conduct routine, practical evaluations that provide information for management and improve program effectiveness. This primer presents a framework for understanding program evaluation and facilitating integration of evaluation throughout the public health and social service systems. The purposes of this primer are to

- summarize the essential elements of program evaluation;
- provide a framework for conducting effective program evaluations;
- clarify the steps in program evaluation;
- review standards for effective program evaluation; and
- address misconceptions regarding the purposes and methods of program evaluation.

***Effective program evaluation is a systematic way to improve and account for public health and social service actions by involving procedures that are useful, feasible, ethical, and accurate. The framework guides public health and social service professionals in their use of program evaluation. It is a practical, nonprescriptive tool, designed to summarize and organize essential elements of program evaluation. The framework comprises steps in program evaluation practice and standards for effective program evaluation. Adhering to the steps and standards of this framework will allow an understanding of each program's context and will improve how program evaluations are conceived and conducted. Furthermore, the framework encourages an approach to evaluation that is integrated with routine program operations. The emphasis is on practical, ongoing evaluation strategies that involve all program stakeholders, not just evaluation experts. Understanding and applying the elements of this framework can be a driving force for planning effective public health and social service strategies, improving existing programs, and demonstrating the results of resource investments.***

## DEFINING KEY CONCEPTS

Throughout this report, the term program is used to describe the object of evaluation, which could be any organized Public health and social service action.

### What Is Outcome Measurement?

Accredit, adjudge, analyze, appraise, appreciate, assay, assess, audit, check, classify, consider, critique, determine, estimate, examine, evaluate, find, gauge, grade, inspect, investigate, judge, measure, monitor, rank, rate, referee, report, review, score, scrutinize, study, test, validate, weigh.

There are countless words used in the world of evaluation and a fair amount of confusion and debate about precisely what each word means. This guidebook uses the phrase "outcome measurement" to describe one approach to exploring the impacts or results of a program and to distinguish outcome measurement from more elaborate or complex types of evaluation.

Outcome measurement is "a systematic way to assess the extent to which a program has achieved its intended results. The main questions addressed are: What has changed in the lives of individuals, families, organizations or the community as a result of this program? Has this program made a difference? How are the lives of program participants better as a result of the program?" (The Evaluation Forum, 2000, p. 9).

During the last 30 years or so, most of the reporting required of groups receiving government or charitable funds has focused on *what staff do*. How many people they serve. How many hours of service they deliver. Outcome measurement asks, and attempts to answer, the question, *So what?*

So what if you provide a family with 10 hours of parent training? Are the parents better able to raise their children? Do they?

So what if you train a client in life skills? Can they effectively perform the activities involved? Do they?

Outcome measurement is most commonly used in the not-for-profit world; a similar phrase, "**performance measurement**," is used more often in the business and government arenas. In essence, they mean the same thing.

"**Compliance monitoring**" is another phrase in use today, referring most often to the contractual arrangements made between an organization and its grantmaker on the use of funds. Compliance monitoring keeps records on what and how much service a program delivers, the clients it serves and how much money it expends in relation to what the organization agreed to with the funder. It may, but does not necessarily, include information on the outcomes of the program. (See our *Managing Your Grant* primer in the Learning Library.)

The term "**evaluation**" is used broadly to cover an entire range of activities, including studies where the steps undertaken can specifically and with more certainty show that the results the program achieves are attributable to it and not to other factors. A typical definition of program evaluation reads: "the systematic application of social research procedures for assessing the conceptualization, design, implementation and utility of health or social interventions."

Evaluation research focuses on "causation," proving that the activities provided through the program are the reason why change occurred for the people receiving the service. This requires considerably more time and effort, and this is the basis for the distinction in this primer between outcome measurement and evaluation:

Outcome measurement will explore what your program provides, what its intended impacts are and whether or not it achieves them. It will *not* prove that the changes that take place are a result of your program.

## INTEGRATING EVALUATION WITH ROUTINE PROGRAM PRACTICE

Evaluation can be tied to routine program operations when the emphasis is on practical, ongoing evaluation that involves all program staff and stakeholders, not just evaluation experts. The practice of evaluation complements program management by gathering necessary information for improving and accounting for program effectiveness. Social service professionals routinely have used evaluation processes when answering questions from concerned persons, consulting partners, making judgments based on feedback, and refining program operations. These evaluation processes, though informal, are adequate for ongoing program assessment to guide small changes in program functions and objectives. However, when the stakes of potential decisions or program changes increase (e.g., when deciding what services to offer), employing evaluation procedures that are explicit, formal, and justifiable to funders becomes important.

### Effective Outcome Measurement Practices

This primer is not meant to make you an expert in evaluation or research methodology. Rather, it will provide you with the basics you need to conduct outcome measurement. Having said this, however, it is important to recognize that the outcome measurement work you undertake (and the outcome-related work of faith and community-based organizations who receive your support and assistance) must meet certain standards. Four program evaluation standards have been developed by The Joint Committee on Standards for Educational Evaluation (1994). They are commonly accepted as guiding principles for any effective evaluation work, including outcome measurement. The standards include:

- § **Utility:** the information generated must serve the needs of the intended users
- § **Feasibility:** the process undertaken must be realistic, prudent, diplomatic and frugal
- § **Propriety:** the evaluators must behave legally, ethically and with due regard for the welfare of those involved and affected
- § **Accuracy:** the findings must reveal and convey technically accurate information

Without attention to these standards, your work and your results will not be credible or useful and, ultimately, will not help you continually improve your program. These standards are discussed in more detail below.

## ASSIGNING VALUE TO PROGRAM ACTIVITIES

Questions regarding values, in contrast with those regarding facts, generally involve three interrelated issues: **merit** (i.e., quality), **worth** (i.e., cost-effectiveness), and **significance** (i.e., importance). If a program is judged to be of merit, other questions might arise regarding whether the program is worth its cost. Also, questions can arise regarding whether even valuable programs contribute important differences. Assigning value and making judgments regarding a program on the basis of evidence requires answering the following questions:

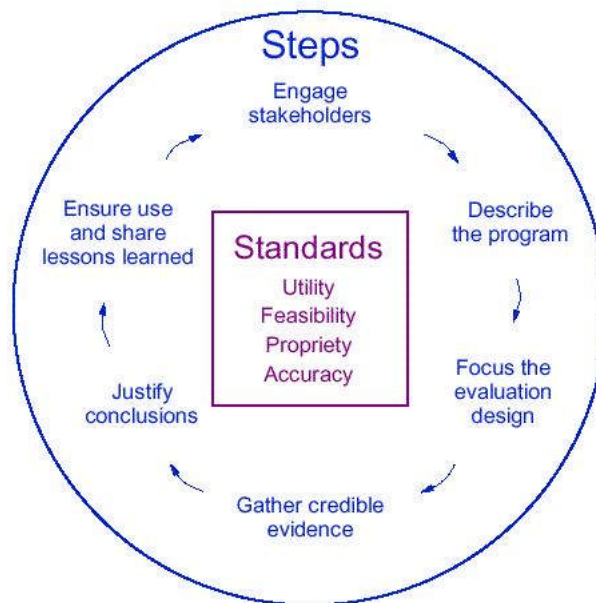
- What will be evaluated? (That is, what is the program and in what context does it exist?)
- What aspects of the program will be considered when judging program performance?
- What standards (i.e., type or level of performance) must be reached for the program to be considered successful?
- What evidence will be used to indicate how the program has performed?

- What conclusions regarding program performance are justified by comparing the available evidence to the selected standards?
- How will the lessons learned from the inquiry be used to improve public health effectiveness?

These questions should be addressed at the beginning of a program and revisited throughout its implementation. The framework described in this report provides a systematic approach for answering these questions.

## A FRAMEWORK FOR PROGRAM EVALUATION

Effective program evaluation is a systematic way to improve and account for public health and social service actions by involving procedures that are useful, feasible, ethical, and accurate. The recommended framework was developed to guide public health professionals in using program evaluation. It is a practical, nonprescriptive tool, designed to summarize and organize the essential elements of program evaluation. The framework comprises steps in evaluation practice and standards for effective evaluation.



The framework is composed of six steps that must be taken in any evaluation. They are starting points for tailoring an evaluation to a particular public health effort at a particular time. Because the steps are all interdependent, they might be encountered in a nonlinear sequence; however, an order exists for fulfilling each -- earlier steps provide the foundation for subsequent progress. Thus, decisions regarding how to execute a step are iterative and should not be finalized until previous steps have been thoroughly addressed. The steps are as follows:

- Step 1: Engage stakeholders.
- Step 2: Describe the program.
- Step 3: Focus the evaluation design.
- Step 4: Gather credible evidence.
- Step 5: Justify conclusions.
- Step 6: Ensure use and share lessons learned.

Adhering to these six steps will facilitate an understanding of a program's context (e.g., the program's history, setting, and organization) and will improve how most evaluations are conceived and conducted.

## STEP 1: ENGAGING STAKEHOLDERS

The evaluation cycle begins by engaging stakeholders (i.e., the persons or organizations having an investment in what will be learned from an evaluation and what will be done with the knowledge). Public health and social service work involves partnerships; therefore, any assessment of a public health or social service program requires considering the value systems of the partners. Stakeholders must be engaged in the inquiry to ensure that their perspectives are understood. When stakeholders are not engaged, an evaluation might not address important elements of a program's objectives, operations, and outcomes. Therefore, evaluation findings might be ignored, criticized, or resisted because the evaluation did not address the stakeholders' concerns or values. After becoming involved, stakeholders help to execute the other steps. Identifying and engaging the following three principal groups of stakeholders are critical:

- those involved in program operations (e.g., sponsors, collaborators, coalition partners, funding officials, administrators, managers, and staff);
- those served or affected by the program (e.g., clients, family members, neighborhood organizations, academic institutions, elected officials, advocacy groups, professional associations, skeptics, opponents, and staff of related or competing organizations); and
- primary users of the evaluation.

***Those Involved in Program Operations.*** Persons or organizations involved in program operations have a stake in how evaluation activities are conducted because the program might be altered as a result of what is learned. Although staff, funding officials, and partners work together on a program, they are not necessarily a single interest group. Subgroups might hold different perspectives and follow alternative agendas; furthermore, because these stakeholders have a professional role in the program, they might perceive program evaluation as an effort to judge them personally. Program evaluation is related to but must be distinguished from personnel evaluation, which operates under different standards.

***Those Served or Affected by the Program.*** Persons or organizations affected by the program, either directly (e.g., by receiving services) or indirectly (e.g., by benefiting from enhanced community assets), should be identified and engaged in the evaluation to the extent possible. Although engaging supporters of a program is natural, individuals who are openly skeptical or antagonistic toward the program also might be important stakeholders to engage. Opposition to a program might stem from differing values regarding what change is needed or how to achieve it. Opening an evaluation to opposing perspectives and enlisting the help of program opponents in the inquiry might be prudent because these efforts can strengthen the evaluation's credibility.

***Primary Users of the Evaluation.*** Primary users of the evaluation are the specific persons who are in a position to do or decide something regarding the program. In practice, primary users will be a subset of all stakeholders identified. A successful evaluation will designate primary users early in its development and maintain frequent interaction with them so that the evaluation addresses their values and satisfies their unique information needs.

The scope and level of stakeholder involvement will vary for each program evaluation. Various activities reflect the requirement to engage stakeholders. For example, stakeholders can be directly involved in designing and conducting the evaluation. Also, they can be kept informed regarding progress of the evaluation through periodic meetings, reports, and other means of communication. Sharing power and resolving conflicts helps avoid overemphasis of values held by any specific stakeholder. Occasionally, stakeholders might be inclined to use their involvement in an evaluation to sabotage, distort, or discredit



the program. Trust among stakeholders is essential; therefore, caution is required for preventing misuse of the evaluation process.

## STEP 2: DESCRIBING THE PROGRAM

Program descriptions convey the mission and objectives of the program being evaluated. Descriptions should be sufficiently detailed to ensure understanding of program goals and strategies. The description should discuss the program's capacity to effect change, its stage of development, and how it fits into the larger organization and community. Program descriptions set the frame of reference for all subsequent decisions in an evaluation. The description enables comparisons with similar programs and facilitates attempts to connect program components to their effects. Moreover, stakeholders might have differing ideas regarding program goals and purposes. Evaluations done without agreement on the program definition are likely to be of limited use. Sometimes, negotiating with stakeholders to formulate a clear and logical description will bring benefits before data are available to evaluate program effectiveness. Aspects to include in a program description are need, expected effects, activities, resources, stage of development, context, and logic model.

**Need.** A statement of need describes the problem or opportunity that the program addresses and implies how the program will respond. Important features for describing a program's need include a) the nature and magnitude of the problem or opportunity, b) which populations are affected, c) whether the need is changing, and d) in what manner the need is changing.

**Expected Effects.** Descriptions of expectations convey what the program must accomplish to be considered successful (i.e., program effects). For most programs, the effects unfold over time; therefore, the descriptions of expectations should be organized by time, ranging from specific (i.e., immediate) to broad (i.e., long-term) consequences. A program's mission, goals, and objectives all represent varying levels of specificity regarding a program's expectations. Also, forethought should be given to anticipate potential unintended consequences of the program.

**Activities.** Describing program activities (i.e., what the program does to effect change) permits specific steps, strategies, or actions to be arrayed in logical sequence. This demonstrates how each program activity relates to another and clarifies the program's hypothesized mechanism or theory of change. Also, program activity descriptions should distinguish the activities that are the direct responsibility of the program from those that are conducted by related programs or partners. External factors that might affect the program's success (e.g., secular trends in the community) should also be noted.

**Resources.** Resources include the time, talent, technology, equipment, information, money, and other assets available to conduct program activities. Program resource descriptions should convey the amount and intensity of program services and highlight situations where a mismatch exists between desired activities and resources available to execute those activities. In addition, economic evaluations require an understanding of all direct and indirect program inputs and costs.

**Stage of Development.** Public health and social service programs mature and change over time; therefore, a program's stage of development reflects its maturity. Programs that have recently received initial authorization and funding will differ from those that have been operating continuously for a decade. The changing maturity of program practice should be considered during the evaluation process. A minimum of three stages of development must be recognized: planning, implementation, and effects. During planning, program activities are untested, and the goal of evaluation is to refine plans. During implementation, program activities are being field-tested and modified; the goal of evaluation is to characterize real, as opposed to ideal, program activities and to improve operations,

perhaps by revising plans. During the last stage, enough time has passed for the program's effects to emerge; the goal of evaluation is to identify and account for both intended and unintended effects.

**Context.** Descriptions of the program's context should include the setting and environmental influences (e.g., history, geography, politics, social and economic conditions, and efforts of related or competing organizations) within which the program operates. Understanding these environmental influences is required to design a context-sensitive evaluation and aid users in interpreting findings accurately and assessing the generalizability of the findings.

**Logic Model.** A logic model describes the sequence of events for bringing about change by synthesizing the main program elements into a picture of how the program is supposed to work. Often, this model is displayed in a flow chart, map, or table to portray the sequence of steps leading to program results. One of the virtues of a logic model is its ability to summarize the program's overall mechanism of change by linking processes (e.g., laboratory diagnosis of disease) to eventual effects (e.g., reduced tuberculosis incidence). The logic model can also display the infrastructure needed to support program operations. Elements that are connected within a logic model might vary but generally include inputs (e.g., trained staff), activities (e.g., identification of cases), outputs (e.g., persons completing treatment), and results ranging from immediate (e.g., curing affected persons) to intermediate (e.g., reduction in tuberculosis rate) to long-term effects (e.g., improvement of population health status). Creating a logic model allows stakeholders to clarify the program's strategies; therefore, the logic model improves and focuses program direction. It also reveals assumptions concerning conditions for program effectiveness and provides a frame of reference for one or more evaluations of the program. A detailed logic model can also strengthen claims of causality and be a basis for estimating the program's effect on endpoints that are not directly measured but are linked in a causal chain supported by prior research. Families of logic models can be created to display a program at different levels of detail, from different perspectives, or for different audiences.

Program descriptions will vary for each evaluation, and various activities reflect the requirement to describe the program (e.g., using multiple sources of information to construct a well-rounded description). The accuracy of a program description can be confirmed by consulting with diverse stakeholders, and reported descriptions of program practice can be checked against direct observation of activities in the field. A narrow program description can be improved by addressing such factors as staff turnover, inadequate resources, political pressures, or strong community participation that might affect program performance.

Sample logic models are included in many of the Program Design Guides in the Rapid Proposal Center, and links to resources on how to develop logic models are included in the Learning Center.

### STEP 3: FOCUSING THE EVALUATION DESIGN

The evaluation must be focused to assess the issues of greatest concern to stakeholders while using time and resources as efficiently as possible. Not all design options are equally well-suited to meeting the information needs of stakeholders. After data collection begins, changing procedures might be difficult or impossible, even if better methods become obvious. A thorough plan anticipates intended uses and creates an evaluation strategy with the greatest chance of being useful, feasible, ethical, and accurate. Among the items to consider when focusing an evaluation are purpose, users, uses, questions, methods, and agreements.

**Purpose.** Articulating an evaluation's purpose (i.e., intent) will prevent premature decision-making regarding how the evaluation should be conducted. Characteristics of the program, particularly its stage of development and context, will influence the evaluation's purpose. Public health and social service

evaluations have four general purposes. The first is to gain insight, which happens, for example, when assessing the feasibility of an innovative approach to practice. Knowledge from such an evaluation provides information concerning the practicality of a new approach, which can be used to design a program that will be tested for its effectiveness. For a developing program, information from prior evaluations can provide the necessary insight to clarify how its activities should be designed to bring about expected changes.

A second purpose for program evaluation is to change practice, which is appropriate in the implementation stage when an established program seeks to describe what it has done and to what extent. Such information can be used to better describe program processes, to improve how the program operates, and to fine-tune the overall program strategy. Evaluations done for this purpose include efforts to improve the quality, effectiveness, or efficiency of program activities.

A third purpose for evaluation is to assess effects. Evaluations done for this purpose examine the relationship between program activities and observed consequences. This type of evaluation is appropriate for mature programs that can define what interventions were delivered to what proportion of the target population. Knowing where to find potential effects can ensure that significant consequences are not overlooked. One set of effects might arise from a direct cause-and-effect relationship to the program. Where these exist, evidence can be found to attribute the effects exclusively to the program. In addition, effects might arise from a causal process involving issues of contribution as well as attribution. For example, if a program's activities are aligned with those of other programs operating in the same setting, certain effects (e.g., the creation of new laws or policies) cannot be attributed solely to one program or another. In such situations, the goal for evaluation is to gather credible evidence that describes each program's contribution in the combined change effort. Establishing accountability for program results is predicated on an ability to conduct evaluations that assess both of these kinds of effects.

A fourth purpose, which applies at any stage of program development, involves using the process of evaluation inquiry to affect those who participate in the inquiry. The logic and systematic reflection required of stakeholders who participate in an evaluation can be a catalyst for self-directed change. An evaluation can be initiated with the intent of generating a positive influence on stakeholders. Such influences might be to supplement the program intervention (e.g., using a follow-up questionnaire to reinforce program messages); empower program participants (e.g., increasing a client's sense of control over program direction); promote staff development (e.g., teaching staff how to collect, analyze, and interpret evidence); contribute to organizational growth (e.g., clarifying how the program relates to the organization's mission); or facilitate social transformation (e.g., advancing a community's struggle for self-determination).

**Users.** Users are the specific persons that will receive evaluation findings. Because intended users directly experience the consequences of inevitable design trade-offs, they should participate in choosing the evaluation focus. User involvement is required for clarifying intended uses, prioritizing questions and methods, and preventing the evaluation from becoming misguided or irrelevant.

**Uses.** Uses are the specific ways in which information generated from the evaluation will be applied. Several uses exist for program evaluation. Stating uses in vague terms that appeal to many stakeholders increases the chances the evaluation will not fully address anyone's needs. Uses should be planned and prioritized with input from stakeholders and with regard for the program's stage of development and current context. All uses must be linked to one or more specific users.

**Questions.** Questions establish boundaries for the evaluation by stating what aspects of the program will be addressed. Creating evaluation questions encourages stakeholders to reveal what they believe the evaluation should answer. Negotiating and prioritizing questions among stakeholders further refines

a viable focus. The question-development phase also might expose differing stakeholder opinions regarding the best unit of analysis. Certain stakeholders might want to study how programs operate together as a system of interventions to effect change within a community. Other stakeholders might have questions concerning the performance of a single program or a local project within a program. Still others might want to concentrate on specific subcomponents or processes of a project. Clear decisions regarding the questions and corresponding units of analysis are needed in subsequent steps of the evaluation to guide method selection and evidence gathering.

**Methods.** The methods for an evaluation are drawn from scientific research options, particularly those developed in the social, behavioral, and health sciences. A classification of design types includes experimental, quasi-experimental, and observational designs. No design is better than another under all circumstances. Evaluation methods should be selected to provide the appropriate information to address stakeholders' questions (i.e., methods should be matched to the primary users, uses, and questions). Experimental designs use random assignment to compare the effect of an intervention with otherwise equivalent groups. Quasi-experimental methods compare nonequivalent groups (e.g., program participants versus those on a waiting list) or use multiple waves of data to set up a comparison (e.g., interrupted time series). Observational methods use comparisons within a group to explain unique features of its members (e.g., comparative case studies or cross-sectional surveys).

The choice of design has implications for what will count as evidence, how that evidence will be gathered, and what kind of claims can be made (including the internal and external validity of conclusions). Also, methodologic decisions clarify how the evaluation will operate (e.g., to what extent program participants will be involved; how information sources will be selected; what data collection instruments will be used; who will collect the data; what data management systems will be needed; and what are the appropriate methods of analysis, synthesis, interpretation, and presentation). Because each method option has its own bias and limitations, evaluations that mix methods are generally more effective. During the course of an evaluation, methods might need to be revised or modified. Also, circumstances that make a particular approach credible and useful can change. For example, the evaluation's intended use can shift from improving a program's current activities to determining whether to expand program services to a new population group. Thus, changing conditions might require alteration or iterative redesign of methods to keep the evaluation on track.

**Agreements.** Agreements summarize the procedures and clarify roles and responsibilities among those who will execute the evaluation plan. Agreements describe how the evaluation plan will be implemented by using available resources (e.g., money, personnel, time, and information). Agreements also state what safeguards are in place to protect human subjects and, where appropriate, what ethical (e.g., institutional review board) or administrative (e.g., paperwork reduction) approvals have been obtained. Elements of an agreement include statements concerning the intended purpose, users, uses, questions, and methods, as well as a summary of the deliverables, time line, and budget. The agreement can include all engaged stakeholders but, at a minimum, it must involve the primary users, any providers of financial or in-kind resources, and those persons who will conduct the evaluation and facilitate its use and dissemination. The formality of an agreement might vary depending on existing stakeholder relationships. An agreement might be a legal contract, a detailed protocol, or a memorandum of understanding. Creating an explicit agreement verifies the mutual understanding needed for a successful evaluation. It also provides a basis for modifying or renegotiating procedures if necessary.

Various activities reflect the requirement to focus the evaluation design. Both supporters and skeptics of the program could be consulted to ensure that the proposed evaluation questions are politically viable (i.e., responsive to the varied positions of interest groups). A menu of potential evaluation uses appropriate for the program's stage of development and context could be circulated among stakeholders to determine which is most compelling. Interviews could be held with specific intended

users to better understand their information needs and time line for action. Resource requirements could be reduced when users are willing to employ more timely but less precise evaluation methods.

## **STEP 4: GATHERING CREDIBLE EVIDENCE**

An evaluation should strive to collect information that will convey a well-rounded picture of the program so that the information is seen as credible by the evaluation's primary users. Information (i.e., evidence) should be perceived by stakeholders as believable and relevant for answering their questions. Such decisions depend on the evaluation questions being posed and the motives for asking them. For certain questions, a stakeholder's standard for credibility might require having the results of a controlled experiment; whereas for another question, a set of systematic observations (e.g., interactions between an outreach worker and community residents) would be the most credible. Consulting specialists in evaluation methodology might be necessary in situations where concern for data quality is high or where serious consequences exist associated with making errors of inference (i.e., concluding that program effects exist when none do, concluding that no program effects exist when in fact they do, or attributing effects to a program that has not been adequately implemented).

Having credible evidence strengthens evaluation judgments and the recommendations that follow from them. Although all types of data have limitations, an evaluation's overall credibility can be improved by using multiple procedures for gathering, analyzing, and interpreting data. Encouraging participation by stakeholders can also enhance perceived credibility. When stakeholders are involved in defining and gathering data that they find credible, they will be more likely to accept the evaluation's conclusions and to act on its recommendations. Aspects of evidence gathering that typically affect perceptions of credibility include indicators, sources, quality, quantity, and logistics.

**Indicators.** Indicators define the program attributes that pertain to the evaluation's focus and questions. Because indicators translate general concepts regarding the program, its context, and its expected effects into specific measures that can be interpreted, they provide a basis for collecting evidence that is valid and reliable for the evaluation's intended uses. Indicators address criteria that will be used to judge the program; therefore, indicators reflect aspects of the program that are meaningful for monitoring. Examples of indicators that can be defined and tracked include measures of program activities (e.g., the program's capacity to deliver services; the participation rate; levels of client satisfaction; the efficiency of resource use; and the amount of intervention exposure) and measures of program effects (e.g., changes in participant behavior, community norms, policies or practices, health status, quality of life, and the settings or environment around the program).

Defining too many indicators can detract from the evaluation's goals; however, multiple indicators are needed for tracking the implementation and effects of a program. One approach to developing multiple indicators is based on the program logic model (developed in the second step of the evaluation). The logic model can be used as a template to define a spectrum of indicators leading from program activities to expected effects. For each step in the model, qualitative/quantitative indicators could be developed to suit the concept in question, the information available, and the planned data uses. Relating indicators to the logic model allows the detection of small changes in performance faster than if a single outcome were the only measure used. Lines of responsibility and accountability are also clarified through this approach because the measures are aligned with each step of the program strategy. Further, this approach results in a set of broad-based measures that reveal how health outcomes are the consequence of intermediate effects of the program. Intangible factors (e.g., service quality, community capacity, or interorganizational relations) that also affect the program can be measured by systematically recording markers of what is said or done when the concept is expressed. During an evaluation, indicators might need to be modified or new ones adopted. Measuring program performance by tracking indicators is only part of an evaluation and must not be confused as a singular

basis for decision-making. Well-documented problems result from using performance indicators as a substitute for completing the evaluation process and reaching fully justified conclusions. An indicator (e.g., a rising rate of adolescent pregnancy) might be assumed to reflect a failing program when, in reality, the indicator might be influenced by changing conditions that are beyond the program's control.

**Sources.** Sources of evidence in an evaluation are the persons, documents, or observations that provide information for the inquiry. More than one source might be used to gather evidence for each indicator to be measured. Selecting multiple sources provides an opportunity to include different perspectives regarding the program and thus enhances the evaluation's credibility. An inside perspective might be understood from internal documents and comments from staff or program managers, whereas clients, neutral observers, or those who do not support the program might provide a different, but equally relevant perspective. Mixing these and other perspectives provides a more comprehensive view of the program. The criteria used for selecting sources should be stated clearly so that users and other stakeholders can interpret the evidence accurately and assess if it might be biased. In addition, some sources are narrative in form and others are numeric. The integration of qualitative and quantitative information can increase the chances that the evidence base will be balanced, thereby meeting the needs and expectations of diverse users. Finally, in certain cases, separate evaluations might be selected as sources for conducting a larger synthesis evaluation.

**Quality.** Quality refers to the appropriateness and integrity of information used in an evaluation. High-quality data are reliable, valid, and informative for their intended use. Well-defined indicators enable easier collection of quality data. Other factors affecting quality include instrument design, data-collection procedures, training of data collectors, source selection, coding, data management, and routine error checking. Obtaining quality data will entail trade-offs (e.g., breadth versus depth) that should be negotiated among stakeholders. Because all data have limitations, the intent of a practical evaluation is to strive for a level of quality that meets the stakeholders' threshold for credibility.

**Quantity.** Quantity refers to the amount of evidence gathered in an evaluation. The amount of information required should be estimated in advance, or where evolving processes are used, criteria should be set for deciding when to stop collecting data. Quantity affects the potential confidence level or precision of the evaluation's conclusions. It also partly determines whether the evaluation will have sufficient power to detect effects. All evidence collected should have a clear, anticipated use. Correspondingly, only a minimal burden should be placed on respondents for providing information.

**Logistics.** Logistics encompass the methods, timing, and physical infrastructure for gathering and handling evidence. Each technique selected for gathering evidence must be suited to the source(s), analysis plan, and strategy for communicating findings. Persons and organizations also have cultural preferences that dictate acceptable ways of asking questions and collecting information, including who would be perceived as an appropriate person to ask the questions. For example, some participants might be willing to discuss their health behavior with a stranger, whereas others are more at ease with someone they know. The procedures for gathering evidence in an evaluation must be aligned with the cultural conditions in each setting of the project and scrutinized to ensure that the privacy and confidentiality of the information and sources are protected.

## **STEP 5: JUSTIFYING CONCLUSIONS**

The evaluation conclusions are justified when they are linked to the evidence gathered and judged against agreed-upon values or standards set by the stakeholders. Stakeholders must agree that conclusions are justified before they will use the evaluation results with confidence. Justifying conclusions on the basis of evidence includes standards, analysis and synthesis, interpretation, judgment, and recommendations.

**Standards.** Standards reflect the values held by stakeholders, and those values provide the basis for forming judgments concerning program performance. Using explicit standards distinguishes evaluation from other approaches to strategic management in which priorities are set without reference to explicit values. In practice, when stakeholders articulate and negotiate their values, these become the standards for judging whether a given program's performance will, for example, be considered successful, adequate, or unsuccessful. An array of value systems might serve as sources of norm-referenced or criterion-referenced standards. When operationalized, these standards establish a comparison by which the program can be judged.

**Analysis and Synthesis.** Analysis and synthesis of an evaluation's findings might detect patterns in evidence, either by isolating important findings (analysis) or by combining sources of information to reach a larger understanding (synthesis). Mixed method evaluations require the separate analysis of each evidence element and a synthesis of all sources for examining patterns of agreement, convergence, or complexity. Deciphering facts from a body of evidence involves deciding how to organize, classify, interrelate, compare, and display information. These decisions are guided by the questions being asked, the types of data available, and by input from stakeholders and primary users.

**Interpretation.** Interpretation is the effort of figuring out what the findings mean and is part of the overall effort to understand the evidence gathered in an evaluation. Uncovering facts regarding a program's performance is not sufficient to draw evaluative conclusions. Evaluation evidence must be interpreted to determine the practical significance of what has been learned. Interpretations draw on information and perspectives that stakeholders bring to the evaluation inquiry and can be strengthened through active participation or interaction.

**Judgments.** Judgments are statements concerning the merit, worth, or significance of the program. They are formed by comparing the findings and interpretations regarding the program against one or more selected standards. Because multiple standards can be applied to a given program, stakeholders might reach different or even conflicting judgments. For example, a program that increases its outreach by 10% from the previous year might be judged positively by program managers who are using the standard of improved performance over time. However, community members might feel that despite improvements, a minimum threshold of access to services has not been reached. Therefore, by using the standard of social equity, their judgment concerning program performance would be negative. Conflicting claims regarding a program's quality, value, or importance often indicate that stakeholders are using different standards for judgment. In the context of an evaluation, such disagreement can be a catalyst for clarifying relevant values and for negotiating the appropriate bases on which the program should be judged.

**Recommendations.** Recommendations are actions for consideration resulting from the evaluation. Forming recommendations is a distinct element of program evaluation that requires information beyond what is necessary to form judgments regarding program performance. Knowing that a program is able to reduce the risk of disease does not translate necessarily into a recommendation to continue the effort, particularly when competing priorities or other effective alternatives exist. Thus, recommendations for continuing, expanding, redesigning, or terminating a program are separate from judgments regarding a program's effectiveness. Making recommendations requires information concerning the context, particularly the organizational context, in which programmatic decisions will be made. Recommendations that lack sufficient evidence or those that are not aligned with stakeholders' values can undermine an evaluation's credibility. By contrast, an evaluation can be strengthened by recommendations that anticipate the political sensitivities of intended users and highlight areas that users can control or influence. Sharing draft recommendations, soliciting reactions from multiple stakeholders, and presenting options instead of directive advice increase the likelihood that recommendations will be relevant and well-received.

Various activities fulfill the requirement for justifying conclusions in an evaluation. Conclusions could be strengthened by a) summarizing the plausible mechanisms of change; b) delineating the temporal sequence between activities and effects; c) searching for alternative explanations and showing why they are unsupported by the evidence; and d) showing that the effects can be repeated. When different but equally well-supported conclusions exist, each could be presented with a summary of its strengths and weaknesses. Creative techniques (e.g., the Delphi process) could be used to establish consensus among stakeholders when assigning value judgments. Techniques for analyzing, synthesizing, and interpreting findings should be agreed on before data collection begins to ensure that all necessary evidence will be available.

## **STEP 6: ENSURING USE AND SHARING LESSONS LEARNED**

Lessons learned in the course of an evaluation do not automatically translate into informed decision-making and appropriate action. Deliberate effort is needed to ensure that the evaluation processes and findings are used and disseminated appropriately. Preparing for use involves strategic thinking and continued vigilance, both of which begin in the earliest stages of stakeholder engagement and continue throughout the evaluation process. Five elements are critical for ensuring use of an evaluation, including design, preparation, feedback, follow-up, and dissemination.

**Design.** Design refers to how the evaluation's questions, methods, and overall processes are constructed. As discussed in the third step of this framework, the design should be organized from the start to achieve intended uses by primary users. Having a clear design that is focused on use helps persons who will conduct the evaluation to know precisely who will do what with the findings and who will benefit from being a part of the evaluation. Furthermore, the process of creating a clear design will highlight ways that stakeholders, through their contributions, can enhance the relevance, credibility, and overall utility of the evaluation.

**Preparation.** Preparation refers to the steps taken to rehearse eventual use of the evaluation findings. The ability to translate new knowledge into appropriate action is a skill that can be strengthened through practice. Building this skill can itself be a useful benefit of the evaluation. Rehearsing how potential findings (particularly negative findings) might affect decision-making will prepare stakeholders for eventually using the evidence. Primary users and other stakeholders could be given a set of hypothetical results and asked to explain what decisions or actions they would make on the basis of this new knowledge. If they indicate that the evidence presented is incomplete and that no action would be taken, this is a sign that the planned evaluation should be modified. Preparing for use also gives stakeholders time to explore positive and negative implications of potential results and time to identify options for program improvement.

**Feedback.** Feedback is the communication that occurs among all parties to the evaluation. Giving and receiving feedback creates an atmosphere of trust among stakeholders; it keeps an evaluation on track by letting those involved stay informed regarding how the evaluation is proceeding. Primary users and other stakeholders have a right to comment on decisions that might affect the likelihood of obtaining useful information. Stakeholder feedback is an integral part of evaluation, particularly for ensuring use. Obtaining feedback can be encouraged by holding periodic discussions during each step of the evaluation process and routinely sharing interim findings, provisional interpretations, and draft reports.

**Follow-Up.** Follow-up refers to the technical and emotional support that users need during the evaluation and after they receive evaluation findings. Because of the effort required, reaching justified conclusions in an evaluation can seem like an end in itself; however, active follow-up might be necessary to remind intended users of their planned use. Follow-up might also be required to prevent lessons learned from becoming lost or ignored in the process of making complex or politically sensitive



decisions. To guard against such oversight, someone involved in the evaluation should serve as an advocate for the evaluation's findings during the decision-making phase. This type of advocacy increases appreciation of what was discovered and what actions are consistent with the findings.

Facilitating use of evaluation findings also carries with it the responsibility for preventing misuse. Evaluation results are always bound by the context in which the evaluation was conducted. However, certain stakeholders might be tempted to take results out of context or to use them for purposes other than those agreed on. For instance, inappropriately generalizing the results from a single case study to make decisions that affect all sites in a national program would constitute misuse of the case study evaluation. Similarly, stakeholders seeking to undermine a program might misuse results by overemphasizing negative findings without giving regard to the program's positive attributes. Active follow-up might help prevent these and other forms of misuse by ensuring that evidence is not misinterpreted and is not applied to questions other than those that were the central focus of the evaluation.

**Dissemination.** Dissemination is the process of communicating either the procedures or the lessons learned from an evaluation to relevant audiences in a timely, unbiased, and consistent fashion. Although documentation of the evaluation is needed, a formal evaluation report is not always the best or even a necessary product. Like other elements of the evaluation, the reporting strategy should be discussed in advance with intended users and other stakeholders. Such consultation ensures that the information needs of relevant audiences will be met. Planning effective communication also requires considering the timing, style, tone, message source, vehicle, and format of information products. Regardless of how communications are constructed, the goal for dissemination is to achieve full disclosure and impartial reporting. A checklist of items to consider when developing evaluation reports includes tailoring the report content for the audience, explaining the focus of the evaluation and its limitations, and listing both the strengths and weaknesses of the evaluation.

**Additional Uses.** Additional uses for evaluation flow from the process of conducting the evaluation; these process uses have value and should be encouraged because they complement the uses of the evaluation findings. Those persons who participate in an evaluation can experience profound changes in thinking and behavior. In particular, when newcomers to evaluation begin to think as evaluators, fundamental shifts in perspective can occur. Evaluation prompts staff to clarify their understanding of program goals. This greater clarity allows staff to function cohesively as a team, focused on a common end. Immersion in the logic, reasoning, and values of evaluation can lead to lasting impacts (e.g., basing decisions on systematic judgments instead of on unfounded assumptions). Additional process uses for evaluation includes defining indicators to discover what matters to decision makers and making outcomes matter by changing the structural reinforcements connected with outcome attainment (e.g., by paying outcome dividends to programs that save money through their prevention efforts). The benefits that arise from these and other process uses provide further rationale for initiating evaluation activities at the beginning of a program.

## STANDARDS FOR EFFECTIVE EVALUATION

Public health and social service professionals will recognize that the basic steps of the framework for program evaluation are part of their routine work. In day-to-day public health and social service practice, stakeholders are consulted; program goals are defined; guiding questions are stated; data are collected, analyzed, and interpreted; judgments are formed; and lessons are shared. Although informal evaluation occurs through routine practice, standards exist to assess whether a set of evaluative activities are well-designed and working to their potential. The Joint Committee on Standards for Educational Evaluation has developed program evaluation standards for this purpose. These

standards, designed to assess evaluations of educational programs, are also relevant for public health and social service programs.

The program evaluation standards make conducting sound and fair evaluations practical. The standards provide practical guidelines to follow when having to decide among evaluation options. The standards help avoid creating an imbalanced evaluation (e.g., one that is accurate and feasible but not useful or one that would be useful and accurate but is infeasible). Furthermore, the standards can be applied while planning an evaluation and throughout its implementation. The Joint Committee is unequivocal in that, "the standards are guiding principles, not mechanical rules. . . . In the end, whether a given standard has been addressed adequately in a particular situation is a matter of judgment"

In the Joint Committee's report, standards are grouped into the following four categories and include a total of 30 specific standards. As described in the report, each category has an associated list of guidelines and common errors, illustrated with applied case examples:

- utility,
- feasibility,
- propriety, and
- accuracy.

### ***Standard 1: Utility***

Utility standards ensure that information needs of evaluation users are satisfied. Seven utility standards address such items as identifying those who will be impacted by the evaluation, the amount and type of information collected, the values used in interpreting evaluation findings, and the clarity and timeliness of evaluation reports.

### ***Standard 2: Feasibility***

Feasibility standards ensure that the evaluation is viable and pragmatic. The three feasibility standards emphasize that the evaluation should employ practical, nondisruptive procedures; that the differing political interests of those involved should be anticipated and acknowledged; and that the use of resources in conducting the evaluation should be prudent and produce valuable findings.

### ***Standard 3: Propriety***

Propriety standards ensure that the evaluation is ethical (i.e., conducted with regard for the rights and interests of those involved and effected). Eight propriety standards address such items as developing protocols and other agreements for guiding the evaluation; protecting the welfare of human subjects; weighing and disclosing findings in a complete and balanced fashion; and addressing any conflicts of interest in an open and fair manner.

### ***Standard 4: Accuracy***

Accuracy standards ensure that the evaluation produces findings that are considered correct. Twelve accuracy standards include such items as describing the program and its context; articulating in detail the purpose and methods of the evaluation; employing systematic procedures to gather valid and reliable information; applying appropriate qualitative or quantitative methods during analysis and synthesis; and producing impartial reports containing conclusions that are justified.

The steps and standards are used together throughout the evaluation process. For each step, a subset of relevant standards should be considered.

## **APPLYING THE FRAMEWORK**

### **CONDUCTING OPTIMAL EVALUATIONS**

Public health professionals can no longer question whether to evaluate their programs; instead, the appropriate questions are

- What is the best way to evaluate?
- What is being learned from the evaluation? And,
- How will lessons learned from evaluations be used to make public health efforts more effective and accountable?

The framework for program evaluation helps answer these questions by guiding its users in selecting evaluation strategies that are useful, feasible, ethical, and accurate. To use the recommended framework in a specific program context requires practice, which builds skill in both the science and art of program evaluation. When applying the framework, the challenge is to devise an optimal -- as opposed to an ideal -- strategy. An optimal strategy is one that accomplishes each step in the framework in a way that accommodates the program context and meets or exceeds all relevant standards.

### **ASSEMBLING AN EVALUATION TEAM**

Harnessing and focusing the efforts of a collaborative group is one approach to conducting an optimal evaluation. A team approach can succeed when a small group of carefully selected persons decides what the evaluation must accomplish and pools resources to implement the plan. Stakeholders might have varying levels of involvement on the team that correspond to their own perspectives, skills, and concerns. A leader must be designated to coordinate the team and maintain continuity throughout the process; thereafter, the steps in evaluation practice guide the selection of team members. For example,

- Those who are diplomatic and have diverse networks can engage other stakeholders and maintain involvement.
- When describing the program, persons are needed who understand the program's history, purpose, and practical operation in the field. In addition, those with group facilitation skills might be asked to help elicit unspoken expectations regarding the program and to expose hidden values that partners bring to the effort. Such facilitators can also help the stakeholders create logic models that describe the program and clarify its pattern of relationships between means and ends.
- Decision makers and others who guide program direction can help focus the evaluation design on questions that address specific users and uses. They can also set logistic parameters for the evaluation's scope, time line, and deliverables.
- Scientists, particularly social and behavioral scientists, can bring expertise to the development of evaluation questions, methods, and evidence gathering strategies. They can also help analyze how a program operates in its organizational or community context.
- Trusted persons who have no particular stake in the evaluation can ensure that participants' values are treated fairly when applying standards, interpreting facts, and reaching justified conclusions.

- Advocates, clear communicators, creative thinkers, and members of the power structure can help ensure that lessons learned from the evaluation influence future decision-making regarding program strategy.

All organizations, even those that are able to find evaluation team members within their own agency, should collaborate with partners and take advantage of community resources when assembling an evaluation team. This strategy increases the available resources and enhances the evaluation's credibility. Furthermore, a diverse team of engaged stakeholders has a greater probability of conducting a culturally competent evaluation (i.e., one that understands and is sensitive to the persons, conditions, and contexts associated with the program). Although challenging for the coordinator and the participants, the collaborative approach is practical because of the benefits it brings (e.g., reduces suspicion and fear, increases awareness and commitment, increases the possibility of achieving objectives, broadens knowledge base, teaches evaluation skills, strengthens partnerships, increases the possibility that findings will be used, and allows for differing perspectives).

## **ADDRESSING COMMON CONCERNS**

Common concerns regarding program evaluation are clarified by using this framework. Evaluations might not be undertaken because they are misperceived as having to be costly. However, the expense of an evaluation is relative; the cost depends on the questions being asked and the level of precision desired for the answers. A simple, low-cost evaluation can deliver valuable results.

Rather than discounting evaluations as time-consuming and tangential to program operations (e.g., left to the end of a program's project period), the framework encourages conducting evaluations from the beginning that are timed strategically to provide necessary feedback to guide action. This makes integrating evaluation with program practice possible.

Another concern centers on the perceived technical demands of designing and conducting an evaluation. Although circumstances exist where controlled environments and elaborate analytic techniques are needed, most public health program evaluations do not require such methods. Instead, the practical approach endorsed by this framework focuses on questions that will improve the program by using context-sensitive methods and analytic techniques that summarize accurately the meaning of qualitative and quantitative information.

Finally, the prospect of evaluation troubles some program staff because they perceive evaluation methods as punitive, exclusionary, or adversarial. The framework encourages an evaluation approach that is designed to be helpful and engages all interested stakeholders in a process that welcomes their participation. Sanctions to be applied, if any, should not result from discovering negative findings, but from failing to use the learning to change for greater effectiveness.

## **EVALUATION TRENDS**

Interest in program improvement and accountability continues to grow in government, private, and nonprofit sectors. The Government Performance and Results Act requires federal agencies to set performance goals and to measure annual results. Nonprofit donor organizations (e.g., United Way) have integrated evaluation into their program activities and now require that grant recipients measure program outcomes. Public-health-oriented foundations (e.g., W.K. Kellogg Foundation) have also begun to emphasize the role of evaluation in their programming. Innovative approaches to staffing program evaluations have also emerged.

## SUMMARY

**E**valuation is the only way to separate programs that promote health and prevent injury, disease, or disability from those that do not; it is a driving force for planning effective public health strategies, improving existing programs, and demonstrating the results of resource investments. Evaluation also focuses attention on the common purpose of public health programs and asks whether the magnitude of investment matches the tasks to be accomplished.

The recommended framework is both a synthesis of existing evaluation practices and a standard for further improvement. It supports a practical approach to evaluation that is based on steps and standards applicable in public health settings. Because the framework is purposefully general, it provides a guide for designing and conducting specific evaluation projects across many different program areas. In addition, the framework can be used as a template to create or enhance program-specific evaluation guidelines that further operationalize the steps and standards in ways that are appropriate for each program. You can see the more specific tips about evaluation in each of our Program Design Guides.

## REFERENCES

Dyal WW. Ten organizational practices of public health: a historical perspective. *Am J Prev Med* 1995;11(6)Suppl 2:6-8.

Koplan JP. CDC sets millennium priorities. *US Medicine* 1999;4-7.

Scriven M. Minimalist theory of evaluation: the least theory that practice requires. *American Journal of Evaluation* 1998;19:57-70.

Shadish WR, Cook TD, Leviton LC. *Foundations of program evaluation: theories of practice*. Newbury Park, CA: Sage Publications, 1991.

Weiss CH. *Evaluation: methods for studying programs and policies*. 2nd ed. Upper Saddle River, NJ: Prentice Hall, 1998.

Worthen BR, Sanders JR, Fitzpatrick, JL. *Program evaluation: alternative approaches and practical guidelines*. 2nd ed. New York, NY: Longman, 1996.

Patton MQ. *Utilization-focused evaluation: the new century text*. 3rd ed. Thousand Oaks, CA: Sage Publications, 1997.

CDC. *Practical evaluation of public health programs*. Atlanta, GA: US Department of Health and Human Services, CDC, Public Health Training Network, 1998; PHTN course no. VC-0017.

Love A. *Internal evaluation: building organizations from within*. Applied social research, vol 24. Newbury Park, CA: Sage Publications, 1991.

Sanders JR. Uses of evaluation as a means toward organizational effectiveness. In: Gray ST, ed. *Leadership IS: a vision of evaluation; a report of learnings from Independent Sector's work on evaluation*. Washington, DC: Independent Sector, 1993.

Shadish WR. Evaluation theory is who we are. *American Journal of Evaluation* 1998;19(1):1-19.

Joint Committee on Standards for Educational Evaluation. *Program evaluation standards: how to assess evaluations of educational programs*. 2nd ed. Thousand Oaks, CA: Sage Publications, 1994.

Joint Committee on Standards for Educational Evaluation. *Personnel evaluation standards: how to assess systems for evaluating educators*. Newbury Park, CA: Sage Publications, 1988.

CDC. *Principles of community engagement*. Atlanta, GA: CDC, Public Health Practice Program Office, 1997.

Mertens DM. Inclusive evaluation: implications of transformative theory for evaluation. *American Journal of Evaluation* 1999;20(1):1-14.

Connell JP, Kubisch AC. Applying a theory of change approach to the evaluation of comprehensive community initiatives: progress, prospects, and problems. In: Fulbright-Anderson K, Kubisch AC, Connell JP, eds. *New approaches to evaluating community initiatives: theory, measurement, and analysis*. Washington, DC: Aspen Institute, 1998.

Chen HT. *Theory driven evaluations*. Newbury Park, CA: Sage Publications, 1990.

U.S. General Accounting Office. *Managing for results: measuring program results that are under limited federal control*. Washington, DC: U.S. General Accounting Office 1998; publication no. GAO/GGD-99-16.

Haddix AC, Teutsch SM, Shaffer PA, Duñet DO, eds. *Prevention effectiveness: a guide to decision analysis and economic evaluation*. New York, NY: Oxford University Press, 1996.

CDC. Framework for assessing the effectiveness of disease and injury prevention. *MMWR* 1992;41(No. RR-3):1-13.

Gold ME, Russell LB, Siegel JE, Weinstein MC. *Cost-effectiveness in health and medicine*. New York: Oxford University Press, 1996.

Eoyang GH, Berkas T. Evaluation in a complex adaptive system. In: Lissack M, Gunz H, eds. *Managing complexity in organizations*. Westport, CT: Quorum Books, 1999.

Rush B, Ogbourne A. Program logic models: expanding their role and structure for program planning and evaluation. *Canadian Journal of Program Evaluation* 1991;6(2):95-106.

W.K. Kellogg Foundation. *W.K. Foundation Evaluation Handbook*. Battle Creek, MI: W.K. Kellogg Foundation, 1998.

Taylor-Powell E, Rossing B, Geran J. *Evaluating collaboratives: reaching the potential*. Madison, WI: University of Wisconsin--Cooperative Extension, 1998.

Poreteous NL, Sheldrick BJ, Stewart PJ. *Program evaluation tool kit: a blueprint for public health management*. Ottawa, Ontario, Canada: Ottawa-Carleton Health Department, Public Health Research, Education and Development Program, 1997.

Weiss CH. Nothing as practical as a good theory: exploring theory-based evaluation for comprehensive community initiatives for families and children. In: Connell JP, Kubisch AC, Schorr LB, Weiss CH, eds. *New approaches to evaluating community initiatives: concepts, methods, and contexts*. Washington, DC: Aspen Institute, 1995.

Weiss CH. How can theory-based evaluation make greater headway? *Evaluation Review* 1997;21:501-24.

Wong-Reiger D, David L. Using program logic models to plan and evaluate education and prevention programs. In: Love AJ, ed. *Evaluation methods sourcebook II*. Ottawa, Ontario, Canada: Canadian Evaluation Society, 1995.



United Way of America. Measuring program outcomes: a practical approach. Alexandria, VA: United Way of America, 1996.

Moyer A, Verhovsek H, Wilson VL. Facilitating the shift to population-based public health programs: innovation through the use of framework and logic model tools. *Can J Public Health* 1997;88(2):95-8.

McLaughlin JA, Jordan GB. Logic models: a tool for telling your program's performance story. *Evaluation and Program Planning* 1999;22(1):65-72.

McEwan KL, Bigelow DA. Using a logic model to focus health services on population health goals. *Canadian Journal of Program Evaluation* 1997;12(1):167-74.

Julian D. Utilization of the logic model as a system level planning and evaluation device. *Evaluation and Program Planning* 1997;20(3):251-7.

Lipsey MW. Theory as method: small theories of treatments. *New Directions for Program Evaluation* 1993;57:5-38.

U.S. General Accounting Office. Designing evaluations. Washington, DC: U.S. General Accounting Office, 1991; publication no. GAO/PEMD-10.1.4.

Taylor-Powell E, Steele S, Douglass M. Planning a program evaluation. Madison, WI: University of Wisconsin Cooperative Extension, 1996.

Fetterman DM, Kaftarian SJ, Wandersman A, eds. Empowerment evaluation: knowledge and tools for self-assessment and accountability. Thousand Oaks, CA: Sage Publications, 1996.

Preskill HS, Torres RT. Evaluative inquiry for learning in organizations. Thousand Oaks, CA: Sage Publications, 1999.

Patton MQ. Toward distinguishing empowerment evaluation and placing it in a larger context. *Evaluation Practice* 1997;18(2):147-63.

Wandersman A, Morrissey E, Davino K, et al. Comprehensive quality programming and accountability: eight essential strategies for implementing successful prevention programs. *Journal of Primary Prevention* 1998;19(1):3-30.

Cousins JB, Whitmore E. Framing participatory evaluation. In: Whitmore E, ed. Understanding and practicing participatory evaluation, vol 80. San Francisco, CA: Jossey-Bass, 1998:5-24.

Bickman L, Rog DJ, eds. Handbook of applied social research methods. Thousand Oaks, CA: Sage Publications, 1998.

Cook TD, Reichardt CS, eds. Qualitative and quantitative methods in evaluation research. Sage research progress series in evaluation, vol 1. Beverly Hills, CA: Sage Publications, 1979.

Patton MQ. Qualitative evaluation and research methods. 2nd ed. Newbury Park, CA: Sage Publications, 1990.

Posavac EJ, Carey RG. Program evaluation: methods and case studies. Englewood Cliffs, NJ: Prentice-Hall, 1980.

Rossi PH, Freeman HE, Lipsey MW. Evaluation: a systematic approach. 6th ed. Thousand Oaks, CA: Sage Publications, 1999.

Trochim WMK. Research methods knowledge base [on-line textbook]. 2nd ed. 1999. Available at <<http://trochim.human.cornell.edu>>. Accessed June 1999.

Boruch RF. Randomized controlled experiments for evaluation and planning. In: Bickman L, Rog DJ, eds. Handbook of applied social research methods. Thousand Oaks, CA: Sage Publications, 1998:161-92.

Cook TD, Campbell DT. Quasi-experimentation: design and analysis issues for field settings. Boston, MA: Houghton Mifflin, 1979.

Reichardt CS, Mark MM. Quasi-experimentation. In: Bickman L, Rob DJ, eds. Handbook of applied social research methods. Thousand Oaks, CA: Sage Publications, 1998:193-228.

Yin RK. Case study research: design and methods. 2nd ed. Applied Social Research Methods Series, vol 5. Thousand Oaks, CA: Sage Publications, 1994.

Yin RK. Abridged version of case study research: design and method. In: Bickman L, Rob DJ, eds. Handbook of applied social research methods. Thousand Oaks, CA: Sage Publications, 1998:229-60.

U.S. General Accounting Office. Case study evaluations. Washington, DC: U.S. General Accounting Office, 1990; publication no. GAO/PEMD-91-10.1.9.

McQueen DV, Anderson LM. What counts as evidence? Issues and debates on evidence relevant to the evaluation of community health promotion programs. In: Rootman I, Goodstadt M, Hyndman B, et al., eds. Evaluation in health promotion: principles and perspectives. Copenhagen, Denmark. World Health Organization (Euro), 1999 (In press).

Green JC, Caracelli V, eds. Advances in mixed-method evaluation: the challenges and benefits of integrating diverse paradigms. New Directions for Program Evaluation, vol 74; San Francisco, CA: Jossey-Bass 1997.

Frechtling J, Sharp L. User-friendly handbook for mixed method evaluations. Arlington, VA: National Science Foundation, 1997.

U.S. General Accounting Office. Evaluation synthesis. Washington, DC: U.S. General Accounting Office, 1992. publication no. GAO/PEMD-10.1.2.

Newman DL, Brown RD. Applied ethics for program evaluation. Thousand Oaks, CA: Sage Publications, 1996.

Fitzpatrick JL, Morris M, eds. Current and emerging ethical challenges in evaluation. New Directions for Program Evaluation, vol 82; San Francisco, CA: Jossey-Bass 1999.

Newcomer K. Using statistics appropriately. In: Wholey J, Hatry H, Newcomer K, eds. Handbook of practical program evaluation. San Francisco, CA: Jossey-Bass, 1994.

Basch CE, Silepcevich EM, Gold RS, Duncan DF, Kolbe LJ. Avoiding type III errors in health education program evaluation: a case study. Health Education Quarterly 1985;12(4):315-31.

Perrin EB, Koshel JJ, eds. Assessment of performance measures for public health, substance abuse, and mental health. Washington, DC: National Academy Press, 1997.

Innes JE. Knowledge and public policy: the search for meaningful indicators. 2nd expanded ed. New Brunswick, NJ: Transaction Publishers, 1990.

McRae D Jr. Policy indicators: links between social science and public debate. Chapel Hill, NC: University of North Carolina Press, 1985.

Institute of Medicine. Improving health in the community: a role for performance monitoring. Durch, JS, Bailey LA, Stoto MA, eds. Washington, DC: National Academy Press, 1997.

Eddy DM. Performance measurement: problems and solutions. Health Aff 1998;17(4):7-25.

Harvard Family Research Project. Performance measurement. Evaluation Exchange 1998;4(1):1-15.

Rugg D. New activities and use of indicators in the evaluation of HIV prevention efforts at CDC. National Alliance of State and Territorial AIDS Directors (NASTAD) HIV Prevention Community Planning Bulletin;1997:2-3.

Center for Substance Abuse Prevention. Measurements in prevention: a manual on selecting and using instruments to evaluate prevention programs. CSAP technical report no. 8 (SMA)93-2041. Rockville, MD: U.S. Department of Health and Human Services, Substance Abuse and Mental Health Services Administration, 1993.

Goodman RM, Speers MA, McLeroy K, et al. Identifying and defining the dimensions of community capacity to provide a basis for measurement. Health Educ Behav 1998;25(3):258-78.

Phillips JJ. Handbook of training evaluation and measurement methods. 3rd ed. Houston, TX: Gulf Publishing Company, 1997.

Knauff EB. What Independent Sector learned from an evaluation of its own hard-to-measure programs. In: Gray ST, ed. Leadership IS: a vision of evaluation; a report of learnings from Independent Sector's work on evaluation. Washington, DC: Independent Sector, 1993.

Perrin B. Effective use and misuse of performance measurement. American Journal of Evaluation 1998;19(3):367-79.

U.S. General Accounting Office. Using statistical sampling. Washington, DC: U.S. General Accounting Office, 1992; publication no. GAO/PEMD-10.1.6.

Henry GT. Practical sampling. In: Bickman L, Rog DJ, eds. Handbook of applied social research methods. Thousand Oaks, CA: Sage Publications, 1998:101-26.

Johnson JC. Selecting ethnographic informants. Qualitative Research Methods Series, vol 22; Thousand Oaks, CA: Sage Publications, 1991,

Steckler A, McLeroy KR, Goodman RM, Bird ST, McCormick L. Toward integrating qualitative and quantitative methods: an introduction. Health Education Quarterly 1992;19(1):191-8.

Greene JC. Qualitative program evaluation: practice and promise. In: Denzin NK, Lincoln YS, eds. Handbook of qualitative research. Thousand Oaks, CA: Sage Publications, 1994.

de Vries H, Weijts W, Dijkstra M, Kok G. Utilization of qualitative and quantitative data for health education program planning, implementation, and evaluation: a spiral approach. Health Education Quarterly 1992;19:101-15.

U.S. General Accounting Office. Prospective evaluation methods: the prospective evaluation synthesis. Washington, DC: U.S. General Accounting Office, 1990; publication no. GAO/PEMD-10.1.10.

Lipsey MW. What can you build with thousands of bricks? Musings on the cumulation of knowledge in program evaluation. New Directions for Evaluation 1997;76:7-23.

Lipsy MW. Design sensitivity: statistical power for applied experimental research. In: Bickman L, Rob DJ, eds. Handbook of applied social research methods. Thousand Oaks, CA: Sage Publications, 1998:39-68.

Siebert JE. Planning ethically responsible research. In: Bickman L, Rob DJ, eds. Handbook of applied social research methods. Thousand Oaks, CA: Sage Publications, 1998:127-56.

Miles MB, Huberman AM. Qualitative data analysis: an expanded sourcebook. 2nd ed. Thousand Oaks, CA: Sage Publications, 1994.

Hennessy M. Evaluation. In: Stroup D, Teutsch S., eds. Statistics in Public Health. New York, NY: Oxford University Press, 1998:193-219.

Henry GT. Graphing data. In: Bickman L, Rog DJ, eds. Handbook of applied social research methods. Thousand Oaks, CA: Sage Publications, 1998:527-56.

Weick KE. Sensemaking in organizations. Thousand Oaks, CA: Sage Publications, 1995.

Rogers PJ, Hough G. Improving the effectiveness of evaluations: making the link to organizational theory. Evaluation and Program Planning 1995;18(4):321-32.

Adler M, Ziglio E. Gazing into the oracle: the Delphi method and its application to social policy and public health. Bristol, PA: Jessica Kingsley Publishers, 1996.

Torres RT, Preskill HS, Piontek ME. Evaluation strategies for communicating and reporting: enhancing learning in organizations. Thousand Oaks, CA: Sage Publications, 1996.

Zaltman G, Barabba VP. Hearing the voice of the market: competitive advantage through creative use of market information. Boston, MA: Harvard Business School Press, 1990.

Shulha LM, Cousins JB. Evaluation use: theory, research, and practice since 1986. *Evaluation Practice* 1997;18(3):195-208.

Weiss CH. Have we learned anything new about the use of evaluation? *American Journal of Evaluation* 1998;19(1):21-33.

Schorr LB. *Common purpose: strengthening families and neighborhoods to rebuild America*. New York, NY: Doubleday, Anchor Books, 1997.

Fawcett SB, Paine-Andrews A, Francisco VT, et al. Evaluating community initiatives for health and development. In: Rootman I, Goodstadt M, Hyndman B, et al., eds. *Evaluation in health promotion: principles and perspectives*. Copenhagen, Denmark: World Health Organization (Euro), 1999 (In press).

Collins J, Rugg D, Kann L, Banspach S, and Kolbe L. Evaluating a national program of school-based HIV prevention. *Evaluation and Program Planning* 1996;19(3):209-18.

Rugg D, Buehler J, Renaud M, et al. Evaluating HIV prevention: a framework for national, state, and local levels. *American Journal of Evaluation* 1999;20(1):35-56.

Greene J, Lincoln Y, Mathison S, Mertens DM, Ryan K. Advantages and challenges of using inclusive evaluation approaches in evaluation practice. *American Journal of Evaluation* 1998;19(1):101-22.

Smith NL. The context of investigations in cross-cultural evaluations. *Studies in Educational Evaluation* 1991;17:3-21.

Bonnet, DG. *An evaluation of year 1 of the American Cancer Society's collaborative evaluation fellows project*. Indianapolis, IN: D. Bonnet Associates, 1999.

National Institute of Standards and Technology. *National Quality Program*. National Institute of Standards and Technology, 1999. Available at <<http://www.quality.nist.gov>>. Accessed June 1999.

National Institute of Standards and Technology. *Baldrige index outperforms S&P 500 for fifth year*. National Institute of Standards and Technology, National Quality Program, 1999. Available at <[http://www.nist.gov/public\\_affairs/releases/n99-02.htm](http://www.nist.gov/public_affairs/releases/n99-02.htm)>. Accessed June 1999.

National Institute of Standards and Technology. *Health care criteria for performance excellence*. Gaithersburg, MD: National Institute of Standards and Technology, Baldrige National Quality Program, 1999. Also available at <<http://www.quality.nist.gov>>. Accessed June 1999.

Public Health Service. *Public health workforce: an agenda for the 21st century: a report of the public health functions project*. Washington, DC: US Department of Health and Human Services, Office of Disease Prevention and Health Promotion, 1997.

CDC. *Handbook for evaluating HIV education*. Atlanta, GA: US Department of Health and Human Services, CDC, National Center for Chronic Disease Prevention and Health Promotion, Division of Adolescent and School Health, 1993.

Thompson, NJ, McClintock HO. Demonstrating your program's worth: a primer on evaluation for programs to prevent unintentional injury. Atlanta: CDC, National Center for Injury Prevention and Control, 1998.

CDC. Guidelines for evaluating surveillance systems. MMWR 1988;37(No. SS-5):1-18.

Fawcett SB, Sterling TD, Paine-Andrews A, et al. Evaluating community efforts to prevent cardiovascular diseases. Atlanta, GA: U.S. Department of Health and Human Services, CDC, National Center for Chronic Disease Prevention and Health Promotion, 1995.

Rootman I, Goodstadt M, Hyndman B, et al., eds. Evaluation in health promotion: principles and perspectives. Copenhagen, Denmark. World Health Organization (Euro), 1999 (In press).

CDCynergy (Version 1.0) [CD ROM]. Atlanta, GA: U.S. Department of Health and Human Services, CDC, Office of Communication, 1998.

Linney JA, Wandersman A. Prevention plus III: assessing alcohol and other drug prevention programs at the school and community level: a four-step guide to useful program assessment. Rockville, MD: U.S. Department of Health and Human Services, Public Health Service, Alcohol, Drug Abuse, and Mental Health Administration, Office for Substance Abuse Prevention, 1991. publication no. (ADM)91-1817.

***NONPROFIT DEVELOPMENT INSTITUTE (NDI)***

***2500 W. 4<sup>TH</sup> STREET  
WILMINGTON, DE 19805***

***(302) 656-1100  
FAX: (302) 656-4600***

[www.nonprofitinstitute.com](http://www.nonprofitinstitute.com)