Step 3: Focus the Evaluation Design

After completing Steps 1 and 2, you and your stakeholders should have a clear understanding of the program and reached consensus. Now your evaluation team will need to focus the evaluation. This includes determining the most important evaluation questions and the appropriate design for the evaluation. Focusing the evaluation is based on the assumption that the entire program does not need to be evaluated at any point in time. Rather, the "right" evaluation of the program depends on what question is being asked, who is asking the question, and what will be done with the information.

Since resources for evaluation are always limited, this chapter provides a series of decision criteria to help you determine the best evaluation focus at any point in time. You will note that these criteria are inspired by the evaluation standards: specifically, utility (who will use the results and what information will be most useful to them) and feasibility (how much time and resources are available for the evaluation).

The logic models developed in the prior step set the stage for determining the best evaluation focus. The approach to evaluation focus in the CDC Evaluation Framework differs slightly from traditional evaluation approaches. In the past, some programs tended to assume all evaluations were "summative" ones, conducted when the program had run its course and intended to answer the question, "Did the program work?" Consequently, a key question was, "Is the program ready for evaluation?"

By contrast, the CDC Framework views evaluation as an ongoing activity over the life of a program that asks, "Is the program working?" Hence, a program is always ready for *some* evaluation. Because the logic model displays the program from inputs through activities/outputs through to the sequence of outcomes from short-term to most distal, it can guide a discussion of what you can expect to achieve at this point in the life of your project. Should you focus on distal outcomes, or only on short- or mid-term ones? Or conversely, does a process evaluation make the most sense right now?

Types of Evaluations

Many different questions can be part of a program evaluation, depending on how long the program has been in existence, who is asking the question, and why the evaluation information is needed. In general, evaluation questions for an existing program[1] fall into one of the following groups:

---

[1] There is another type of evaluation—"formative" evaluation—where the purpose of the evaluation is to gain insight into the nature of the problem so that you can "formulate" a program or intervention to address it. While many steps of the Framework will be helpful for formative evaluation, the emphasis in this manual is on instances wherein the details of the program/intervention are already known even though it may not yet have been implemented.

Implementation/Process

Implementation evaluations (more commonly called "process evaluations") document whether a program has been implemented as intended—"implementation fidelity"—and why or why not? In conducting process evaluations, you might examine whether the activities are taking place, who is conducting the activities, who is reached through the activities, and whether sufficient inputs have been allocated or mobilized. Process evaluation is important to help you distinguish the causes of poor program performance—was the program a bad idea, or was it a good idea that could not reach the standard for implementation that you set? In all cases, process evaluations measure whether actual program performance was faithful to some initial plan. This might include contrasting actual and planned performance on all or some of the following:

- The locale where services or programs are provided (e.g., rural, urban)
- The number of people receiving services
- The economic status and racial/ethnic background of people receiving services
- The quality of services
- The actual events that occur while the services are delivered
- The amount of money the project is using
- The direct and in-kind funding for services
- The staffing for services or programs
- The number of activities and meetings
- The number of training sessions conducted

When evaluation resources are limited, only the most important issues of implementation fidelity can be included. Here are some "usual suspects" that compromise implementation fidelity and should be considered for inclusion in the process evaluation portion of the evaluation focus:

- **Transfers of Accountability:** Where a program's activities cannot produce the intended outcomes unless some other person or organization takes appropriate action, there is a transfer of accountability.
- **Dosage:** The intended outcomes of program activities (e.g., training, case management, counseling) may presume a threshold level of participation or exposure to the intervention.
- **Access:** Where intended outcomes require not only an increase in consumer demand but also an increase in supply of services to meet it, then the process evaluation might include measures of access.
- **Staff Competency:** The intended outcomes may presume well-designed program activities that are delivered by staff who are not only technically competent but also are matched appropriately with the target audience. Measures of the match of staff and target audience might be included in the process evaluation.

Our childhood lead poisoning logic model illustrates many of these potential process issues. Reducing EBLL presumes the house will be cleaned, medical care referrals will be fulfilled, and specialty medical care will be provided. All of these are transfers of accountability beyond the program to the housing authority, the parent, and the provider, respectively. For provider training to achieve its outcomes, it may presume completion of a three-session curriculum, which is a dosage issue. Case management results in medical referrals, but it presumes adequate access to specialty

medical providers. And because lead poisoning tends to disproportionately affect children in low-income urban neighborhoods, many program activities presume cultural competence of the caregiving staff. Each of these components might be included in a process evaluation of a childhood lead poisoning prevention program.

Effectiveness/Outcome

Outcome evaluations assess progress on the sequence of outcomes that the program is to address. Programs often describe this sequence using terms like short-term, intermediate, and long-term outcomes, or proximal (close to the intervention) or distal (distant from the intervention). Depending on the stage of development of the program and the purpose of the evaluation, outcome evaluations may include any or all of the outcomes in the sequence, including

- Changes in people's attitudes and beliefs
- Changes in risk or protective behaviors
- Changes in the environment, including public and private policies, formal and informal enforcement of regulations, and influence of social norms and other societal forces
- Changes in trends in morbidity and mortality.

While process and outcome evaluations are the most common, there are several other types of evaluation questions that are central to a specific program evaluation. These include the following:
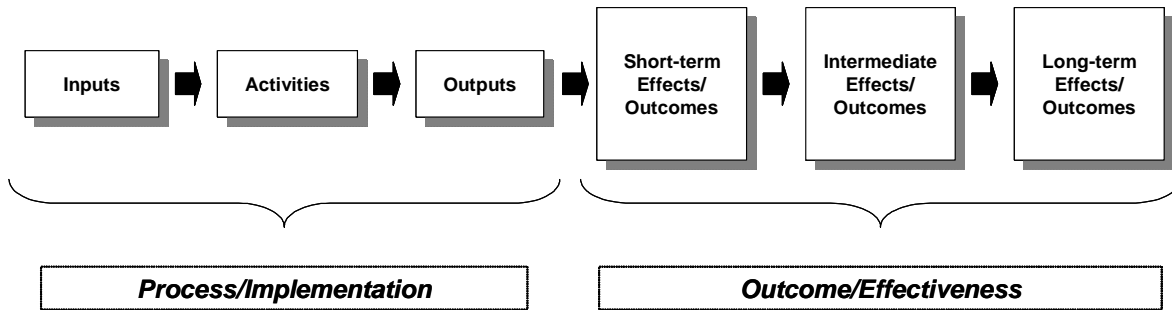
Efficiency: Are your program's activities being produced with minimal use of resources such as budget and staff time? What is the volume of outputs produced by the resources devoted to your program?

Cost-Effectiveness: Does the value or benefit of your program's outcomes exceed the cost of producing them?
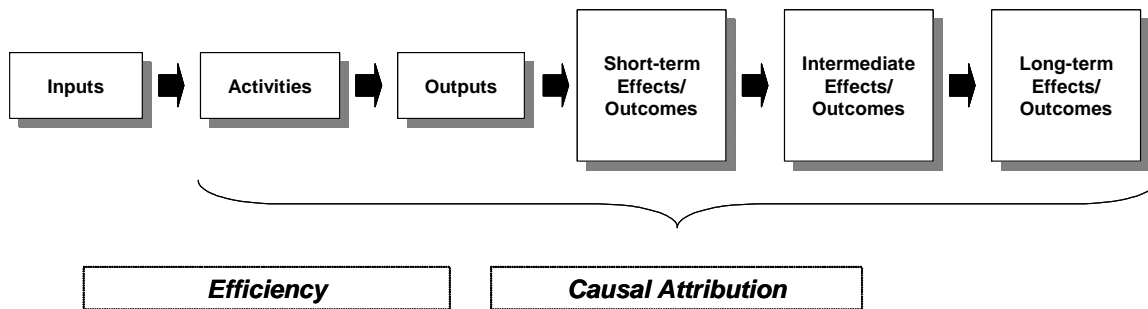
Attribution: Can the outcomes that are being produced be shown to be related to your program, as opposed to other things that are going on at the same time?

All of these types of evaluation questions relate to some part, but not all, of the logic model. Exhibits 3.1 and 3.2 show where in the logic model each type of evaluation would focus. Implementation evaluations would focus on the inputs, activities, and outputs boxes and not be concerned with performance on outcomes. Effectiveness evaluations would do the opposite—focusing on some or all *outcome boxes*, but not necessarily on the activities that produced them. Efficiency evaluations care about the *arrows linking inputs to activities/outputs*—how much output is produced for a given level of inputs/resources. Attribution would focus on the *arrows between specific activities/outputs and specific outcomes*—whether progress on the outcome is related to the specific activity/output.

Exhibit 3.1
Evaluation Domains — Boxes

| Inputs | → | Activities | → | Outputs | → | Short-term Effects/ Outcomes | → | Intermediate Effects/ Outcomes | → | Long-term Effects/ Outcomes |

Process/Implementation   *Outcome/Effectiveness*

Exhibit 3.2
Evaluation Domains — Arrows

| Inputs | → | Activities | → | Outputs | → | Short-term Effects/ Outcomes | → | Intermediate Effects/ Outcomes | → | Long-term Effects/ Outcomes |

Efficiency   *Causal Attribution*

Determining the Evaluation Focus

Determining the correct evaluation focus is a case-by-case decision. As noted, several guidelines inspired by the "utility" and "feasibility" evaluation standards can help determine the best focus.

Utility Considerations

1) **What is the purpose of the evaluation?**
   Purpose refers to the general intent of the evaluation. A clear purpose serves as the basis for the evaluation questions, design, and methods. Some common purposes:

   - Gain new knowledge about program activities
   - Improve or fine-tune existing program operations (e.g., program processes or strategies)
   - Determine the effects of a program by providing evidence concerning the program's contributions to a long-term goal
   - Affect program participants by acting as a catalyst for self-directed change (e.g., teaching)

2) Who will use the evaluation results?

Users are the individuals or organizations that will employ the evaluation findings in some way. The users will likely have been identified during Step 1 during the process of engaging stakeholders. In this step, you need to secure their input into the design of the evaluation and the selection of evaluation questions. Support from the intended users will increase the likelihood that the evaluation results will be used for program improvement.

3) How will they use the evaluation results?

Uses describe what will be done with what is learned from the evaluation, and many insights on use will have been identified in Step 1. Information collected may have varying uses, which should be described in detail when designing the evaluation. Some examples of uses of evaluation information:

- To document the level of success in achieving objectives
- To identify areas of the program that need improvement
- To decide how to allocate resources
- To mobilize community support
- To redistribute or expand the locations where the intervention is carried out
- To improve the content of the program's materials
- To focus program resources on a specific population
- To solicit more funds or additional partners

4) What do *other* key stakeholders need from the evaluation?

Of course, the most important stakeholders are those who are requesting or who will use the evaluation results. Nevertheless, in Step 1, you may also have identified stakeholders who, while they are not the users of the findings of the current evaluation, have key questions that may need to be addressed in the evaluation to keep them engaged. For example, a particular stakeholder may always be concerned about costs, disparities, or attribution. If so, and if that stakeholder is important long-term to credibility, implementation, or funding, then you may need to add those questions to your evaluation focus.

Feasibility Considerations

The first four questions help identify the most useful focus of the evaluation, but you must also determine whether it is a realistic/feasible one. Three questions provide a "reality check" on our desired focus:

5) What is the stage of development of the program?

During Step 2, you will have identified the program's stage of development. As noted, there are roughly three stages in program development: planning, implementation, and maintenance. These stages suggest different focuses. In the planning stage, a truly formative evaluation—who is your target, how do you reach them, how much will it cost—may be the most appropriate focus. An evaluation that included outcomes would make little sense at this stage. Conversely, an evaluation of a program in maintenance stage would need to include some measurement of progress on outcomes, even if it also included measurement of implementation.

6) How intensive is the program?

Some programs are wide-ranging and multifaceted. Others may use only one approach to address a large problem. Some programs provide extensive exposure ("dose") of the program, while others involve participants quickly and superficially. Simple or superficial programs, while potentially useful, cannot realistically be expected to make significant contributions to distal outcomes of a larger program, even when they are fully operational.

7) What are relevant resource and logistical considerations?
Resources and logistics may influence the decision about evaluation focus. Some outcomes are quicker, easier, and cheaper to measure, while others may not be measurable at all. In the short run, at least, these facts may tilt the decision about evaluation focus toward some outcomes as opposed to others.

Early identification of any inconsistencies between "utility" and "feasibility" is an important purpose of the evaluation focus step. For evaluation results to be used, the focus must include questions that matter to those who will implement or otherwise use the results. But we must also ensure a "meeting of the minds" on what is a realistic focus for program evaluation at any point in time.

The affordable housing example shows how the desired focus might be constrained by "reality." The elaborated logic model was important in this case because it clarified that, while program staff were focused on production of new houses, important stakeholders like community-based organizations and faith-based donors were committed to more distal outcomes such as changes in life outcomes of families or on the outcomes of outside investment in the community. The model led to a discussion of reasonableness of expectations and, in the end, to expanded evaluation indicators that included some of the more distal outcomes, but also to a greater appreciation by stakeholders of the intermediate milestones on the way to their preferred outcomes.

Are You Ready to Evaluate Outcomes?

While it is understood that the evaluation focus of the program will shift over time, here are some handy decision rules to decide whether it is time to shift the evaluation focus toward an emphasis on program outcomes:

- **Sustainability:** Political and financial will exists to sustain the intervention while the evaluation is conducted.

- **Fidelity:** Actual intervention implementation matches intended implementation. Erratic implementation makes it difficult to know what "version" of the intervention was implemented and, therefore, which version produced the outcomes.

- **Stability:** Intervention is not likely to change during the evaluation. Changes to the intervention over time will confound understanding of which aspects of the intervention caused the outcomes.

- **Reach:** Intervention reaches a sufficiently large number of clients (sample size) to employ the proposed data analysis. For example, the number of clients needed may vary with the magnitude of the change expected in the variables of interest (i.e., effect size) and the power needed for statistical purposes.

- **Dosage:** Clients have sufficient exposure to the intervention to result in the intended outcomes. Interventions with limited client contact are less likely to result in measurable outcomes as compared to interventions that provide more in-depth intervention with clients.

Illustrating Evaluation Focus Decisions

Because the appropriate evaluation focus is case-specific, let's apply these focus issues to a few different evaluation scenarios for the CLPP program. Think about two scenarios and how evaluation focus might differ for each.

- **Scenario 1**

  At the 1-year mark, a neighboring community would like to adopt your program but wonders, "What are we in for?" Here you might determine that questions of efficiency and implementation are central to the evaluation. You would likely conclude this is a realistic focus, given the stage of development and the intensity of the program. Questions about outcomes would be premature.

- **Scenario 2**

  At the 5-year mark, the auditing branch of your government funder wants to know, "Did you spend our money well?" Clearly, this requires a much more comprehensive evaluation, and would entail consideration of efficiency, effectiveness, possibly implementation, and cost-effectiveness. It is not clear, without more discussion with the stakeholder, whether research studies to determine causal attribution are also implied. Is this a realistic focus? At year 5, probably yes. The program is a significant investment in resources and has been in existence for enough time to expect some more distal outcomes to have occurred.

Note that in either scenario, you must also consider questions of interest to key stakeholders who are not necessarily intended users of the results of the current evaluation. Here those were defined to be advocates, who are concerned that families not be blamed for lead poisoning in their children, and housing authority staff, who are concerned that amelioration include estimates of costs and identification of less costly methods of lead reduction in homes. By year 5, these look like reasonable questions to include in the evaluation focus. At year 1, stakeholders might need assurance that you care about their questions, even if you cannot address them with this early evaluation.

Defining the Specific Evaluation Questions

These focus criteria just discussed identify the components of the logic model that are to be included in the evaluation focus, i.e., these activities, but not these; these outcomes, but not these. At this point, you convert the components of your focus into specific questions, i.e., implementation, effectiveness, efficiency, and attribution. Were my activities implemented as planned? Did my intended outcomes occur? Were the outcomes due to my activities as opposed to something else? If the outcomes occurred at some but not all sites, what barriers existed at less successful locations and what factors were related to success? At what cost were my activities implemented and my outcomes achieved?

Deciding On the Evaluation Design

Besides determining the evaluation focus and specific evaluation questions, at this point you also need to determine the appropriate evaluation design. There are many types of evaluation designs. Of chief interest in choosing the evaluation design is whether you are being asked to monitor progress on outcomes or whether you are also asked to show "attribution"—that progress on outcomes is related to your program efforts. These "attribution" questions may more appropriately be viewed as "research" as opposed to "program evaluation" depending on the level of scrutiny with which they are being asked.

Three general types of research designs are commonly recognized: experimental, quasi-experimental, and non-experimental/observational. Traditional program evaluation typically uses the third type, but all three are presented here because, over the life of the program, traditional evaluation approaches may need to be supplemented with other studies that look more like research.

Experimental designs use random assignment to compare the outcome of an intervention on one or more groups with an equivalent group or groups that did not receive the intervention. For example, a you could select a group of similar schools, and then randomly assign some schools to receive a prevention curriculum and other schools to serve as controls. All schools have the same chance of being selected as an intervention or control school. Because of the random assignment, you reduce the chances that the control and intervention schools vary in any way that could influence differences in program outcomes. This allows you to attribute change in outcomes to your program. For example, if the students in the intervention schools delayed onset or risk behavior longer than students in the control schools, you could attribute the success to your program.

However, in community settings it is hard, or sometimes even unethical, to have a true control group. While there are some solutions that preserve the integrity of experimental design, another option is to use a *quasi-experimental design*. These designs make comparisons between nonequivalent groups and do not involve random assignment to intervention and control groups. An example would be to assess adults' beliefs about the harmful outcomes of environmental tobacco smoke (ETS) in two communities, then conduct a media campaign in one of the communities. After the campaign, you would reassess the adults and expect to find a higher percentage of adults believing ETS is harmful in the community that received the media campaign. Critics could argue that other differences between the two communities caused the changes in beliefs, so it is important to document that the intervention and comparison groups are similar on key factors such as population demographics and related current or historical events.

Related to quasi-experimental design, comparing outcomes/outcome data among states and between one state and the nation as a whole are common and important ways to evaluate public health efforts. Such comparisons will help you establish meaningful benchmarks for progress. States can also compare their progress with that of states with a similar investment in their area of public health, or they can contrast their outcomes with the results that could be expected if their programs were similar to those of states with a larger investment.

Comparison data are also useful for measuring indicators in anticipation of new or expanding programs. For example, noting a "lack of change" in key indicators over time prior to program implementation helps demonstrate the need for your program and highlights the comparative

progress of states with comprehensive public health programs already in place. A lack of change in indicators may continue for several years and is useful as a justification for greater investment in evidence-based, well-funded, and more comprehensive programs. There are many opportunities for between-state comparisons, which can be highlighted with time–series analyses. For example, questions on many of the larger national surveillance systems have not changed in several years, so you can make comparisons with other states and over time, using specific indicators. Program managers are encouraged to collaborate with state epidemiologists, surveillance coordinators, and statisticians to make state and national comparisons an important component of your evaluation.

Observational designs are common in program evaluation. These include, but are not limited to, time–series analysis, cross-sectional surveys, and case studies. Periodic cross-sectional surveys (e.g.., the YTS or BRFSS) can inform your evaluation. Case studies may be particularly appropriate for assessing changes in public health capacity in disparate population groups. Case studies are often applicable when the program is unique, when an existing program is used in a different setting, when a unique outcome is being assessed, or when an environment is especially unpredictable. Case studies can also allow for an exploration of community characteristics and how these may influence program implementation, as well as identifying barriers to and facilitators of change.

This issue of "causal attribution," while often a central research question, may or may not need to supplement traditional program evaluation. The field of public health is under increasing pressure to demonstrate that programs are worthwhile, effective, and efficient. During the last two decades, knowledge and understanding about how to evaluate complex programs have increased significantly. Nevertheless, because programs are so complex, these traditional research designs described here may not be a good choice. As the World Health Organization notes, "the use of randomized control trials to evaluate health promotion initiatives is, in most cases, inappropriate, misleading, and unnecessarily expensive."[2]

Therefore, before choosing experimental or even quasi-experimental designs to supplement more traditional program evaluation, consider the appropriateness and feasibility of less traditional designs (e.g., simple before–after [pretest–posttest] or posttest-only designs). Depending on your program's objectives and the intended use(s) for the evaluation findings, these designs may be more suitable for measuring progress toward achieving program goals. Even when there is desire or need to "prove" that the program was responsible for progress on outcomes, traditional research designs may not be the only or best alternative. Depending on how rigorous the proof needs to be, proximity in time between the implementation of the program and the progress on outcomes, or systematic elimination of other alternative explanations may be enough to persuade key stakeholders that the program is making a contribution. While these design alternatives often cost less and require less time, keep in mind that saving time and money should not be the main criterion when selecting an evaluation design. It is important to choose a design that will measure what you need to measure and that will meet both your immediate and long-term needs.

Another alternative to experimental and quasi-experimental models is a goal-based evaluation model, which uses predetermined program goals and the underlying program theory as the standards for evaluation, thus holding the program accountable to prior expectations. The CDC Framework's emphasis on program description and the construction of a logic model sets the stage for strong goal-

---

[2] WHO European Working Group on Health Promotion Evaluation. op cit.

based evaluations of programs. In such cases, evaluation planning focuses on the activities; outputs; and short-term, intermediate, and long-term outcomes outlined in a program logic model to direct the measurement activities.

The design you select influences the timing of data collection, how you analyze the data, and the types of conclusions you can make from your findings. A collaborative approach to focusing the evaluation provides a practical way to better ensure the appropriateness and utility of your evaluation design.

Standards for Step 3
Focus the Evaluation Design

Standard	Questions
Utility	• What is the purpose of the evaluation? • Who will use the evaluation results and how will they use them? • What are special needs of any other stakeholders that must be addressed?
Feasibility	• What is the program's stage of development? • How intense is the program? • How measurable are the components in the proposed focus?
Propriety	• Will the focus and design adequately detect any unintended consequences? • Will the focus and design include examination of the experience of those who are affected by the program?
Accuracy	• Is the focus broad enough to detect success or failure of the program? • Is the design the right one to respond to the questions—such as attribution—that are being asked by stakeholders?

Checklist for Focusing the Evaluation Design

☐ Define the purpose(s) and user(s) of your evaluation.

☐ Identify the use(s) of the evaluation results.

☐ Consider stage of development, program intensity, and logistics and resources.

☐ Determine the components of your logic model that should be part of the focus given these "utility and "feasibility" considerations.

☐ Formulate the evaluation questions to be asked of the program components in your focus, i.e., implementation, effectiveness, efficiency, and attribution questions.

☐ Review evaluation questions with stakeholders, program managers, and program staff.

☐ Review options for the evaluation design, making sure that the design *fits* the evaluation questions.

Worksheet 3A
Focusing the Evaluation in the Logic Model

#	If this is the situation …	Then these are the parts of the logic model, I would include in my evaluation focus:
1	Who is asking evaluation questions of the program?	
2	Who will use the evaluation results and for what purpose?	
3	In Step 1, did we identify interests of other stakeholders that we must take into account?	

Worksheet 3B
"Reality Checking" the Evaluation Focus

#	If this is my answer to these questions…	Then I would conclude the questions in my evaluation focus are/are not reasonable ones to ask right now.
1	How long has the intervention been underway?	
2	How intensive/ambitious is the intervention? Multi-faceted effort or simple intervention?	
3	How much resources (time and money) are able to be devoted to evaluation of this effort?	

Step 4: Gather Credible Evidence

Now that you have developed a logic model, chosen an evaluation focus, and selected your evaluation questions, your next task is to gather the evidence. The gathering of evidence for an evaluation resembles the gathering of evidence for any research or data-oriented project, with a few exceptions noted below.

What's Involved in Gathering Evidence?

Evidence gathering must include consideration of each of the following:

- Indicators
- Sources of evidence/methods of data collection
- Quality
- Quantity
- Logistics

Developing Indicators

Because the components of our programs are often expressed in global or abstract terms, indicators are specific, observable, and measurable statements that help define exactly what we mean or are looking for. For example, the CLPP model includes global statements such as "Children receive medical treatment" or "Families adopt in-home techniques." The medical treatment indicator might specify the type of medical treatment, the duration, or perhaps the adherence to the regimen. Likewise, the family indicator might indicate the in-home techniques or the intensity or duration of their adoption. For example, "Families with EBLL children clean all window sills and floors with the designated cleaning solution each week" or "Families serve leafy green vegetables at three or more meals per week." *Outcome indicators* such as these indicators provide clearer definitions of the global statement and help guide the selection of data collection methods and the content of data collection instruments.

The activities in your focus may *also* include global statements such as "good coalition," "culturally competent training," and "appropriate quality patient care." These activities would benefit from elaboration into indicators, often called "*process indicators*." What does "good" mean, what does "quality" or "appropriate" mean?

Keep the following tips in mind when selecting your indicators:

- Indicators can be developed for activities (process indicators) and/or for outcomes (outcome indicators).[3]
- There can be more than one indicator for each activity or outcome.

---

[3] Note that if you are developing your evaluation after completing an evaluation plan, you may already have developed process or outcome *objectives*. If the objectives were written to be specific, measurable, action-oriented, realistic, and time-bound (so-called "SMART" objectives), then they may serve as indicators as well.

- The indicator must be focused and must measure an important dimension of the activity or outcome.
- The indicator must be clear and specific in terms of what it will measure.
- The change measured by the indicator should represent progress toward implementing the activity or achieving the outcome.

Consider CDC's immunization program, for example. The table below lists the components of the logic model that were included in our focus in Step 3. Then each of these components has been defined in one or more indicators.

Table 4.1
Provider Immunization Program:
Indicators for Program Component in Our Evaluation Focus

Program Component	Indicator(s)
Provider training	A series of 3 trainings will be conducted in all 4 regions of the state
Nurse educator LHD presentations	Nurse educators will make presentations to 10 largest local health departments (LHDs)
Physicians peer ed rounds	Physicians will host peer ed rounds at 10 largest hospitals
Providers attend trainings and rounds	Trainings will be well attended and reflect good mix of specialties and geographic representation
Providers receive and use tool kits	50%+ of providers who receive tool kit will report use of it (or "call to action" cards will be received from 25% of all providers receiving tool kit)
LHD nurses conduct private provider consults	Trained nurses in LHDs will conduct provider consults with largest provider practices in county
Provider KAB increases	Providers show increases in knowledge, attitudes, and beliefs (KAB) on selected key immunization items
Provider motivation increases	Provider intent to immunize increases

You may need to develop your own indicators or you may be able to draw on existing indicators developed by others. Some large CDC programs have developed indicator inventories that are tied to major activities and outcomes for the program. Advantages of these indicator inventories:

- They may have been pre-tested for "relevance" and accuracy.
- They define the best data sources for collecting the indicator.
- There are often many potential indicators for each activity or outcome, ensuring that at least one will be appropriate for your program.
- Because many programs are using the same indicator(s), you can compare performance across programs or even construct a national summary of performance.

Selecting Data Collection Methods and Sources

Now that you have determined the activities and outcomes you want to measure and the indicators you will use to measure progress on them, you need to select data collection methods and sources from which to gather information on your indicators.

A key decision is whether there are existing data sources—*secondary* data collection—to measure your indicators or whether you need to collect new data—*primary* data collection.

Depending on your evaluation questions and indicators, some secondary data sources may be appropriate data collection sources. Some existing data sources that often come into play in measuring outcomes of public health programs:

- Current Population Survey and other U.S. Census files
- Behavioral Risk Factor Surveillance System (BRFSS)
- Youth Risk Behavior Survey (YRBS)
- Pregnancy Risk Assessment Monitoring System (PRAMS)
- Cancer registries
- State vital statistics
- Various surveillance databases
- National Health Interview Survey (NHIS)

Before using secondary data sources, ensure that they meet your needs. Although large ongoing surveillance systems have the advantages of collecting data routinely and having existing resources and infrastructure, some of them (e.g., Current Population Survey [CPS]) have little flexibility with regard to the questions asked in the survey, making it nearly impossible to use these systems to collect the special data you may need for your evaluation. By contrast, other surveys such as BRFSS or PRAMS are more flexible. For example, you might be able to add program-specific questions, or you might expand the sample size for certain geographic areas or target populations, allowing for more accurate estimates in smaller populations.

The most common primary data collection methods also fall into several broad categories. Among the most common are:

- Surveys, including personal interviews, telephone, or instruments completed in person or received through the mail or e-mail
- Group discussions/focus groups
- Observation
- Document review, such as medical records, but also diaries, logs, minutes of meetings, etc.

Choosing the "right" method from the many secondary and primary data collection choices must consider both the **context** in which it is asked (How much money can be devoted to collection and measurement? How soon are results needed? Are there ethical considerations?) and the **content** of the question (Is it a sensitive issue? Is it about a behavior that is observable? Is it something the respondent is likely to know?).

Some methods yield qualitative data and some yield quantitative data. If the question involves an abstract concept or one where measurement is poor, using multiple methods is often helpful. Insights from stakeholder discussions in Step 1 and the clarity on purpose/user/use obtained in Step 3 will usually help direct the choice of sources and methods. For example, stakeholders may know which methods will work best with some intended respondents and/or have a strong bias toward quantitative or qualitative data collection that must be honored if the results are to be credible. More importantly, the purpose and use/user may dictate the need for valid, reliable data that will withstand close scrutiny or may allow for less rigorous data collection that can direct managers.

Each method comes with advantages and disadvantages depending on the context and content of the data collection (see Table 4.2).

Table 4.2
Advantages and Disadvantages of Various Survey Methods

Method	Advantages	Disadvantages
Personal interviews	• Least selection bias: can interview people without telephones—even homeless people. • Greatest response rate: people are most likely to agree to be surveyed when asked face to face. • Visual materials may be used.	• Most costly: requires trained interviewers and travel time and costs. • Least anonymity: therefore, most likely that respondents will shade their responses toward what they believe is socially acceptable.
Telephone interviews	• Most rapid method. • Most potential to control the quality of the interview: interviewers remain in one place, so supervisors can oversee their work. • Easy to select telephone numbers at random. • Less expensive than personal interviews. • Better response rate than for mailed surveys.	• Most selection bias: omits homeless people and people without telephones. • Less anonymity for respondents than for those completing instruments in private. • As with personal interviews, requires a trained interviewer.
Instruments to be completed by respondent	• Most anonymity: therefore, least bias toward socially acceptable responses. • Cost per respondent varies with response rate: the higher the response rate, the lower the cost per respondent. • Less selection bias than with telephone interviews.	• Least control over quality of data. • Dependent on respondent's reading level. • Mailed instruments have lowest response rate. • Surveys using mailed instruments take the most time to complete because such instruments require time in the mail and time for respondent to complete.

The text box below lists possible sources of information for evaluations clustered in three broad categories: people, observations, and documents.

When choosing data collection methods and sources, select those that meet your project's needs. Try to avoid choosing a data method/source that may be familiar or popular but does not necessarily answer your questions. Keep in mind that budget issues alone should not drive your evaluation planning efforts.

The four evaluation standards can help you reduce the enormous number of data collection options to a more manageable number that best meet your data collection situation. Here is a checklist of issues — based on the evaluation standards — that will help you choose appropriately:

Utility

- Purpose and use of data collection: Do you seek a "point in time" determination of a behavior, or to examine the range and variety or experiences, or to tell an in-depth story?
- Users of data collection: Will some methods make the data more credible with skeptics or key users than others?

Feasibility

- Resources available: Which methods can you afford?
- Time: How long until the results are needed?
- Frequency: How often do you need the data?
- Your background: Are you trained in the method, or will you need help from an outside consultant?

Propriety

- Characteristics of the respondents: Will issues such as literacy or language make some methods preferable to others?
- Degree of intrusion to program/participants: Will the data collection method disrupt the program or be seen as intrusive by participants?
- Other ethical issues: Are there issues of confidentiality or safety of the respondent in seeking answers to questions on this issue?

Accuracy

- Nature of the issue: Is it about a behavior that is observable?
- Sensitivity of the issue: How open and honest will respondents be in responding to the questions on this issue?
- Respondent knowledge: Is it something the respondent is likely to know?

Using Multiple Methods and Mixed Methods

Sometimes a single method is not sufficient to accurately measure an activity or outcome because the thing being measured is complex and/or the data method/source does not yield data that are reliable or accurate enough. Employing multiple methods (sometimes called "triangulation") helps increase the accuracy of the measurement and the certainty of your conclusions when the various methods yield similar results. Mixed data collection methods refers to gathering both quantitative and qualitative data. Mixed methods can be used sequentially, when one method is used to prepare for the use of another, or concurrently, when both methods are used in parallel. An example of sequential use of mixed methods is when focus groups (qualitative) are used to develop a survey instrument (quantitative), and then personal interviews (qualitative and quantitative) are conducted to investigate issues that arose during coding or interpretation of survey data. An example of concurrent use of mixed methods would be using focus groups or open-ended personal interviews to help affirm the response validity of a quantitative survey.

Different methods reveal different aspects of the program. Consider some interventions related to tobacco control:

- You might include a group assessment of a school-based tobacco control program to hear the group's viewpoint, as well as individual student interviews to get a range of opinions.
- You might conduct a survey of all legislators in a state to gauge their interest in managed care support of cessation services and products, and you might also interview certain legislators individually to question them in greater detail.

- You might conduct a focus group with community leaders to assess their attitudes regarding tobacco industry support of cultural and community activities. You might follow the focus group with individual structured or semistructured interviews with the same participants.

When the outcomes under investigation are very abstract or no one quality data source exists, combining methods maximizes the strengths and minimizes the limitations of each method. Using multiple or mixed methods can increase the cross-checks on different subsets of findings and generate increased stakeholder confidence in the overall findings.

Illustrations from Cases

Consider the provider immunization education and the childhood lead poisoning examples. Table 4.3 presents data collection methods/sources for each of the indicators presented earlier for the provider immunization education program. Table 4.4 shows both the indicators and the data sources for key components of the CLPP effort presented earlier. Note in both cases that the methods/sources can vary widely and that in some cases multiple methods will be used and synthesized.

Table 4.3
Provider Immunization Education Program:
Data Collection Methods and Sources for Indicators

Indicator(s)	Data Collection Methods/Sources
A series of 3 trainings will be conducted in all 4 regions of the state	Training logs
Nurse educators will make presentations to 10 largest local health departments (LHDs)	Training logs
Physicians will host peer ed rounds at 10 largest hospitals	Training logs
Trainings will be well-attended and reflect good mix of specialties and geographic representation	Registration information
50%+ of providers who receive tool kit will report use of it (or "call to action" cards will be received from 25% of all providers receiving tool kit)	Survey of providers Analysis/count of call-to-action cards
Trained nurses in LHDs will conduct provider consults with largest provider practices in county	Survey of nurses, survey of providers, or training logs
Providers show increases in knowledge, attitudes, and beliefs (KAB) on selected key immunization items	Survey of providers, or focus groups, or intercepts
Provider intent to immunize increases	Survey of providers, or focus groups, or intercepts

Table 4.4
CLPP: Indicators and Data Collection Methods/Sources

Logic Model Element	Indicator(s)	Data Source(s) and Method(s)
Outreach	High-risk children and families in the district have been reached with relevant information	Logs of direct mail and health fair contacts

Demographic algorithm

Geographic Information System (GIS) algorithm |
Screening	High-risk children have completed initial and follow-up screening	Logs and lab data
Environment assessment	Environments of all children over EBLL threshold have been assessed for lead poisoning	Logs of environmental health staff
Case management	All children over EBLL threshold have a case management plan including social, medical, and environmental components	Case file of EBLL child
Family training	Families of all children over EBLL threshold have received training on household behaviors to reduce EBLL	Logs of case managers

Survey of families |
| "Leaded" houses referred | All houses of EBLL children with evidence of lead have been referred to housing authority | Logs and case files |
| "Leaded" houses cleaned | All referred houses have been cleaned up | Follow-up assessment by environmental health staff

Logs of housing authority |

Quality of Data

A quality evaluation produces data that are reliable, valid, and informative. An evaluation is reliable to the extent that it repeatedly produces the same results, and it is valid if it measures what it is intended to measure. The advantage of using existing data sources such as the BRFSS, YRBS, or PRAMS is that they have been pretested and designed to produce valid and reliable data. If you are designing your own evaluation tools, you should be aware of the factors that influence data quality:

- The design of the data collection instrument and how questions are worded
- The data collection procedures
- Training of data collectors
- The selection of data sources
- How the data are coded
- Data management
- Routine error checking as part of data quality control

A key way to enhance quality of primary data collection is through a pretest. The pretest need not be elaborate but should be extensive enough to determine issues of logistics of data collection or intelligibility of instruments prior to rollout. Obtaining quality data involves trade-offs (i.e., breadth vs. depth). Thus, you and stakeholders must decide at the beginning of the evaluation process what level of quality is necessary to meet stakeholders' standards for accuracy and credibility.

Quantity of Data

You will also need to determine the amount of data you want to collect during the evaluation. There are cases where you will need data of the highest validity and reliability, especially when traditional program evaluation is being supplemented with research studies. But there are other instances where the insights from a few cases or a convenience sample may be appropriate. If you use secondary data sources, many issues related to quality of data—such as sample size—have already been determined. If you are designing your own data collection tool and your examination of your program includes research as well as evaluation questions, the quantity of data you need to collect (i.e., sample sizes) will vary with the level of detail and the types of comparisons you hope to make. You will also need to determine the jurisdictional level for which you are gathering the data (e.g., state, county, region, congressional district). Counties often appreciate and want county-level estimates; however, this usually means larger sample sizes and more expense. Finally, consider the size of the change you are trying to detect. In general, detecting small amounts of change requires larger sample sizes. For example, detecting a 5% increase would require a larger sample size than detecting a 10% increase. You may need the help of a statistician to determine adequate sample size.

Logistics and Protocols

Logistics are the methods, timing, and physical infrastructure for gathering and handling evidence. People and organizations have cultural preferences that dictate acceptable ways of asking questions and collecting information, and influence who is perceived as an appropriate person to ask the questions (i.e., someone known within the community versus a stranger from a local health agency). The techniques used to gather evidence in an evaluation must be in keeping with a given community's cultural norms. Data collection procedures should also protect confidentiality.

In outlining procedures for collecting the evaluation data, consider these issues:

- When will you collect the data? You will need to determine when (and at what intervals) it is most appropriate to collect the information. If you are measuring whether your objectives have been met, your objectives will provide guidance as to when to collect certain data. If you are evaluating specific program interventions, you might want to obtain information from participants before they begin the program, upon completion of the program, and several months after the program. If you are assessing the effects of a community campaign, you might want to assess community knowledge, attitudes, and behaviors among your target audience before and after the campaign.
- Who will be considered a participant in the evaluation? Are you targeting a relatively specific group (African-American young people), or are you assessing trends among a more

general population (all women of childbearing age)?

- Are you going to collect data from all participants or a sample? Some programs are community-based, and surveying a sample of the population participating in such programs is appropriate. However, if you have a small number of participants (such as students exposed to a curriculum in two schools), you may want to survey all participants.
- Who will collect the information? Are those collecting the data trained and trained consistently? Will the data collectors uniformly gather and record information? Your data collectors will need to be trained to ensure that they all collect information in the same way and without introducing bias. Preferably, interviewers should be trained together and by the same person.
- How will the security and confidentiality of the information be maintained? It is important to ensure the privacy and confidentiality of the evaluation participants. You can do this by collecting information anonymously and making sure you keep data stored in a locked and secure place.
- If your examination of your program includes research as well as evaluation studies: Do you need approval from an institutional review board (IRB) before collecting the data? What will be your informed consent procedures?

You may already have answered some of these questions while selecting your data sources and methods.

Agreements: Affirming Roles and Responsibilities

Agreements summarize the evaluation procedures, clarify everyone's role and responsibilities, and describe how the evaluation procedures will be implemented. Elements of an agreement include statements concerning the intended users, uses, purpose, questions, design, and methods, as well as a summary of the deliverables, timeline, and budget. An agreement might be a legal contract, a memorandum of understanding, or a detailed protocol. Creating an agreement establishes a mutual understanding of the activities associated with the evaluation. It also provides a basis for modification if necessary.

Standards for Step 4: Gather Credible Evidence

Standard	Questions
Utility	- Have key stakeholders been consulted who can assist with access to respondents? - Are methods and sources appropriate to the intended purpose and use of the data? - Have key stakeholders been consulted to ensure there are no preferences for or obstacles to selected methods or sources? - Are there specific methods or sources that will enhance the credibility of the data with key user and stakeholders?
Feasibility	- Can the data methods and sources be implemented within the time and budget for the project? - Does the evaluation team have the expertise to implement the chosen methods? - Are the methods and sources consistent with the culture and characteristics of the respondents, such as language and literacy level? - Are logistics and protocols realistic given the time and resources that can be devoted to data collection?
Propriety	- Will data collection be unduly disruptive? - Are there issues of safety of respondents or confidentiality that must be addressed? - Are the methods and sources appropriate to the culture and characteristics of the respondents—will they understand what they are being asked?
Accuracy	- Are appropriate QA procedures in place to ensure quality of data collection? - Are enough data being collected,—i.e., to support chosen confidence levels or statistical power? - Are methods and sources consistent with the nature of the problem, the sensitivity of the issue, and the knowledge level of the respondents?

Checklist for Gathering Credible Evidence

☐ Identify indicators for activities and outcomes in the evaluation focus.

☐ Determine whether existing indicators will suffice or whether new ones must be developed.

☐ Consider the range of data sources and choose the most appropriate one.

☐ Consider the range of data collection methods and choose those best suited to your context and content.

☐ Pilot test new instruments to identify and/or control sources of error.

☐ Consider a mixed-method approach to data collection.

☐ Consider quality and quantity issues in data collection.

☐ Develop a detailed protocol for data collection.

Worksheet 4A
Evaluation Questions, Indicators, and Data Collection Methods/Sources

	Logic Model Components in Evaluation Focus	Indicator(s) or Evaluation Questions	Data Method(s)/Source(s)
1			
2			
3			
4			
5			
6			
7			
8			
9			
10			

Worksheet 4B
Data Collection Logistics

	Data Collection Method/Source	From whom will these data be collected	By whom will these data be collected and when	Security or confidentiality steps
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				