

The logo for the Centers for Disease Control and Prevention (CDC), featuring the letters "CDC" in a white serif font inside a blue square with a white grid pattern.The logo for the Morbidity and Mortality Weekly Report (MMWR), featuring the letters "MMWR" in a white bold sans-serif font inside a blue rounded rectangle.

Recommendations and Reports

September 17, 1999 / 48(RR11);1-40

Framework for Program Evaluation in Public Health

FRAMEWORK FOR PROGRAM EVALUATION IN PUBLIC HEALTH

Effective program evaluation is a systematic way to improve and account for public health 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.

The second element of the framework is a set of 30 standards for assessing the quality of evaluation activities, organized into the following four groups:

- Standard 1: utility,
- Standard 2: feasibility,
- Standard 3: propriety, and
- Standard 4: accuracy.

These standards, adopted from the Joint Committee on Standards for Educational Evaluation, answer the question, "Will this evaluation be effective?" and are recommended as criteria for judging the quality of program evaluation efforts in public health. The remainder of this report discusses each step, its subpoints, and the standards that govern effective program evaluation.

Steps in Program Evaluation

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 work involves partnerships; therefore, any assessment of a public health 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 benefitting 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 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.

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 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 disease) 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 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 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 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.