

Guidelines for Evaluating an EPA Partnership Program (Interim)



The National Center for Environmental Innovation wishes to thank the members of the Partnership Programs Evaluation Guidelines Workgroup and other contributors at EPA and outside of EPA who assisted in this effort.

This document was developed for use by EPA managers and staff, as well as their program evaluation contractors, as they consider program evaluation for Partnership Programs.

United States Environmental Protection Agency
National Center for Environmental Innovation (I807T)
Washington, D.C.
March 2009



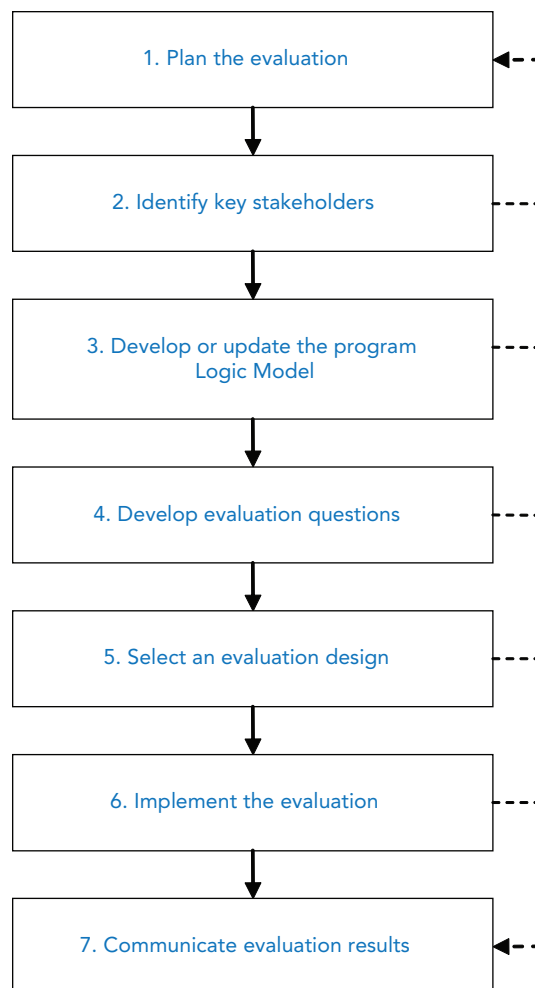
Executive Summary

The U.S. Environmental Protection Agency (EPA) defines its Partnership Programs as those programs designed to proactively target and motivate external parties to take specific, voluntary environmental actions. They do not compel these actions through legal means, but rather serve as a leadership and decision-making authority for the partners.

EPA Partnership Programs vary greatly in style, type, and function; however, they all share the need to demonstrate that they are achieving environmental results and supporting EPA's mission. The Agency therefore identified a need for program evaluation guidelines specific to Partnership Programs. These guidelines should be used in conjunction with *Guidelines for Designing EPA Partnership Programs*, *Guidelines for Marketing EPA Partnership Programs*, and *Guidelines for Measuring the Performance of EPA Partnership Programs*.

Why Evaluate?

Stakeholders are increasingly interested in ensuring that EPA Partnership Programs are adequately evaluated, to determine whether they are well designed and effective. Program evaluation is important for learning about programs



and improving them. Evaluations can produce the data needed to respond to and answer key management questions and accountability demands, identifying why a program has or has not met its goals.

Program Evaluation's Role in Performance Management

Program evaluation is one distinct tool in the “performance management suite,” building upon logic modeling and performance measurement. Program evaluation provides a systematic assessment of program elements by drawing conclusions about the effectiveness of a program's design, implementation, and impacts.

How To Use These Guidelines

At its most sophisticated level, program evaluation can be a very complex discipline, with practitioners devoting entire careers to narrow aspects of the field. These guidelines do not assume that you are such an expert, nor do they aim to make you one. They are intended to introduce the novice to the world of program evaluation. These guidelines walk you through a seven-chapter framework for how to design and conduct an evaluation for an individual Partnership Program. This framework will enable you to work more effectively with a contractor or evaluation expert.

Steps of an Evaluation

1. Plan the Evaluation

When planning a program evaluation:

- Choose the right evaluation for your Partnership Program by determining whether you will conduct a design, process, outcome, or impact evaluation.

- Decide whether the evaluation should be internal (i.e., conducted by EPA staff and supporting contractors) or external (i.e., conducted by third-party evaluators who will operate at an “arms length” from your program).
- Budget for an evaluation by considering the relevant fiscal and resource constraints.
- Anticipate potential data limitations and stakeholder concerns by planning to address limitations in current data sources, barriers to collecting new data, and potential stakeholder concerns.

2. Identify Key Stakeholders

A stakeholder is broadly defined as any person or group who has an interest in the program being evaluated or in the results of the evaluation. Your Partnership Program should incorporate a variety of stakeholder perspectives in the planning and implementation stages of the evaluation. This inclusiveness will provide many benefits, including fostering a greater commitment to the evaluation process, ensuring that the evaluation is properly targeted, and increasing the chances that evaluation results are implemented. Key steps include:

- Identifying relevant stakeholder groups and determining the appropriate level of involvement for each group.
- Incorporating a variety of perspectives by considering people or organizations involved in the program operations, people or organizations affected by your program, primary intended users of the evaluation results, and Agency planners. Try to identify a program staff person or other individual with knowledge of the program who will ask tough,

critical questions about your program and evaluation process.

- Choosing how to involve your stakeholders by using a method, or combination of methods, that works for the people in the group (e.g., face-to-face meetings, conference calls, electronic communications).
- Developing a stakeholder involvement plan, that is as formal or informal as the situation warrants.

3. Develop or Update the Program Logic Model

A logic model is a diagram and text that shows the relationship between your program's work and its desired results. Having a clear picture of your program is essential to conducting a quality evaluation because it helps ensure that you are evaluating the right aspects of your program and asking the questions that will be most helpful. A logic model has seven basic program elements:

1. **Resources/Inputs**—What you have to run your program (e.g., people, dollars).
2. **Activities**—What your program does.
3. **Outputs**—The products/services your program produces or delivers.
4. **Target Decision-Makers**—Those groups whose behavior your program aims to affect.
5. **Short-Term Outcomes**—Changes in target decision-makers' knowledge, attitude, or skills.
6. **Intermediate-Term Outcomes**—Changes in target decision-makers' behavior, practices, or decisions.

7. **Long-Term Outcomes**—Changes in public health and/or the environment as a result of your program.

Also included in logic models are external influences (i.e., factors beyond your control), such as state programs that mandate or encourage the same behavioral changes as your program and other circumstances (positive or negative) that can affect how the program operates. Logic models also often include assumptions you currently have about your program (e.g., using water efficiently will extend the useful life of existing water and wastewater infrastructure). The [Guidelines for Measuring EPA Partnership Programs](#) (June 2006) includes an exercise to help you through the process of developing a logic model.

4. Develop Evaluation Questions

Evaluation questions are the broad questions that the evaluation is designed to answer. Evaluation questions delve into the reasons behind program accomplishments and seek to answer whether current operations are sufficient to achieve long-term goals. Good evaluation questions are important because they articulate the issues and concerns of stakeholders, examine how the program ought to work and its intended outcomes, and frame the scope of the evaluation. Typical EPA program evaluations include three to eight evaluation questions. The following five steps should aid evaluators in designing evaluation questions:

1. Review the purpose and objectives of the program and the evaluation.
2. Review the logic model and identify what aspects of your program you wish to evaluate.

3. Consult with stakeholders and conduct a brief literature search for studies on programs similar to yours.
4. Generate a potential list of questions.
5. Group questions by themes or categories (e.g., resource questions, process questions, outcome questions).

Evaluation questions drive the evaluation design, measurement selection, information collection, and reporting.

5. Select an Evaluation Design

Selection of an evaluation design involves being prepared to give your stakeholders thoughtful responses to questions related to the rigor and appropriateness of the program evaluation design:

1. Is the evaluation design appropriate to answer the evaluation question(s)? Is a process evaluation design most desirable, or are outcome and impact evaluations designs?
2. Are the data you are collecting to represent performance elements measuring what they are supposed to measure? Are the data valid?
3. Is your measurement of the resources, activities, outputs, and outcomes repeatable and likely to yield the same results if undertaken by another evaluator? Are the data reliable?
4. Do you have the money, staff time, and stakeholder buy-in that you need to answer your program evaluation question(s)? Is the evaluation design feasible?

5. Can the information collected through your evaluation be acted upon by program staff? Is the evaluation design functional?

Selecting an evaluation design also involves considering whether existing (secondary) data will be sufficient, whether new (primary) data will need to be collected to address your evaluation questions, or whether you will need both. If you require the collection of primary data, you might need to give ample time to and consideration of the Information Collection Request process imposed by the Paperwork Reduction Act and administered by the Office of Management and Budget.

The design phase of a program evaluation is a highly iterative process; although this chapter gives a linear description of the design phase, you and your evaluator are likely to revisit various issues several times.

6. Implement the Evaluation

Generally, this is the stage where an individual who has technical expertise in program evaluation becomes the leader of the evaluation. This expert evaluator works independently to ensure objectivity, so program staff and stakeholder involvement in this particular stage of the evaluation might be minimal.

Implementing the evaluation involves consulting with the program staff and managers to ensure that the design is, in practice:

- Yielding the appropriate data to address the evaluation questions.
- Pilot-testing procedures.
- Considering the results of expert review of the evaluation design (if applicable and appropriate).

- Undertaking the data analyses.
- Sharing preliminary results as a quality-assurance check.
- Ensuring that the data and data analysis are being reported in an objective and unbiased manner.

7. Communicate Evaluation Results

Careful consideration of your Partnership Program's stakeholders will influence how to best organize and deliver evaluation reports and briefings. Keep in mind that the results have three basic elements: 1) findings, 2) conclusions, and 3) recommendations.

- **Findings** refer to the raw data and summary analyses obtained during the program evaluation effort. Because the findings are a part of the data analysis process, the evaluator should have the primary responsibility for communicating findings to the program staff and management (in verbal or written form). The expert evaluator often delivers the findings to the Partnership Program in the form of a draft report or draft briefing.
- **Conclusions** represent the interpretation of the findings, given the context and specific operations of your Partnership Program. Your evaluator may undertake an appropriate analysis of the data and may independently derive some initial interpretations of what these data suggest; however, you and others closely linked to the program should have an opportunity to provide comments based on a draft report, to suggest ways to refine or contextualize the interpretation of the findings. This same process applies even if you have commissioned an independent, third-party evaluation, because a strong external evaluator should ensure that the

presented conclusions are sound, relevant, and useful.

- **Recommendations** are based on the sound findings and conclusions of your evaluation. A strong evaluator will understand that framing recommendations is an iterative process that should involve obtaining feedback from Partnership Program managers, staff, and key stakeholders. Again, this same process applies even if you have commissioned an independent, third-party evaluation, though in this case the external evaluator will make the key judgments about the report's final recommendations. Your involvement in the development of recommendations is important because, to get the most value out of your evaluation, you should be prepared to implement some or all of the recommendations. Implementation of recommendations and the resulting improvements to your program are some of the greatest sources of value added to programs by the evaluation process.

You must tailor presentations of evaluation results to the specific needs of your stakeholders, who might or might not be satisfied by a lengthy report. Key questions you and your evaluator should ask in presenting results are:

- Which evaluation questions are most relevant to these stakeholders?
- How do the stakeholders like to receive information?
- How much detail do the stakeholders want?
- Are the stakeholders likely to read an entire report?

Based on the answers to these questions, in addition to a full-length report, you can opt for one or more of the following reporting formats depending on the needs of each stakeholder group:

- A shortened version of the evaluation report for broad distribution.
- A one- or two-page executive summary of key results and conclusions.
- A PowerPoint briefing on the evaluation reports.

If you have any questions or would like additional information about Partnership Programs in general, please contact Stephan Sylvan, Partnership Program Coordinator (sylvan.stephan@epa.gov). If you have any questions or would like additional information about these guidelines specifically, please contact Terell Lasane, Social Scientist (lasane.terell@epa.gov). Both are based in EPA's Office of Policy, Economics, and Innovation, National Center for Environmental Innovation.

Contents

Introduction	1
Why Evaluate?	2
Program Evaluation's Role in Performance Management.	2
Who Should Use These Guidelines?	4
How To Use These Guidelines	4
Guidelines Roadmap	5
Chapter 1: Plan the Evaluation	6
Choosing the Right Evaluation for Your Partnership Program	6
Deciding Whether to Conduct an Internal or External Evaluation	9
Budgeting for an Evaluation	10
Anticipating Potential Data Limitations and Stakeholder Concerns	11
Chapter 2: Identify Key Stakeholders.	15
Who Should Be Involved in Evaluations of Partnership Programs?	15
Identifying Relevant Stakeholders.	16
Involving Stakeholders.	16
Incorporating a Variety of Perspectives	18
Chapter 3: Develop or Update the Program Logic Model.	19
Why Is a Logic Model Important for Program Evaluation?	19
Logic Model Elements	20
Chapter 4: Develop Evaluation Questions	23

Chapter 5: Select an Evaluation Design.26
The Foundations of Good Program Evaluation Design27
Assessing the Data Needs for the Evaluation Design28
Primary Data Collection Challenges32
Choosing an Evaluation Methodology33
Expert Review of the Evaluation Design.35
Chapter 6: Implement the Evaluation.37
Pilot Testing the Evaluation37
Protocols for Collecting and Housing Evaluation Data.38
Data Analysis38
Chapter 7: Communicate Evaluation Results41
Presenting Results43
Appendices45
Appendix A: Glossary.46
Appendix B: Evaluation Resources51
Appendix C: Case Study.55

Introduction

EPA Partnership Programs are some of many tools the U.S. Environmental Protection Agency (EPA) uses to protect public health and the environment. These programs build upon a rich tradition of EPA working collaboratively with others to find innovative solutions to environmental challenges. Whether promoting environmental improvements complementary to or beyond those required by regulation, or functioning in the absence of regulation, EPA Partnership Programs proactively target and motivate external parties to take specific environmental action steps on a voluntary basis with EPA in a leadership and decision-making role. They do not compel this action through legal means. These programs vary greatly in style, type, and function; however, they all share the need to demonstrate that they are achieving environmental results and supporting EPA's mission. Thus, EPA identified a need for program evaluation guidelines specific to its Partnership Programs.

These guidelines offer a general overview of standard program evaluation methods and techniques but also contain information tailored to the unique challenges faced by EPA Partnership Programs.

The goal of these guidelines is to provide a clear, practical, and useful guide for EPA Partnership Program managers and staff. They will prepare EPA Partnership Program managers and staff to work effectively with expert evaluators who have technical knowledge of and practical experience with program evaluation. These expert evaluators (often contractors, but also EPA staff) work during the evaluation process to define key terms, clarify steps, and identify issues that may affect the quality of the evaluation. EPA encourages program managers and staff to share these guidelines with their expert evaluators and program stakeholders so that all parties share a common starting point and understanding of program evaluation in the context of EPA Partnership Programs.

These program evaluation guidelines are part of a suite of guidelines for EPA Partnership Programs, including *Guidelines for Designing EPA Partnership Programs*, *Guidelines for Marketing EPA Partnership Programs*, and *Guidelines for Measuring the Performance of EPA Partnership Programs* (all available at [intranet.epa.gov/partners](https://www.epa.gov/partners/intranet.epa.gov/partners)). In particular, these program evaluation guidelines build on the *Guidelines for Measuring the Performance of EPA Partnership Programs*.

Why Evaluate?

Some argue that program evaluation is too time consuming, too onerous, and too costly for EPA Partnership Programs. In fact, the failure to evaluate your program can be more costly in the long run. Program evaluation results can illustrate that EPA Partnership Programs are making a difference, are effective and efficient, provide customer satisfaction, offer benefits that outweigh program costs, and merit continued funding. If evaluation results show that your program needs improvements, this information can help decision-makers determine where adjustments should be made to ensure future success. Reasons for evaluating EPA Partnership Programs include:

- **Providing data to stakeholders:** Program evaluations provide valuable information to EPA Partnership Program managers and staff, EPA senior management, target decision-makers, program participants, and other external stakeholders.
- **Improving the program:** Program evaluations can help identify when program goals have been met and whether changes need to be made (in activities or allocation of resources) to meet program goals.
- **Informing policy and funding decisions:** By helping EPA understand the role of an individual Partnership Program, in its broader policy toolbox, program evaluations help EPA senior management allocate resources and set priorities among programs. EPA Partnership Programs that are able to demonstrate a link between program activities and outcomes through objective evaluation are more likely to receive continued support.

Program evaluation helps EPA respond to the [Government Performance and Results Act \(GPRA\)](#), the [Program Assessment Rating Tool \(PART\)](#), and [Executive Order 13450: Improving Government Program Performance](#).

Because of the increased number and prominence of EPA Partnership Programs, stakeholders are increasingly interested in ensuring that these programs are adequately evaluated, to determine whether they are well designed and effective. Program evaluation is important for learning about programs and improving them. Evaluations can produce data needed to respond to and answer key management questions and accountability demands, identifying why a program has or has not met its goals. Program evaluation helps EPA respond to the Government Performance and Results Act ([GPRA](#)), the Program Assessment Rating Tool ([PART](#)), and [Executive Order 13450: Improving Government Program Performance](#).

Program Evaluation's Role in Performance Management

Program evaluation is one component of a performance management system. Performance management systems include logic models, performance measurement, and program evaluation, as illustrated on the following page. Together, performance management activities ensure that Partnership Programs are meeting their goals in an effective and efficient manner. This guide focuses on program evaluation, one component of a performance management system.



Logic modeling, performance measurement, and program evaluation work in a dynamic system. The logic model provides a framework that will help you clearly understand and communicate how your program's activities, outputs, and outcomes connect to your long-term goals. Performance measurement involves ongoing monitoring and reporting of the program progress and accomplishments. Program evaluation builds on these as a formal assessment that examines and draws conclusions about the effectiveness of a program's design, implementation, and impacts.

The *Guidelines for Measuring the Performance of EPA Partnership Programs* cover logic modeling and performance measurement, which are important concepts to understand fully before undertaking a program evaluation.

Because program evaluation uses performance measurement data to assess why results are

occurring, information collected for performance measurement is an important component of program evaluation. If your program has not identified or collected performance data, you must include this task as part of your evaluation process. The program logic model, described in [Chapter 3](#) (as well as in [Chapter 4](#) of the performance measurement guidelines), will help to identify potential measures. If you have already developed a logic model for your program, you do not need to develop a different one for the evaluation. Instead, you should regularly review your existing logic model and make any necessary updates or revisions.

Who Should Use These Guidelines?

Not everyone at EPA is, or is expected to be, an expert in program evaluation. Many people are evaluation users; they have limited knowledge of program evaluation but benefit from

Other Evaluation Resources

Appendix B of these guidelines presents a variety of resources for you to tap as you plan for, design, and carry out evaluations.

The most basic resource is EPA's Evaluation Support Division (ESD), located in the Office of Policy, Economics, and Innovation (OPEI). ESD is EPA's source of in-house evaluation expertise, providing training, technical assistance, and evaluation support to EPA and its partners.

- www.epa.gov/evaluate or intranet.epa.gov/evaluate
- Yvonne Watson, 202-566-2239; watson.yvonne@epa.gov

Performance Measurement vs. Program Evaluation

Imagine you just bought a new car—your pride and joy. Both the salesperson and the owner's manual indicate your car should get 30 miles per gallon of gas. Well, it has been six months, and you have kept meticulous records. You notice your car has only managed to get 20 miles a gallon. What do you do? You take the car back to the dealership and ask the mechanic to find out why the car is not meeting the specified performance standard. The mechanic finds a problem with the engine, fixes it, and you drive off with a better functioning car.

The gas mileage records are the performance measurement part of the equation, and the mechanic's diagnosis is the program evaluation. This scenario is an analogy of the differences and relationships between these two tools as applied to environmental programs.

and see the value of evaluations and might be called on to participate in the evaluation process occasionally. Others are evaluation practitioners, with an in-depth knowledge of program evaluation and capable of advising, managing, or conducting evaluations. Although evaluation practitioners are generally capable of planning and managing an evaluation without external aid, they may need to seek assistance from others on the actual conduct of evaluations because of time and/or resource constraints. A further subset of evaluation practitioners is evaluation experts, who Partnership Programs can access for advice on advanced concepts and techniques.

We developed these guidelines primarily for evaluation users (i.e., most EPA Partnership Program managers and staff). As users:

- **Program managers** are responsible for determining whether their programs should be evaluated and when an evaluation should take place. Although managers need not have the technical expertise to conduct an evaluation, knowledge of the basic steps in the evaluation process will help inform decisions that must be made when commissioning

evaluations and using evaluation findings to make management decisions.

- **Program staff** are responsible for leading or participating in the program evaluation. They will benefit from having a basic understanding of the program evaluation concepts and techniques that they may encounter during an evaluation. This background will allow them to be able to “speak the same language” as the seasoned evaluators on their team.

How To Use These Guidelines

At its most sophisticated level, program evaluation can be a very complex discipline with practitioners devoting entire careers to narrow aspects of the field. These guidelines do not assume that you are such an expert, nor do they aim to make you one. They are intended to introduce the novice to the world of program evaluation and walk you through a step-by-step framework for how to design and conduct an evaluation for an individual Partnership Program that will enable you to work more effectively with a contractor or evaluation expert. We have included actual examples of Partnership Programs to help illustrate the concepts

described. Partnership program managers and staff should use these guidelines in conjunction with *Guidelines for Designing EPA Partnership Programs*, *Guidelines for Marketing EPA Partnership Programs*, and *Guidelines for Measuring the Performance of EPA Partnership Programs*.

Guidelines Roadmap

Before starting a program evaluation, you should become familiar with the key steps in the process. These guidelines are organized in seven chapters that reflect each of these steps. While the framework appears to be linear and sequential, you and your evaluator are likely to revisit one or more of these steps.

- [Chapter 1: Plan the Evaluation](#)
- [Chapter 2: Identify Key Stakeholders](#)
- [Chapter 3: Develop or Update the Program Logic Model](#)
- [Chapter 4: Develop Evaluation Questions](#)
- [Chapter 5: Select an Evaluation Design](#)
- [Chapter 6: Implement the Evaluation](#)
- [Chapter 7: Communicate Evaluation Results](#)

Three appendices are also included in these guidelines:

- [Appendix A: Glossary](#)
- [Appendix B: Evaluation Resources](#)
- [Appendix C: Case Study \(of an EPA Partnership Program's experience with program evaluation\)](#)

A Case Study of an EPA Partnership Program's Evaluation Experience: Hospitals for a Healthy Environment (H2E)

To show how an actual EPA Partnership Program handles the evaluation process described in these guidelines, we traced the experience of a program evaluation for Hospitals for a Healthy Environment (H2E), completed in 2006. The program evaluation process for H2E was typical but not always straightforward. At the end of each chapter, we give short vignettes from H2E's program evaluation experiences. A more detailed case study appears in Appendix C.

Chapter 1:

Plan the Evaluation

Four key considerations frame how you plan for an evaluation:

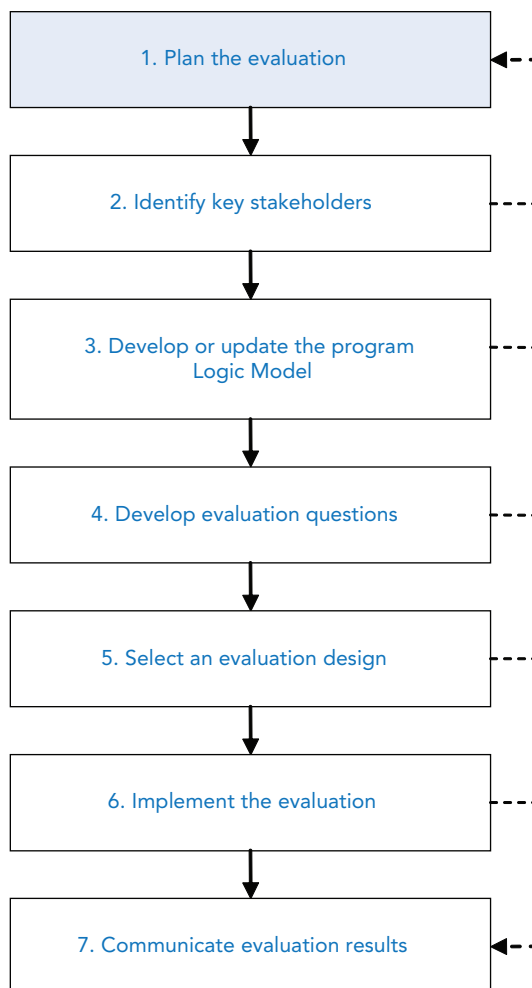
This chapter will help you:

- Choose the right evaluation for your program.
- Decide whether to conduct an internal or external evaluation.
- Budget for an evaluation.
- Anticipate potential data limitations and stakeholder concerns.

If evaluation planning is incorporated into the design of a program, evaluation costs can be far lower and the quality of the final evaluation much higher. Adding an evaluation after a program is in operation can result in higher costs, fewer options, and decreased capacity to obtain good answers to important program questions.

Choosing the Right Evaluation for Your Partnership Program

Program evaluations help assess effectiveness and lead to recommendations for changes at all stages of a program's development. The type of program evaluation you do should align with



the program's maturity and be driven by your purpose for conducting the evaluation and the questions that you want to answer.

- **Design evaluation** seeks to assess whether the program will operate as planned. It should be conducted during the program design process. Evaluating a program's design can be very helpful for developing an effective Partnership Program if: 1) program goals are less clearly defined, 2) only a few staff members were charged with developing the program, or 3) uncertainties exist about a program's intended activities. On the other hand, evaluating a program's design might not be necessary if you have a robust, inclusive, and clear program development process.
- **Process evaluation** is typically a check to determine if all essential program elements are in place and operating successfully. This type of evaluation is often conducted once a program is up and running. Process evaluations can also be used to analyze mature programs under some circumstances, such as when you are considering changing the mechanics of the program or if you want to assess whether the program is operating as effectively as possible. Evaluating a program's process usually is not necessary in the early stages of a Partnership Program if 1) early indicators show that the program is being implemented according to plan, and 2) program managers and stakeholders are confident that a program's implementation is on target.
- **Outcome evaluation** looks at programs that have been up and running long enough to show results and assesses their success in

Tip: *The type of program evaluation you do should align with the program's maturity and be driven by your purpose for conducting the evaluation and the questions that you want to answer.*

reaching their stated goals. Program outcomes can be demonstrated by measuring the *correlations* that exist between program activities and outcomes after you have controlled for all of the other plausible explanations that could influence the results you observe. This process is sometimes referred to as measuring *contribution* (a concept discussed in detail in Chapter 5).

Correlation does not imply causation, however. Outcome evaluation can tell you that your program likely had an effect on the outcome, but to confidently demonstrate that your program has caused the results you observe, you would need to conduct an impact evaluation. Outcome evaluations are appropriate when baseline and post-baseline data sets are available or could be developed. Outcome evaluations can also be undertaken if you are interested in determining the role, if any, context plays or if your program is producing unintended outcomes. Outcome evaluations are not appropriate, however, when the program is too new to have produced measurable results.

- **Impact evaluation** is a subset of outcome evaluation that focuses on assessing the causal links between program activities and outcomes. This is achieved by comparing the observed outcomes with an estimate of what would have happened in the absence of the program. While an outcome evaluation is

only able to identify that goals have been met, an impact evaluation identifies the reason that the goals have been met and that results would not have been achieved without the program. This process is sometimes referred to as measuring *attribution* (a concept discussed in detail in Chapter 5).

Impact evaluations can be conducted at two phases in a program's lifecycle. First, they can be conducted as part of the piloting stage to determine if a particular partnership ap-

proach should be expanded into a full-scale program. Second, they can be conducted on mature programs to determine whether a Partnership Program is having the intended behavior change and/or environmental result. Causal claims in the purest sense can only be made when a program is subjected to a randomized control trial (RCT).

Four Types of Program Evaluation

Type	When to Use	What It Shows	Why It Is Useful
Design Evaluation	During program development	Identifies needs that the program should address (e.g., is the program's approach conceptually sound?)	Informs program design and increases the likelihood of success
Process Evaluation	As needed after the program development stage	How all essential program elements are in place and operating (e.g., how will are the program's activities being implemented?)	Allows program managers to check how program plans are being implemented
Outcome Evaluation	After program has been implemented for a reasonable period of time	The extent to which a program has demonstrated success in reaching its stated short-term and intermediate outcomes after you have ruled out other plausible rival factors that may have produced program results (e.g., to what extent is the program meeting its short and intermediate term goals?)	Provides evidence of program accomplishments and short-term effects of program activities
Impact Evaluation	Both during the pilot stage and with mature programs	Causal relationship between program activities and outcomes (e.g., did the program's activities cause its long-term goals to occur?)	Provides evidence that the program, and not outside factors, has led to the desired effects

Deciding Whether to Conduct an Internal or External Evaluation

An internal evaluation is conducted by EPA staff or led by EPA staff with the support of contractors who regularly support evaluations at EPA. An external evaluation is conducted by an independent third party, such as an academic or other institution, that operates “at arms length” from the program, even if the evaluation is commissioned or funded by EPA.

Internal evaluators typically have a greater understanding of EPA operations and culture, have ongoing contact with EPA, and are more likely to have greater access to decision-makers. A Partnership Program conducting an internal evaluation might hire a contractor to act as the evaluator to help with the technical aspects of an evaluation, but the program staff retains ongoing control over the evaluation’s planning, design, and implementation. Often internal evaluations cost less than external evaluations.

Internal evaluations can be perceived to be less credible than evaluations conducted by an objective, independent third party. Therefore, you may need to take steps to increase credibility and mitigate against bias when conducting internal evaluations, such as conducting an expert review of the evaluation methodology

and findings. An expert review involves commissioning program evaluation experts who are not otherwise involved with your program or the evaluation to provide an impartial assessment of the evaluation methodology, analysis, and conclusions. Alternatively, you could convene an evaluation advisory group to provide input throughout the evaluation. An advisory group could include individuals from within and outside EPA who have expertise relevant to the program and/or to evaluation.

When conducting an external evaluation, the program staff has less involvement in evaluation design and implementation. You should seriously consider conducting an external evaluation when issues of objectivity are paramount. Objectivity might have greater importance in a variety of situations that are not necessarily unique to EPA Partnership Programs, such as accountability demands from Congress or the Office of Management and Budget (OMB). Furthermore, using an external evaluator can be an especially useful way to allay stakeholder fears when trust is an issue and is useful for programs that find themselves in a defensive posture due to repeated criticism and heightened scrutiny. Finally, gaining a fresh perspective from someone with experience evaluating many different programs can be helpful.

A CONTRACTOR FOR AN INTERNAL EVALUATION SHOULD:

- Document potential real and perceived conflicts of interest for transparency.
- Work closely with the program staff to design the evaluation; they will expect to “weigh-in” on key design decisions.

A CONTRACTOR FOR AN EXTERNAL EVALUATION SHOULD:

- Take visible steps to avoid real and perceived conflicts of interest throughout the process.
- Consult program staff to design the evaluation but independently make key design and reporting decisions.

Working With a Program Evaluation Contractor

Use these tips for working with a program evaluation contractor:

- Select contractors that have experience in the subject matter of the program being evaluated and technical evaluation expertise.
- Choose a contract vehicle that allows uninterrupted service and access to contractors with evaluation expertise.
- Work with the contractor to facilitate data collection from internal and external evaluation stakeholders. This step can cut the cost of an evaluation greatly, increase the response rate, and reduce the frustration of program participants.
- Promote the active involvement of the Partnership Program staff. Doing so will lead to a better report that is more likely to meet the needs of the program with recommendations that are more likely to be implemented.
- Have an explicit and documented agreement with the contractor about steps that will be taken to ensure objectivity (e.g., peer review).
- Be clear about who will make final decisions about how the program and the contractor will share information about the evaluation process, draft evaluation products, and final evaluation reports or briefings.

Budgeting for an Evaluation

Conducting an evaluation can take considerable time and incur significant expense. Budgets required for evaluations vary widely, depending on the scope and scale of the program, the type and complexity of evaluation questions, the evaluation design, and the availability of existing data (the Evaluation Support Division [ESD] and other agency evaluation practitioners can help you estimate a budget based on your program's unique evaluation goals).

Whether you choose to conduct an internal or external evaluation will depend on your reasoning for conducting the evaluation. Among the factors to consider in making the decision are cost, knowledge of program operations and culture, perceived objectivity, and accountability.

The size and scale of your Partnership Program is likely to drive many of your budgeting consid-

erations. For example, large programs with multiple partners might require designs that allow for a comparison of data from unique subgroups involved in the program's efforts. Some Partnership Programs might be able to take advantage of already existing data; costs of using preexisting data can vary, but sometimes data can be accessed quickly and at a relatively low cost.

If you need to collect new data you should budget additional time and money. The more complicated the data collection and analysis, the more expensive the evaluation will be. A qualitative analysis based on interview or focus group data, for example, can be very time consuming and expensive to conduct. A smaller budget will limit the sophistication of any new data collection methods and the statistical analyses you can conduct.

As we point out throughout this document, however, there are several ways you can answer

your evaluation questions. These alternate design options may fit within your time and fiscal constraints while still providing information useful for your program.

Finally, you should ensure that you have management buy-in to authorize the reallocation of internal resources (i.e., time, funding, staff time) to support the evaluation effort.

Anticipating Potential Data Limitations and Stakeholder Concerns

You should be aware of potential challenges that EPA Partnership Programs often face related to program evaluation. These include limitations in identifying existing data resources, barriers to collecting new data, and how to address stakeholder concerns. These barriers are typical to all program evaluations, but anticipating them up-front can help you prepare for and overcome them. In the following sections, we describe these challenges in more detail and provide tips for addressing them.

Identifying Existing Data Sources

Ideally, your program should have been collecting performance measurement data since it began, and those data can be easily used to evaluate the program. As discussed in more detail in [Chapter 5](#), however, you might discover that you do not have the right type of data needed to conduct the evaluation. If this is the case for your program, first look to see if the data you need were already collected by another source, such as studies and reports by other organizations (e.g., the Government Accountability Office [GAO], EPA's Office of Inspector General). You and your evaluator can also use information from a readily available

source such as a public database or company reports. A surprising amount of data is collected on thousands of topics, and the key is often simply knowing where to look and being persistent. Be aware of how the data are collected, however, and that the organizations collecting the data might define terms differently than you do. These issues can affect data quality and validity, as discussed in more detail in [Chapter 5](#).

Collecting New Data

In some cases, existing data sources might be inadequate for your evaluation needs or have quality issues that cannot be overcome. In this scenario, you will need to develop new data. One approach to data collection is to research Partnership Programs that have previously been evaluated to identify examples of the types of data gathered and to determine how these programs handled similar challenges. Another approach is to convene a group of experts to obtain ideas on potential data sources. You might be able to add questions to existing surveys other agencies, organizations, or research-oriented groups are conducting.

When you are ready to collect new data, you might be required by the Paperwork Reduction Act to obtain an Information Collection Request (ICR). [Chapter 5](#) goes into greater depth on navigating the ICR process and the [Guidelines for Measuring the Performance of EPA Partnership Programs](#) also contains detailed information on data collection.

Stakeholder Concerns

Several classes of stakeholders have particular concerns you will need to address proactively throughout the evaluation process.

EPA Stakeholder Concerns. First, you must anticipate the concerns of the stakeholders most closely involved in the program: Partnership Program staff, managers, and EPA senior management. Apprehension about program evaluation is not unique to EPA Partnership Programs. Program evaluation is often associated with external accountability demands. The program staff might feel pressured to show results, yet often feel unprepared for program evaluations. The table that follows presents common concerns and responses to consider.

Target Audience Concerns. The target audience of the program might be apprehensive about evaluation as well. To address their concerns you must discuss the goals and purpose of the evaluation with program participants and emphasize that the objective is to improve program function. Provide clear information to participants on:

- How the evaluation results will be used.
- The level of data transparency (e.g., whether individual participant data will be identified in the evaluation report or if the data will be aggregated up across participants in a way that preserves confidentiality).
- How confidential business information will be treated (if applicable).
- In addition, consider these ideas for involving the program's target audience in the evaluation process:

- o Involve stakeholders as you develop your key evaluation questions (discussed in [Chapter 4](#)).
- o Continue to involve a smaller subset of program participants and staff throughout the course of the evaluation, to help address concerns about the evaluation and increase the extent and reliability of any new information collected (discussed in [Chapter 4](#) and [Chapter 5](#)).
- o Consider ways to minimize data collection burdens faced by participants and staff throughout the course of the evaluation by making the best use of existing data and only asking questions that are relevant to evaluation objectives (discussed in [Chapter 5](#)).
- o Provide participants with timely results and feedback (discussed in [Chapter 6](#) and [Chapter 7](#)).

Public Accountability Concerns. Finally, governmental oversight bodies and key public stakeholders often look to program evaluation as a means of verifying that programs are achieving their intended long-term goals and thus using taxpayer money effectively. Recently, some parties have claimed that impact evaluations, because they are the only type of evaluation design capable of making true causal links between programs and their long-term goals, are the only type of evaluations worth conducting. Consequently, EPA Partnership Programs are under increasing pressure to conduct impact evaluations. Although impact evaluations—which by design, demonstrate a program's definitive causal effect—should be undertaken whenever it is possible to do so, program staff, managers, and stakeholders should understand

Evaluation Concerns and Responses to Consider

Concerns	Responses
Our program is different from other federal programs and other programs at EPA.	It is true that environmental program evaluation is a relatively new subfield, but EPA does have a growing track record of program evaluation for Partnership Programs (see case study in Appendix C). Many federal agencies with similarly far-reaching and ambitious missions (e.g., education, public health) have developed a culture of evaluation that has worked to improve public policy. We also recognize that Partnership Programs represent a unique subset of EPA programs, and that is why we have developed these guidelines to help you.
Evaluation costs too much.	Program evaluation does put demands on limited resources, but demonstrating your program's environmental results could lead to maintaining or increasing budgets in the future. Depending on the type of evaluation you conduct, program evaluations can be scaled to meet most budgets (see Chapter 1 and Chapter 5), even those of small Partnership Programs, but it is critical to be proactive about managing evaluation costs and recognizing tradeoffs.
We don't have the time to evaluate.	A well-managed evaluation process recognizes staff time as a resource and aims to minimize time demands on program staff. A process evaluation can also help to identify areas of inefficiency in even the most high-achieving programs, freeing up staff time in the future.
The evaluation process will take too long.	Lengthy evaluations are not the norm. Evaluations can be designed and paced realistically to respond to the timeframes facing your program. A discussion of the evaluation's schedule should occur early on so that you can account for relevant timeframes.
Our program doesn't need to be evaluated.	It is difficult to assess and communicate program performance in the absence of evaluation. Beyond telling you if the program is having a positive impact, an evaluation can reveal information that is helpful even to the most successful programs, such as pinpointing underused resources and potential areas of increased efficiency.
We don't know how to evaluate.	No one expects you to become an expert when your program undergoes an evaluation. All that is needed is a basic understanding of the evaluation process, as laid out in these guidelines. A variety of resources are available when you need technical help (see Appendix B).
Our program is not ready for evaluation.	Consider program design issues, program process issues, and the intended outcomes of your program. As you consider the management issues that most affect your program, you will find that distinct evaluation approaches are applicable to the maturity of your program, the effectiveness of your operations and for assessing your program's outcomes. If your program is early in its development, you may benefit from a program design or process evaluation, whereas older programs may find an outcome or impact evaluation most useful.
Evaluation is unnecessary—GPRA, PART, or an IG review will suffice.	The Government Performance and Results Act (GPRA) is focused on performance measurement, not evaluation. The Program Assessment Rating Tool (PART) emphasizes conducting evaluation prior to a PART review. PART is not an evaluation but relies on evaluations that have already been conducted. Inspector General (IG) reviews vary in structure but do not constitute program evaluation. In particular, IG audits assess whether proper procedures are in place, and not whether the program design is effective. While program evaluation can help you respond to these accountability demands, these mechanisms are <u>not</u> the same as program evaluations.

that demonstrating a program's causal effect through a rigorous impact evaluation often cannot be realistically achieved without a substantial (and often overwhelming) investment.

As stated earlier, impact evaluations are most easily undertaken when the evaluation approach has been written into a program's design. Undertaking an impact evaluation subsequent to a program's implementation can be considerably more challenging. Principal barriers to the conduct of impact evaluations are: 1) fiscal and staffing limitations, 2) the inability of programs

to control the external factors that work in tandem with programs to produce long-term environmental outcomes, 3) the role of Partnership Programs as one of several approaches used to achieve the Agency's mission, and 4) the difficulty of collecting data from non-participants (necessary to form control groups). Further, questions of impact are not the only questions of value to programs. We strongly advise that you make preliminary consultations with expert evaluators and program stakeholders to determine what type of evaluation design is the most viable and useful option for your program.

Planning the Evaluation: The H2E Experience

EPA launched H2E in 1998 to advance waste reduction and pollution prevention efforts in hospitals across the country. The program's goals included: 1) virtually eliminating mercury-containing waste, 2) reducing the overall volume of both regulated and non-regulated waste, and 3) identifying hazardous substances for pollution prevention and waste reduction opportunities.

By 2004, H2E managers and staff wanted to better understand whether and how program activities were leading to environmental results (e.g., were H2E's Partnership Program activities directly leading to reductions in mercury in the environment?) They decided that a program evaluation would be one way to answer this question. H2E staff submitted a proposal to EPA's annual Program Evaluation Competition in 2004 to access the funding and expertise to conduct an internal evaluation. The competition provided H2E with partial funding, a contractor with evaluation expertise, and an EPA staff person with evaluation expertise to manage the evaluation contractor's work.

During the initial planning phase, H2E asked the evaluation contractor to design an impact evaluation. H2E used an ICR to collect the available data from its partners; however, the evaluation contractor soon advised H2E staff that the data that were available would not work for an impact evaluation because they were incomplete and represented only a small percentage of partners. In addition, the cost of designing and implementing an impact evaluation would be prohibitively expensive and time-consuming. After consulting with the evaluation contractor and stakeholders, H2E staff determined that an outcome evaluation was a better fit for the program; it would provide information that was most useful to the program, worked with readily available data, and could be completed within a reasonable budget and timeframe.

Chapter 2:

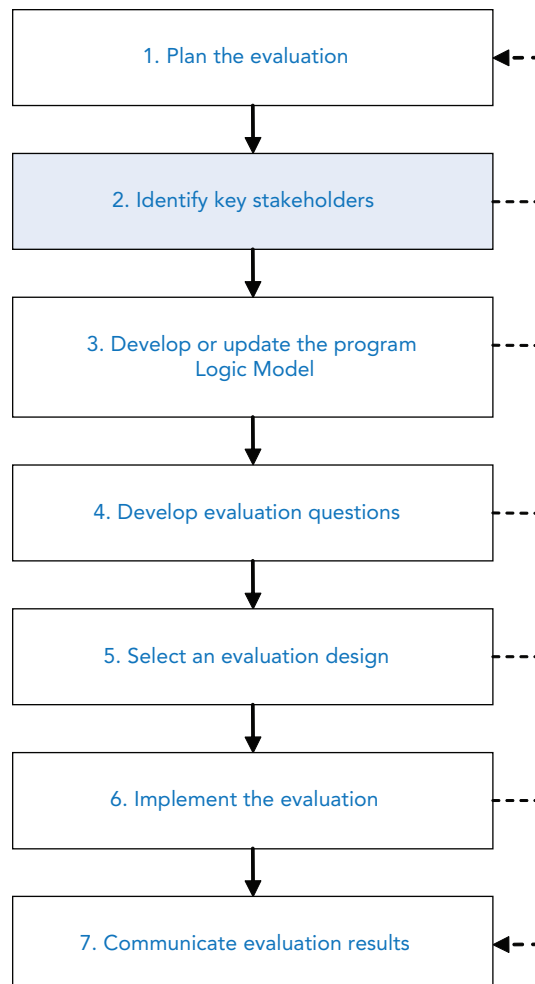
Identify Key Stakeholders

Who Should Be Involved in Evaluations of Partnership Programs?

A key step in evaluating a program is identifying stakeholders and developing a stakeholder involvement plan. This plan can be as formal or informal as the situation warrants. In these guidelines, a stakeholder is broadly defined as any person or group who has an interest in the program being evaluated or in the results of the evaluation. Incorporating a variety of stakeholder perspectives in the planning and implementation stages of your evaluation will provide many benefits, including:

- Fostering a greater commitment to the evaluation process.
- Ensuring that the evaluation is properly targeted.
- Increasing the chances that evaluation results are implemented.

To foster the desired level of cooperation, you should first identify relevant stakeholder groups and then determine the appropriate level of involvement for each group. The remainder of this chapter discusses these steps in more detail.



Identifying Relevant Stakeholders

Identify and engage the following principal groups of internal and external stakeholders:

- **People or organizations involved in program operations** such as designing and implementing the program and collecting performance information. These entities could include program participants, sponsors, collaborators, coalition partners, funding officials, administrators, and program managers and staff.
- **People or organizations served or affected by the program**, which might include the program's target audience, academic institutions, elected and appointed officials, advocacy groups, and community residents.
- **Primary intended users of the evaluation results**—the individuals in a position to decide and/or take action with evaluation results, such as program managers and upper management. This group should not be confused with primary intended users of the program itself, although some overlap can occur.

- **Agency planners**, such as key regional and program office liaisons who support all aspects of planning and accountability.

Involving Stakeholders

Involving principal stakeholders in the evaluation from the beginning is important for fostering their commitment to the evaluation design and, ultimately, the evaluation findings and recommendations. To involve stakeholders, you can use face-to-face meetings, conference calls, and/or electronic communications. Choose a method or combination of methods that works best for the people in the group.

Continued feedback from stakeholders throughout the evaluation process will help to ensure that the evaluation remains on track to produce useful results. The scope and level of stakeholder involvement will vary for each program evaluation and stakeholder group, however, and keeping the size of the group manageable is important. Following are suggestions for involving relevant stakeholders.

Your Core Evaluation Team

Although several individuals will be stakeholders in the evaluation outcome, you should narrow your working group in order to have a manageable team that will be actively engaged throughout the evaluation process. Core members of this team should represent:

- **The Client:** You and one or two other individuals from the EPA Partnership Program that is the focus of the evaluation and will use the evaluation results.
- **Stakeholders:** Individuals with a vested interest in the program (the focus of the present chapter).
- **The Evaluator:** The individual(s) who carry out the evaluation. (As described in Chapter 1, the evaluators can be internal or external.)

Stakeholder involvement in program evaluation is often iterative. You should expect your expert evaluator to work closely with you on managing stakeholder involvement throughout the program evaluation process.

Planning Stage

Before you begin designing the evaluation, make sure that all participating stakeholders understand the purpose of the evaluation and the proposed process: have a concrete conversation with all parties, laying out all obligations and expectations of each party (including informal and implicit expectations). Any conflicts of interest should be addressed openly at this stage, so as not to compromise the reliability and credibility of the evaluation process and results.

Design Stage

When you and your evaluator are ready to begin designing the evaluation, involving as many stakeholders in the initial discussions as possible is essential. Continue to consult and negotiate with stakeholders as you design the evaluation, including soliciting their reactions to the program logic model (Chapter 3) and evaluation questions (Chapter 4). You should also consult and negotiate with stakeholders to come to agreement on key data (e.g., including how to select measures, how to measure program impacts, how to set a baseline and use baseline data, and how to ensure data quality throughout the evaluation process).

Implementation Stage

From the wider group of stakeholders that you consulted during the evaluation design phase, select a manageable subset of stakeholder rep-

resentatives to join your core evaluation team or task force to help make ongoing decisions about the evaluation. Continued use of this team throughout the evaluation process will help keep the evaluation focused, help to allay concerns, and increase the quantity and quality of information collected.

You and your evaluator can also consider implementing a full participatory evaluation, which involves stakeholders in all aspects of the evaluation, including design, data collection, and analysis. A fully participatory evaluation will help you and your evaluator to:

- Select appropriate evaluation methodologies.
- Develop evaluation questions that are grounded in the perceptions and experiences of clients.
- Overcome resistance to evaluation by participants and staff.
- Foster a greater understanding of the evaluation among stakeholders.

A full participatory evaluation is not a good fit for every Partnership Program, however, as evaluations of this type requires an additional investment of time and resources to facilitate. You and your evaluator might choose instead to elicit broad stakeholder input only at key points, consider this input carefully, and be transparent in decision-making. Key points include developing or reviewing the program logic model, formulating evaluation questions, developing the evaluation methodology, reviewing the draft evaluation report, and disseminating findings.

Incorporating a Variety of Perspectives

In addition to the principal groups of stakeholders, consider inviting someone to play the role of “devil’s advocate.” A skeptic, or someone in the core evaluation team who will challenge your assumptions, can strengthen an evaluation’s credibility by ensuring that all decisions and assumptions are thoroughly examined. Try to identify a program staff person or other individual with knowledge of the program who will ask tough, critical questions about the program and evaluation process, or someone on the core evaluation team can play this role.

Above all, remember that the goal of the evaluation is to produce findings that can be used to improve the program. Common sense dictates that an evaluation process involving the individuals involved in the program will produce findings that are relevant and useful. You should, therefore, plan, conduct, and report the evaluation in a way that incorporates stakeholders’ views and encourages their feedback, thereby increasing the likelihood that key stakeholders will act upon findings.

Identifying Key Stakeholders: The H2E Experience

H2E staff identified EPA program managers, team leaders, and program staff as key stakeholders to be consulted during the evaluation process. H2E staff also identified key external partners (e.g., major trade associations, participating hospitals). These internal and external stakeholders participated to varying degrees, from occasional consultation on evaluation design and comments on draft documents to ongoing involvement in data collection and report drafting.

H2E’s core evaluation team included the program staff lead, evaluation contractor, and EPA evaluation expert. The evaluation contractor served as the team’s “skeptic,” asking those closely involved with the program to explain their assumptions about program activities and measurable outcomes. By regularly consulting with a diversity of stakeholders, H2E’s core evaluation team was able to gain assistance with data collection and sustain buy-in throughout the evaluation process.

Chapter 3:

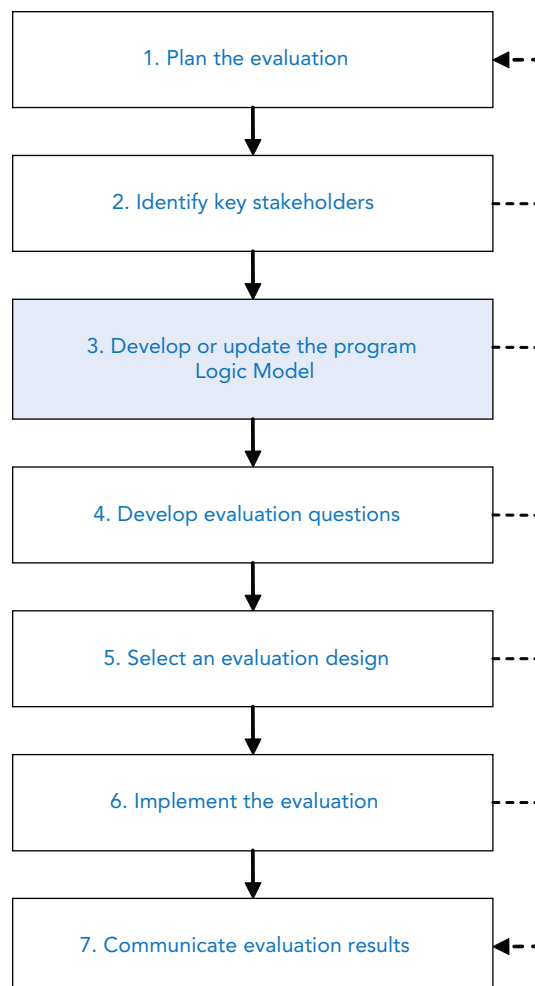
Develop or Update the Program Logic Model

Why Is a Logic Model Important for Program Evaluation?

A logic model is a diagram and text that shows the relationship between your program's work and its desired results. Every program has resources, activities, outputs, target decision-makers, and desired outcomes; a logic model describes the logical (causal) relationships among these program elements.

Understanding your program clearly is essential for conducting a quality evaluation, as it helps to ensure that you are measuring the right indicators from your program, evaluating the right aspects of your program, and asking the right questions about your program.

If your program is already collecting performance information, someone might have previously constructed a logic model. Whether reviewing an existing logic model or creating a new one, accurately characterizing the program through logic modeling is important because it ensures that program managers, contractors, and other stakeholders involved in designing the evaluation fully understand the Partnership Program.



These guidelines provide a simple approach to logic modeling, but other more complex logic model approaches could be used by EPA Partnership Programs. The logic model terms and definitions described here provide a basic framework that can be used across the variety of logic model approaches, however.

Logic Model Elements

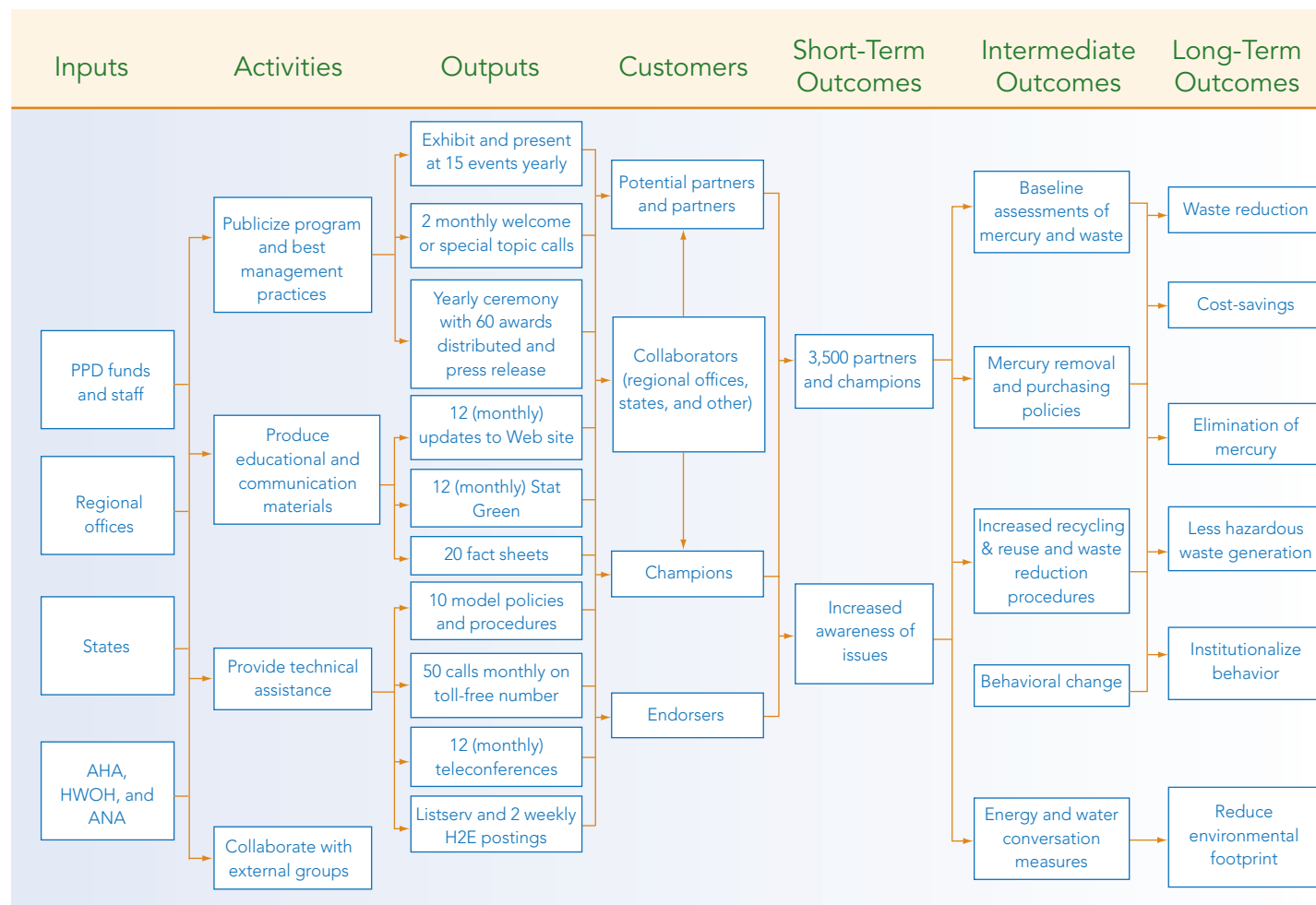
A logic model has seven basic program elements:

1. **Resources/Inputs**—What you have to run your program (e.g., people and dollars).
2. **Activities**—What your program does.
3. **Outputs**—The products/services your program produces or delivers.
4. **Target Decision-Makers**—Those groups whose behavior your program aims to affect.
5. **Short-Term Outcomes**—Changes in target decision-makers' knowledge, attitude, or skills.
6. **Intermediate-Term Outcomes**—Changes in the target decision-makers' behavior, practices, or decisions.
7. **Long-Term Outcomes**—Changes in public health and/or the environment as a result of your program.

EPA's Evaluation Support Division (ESD) offers periodic logic model training and can provide you with assistance in developing or revising a logic model for your program. In addition, presentations on how to develop a logic model are available online: www.epa.gov/evaluate/training.htm.

Also included in logic models are external influences (i.e., factors beyond your control), such as state programs that mandate or encourage the same behavioral changes as your program and other circumstances (positive or negative) that can affect how the program operates. Logic models also often include assumptions you currently have about your program (e.g., using water efficiently will extend the useful life of our existing water and wastewater infrastructure). The following figure is an example of what a Partnership Program logic model might look like. Boxes and arrows represent the logical connection between the separate program elements. Exercise 2 in the *Guidelines for Measuring the Performance of EPA Partnership Programs* includes a guide to help you through the process of developing a logic model for your program.

Logic Model for Hospitals for a Healthy Environment (H2E) Program (August 23, 2005)

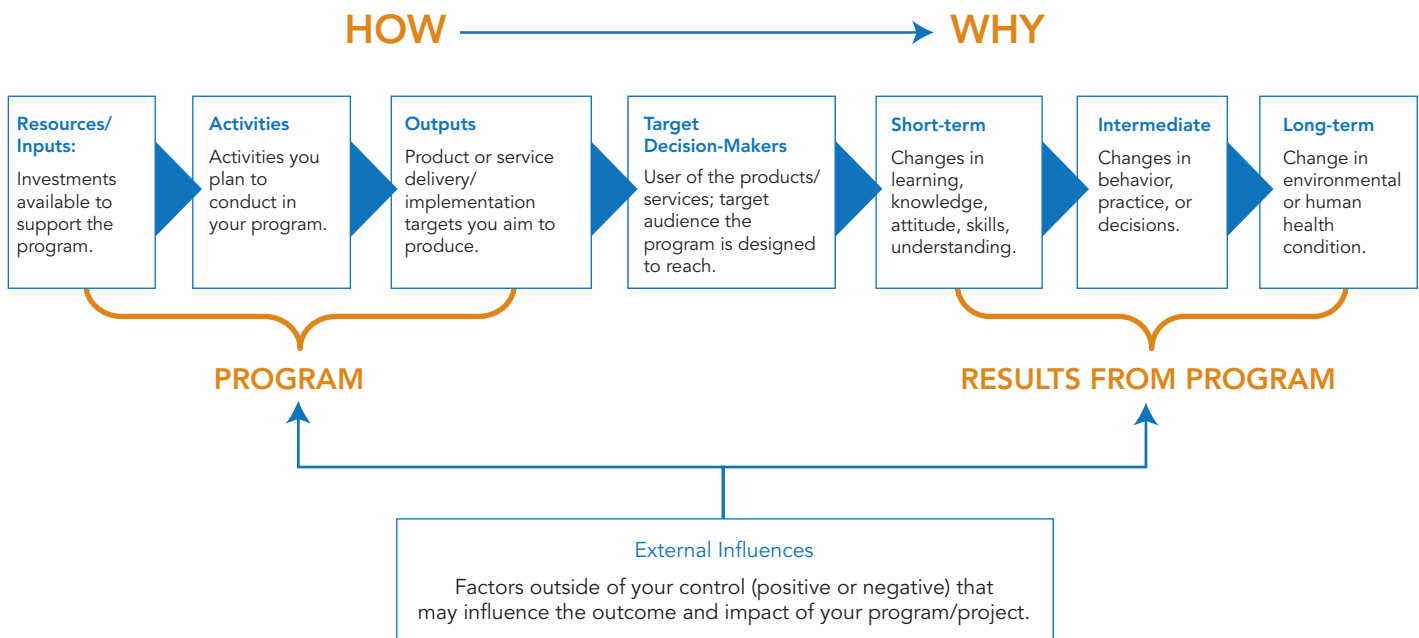


Abbreviations:

H2E: Hospitals for a Healthy Environment
PPD: Pollution Prevention Division

AHA: American Hospital Association
HWOH: Health Care Without Harm

ANA: American Nurses Association



Developing the Program Logic Model: The H2E Experience

EPA did not use logic models regularly until quite recently. In 2004, when H2E decided to undertake a program evaluation, the Partnership Program did not have a logic model. H2E proceeded to develop a logic model by involving all key internal and external stakeholder groups, allowing different stakeholders to see how others conceptualized the Partnership Program. This activity helped to build a broad consensus about: 1) major elements of the program (e.g., inputs, activities, and outputs); 2) expected program results (especially the short-term and intermediate outcomes), and 3) major influences on program results that fell outside of H2E's direct control. The logic model also helped the core evaluation team to clarify stakeholder concerns about conducting a program evaluation.

H2E managers and staff used the logic model process to develop a clearer picture of the links between the program's elements and expected results. This process helped the core evaluation team prioritize among a wide range of potential evaluation questions, select the program evaluation's design, and communicate the results.

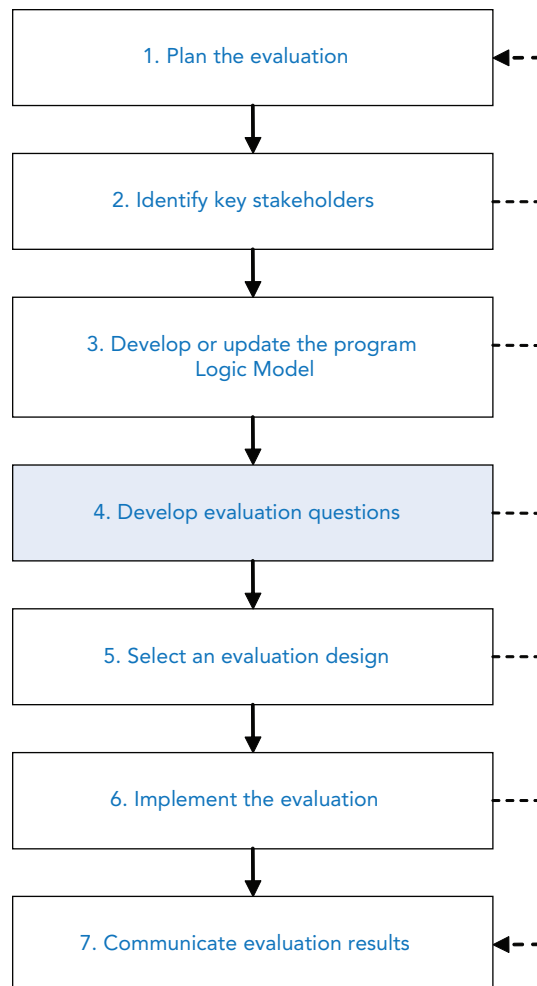
Chapter 4:

Develop Evaluation Questions

Evaluation questions are the broad questions that the evaluation is designed to answer. They are often inspired by or build upon existing performance measures, but they differ from performance measures in several ways.

Performance measures are used to gather data on your program's day-to-day activities and outputs. In contrast, evaluation questions delve more deeply into the reasons behind program accomplishments and seek to answer whether current operations are sufficient to achieve long-term goals. Good evaluation questions are important because they articulate the issues and concerns of stakeholders, examine how the program is expected to work and its intended outcomes, and frame the scope of the evaluation.

While interview, focus group, or survey questions are specific data collection tools that are used to gather information from participants that will be used to address the larger evaluation, evaluation questions specify the overall questions the study seeks to answer.



Your logic model is an excellent place for you and your evaluator to start the process of determining what questions you will answer in your evaluation. Each of the elements in a logic model can be thought of as an evaluation question, such as those questions produced by the logic model shown in the final row of the following table.

Typical EPA program evaluations use three to eight evaluation questions. By working with the program logic model and engaging relevant stakeholders, you and your evaluator can develop the key evaluation questions. The following five steps should aid evaluators in the process of designing evaluation questions:

1. Review the purpose and objectives of the program and the evaluation.
2. Review the logic model and identify what aspects of your program you wish to evaluate.
3. Consult with stakeholders and conduct a brief literature search for studies on programs similar to yours.
4. Generate a potential list of questions.
5. Group questions by themes or categories (e.g., resource questions, process questions, outcome questions).

Logic Model and Evaluation Questions Mapping Example

	Resources	Activities	Outputs	Target Decision-Makers Reached	Short-Term Outcomes	Intermediate Outcomes	Long-term Outcomes
Logic Model Elements	\$100,000 2 FTEs	Develop workbooks Develop Web site and marketing materials Develop technical assistance program	Workbook in Spanish and English Web site Information packet Onsite visits	Sector trade associations Plant managers	Participants learn about the program and chemical substitutions through training Trade associations sign memoranda of understanding and advocate for member participation	Plant managers use greener chemicals	Reduced risk to the environment and human health
Evaluation Questions	Are resources sufficient to affect desired change?	Are activities in line with program goals?	Are target decision-makers aware of outputs? Is the program being delivered as intended to target decision-makers?		Is the program effective in educating target decision-makers?	Are the desired program outcomes obtained? Did the program cause the outcomes?	(Because it is very difficult to measure long-term outcomes directly, we use questions about intermediate outcomes as proxies)

When you review your chosen evaluation questions, you and your evaluator should make sure that they will be effective in measuring progress toward program goals and against identified baselines. When finalizing your evaluation questions consider the following:

- Are the questions framed so that the answers are measurable in a high quality and feasible way?
- Are the questions relevant, important, and useful for informing program management or policy decisions?
- Are the primary questions of all of the key stakeholders represented?

Defining evaluation questions carefully at the beginning of an evaluation is important, as they will drive the evaluation design, measurement selection, information collection, and reporting.

Developing Evaluation Questions: The H2E Experience

H2E's core evaluation team used the logic modeling process to identify evaluation questions but generated too many questions to answer with one program evaluation. The next step was to prioritize questions.

The core evaluation team considered the balance among practical constraints (such as data necessary to answer questions), resources (such as time), and programmatic priorities (the information the program could use immediately to make key decisions). H2E's core evaluation team decided that the program evaluation should focus on four questions that could be traced along the logic model: 1) In what types of environmental activities are H2E partner hospitals engaged? 2) How can H2E be improved in terms of the services it offers? 3) How satisfied are H2E partners with the key elements of the program? 4) What measurable environmental outcomes can H2E partner hospitals show? The fourth question became the heart of H2E's outcome evaluation, but questions 1 through 3 were also essential because they helped illustrate the logical links between the program activities and the outcomes observed.

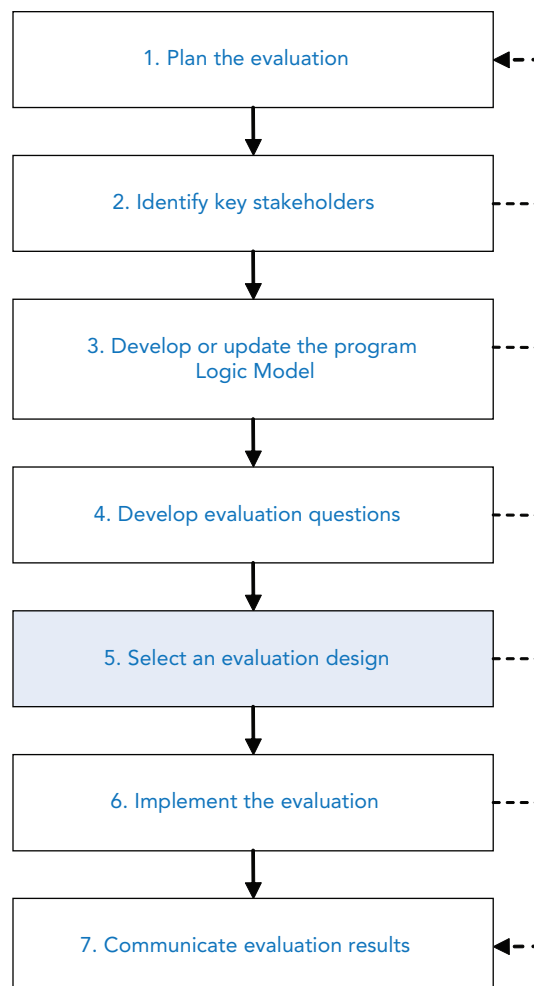
Chapter 5:

Select an Evaluation Design

Once you and your evaluator have reviewed your logic model and evaluation questions, consider the following issues to help choose the right design:

- What is the overarching question your Partnership Program needs to answer?
- Where is your Partnership Program in its life cycle?
- What do you hope to show with the results obtained from the evaluation?
- What additional technical evaluation expertise will you need to carry out the evaluation as designed?

The issues above overlap with those raised in Chapter 1: because the program evaluation process is typically iterative as it proceeds through the planning, design, and implementation steps. At this stage, you should revisit your *overarching* evaluation and determine if you will be conducting a design, process, outcome, or impact evaluation (each described in detail in [Chapter 1](#)).



The Foundations of Good Program Evaluation Design

When your Partnership Program communicates with key stakeholders about the implementation and results of a program evaluation, you and your evaluator will likely be asked questions related to the rigor and appropriateness of the program evaluation design. You and your evaluator should have a thoughtful response to these types of questions:

1. Is the evaluation design appropriate to answer the evaluation question(s) posed? Is a process evaluation design most desirable, or are outcome and impact evaluations designs?
2. Are the data you are collecting to represent performance elements measuring what they are supposed to measure? Are the data valid?
3. Is your measurement of the resources, activities, outputs, and outcomes repeatable and likely to yield the same results if undertaken by another evaluator? Are the data reliable?
4. Do you have the money, staff time, and stakeholder buy-in that you need to answer your program evaluation question(s)? Is the evaluation design feasible?
5. Can the information collected through your evaluation be acted upon by program staff? Is the evaluation design functional?

Clarifying how the program evaluation design handles validity, reliability, feasibility, and functionality will help you and your evaluator prepare for the scrutiny of external reviewers and yield results that will more accurately reflect your program's performance, ultimately leading to high-quality recommendations on which your program can act.

To ensure that the program evaluation design addresses validity, reliability, and feasibility, a good program evaluator will consult the relevant technical and program evaluation literature. A technical literature review involves consulting published information on how the Partnership Program operates. Additionally, a review of relevant program evaluation literature will focus on past program evaluations of programs with similarities to your program. The documentation of this review can be as simple as a bibliography in the report or as complex as a detailed stand-alone document. Regardless of its length, the literature review should be made available to internal and external stakeholders to increase the transparency of the program evaluation process and assist in validating the program evaluations findings, conclusions, and recommendations.

Much of the discussion surrounding the quality of a program evaluation involves the concept of rigor. Because well-designed outcome and impact evaluations are better able to determine a direct causal link between a program's activities and a program's results than other evaluation types, these evaluations are frequently associated with greater design rigor. In spite of this, an impact evaluation design is not necessarily more rigorous than a process evaluation design. The rigor of a program evaluation is not determined solely by the type of evaluation that you undertake but instead by the overall evaluation design and implementation (for more about implementation, please see Chapter 6).

The design phase of a program evaluation is a highly iterative process; while this chapter gives a linear description of the design phase, you and your evaluator are likely to revisit various issues several times. Decisions about data needs, how

those data can be collected, and the evaluation methodology will all inform the overall design. Your approach to engaging stakeholders (e.g., the members of your core evaluation team and other interested parties) will influence how iterative this phase becomes.

Assessing the Data Needs for the Evaluation Design

You should consider the several classes of data needs when planning your evaluation design.

1) Type of claims your program is expected to address: attribution or contribution.

Attribution involves making claims about the causal links between program activities and outcomes, achieved by comparing the observed outcomes with an estimate of what would have happened in the absence of the program. Partnership Programs, like other EPA programs, often have a difficult time demonstrating attribution because the program itself is often only one of a variety of factors that influence partners' environmental decision-making.

Contribution, in contrast to attribution, involves measuring the correlations that exist between program activities and outcomes after you have controlled for all of the other plausible explanations that might influence the results you observe. Contribution can tell you that your program likely had an effect on the outcome but cannot confidently demonstrate that your program alone has caused the results observed.

Demonstrating attribution should not be thought of as inherently better than demonstrat-

ing contribution; instead, it is simply a matter of what is needed by the program.

2) The use of original primary data or existing secondary data.

Primary data are data collected first-hand by your Partnership Program, whereas secondary data are data gathered from existing sources that have been collected by others for reasons independent of your evaluation.

The assessment of your data needs should follow in three broad steps:

- Review the primary data that your program already collects for existing performance measurement reporting and see if these measures can be used to address your evaluation questions.
- Search for sources of secondary data that others are collecting and that will appropriately serve your evaluation needs.
- If needed, plan a primary data collection specifically for the purpose of the evaluation.

3) The form of data you require: qualitative or quantitative data.

Data form will shape your later analyses and the degree to which you can generalize your findings. **Qualitative data** are often in-depth collections of information gathered through observations, focus groups, interviews, document reviews, and photographs. They are non-numerical and are classified into discrete categories for analysis. In contrast, **quantitative data** are usually collected through reports, tests, surveys, and existing databases. They are numerical measures of your program (e.g., pounds of emissions) that are usually summarized to present

general trends that characterize the sample from which these data are drawn. The decision to use qualitative or quantitative data is not an either/or proposition. Instead, you should consider which form of data is most useful (given the evaluation question and context). In many cases, collecting both qualitative and quantitative data in the same evaluation will present the most complete picture of your program. As you are designing your evaluation, consult with your evaluator on which type of data will best suit your evaluation needs.

Planning ahead in regard to data collection can reduce the costs of conducting a program evaluation and increase the quality. Early data collection improves the likelihood that you have access to baseline data, and by planning to evaluate early on, you can ensure that your program's performance measures are collecting the type and quality of data that you need. Your evaluator should assess your program's performance measurement data by asking you the following questions:

- Are the data complete and of high quality? Can you be sure that pieces of data are not missing due to inconsistent recordkeeping, systematic omissions in data, or other irregularities?
- Are your measurement tools a valid assessment of the program elements you are investigating with your evaluation questions?
- Are the data collection techniques reliable enough to render the same results if they were independently collected by someone else?
- Are the data gathered in a way that allows them to be used to answer any of the evaluation questions (e.g., are comparable data available from program non-participants)?

If you find yourself answering “no” to any of these questions, you should consider collecting additional data.

Quality Assurance Project Plans

Regardless of the form of your data (qualitative or quantitative, primary or secondary), you should ensure that the data have been subjected to a quality assurance project plan (QAPP) review. Specifically, the QAPP will describe the purpose of the evaluation, the methodology used to collect data for the report, how and where data for the evaluation were collected, why the particular data collection method was chosen, how the data will be used and by whom, how the resulting evaluation report will be used and by whom, and what the limitations are of the data collected. www.epa.gov/quality/qual_sys.html

The *Guidelines for Measuring the Performance of EPA Partnership Programs* present a more detailed guide to data collection. The table that follows describes a number of data collection methods used for program evaluation and the relative advantages and challenges associated with each. You and your evaluator should weigh the benefits and costs of each before selecting a data collection method. Using these methods to collect data can be more complex than it appears at first glance. Poorly collected data can undermine your evaluation's usefulness and credibility. Before undertaking any of these methods, consult with someone experienced in your chosen method.

Data Collection Methods

Method	Overall Purpose	Advantages	Challenges	Form of Data
<ul style="list-style-type: none"> • Direct Monitoring 	<ul style="list-style-type: none"> • To measure environmental indicators or emissions (e.g., pounds of waste, ambient air quality) to assess degree to which changes are occurring 	<ul style="list-style-type: none"> • Can provide evidence of program impact and yield information useful for accountability purposes • Shows whether the program is accomplishing its primary goal – environmental improvement 	<ul style="list-style-type: none"> • Might reveal changes in indicators only over periods of many years; might not be very sensitive to annual changes for annual reporting • Is time consuming because it takes time to obtain data and see trends in the results • Might make it difficult to attribute environmental results to program activities • Is costly if not normally collected • Requires that quality of secondary data be ensured 	<ul style="list-style-type: none"> • Quantitative
<ul style="list-style-type: none"> • Interviews 	<ul style="list-style-type: none"> • To fully understand someone's impressions or experiences, or learn more about their answers to questionnaires 	<ul style="list-style-type: none"> • Provide a full range and depth of information • Allow for development of relationship with respondent • Can be flexible 	<ul style="list-style-type: none"> • Are time consuming/costly • Produce results that can be hard to compare • Can produce biased responses depending on the interviewer's technique • Can produce inaccurate results if respondent recall is inaccurate • Might require an Information Collection Request (ICR) 	<ul style="list-style-type: none"> • Qualitative or quantitative
<ul style="list-style-type: none"> • Focus Groups 	<ul style="list-style-type: none"> • To explore a topic in depth through group discussion 	<ul style="list-style-type: none"> • Quickly and reliably capture common impressions • Can be an efficient way to get a greater range and depth of information in a short time • Can convey key information about programs 	<ul style="list-style-type: none"> • Can be difficult to analyze • Can involve a group dynamic that may affect responses • Need a good facilitator • Can be difficult to schedule • Can produce inaccurate results if respondent recall is inaccurate • Might require an Information Collection Request (ICR) 	<ul style="list-style-type: none"> • Qualitative

Data Collection Methods (continued)

Method	Overall Purpose	Advantages	Challenges	Form of Data
<ul style="list-style-type: none"> • Direct Observation of Behavior and Program Process 	<ul style="list-style-type: none"> • To gather information about how a program actually operates, particularly about processes 	<ul style="list-style-type: none"> • Allow events to be witnessed in real-time • Allow events to be observed within a context • Provide possible insight into personal behavior and motives 	<ul style="list-style-type: none"> • Can be difficult to interpret • Are time consuming • When observers are present, can influence behaviors of program participants 	<ul style="list-style-type: none"> • Qualitative or quantitative
<ul style="list-style-type: none"> • Surveys, Checklists 	<ul style="list-style-type: none"> • To collect answers to pre-determined questions from a large number of respondents, often for statistical analysis 	<ul style="list-style-type: none"> • Can be completed anonymously • Are inexpensive to administer to many people • Are easy to compare and analyze • Can produce a lot of data • With a representative sample, can produce results that can be extrapolated to wider population • Can partner with other programs, academic institutions, federal partners, and trade associations to share existing instruments and data sets 	<ul style="list-style-type: none"> • Can bias responses, depending on wording; might not provide full story • Are impersonal • Can produce inaccurate results if respondent recall or feedback is inaccurate • Might require sampling expert, which can be costly • Might require an Information Collection Request (ICR) 	<ul style="list-style-type: none"> • Quantitative
<ul style="list-style-type: none"> • Document Reviews 	<ul style="list-style-type: none"> • To provide an impression of program operations through the review of existing program documentation 	<ul style="list-style-type: none"> • Gather historical information • Don't interrupt program or client's routine in program • Collects information that already exists 	<ul style="list-style-type: none"> • Are time consuming • Might provide incomplete information • Contain already-existing data only • Might be incomplete if access to some documents is restricted 	<ul style="list-style-type: none"> • Qualitative or quantitative
<ul style="list-style-type: none"> • Case Studies 	<ul style="list-style-type: none"> • To provide a comprehensive look at one or two elements or an entire program 	<ul style="list-style-type: none"> • Can provide full depiction of program operation • Can be a powerful means through which to portray the program 	<ul style="list-style-type: none"> • Are usually quite time consuming • Focus on one or two elements fundamental to program and give a deep, but not broad, view 	<ul style="list-style-type: none"> • Qualitative and occasionally quantitative

Primary Data Collection Challenges

The basic nature of EPA Partnership Programs creates several challenges for collecting primary data for program evaluation.

Data Needs Versus Data Collection

Techniques. EPA Partnership Programs must always balance obtaining data of sufficient quality to demonstrate useful results with not overburdening the partners from whom you would solicit the data. Though you and your evaluator must gather high-quality data, the requirements cannot be too onerous for partners. Any approach to primary data collection must consider the “tipping point” where the data collection itself becomes a disincentive to participation in your program. Additionally, obtaining data from non-participants is often difficult, which creates a major barrier to the design of control groups. Your evaluator can help you brainstorm possible sources for data on non-participants and evaluations designs without control groups.

Information Collection Requests. Another barrier worthy of particular note is the Information Collection Request (ICR). According to the Paperwork Reduction Act, ICRs must be granted by the Office of Management and Budget (OMB) before a federal agency collects the same or similar information from 10 or more non-federal parties. ICRs describe the information to be collected, give the reason why the information is needed, and estimate the time and cost to the public to answer the request. In ideal situations, OMB processes ICRs within six months of receipt; however, the ICR process can take a year or more to complete. If you and your evaluator anticipate needing to collect original data from outside the federal

government, you should begin this process very early in your evaluation planning. Currently ESD is working to develop resources to aid programs in navigating the ICR process to minimize the time for the review to be completed.

Before embarking on the ICR process, consider the following strategies for collecting new data that do not require obtaining an ICR (although the nature of some of the data you require might still make an ICR necessary):

- Develop strategies for collecting the same data from nine or fewer entities. For example, plan to ask different interview and survey questions to different respondents to allow for the participation of more than nine individuals.
- Identify third-party organizations that might be interested in collecting some of the data that you need for their own purposes. For example, the American Hospital Association conducted a survey of its members that EPA used as a data source for the evaluation of Hospitals for a Healthy Environment. **IMPORTANT:** you cannot ask third parties to collect data to support an EPA evaluation without triggering the ICR requirement; the third party must have an interest beyond the EPA evaluation for collecting the data.
- Explore EPA experts' access to scientific, technical, and economic data (e.g., Toxic Release Inventory, Risk-Screening Environmental Indicators, Inventory Update Rule Amendments, Dun and Bradstreet, Census Bureau, Energy Information Administration) and their availability to conduct data analyses.
- Evaluate the possibility of collaborating with a related evaluation effort on data collection,

especially other programs that have already received an ICR or plan to file an ICR (see box at right for more information on the ICR process).

- Explore the availability of existing EPA ICRs that might apply to your evaluation questions, such as [EPA's Customer Service ICR](#).
- Consider collecting data from federal sources. An ICR is not required if you survey federal employees.
- Consider all of the government agencies, academic institutions, other research organizations, professional associations, trade associations, and other groups that might share data they have collected that will serve your program's needs.
- Consider teaming up with another EPA program that needs to collect data from similar enterprises or sources and which might be willing to share the expense and effort to pursue an ICR.

Choosing an Evaluation Methodology

When a Partnership Program communicates with key stakeholders about the implementation and results of a program evaluation, the important question that will be asked is, "What is your program evaluation methodology?"

Your evaluator should be able to give the detailed technical answer to this question. As a Partnership Program manager or staff person, you do not need to be fully conversant on the technical aspects of design methodology, but you should be able to identify the defining characteristics and strengths and limitations of each of three broad classes of evaluation methodologies: non-experimental, quasi-experimental, and true experimental.

Tips When Filing Your Own ICR:

- Start the process as early as possible.
- Identify examples of similar programs that have received similar data collection clearance, and provide the examples to OMB.
- Look for examples of successful and pending ICR packages for projects similar to yours and read these as potential models for your own ICR. One way to locate these is through the General Services Administration site: www.reginfo.gov/public/do/PRAMain.
- Build future evaluation considerations into any program ICRs filed to avoid needing to file more than one. For example, new EPA Partnership Programs can file an ICR early on to cover planned performance measurement and future evaluation needs.

For more information or assistance with the ICR process, see www.epa.gov/icr.

I) Non-experimental designs are generally best suited to answering design and process questions (e.g., What are the inputs available for this program? Are the activities leading to customer satisfaction?). Non-experimental designs do not include comparison groups of individuals or groups not participating in the program. In fact, many of these designs involve no inherent comparison groups. Non-experimental designs involve measuring various elements of a logic model and describing these elements, rather than correlating them to other elements in the logic model. These designs can yield qualitative or quantitative data and are the most common in evaluations of EPA Partnership Programs.

Non-Experimental Design Example: A Partnership Program hires an independent evaluator to conduct an evaluation. Six months after the Agency rolls out the Partnership Program, the evaluator measures the air quality in the areas served by the program participants. The evaluator determines that air quality improved. The evaluator had no baseline or control group against which to compare the program's data; however, in assessing trends in the air quality data, and with a systematic consideration of other factors that could have produced the change, the evaluator could conclude that the Partnership Program worked to improve air quality.

2) Quasi-experimental designs are usually employed to answer questions of program outcome; they often compare outcomes of program participants with non-participants that have not been randomly selected. Alternately, a quasi-experiment might measure the results of a program before and after a particular intervention has taken place to see if the time-related changes can be linked to the program's interventions. Achieving the perfect equivalence between the groups being compared is often difficult because of uncontrolled factors such as spillover effects (see the text box on the following page). Instead, quasi-experimental designs demonstrate causal impact by ruling out other plausible explanations through rigorous measurement and control. Data generated through quasi-experimental methods are typically quantitative.

Quasi-Experimental Design Example: A Partnership Program hires an independent evaluator to conduct an evaluation. The evaluator collects air quality ratings from partner dry cleaners for the five years prior to program implementation—this shows the evaluator previous trends and provides a baseline. Six months after the Agency rolls out the EPA Partnership Program, the evaluator measures the air quality in the areas served by the partner dry cleaners and compared these data to the data from the previous five years. Based on the trends and changes from the baseline, the evaluator determines that air quality measurably and significantly improved after the Agency implemented the Partnership Program. The evaluator concluded that the Partnership Program worked to improve air quality.

Natural Experiments

You might get lucky and be able to use a quasi-experimental method known as a “natural experiment.” You are best able to capitalize on this scenario if, as a part of your program design, you identify one group that is receiving a particular program benefit and another that is not. Such intentional comparisons can be only achieved if the two groups are not systematically different on a dimension that might affect program outcomes and if any such pre-existing differences between the two groups can be reliably assessed. You should actively seek opportunities to compare similar groups who are program participants or non-participants in order to apply a “natural” group design.

The Best Workplaces for Commuters evaluation used a natural experiment—comparing individuals who joined the program with those who did not—to support its claims of effectiveness (www.bestworkplaces.cutr.usf.edu/pdf/evaluation-survey-findings-2005.pdf).

3) True experimental designs (alternately referred to as randomized control trials, or RCTs) involve the random assignment of potential program participants to either participate in or be excluded from the Partnership Program. These studies enable measurement of the causal impact and yield quantitative data that are analyzed for differences in environmental results between groups based on program participation. True experiments can be used in evaluating Partnership Programs when clearly defined interventions can be manipulated and uniformly

administered; when there is no possibility that treatment will spill over to control groups (those for whom a program's intervention is not intended); and when it is ethical and feasible to deny a program's services to a particular group. RCTs have been labelled as the "gold standard" for program evaluation; however, because of the caveats just described, true experimental designs are more a theoretical ideal than a practical reality for most programs, making the demonstration of statistically significant impact very difficult for EPA Partnership Programs. The manipulation of a particular program's benefits, which would be central to the design of a RCT on a Partnership Program, runs counter to the

True Experimental Design Example: A Partnership Program hires an independent evaluator to conduct an evaluation. *The first step in the evaluation is to work with EPA to identify a pool of 12 possible communities where the Partnership Program could be implemented. All communities have similar demographic, ecological, economic, and sociological characteristics. EPA, with the support of the evaluator, randomly assigns six sites to be a comparison group designated as areas served by non-participants. The evaluator collects air quality monitoring data for the five years prior to the program implementation from all 12 sites. As the study progresses, the evaluator collects data on program implementation from participants to determine if the program is being applied as designed. The evaluator also collected process data from non-participants*

After six months, the evaluator measures the air quality in the areas served by the participants and compares the data to the five-year data. In addition, the evaluator compares the areas served by participants to air quality in the non-participant comparison sites. The air quality in areas served by the Partnership Program is significantly better than the pre-assessment trends and is significantly better than the air quality from non-participant comparison sites. The evaluator determines that air quality has improved after the implementation of and due to the Partnership Program.

The Spillover Effect

The spillover effect occurs when participants of Partnership Program share knowledge or technologies gained through participation in the program with non-participants. This effect is quite common to Partnership Programs, and it is desirable because the transfer of technology and knowledge and best practices can lead to environmental improvements from non-participants as well as participants.

The spillover effect can pose a challenge to program evaluators, however, in determining causality when non-participants gain the same knowledge as program participants, indirectly and not within measurable circles.

Analyzing spillover effects can be particularly fruitful for sector-based programs. The Coal Combustion Partnership is one example of an EPA Partnership Program that has analyzed spillover effects. (See *Evaluating Voluntary Programs with Spillovers: The Case of Coal Combustion Products Partnership*).

spirit of spillover, or the sharing of a program's goals and philosophy, that Partnership Programs both espouse and encourage.

Quasi-experimental and experimental designs can be very complex to implement unless the capacity to conduct them has been a central part of the program's initial design. As the complexity of your evaluation methodology increases, so too will the resources (money, time and buy-in) required. Therefore, you and your evaluator should regularly check in throughout the program evaluation design selection phase to ensure that the evaluation methodology selected can be supported by your available resources. You and your evaluator might determine that a particular evaluation question cannot be sufficiently answered with the evaluation design options available to you. In such instances, you and your evaluator might want to revisit the logic model to see if you can determine another important evaluation question that fits within your resource capacity.

Expert Review of the Evaluation Design

A final step that you should consider before implementing your evaluation is an external expert review of the evaluation design selected. These reviews will help ensure the actual and perceived quality and credibility of your evaluation. Before commissioning a review of your design, you should carefully consider the technical expertise of the intended audience, the availability of resources and time, and the function of the evaluation's results. Not all evaluations need to undergo an external review before the implementation is underway.

Selecting the Evaluation Design: The H2E Experience

The centerpiece of H2E's outcome evaluation was a quasi-experimental design to compare the behavior of program participants to the behavior of non-participants in terms of implementing actions that would eliminate mercury. Answering the question "What measurable environmental outcomes can H2E partner hospitals show?" relied on primary data collected about H2E participants' self-reported actions to eliminate mercury-containing waste. H2E did not collect these data directly, however; the program was able to access information from a trade association. Because this trade association was collecting this information through its own survey of its members, H2E did not need to have an ICR for this data collection.

In addition, H2E gained access to secondary data from EPA program offices about mercury-containing waste materials at medical waste incinerators and municipal landfills. These data were used to shed light on national trends in the level of mercury-containing waste, though it was not possible to isolate the direct causal impact of H2E on these national data.

H2E did collect data to answer three other questions that would support the results of the outcome question. A telephone survey conducted by the evaluation contractor gathered primary data on customer satisfaction. H2E obtained OMB approval for the telephone survey through EPA's generic customer service survey ICR, which minimized time and paperwork.

Chapter 6:

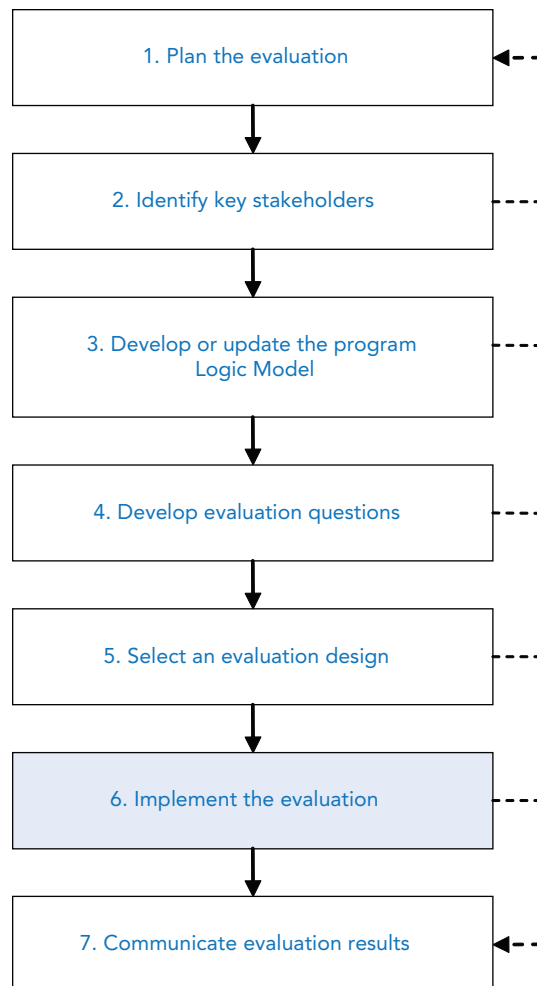
Implement the Evaluation

After you have settled on your evaluation questions and evaluation design, you are ready to implement the evaluation. At this important juncture, you should step back and let your evaluator carry out the program evaluation; however, a few key areas of implementation require your involvement

Your involvement in the implementation phase might be limited to ensuring that your evaluator has employed proper pilot-testing/field testing procedures for example. In fact, whether you are conducting an internal or external evaluation, your periodic check-ins will ensure that the method used is yielding data that will allow you to answer your evaluation questions. Informing participants about the importance of the evaluation and encouraging them to participate in the data collection conducted by the evaluator is another way to be involved.

Pilot Testing the Evaluation

Pilot testing should take place prior to the full implementation of your evaluation. A pilot test involves testing particular tools or components of the evaluation, in a limited capacity, with a small number of informed respondents who



can provide feedback on the usefulness of the approach; for example, you should encourage your evaluator to test a draft of interview questions/survey questions with two to four people who represent (or are similar to) the people from whom the evaluation will ultimately be collecting data. Your evaluator might also want to pilot-test the sampling and data entry processes, particularly if different people will be collecting and/or entering the information. Your evaluator might also want to revise the data collection instrument or processes based on the comments of the pilot respondents or trial runs at data collection.

Once you and your evaluator are confident about and comfortable with the tools and processes you have selected, your evaluator should proceed to full implementation of the evaluation design.

Protocols for Collecting and Housing Evaluation Data

You and your evaluator should agree to protocols for collecting and housing data during and after the implementation of your program evaluation. Issues to consider include:

- What form will my data take (e.g., text or numbers)?
- How much information will be collected, how often, and for how long?
- Do I anticipate that my data collection needs will grow or diminish in the future?
- What capabilities am I looking for in my data management system (e.g., a place to input and store data, software that will enable the analysis of quantitative or qualitative data)?

- What data management systems for the program currently exist? Could they fulfill my needs or be adapted to meet our needs?
- Who will need to have access to the data (e.g., EPA staff, the public)?

Data Analysis

Once the pilot testing and data collection are complete, you and your evaluator must analyze and interpret the information and reach judgments about the findings for your program. The analysis will vary depending on the data collected (quantitative or qualitative; primary or secondary) and the purpose of the evaluation.

Quantitative Data: Often, quantitative data are collected and organized with the intent of being statistically analyzed; however, some important limitations on statistical analysis sometimes affect a Partnership Program's ability to conduct a valid statistical analysis. The most common barrier is small sample sizes, which lead to low statistical power, or a low probability of observing a statistically significant effect. Due to the number of Partnership Program participants, sample sizes are often small. Your evaluator can help you brainstorm ways to overcome this barrier that will enable you to draw inferences about causation or correlation.

If you are conducting an impact evaluation and have sufficient data, then you can evaluate the extent to which the relationship between your program and a change you have observed is statistically significant. These tests generally involve examining the relationship between **dependent variables** and **independent variables**.

Dependent variables are those aspects of your program that are subjected to performance measurement and are the central focus of your evaluation efforts. In some focused way, you are examining the degree to which your program produces the desired outcome or result that is captured with a particular part of your program's logic model. This could be a measure you are trying to influence with your program. For example, did emissions decrease or did the environment otherwise improve? Independent variables are those measured aspects of your program that you believe might have *caused* the observed change, such as the activities of the Partnership Program. Sometimes, you will collect data that will give you a sense of whether your program can reasonably (within the rules of statistical probability) conclude that there is a relationship between the dependent and independent variables. In other words, is your outcome unlikely to have resulted by chance (i.e., is this relationship statistically significant). In a number of other cases, you may describe that a certain element of your program has been produced by another element based on logic and reasoning that cannot be subjected to formal statistical tests but that reasonably follow from other systematic methods. When working with your evaluator, be sure to ask:

- What types of analyses do our data support?
- What do the results tell us?
- How confident are you in the results? Are the results statistically significant?
- What do the results allow us to say about the relationship between the variables?
- Are there any findings that we predicted that the findings *do not* support?

Even if your quantitative data do not support an analysis of statistical significance, they still may be systematically analyzed in order to observe trends. At a minimum, your evaluator should also provide descriptive statistics such as means and medians, ranges, and quartiles, as appropriate.

Qualitative data: Data collected from interviews, surveys, focus groups, and other means should be categorized and organized in a manner that supports analysis. One helpful practice is to code the data by category. Coding makes it easier to search the data, make comparisons, and identify any patterns that require further investigation. Placing the information in a database will allow you or your evaluator to efficiently organize the data by question or respondent and allow you to see important themes and trends. A database will also help with simple quantitative analyses such as the number of respondents who provided a certain reply. The evaluator should also provide numeric breakdown, as appropriate; for example, the percentage breakdown of various responses to a specific interview or survey question.

Your evaluator should have the technical expertise to undertake a proper content analysis for qualitative data or a statistical analysis for quantitative data. You also play an important role in this analysis. You should be available to answer questions that enable the evaluator to identify and investigate potential data problems or other anomalies as they arise, give the evaluator feedback on what data analysis will meet the needs of your audience, and help provide context and insights during interpretation of the findings, including possible explanations for counterintuitive results.

Based on your expertise and familiarity with the program, you can provide important insight into how the findings are interpreted and what program changes might be needed to respond to the findings. Merely because some relationships are seen as statistically significant does not mean that they are meaningful with regard to your program. The reverse is also true. You need

to carefully review all results and determine which are meaningful and should guide possible changes in your program. You and your evaluator should work together to make sure that the data analysis is transparent and that results are communicated effectively to the intended audience.

Implementing the Evaluation: The H2E Experience

In contrast to the previous steps, H2E staff had a less direct role in the implementation stage of the evaluation process. During the data collection process, H2E staff worked with the evaluation contractor to address data collection challenges and served as liaisons between the evaluation contractor and partners. H2E staff helped the contractor identify the key materials for the document review. H2E staff worked with OMB to gain approval of the telephone survey via the generic customer service ICR. The evaluation contractor took the lead role in analyzing the data but did conduct regular check-ins with the other members of the core evaluation team. During these check-ins, the contractor asked program staff for their reactions to preliminary results and checked to make sure the work stayed on schedule.

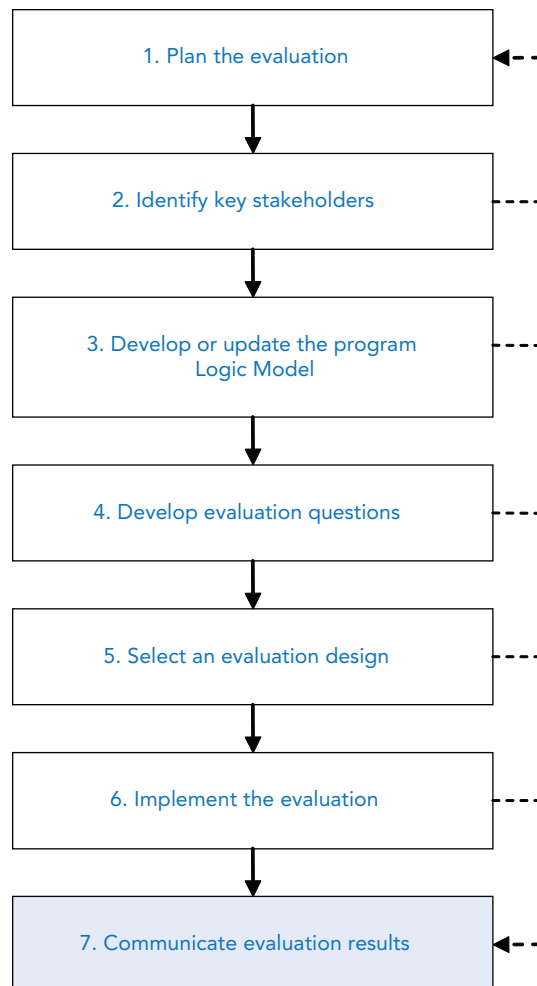
Chapter 7:

Communicate Evaluation Results

Although communicating your results is the final step in the evaluation process, you and your evaluator should start planning early for this important step. As discussed in [Chapter 6](#), when implementing the evaluation, your evaluator will take primary responsibility for collecting and analyzing the data; however, the process of communicating evaluation results requires continual collaboration between the evaluator and Partnership Program staff.

Careful consideration of your Partnership Program's stakeholders will influence how to best organize and deliver evaluation reports and briefings. The results have three basic elements: findings, conclusions, and recommendations.

Data collected during the implementation of the project will yield **findings**. Findings refer to the raw data and summary analyses. Because the findings are a part of the data analysis process, the evaluator should retain the primary responsibility for communicating findings to the program staff and management (in verbal or written form). Evaluators often deliver findings to the Partnership Program in a draft report or draft briefing.



Conclusions represent the interpretation of the findings, given the context and specific operations of your Partnership Program. Your evaluator might undertake an appropriate analysis of the data and might independently derive some initial interpretations; however, you and others closely linked to the program should have an opportunity to provide comments based on a draft report, in order to suggest ways to refine or contextualize the interpretation of the findings. This same process applies even if you have commissioned an independent, third-party evaluation, because a strong external evaluator will want to ensure that the presented conclusions are sound, relevant, and useful.

Regardless of the design or data collection employed, there will be some limitations to the explanatory power of any methodology used. Make sure that your evaluator has clearly pointed out the limitations of the findings, based on the design selected, when framing and reporting conclusions from the evaluation.

Recommendations are based on the findings and conclusions of your evaluation. A strong evaluator will understand that framing recommendations is an iterative process that should involve obtaining feedback from Partnership Program managers, staff, and key stakeholders. Again, this same process applies even if you have commissioned an independent, third-party evaluation, although in this case the external evaluator will make the key judgments about the report's final recommendations. Your involvement in the development of recommendations is important; to get the most value out of your evaluation, you should be prepared to implement some or all of the recommendations. Implementing the recommendations and the

resulting improvements to your program is one of the greatest sources of value to programs from the evaluation process.

Although you will commission an evaluation expert to conduct an objective, independent analysis, preliminary results and draft reports should be shared with core evaluation team members (at a minimum) for their feedback. Those who are directly involved in the program's activities are likely to have a critical role in helping to make sense of draft findings and make suggestions to the evaluator during the development of conclusions and recommendations. The evaluator will often also consult the published literature and experts in the area to make sure recommendations are objective, informed, and appropriate.

Throughout the program evaluation process, your evaluator should share the “evolving story” that is emerging from the data, when appropriate (i.e., without jeopardizing data validity and the evaluation's objectivity). In turn, the Partnership Program must keep the evaluator apprised of cultural and political sensitivities that could influence the form and format of how the results are presented. There should be no “surprises” when the final report is delivered, whether by an internal or external evaluator.

Despite the collaborative process that unfolds throughout the evaluation process and the need for active discussion of the findings, conclusions, and recommendations, the evaluator should take the lead on developing conclusions, recommendations and drafting the final report. Granting this autonomy to your evaluator will help ensure that the report is objective and is not unduly influenced by the vested interests and stakeholders

who might be affected—directly or indirectly—by the findings. This autonomy will also make the evaluation less vulnerable to the criticisms of external reviewers, who may be skeptical of the subjectivity and self-serving interpretations of those who work closely with the program.

Presenting Results

You and your evaluator should work closely to determine the level of detail and format of the draft report. You must tailor presentations of evaluation results to the specific needs of your stakeholders, which might or might not be satisfied by a lengthy report. Key questions you and your evaluator should ask in presenting results are:

- What evaluation questions are most relevant to these stakeholders?
- How do they prefer to receive information?
- How much detail do they want?
- Are they likely to read an entire report?

Based on the answers to these questions, in addition to a full-length report, you can opt for one or more of the following reporting formats depending on the needs of each stakeholder group:

- A shortened version of the evaluation report for broad distribution
- A one- or two-page executive summary of key results and conclusions
- A PowerPoint briefing

ESD's Report Formatting and Presentation Guidelines presents additional information on evaluation report development.

The applicability and relevance of your results will be strengthened by the degree to which

you tie your findings directly to the evaluation questions and back to the logic model. Organizing your findings and recommendations in this manner will ensure that you have collected and are reporting on the key questions that the evaluation was designed to answer. Here are some tips to assist you in applying the findings of your program evaluation:

- Consider whether the results provide support for or challenge the linkages you expected to see in your logic model. Work with program staff and your evaluator to consider a reasonable set of explanations for the results obtained.
- Consult the literature to see if these results are consistent with findings published and presented on similar programs.
- Work with technical experts and program personnel to develop evidence-based explanations to interpret your results.
- If you did not get the results expected, develop a set of possible explanations that might explain your counterintuitive findings.

Questions to Ask About Your Results

- Do the results make sense?
- Do the results provide answers to evaluation questions?
- Can the evaluation results be attributed to the program?
- What are some possible explanations for findings that are surprising?
- Have we missed other indicators or confounding variables?
- How will the results help you identify actions to improve the program?

- Consult with stakeholders and external experts to develop a list of actionable items that can inform your management decisions; these items might be used later used to frame recommendations.
- Consider any methodological deficits of your evaluation strategy and consider design shortcomings when applying the results to your program management directives.
- Make sure that your results are transparent and that you share expected as well as counterintuitive results. Do not suppress findings. Obtaining results inconsistent with your logic model does not necessarily suggest that the core goals of your program are not worth pursuing, and including such findings will boost the integrity of your report.
- Suggest future evaluations that should follow from the current evaluation effort.
- Build the means for future evaluations into your program infrastructure (e.g., reliable record-keeping, accessible storage of data, valid measurement of baselines for new program activities) so that future program evaluations will have the advantage of having useful records to answer evaluation questions.

Checklist for Reporting Results and Conclusions (Yes or No)

Linkage of results to logic model is clear	
Conclusions and results are clearly presented and address key evaluation questions	
Clear discussion of next steps is included	
Stakeholders have participated in decisions concerning outreach method	
Stakeholders are provided with opportunity for comment before evaluation is finalized	

Communicating the Results: The H2E Experience

H2E's core evaluation team began communicating the initial findings with EPA stakeholders through internal briefings. The evaluation contractor took the lead in synthesizing input from these briefings and worked collaboratively with the rest of the H2E core evaluation team to draft conclusions from these findings. Finally, based on these conclusions, H2E's core evaluation team developed a series of recommendations, which the evaluation contractor summarized in a draft of the final report. The core evaluation team communicated with H2E's external partners through briefings and other meetings about the results before finishing the final report.

The evaluation contractor delivered a [final report](#) with several technical chapters and appendices that gave details about data sources, methodology, and other key aspects of the evaluation process. This report shared important insights into the limitations of the evaluation design and data collection and measurement challenges.

The executive summary played a key role in communicating the results of the evaluation because of its brevity. The H2E staff also developed talking points for briefings and fact sheets that highlighted the most important points for various audiences.

H2E managers and staff then used the evaluation results to help to determine EPA's role in the future of H2E: in 2006, this Partnership Program was "spun off" to become an independent nonprofit organization.

Appendices

Appendix A: Glossary

Activities: The actions you do to conduct your program. Examples of Partnership Program activities are developing and maintaining a program Web site, offering trainings, and establishing relationships with partners.

Attribution: The assertion that certain events or conditions were, to some extent, caused or influenced by other events or conditions. In program evaluation, attribution means a causal link can be made between a specific outcome and the actions and outputs of the program.

Baseline Data: Initial information on a program or program components collected prior to receipt of services or participation activities. Baseline data provide a frame of reference for the change that you want the Partnership Program to initiate. These data represent the current state of the environment, community, or sector before your program begins. Baseline data can also approximate what environmental results might have been in absence of the program.

Conclusions: The interpretation of the evaluation findings, given the context and specific operations of your Partnership Program.

Confounding Variable: A variable that is combined with your program's activities in such a way that your program's unique effects cannot be validly determined.

Contribution: The assertion that a program is statistically correlated with subsequent events or conditions, even after you have accounted for non-program factors also associated with the same events and conditions.

Control Group: A group whose characteristics are similar to those of the program but which did not receive the program services, products, or activities being evaluated. Collecting and comparing the same information for program participants and non-participants enables evaluators to assess the effect of program activities.

Customers: See "Target Decision-Makers"

Dependent Variable: The variable that represents what you are trying to influence with your program. It answers the question "what do I observe" (e.g., environmental results).

Evaluation Methodology: The methods, procedures, and techniques used to collect and analyze information for the evaluation.

Evaluation Practitioners: Those individuals that typically have significant evaluation knowledge and are generally capable of planning and managing an evaluation without external assistance. Evaluation practitioners might occasionally need to seek advice on advanced methodologies from outside experts or the Evaluation Support Division.

Evaluation Questions: The broad questions the evaluation is designed to answer and the bridge between the description of how a program is intended to operate and the data necessary to support claims about program success.

Evaluation Users: Most EPA Partnership Program managers and staff, who often have limited knowledge of program evaluation but benefit from and see the value of evaluations. From time to time, evaluation users might be called upon to participate in the evaluation process.

Expert Review: An impartial assessment of the evaluation methodology by experts who are not otherwise involved with the program or the evaluation; a form of peer review. EPA's Peer Review Handbook outlines requirements for Peer Review of major scientific and technical work products, provides useful tips to managing expert reviews.

External Evaluation: Development and implementation of the evaluation methodology by an independent third party, such as an academic institution or other group.

External Influences: Positive or negative factors beyond your control that can affect the ability of your program to reach its desired outcomes.

Feasibility: The extent to which an evaluation design is practical, including having an adequate budget, data collection and analysis capacity, staff time, and stakeholder buy-in required to answer evaluation questions.

Findings: The raw data and summary analyses obtained from the respondents in a program evaluation effort.

Functionality: The extent to which information collected through the evaluation process can be acted upon by program staff.

Impact Evaluation: Focuses on questions of program causality; allows claims to be made with some degree of certainty about the link between the program and outcomes; assesses the net effect of a program by comparing program outcomes with an estimate of what would have happened in the absence of the program.

Independent Variable: The variable that represents the hypothesized cause (e.g., Partnership Program activities) of the observations during the evaluation.

Indicator: Measure, usually quantitative, that provides information on program performance and evidence of a change in the "state or condition" of the system.

Information Collection Request (ICR):

A set of documents that describe reporting, recordkeeping, survey, or other information collection requirements imposed on the public by federal agencies. Each request must be sent to and approved by the Office of Management and Budget before a collection begins. The ICR provides an overview of the collection and an estimate of the cost and time for the public to respond. The public may view an ICR and submit comments on the ICR.

Internal Evaluation: Conducted by staff members within the program being studied, typically EPA staff and/or by EPA staff and contractors who regularly support evaluation at EPA.

Intermediate-Term Outcomes: Changes in behavior that are broader in scope than short-term outcomes; often build upon the progress achieved in the short-term.

Logic Model: A diagram with text that describes and illustrates the components of a program and the causal relationships among program elements and the problems they are intended to solve, thus defining measurement of success. Essentially, a logic model visually represents what a program does and how it intends to accomplish its goals.

Long-Term Outcomes: The overarching goals of the program, such as changes in environmental or human health conditions.

Mean: A measure of central tendency sometimes referred to as the average; the sum of the values divided by the number of values.

Median: A measure of central tendency; the number separating the upper and lower halves of a sample. The median can be found by ordering the numbers from lowest to highest and finding the middle number.

Natural Experiment: Situations that approximate a controlled experiment; that is, have “natural” comparison and treatment groups. This scenario provides evaluators with the opportunity to compare program participants with a group that is not receiving the program offered. Natural experiments are not randomized, however, and therefore strong causal claims of direct impact cannot be made and evidence is required to show that the comparison group is a reasonable approximation of an experimental control group.

Non-Experimental Design: A research design in which the evaluator is able to describe what has occurred but is not able to control or manipulate the provision of the treatment to participants as in a true experimental design or approximate control using strong quasi-experimental methods.

Outcome Evaluation: Assesses a mature program’s success in reaching its stated goals; the most common type of evaluation conducted for EPA programs. It focuses on outputs and outcomes (including unintended effects) to judge program effectiveness but can also assess program process to understand how outcomes are produced. Often, outcome evaluations are appropriate only when at least baseline and post-baseline data sets are available or could be developed.

Outputs: The immediate products that result from activities, often used to measure short-term progress.

Participatory Evaluation: Involves stakeholders in all aspects of the evaluation, including design, data collection, analysis, and communication of findings.

Partnership Program: Designed to proactively target and motivate external parties to take specific actions that improve human health and the environment. EPA does not compel external partners by law to take these actions and serves in a leadership role and has decision-making authority.

Partnership Program Manager: Responsible for determining what programs should be evaluated and when these evaluations should take place. Managers do not necessarily need to have the technical expertise to conduct an evaluation but should be aware of the basic structure of the evaluation process so they can make informed decisions when commissioning evaluations and using evaluation findings to make management decisions.

Partnership Program Staff: Responsible for leading or participating in the program evaluation; typically have limited experience with the technical aspects of program evaluation. Knowledge of basic program evaluation techniques they might encounter would be useful to them when working with seasoned evaluators, allowing them to be able to “speak the same language” as evaluation experts.

Performance Measure: An objective metric used to gauge program performance in achieving objectives and goals. Performance measures can address the type or level of program activities conducted (process), the direct products and services delivered by a program (outputs), or the results of those products and services (outcomes).

Performance Measurement: The ongoing monitoring and reporting of program accomplishments, particularly progress toward pre-established goals.

Primary Data: Data collected “first-hand” by your Partnership Program specifically for the evaluation.

Process Evaluation: This form of evaluation assesses the extent to which a program is operating as it was intended. Process evaluations are typically a check to see if all essential program elements are in place and operating successfully. Process evaluations can also be used to analyze mature programs under some circumstances, such as when you are considering changing the mechanics of the program.

Program Design Evaluation: Most appropriately conducted during program development; can be very helpful when staff have been charged with developing a new program. Program design evaluations provide a means for programs to evaluate the strategies and approaches that are most useful for a program to achieve its goals.

Program Evaluation: Systematic study that uses objective measurement and analysis to answer specific questions about how well a program is working to achieve its outcomes and why. Evaluation has several distinguishing characteristics relating to focus, methodology, and function. Evaluation 1) assesses the effectiveness of an ongoing program in achieving its objectives, 2) relies on the standards of project design to distinguish a program's effects from those of other forces, and 3) aims to improve programs by modifying current operations.

Qualitative Data: Describe the attributes or properties of a program's activities, outputs, or outcomes. Data can be difficult to measure, count, or express in numerical terms; therefore, data are sometimes converted into a form that enables summarization through a systematic process (e.g., content analysis, behavioral coding). Qualitative data are often initially unstructured and contain a high degree of subjectivity, such as free responses to open-ended questions. Various methods can be used constrain subjectivity of qualitative data, including analytical methods that use quantitative approaches.

Quality Assurance Project Plan (QAPP):

Describes the purpose of the Partnership Program evaluation, the methodology used to collect data for the report, how and where data for the evaluation were collected, why the particular data collection method was chosen, how the data will be used and by whom, how the resulting evaluation report will be used and by whom and, what are the limitations of data collected.

Quantitative Data: Can be expressed in numerical terms, counted, or compared on a scale. Measurement units (e.g., feet and inches) are associated with quantitative data.

Quartile: The three data points that divide a data set into four equal parts.

Quasi-Experimental Design: A research design with some, but not all, of the characteristics of an experimental design. Like randomized control trials (see below), these evaluations assess the differences that result from participation in program activities and the result that would have occurred without participation. The control activity (comparison group) is not randomly assigned, however. Instead, a comparison group is developed or identified through non-random means, and systematic methods are used to rule out confounding factors other than the program that could produce or mask differences between the program and non-program groups.

Randomized Control Trial (RCT): A true experimental study that is characterized by random assignment to program treatments (at least one group receives the goods or services offered by a program and at least one group—a control group—does not). Both groups are measured post-treatment. The random assignment enables the evaluator to assert with confidence that no other factors other than the program produced the outcomes achieved with the program.

Range: The difference between the highest and lowest value in a sample.

Recommendations: Suggestions for the Partnership Program based on the evaluation's findings and conclusions.

Reliability: The extent to which a measurement instrument yields consistent, stable, and uniform results over repeated observations or measurements under the same conditions.

Resources: The basic inputs of funds, staffing, and knowledge dedicated to the program.

Secondary Data: Data taken from existing sources and re-analyzed for a different purpose.

Short-Term Outcomes: The changes in awareness, attitudes, understanding, knowledge, or skills resulting from program outputs.

Spillover Effects: Environmental improvements by non-participants due to transfers of attitudes, beliefs, knowledge, or technology from program participants.

Stakeholder: Any person or group that has an interest in the program being evaluated or in the results of the evaluation.

Stakeholder Involvement Plan: A plan to identify relevant stakeholder groups to determine the appropriate level of involvement for each group and engage each group in the evaluation accordingly.

Targets: Improved level of performance needed to achieve stated goals.

Target Decision-Makers: The groups and individuals targeted by program activities and outputs, also known as the target audience or program participants.

True Experimental Design: A research design in which the researcher has control over the selection of participants in the study, and these participants are randomly assigned to treatment and control groups. See “Randomized Control Trial.”

Validity: The extent to which a data collection technique accurately measures what it is supposed to measure.

Appendix B: Evaluation Resources

Selected Evaluations of EPA Partnership Programs

The evaluations listed as follows represent a sample of individual EPA Partnership Programs that have conducted program evaluations. Full copies of some of these evaluation reports can be furnished upon request to EPA staff.

- *Do Employee Commuter Benefits Reduce Vehicle Emissions and Fuel Consumption? Results of the Fall 2004 Best Workplaces for Commuters Survey* (<http://www.bestworkplaces.cutr.usf.edu/pdf/evaluation-survey-findings-2005.pdf>): This impact evaluation involved measuring the benefits of the Best Workplaces for Commuters Partnership Program.
- *Evaluating Voluntary Programs With Spillovers: The Case of Coal Combustion Products Partnership (C2P2)* ([http://yosemite.epa.gov/ee/epa/eed.nsf/ffb05b5f4a2cf40985256d2d00740681/fla5438303eaa5b08525751b00690389/\\$FILE/2008-12.pdf](http://yosemite.epa.gov/ee/epa/eed.nsf/ffb05b5f4a2cf40985256d2d00740681/fla5438303eaa5b08525751b00690389/$FILE/2008-12.pdf)): This outcome evaluation measured the outcomes of participants and non-participants in the C2P2 Partnership Program.
- *Community Based Environmental Protection (CBEP)* (<http://www.epa.gov/evaluate/cbep1999.pdf>): In this process evaluation, the program sought to identify the factors that contributed to the success or failure of EPA-led CBEP projects.
- *Evaluating the Hospitals for a Healthy Environment (H2E) Program's Partner Hospitals' Environmental Improvements* (http://intranet.epa.gov/evaluate/capacity_building/opptsfinal.pdf): This outcome evaluation determined the level of success that the H2E program has reached in achieving its program goals.
- *Measuring the Effectiveness of EPA's Indoor Air Quality Tools for Schools (IAQ TFS) Program Appendix* (<http://intranet.epa.gov/evaluate/pdfs/IAQ%20TfS%20FINAL%20REPORT.pdf>): This evaluation, with process, outcome, and impact elements, enabled the IAQ TFS Program estimate its impacts through field data, help define better measures of program outcomes, and provide insight(s) into the effectiveness of the overall approach in helping to meet EPA's clean air goals.
- *National Environmental Performance Track - Evaluating New England Performance Track Facility Members' Environmental Performance and Impact on New England's Environment* (http://intranet.epa.gov/evaluate/capacity_building/r1pt03.pdf): This evaluation, containing design evaluation and outcome evaluation elements, assessed the extent to which Performance Track in New England is operating according to its program theory and stated outcome goals.
- *Results Evaluation of the RCC (Resource Conservation Challenge) Schools Chemical Cleanout Campaign* (http://intranet.epa.gov/evaluate/capacity_building/sc3result.pdf): This outcome evaluation helped identify successful projects and provide valuable information to define how best to work with schools to ensure a healthy and safe school environment.

ESD Program Evaluation Resources

- What Is Program Evaluation and Performance Measurement? (<http://intranet.epa.gov/evaluate/overview/whatis.htm>)
- ESD resources and tools (<http://intranet.epa.gov/evaluate/resources/tools.htm>): These tools will help you throughout the program evaluation process from the planning stage to the communication of evaluation results. Of these tools, the following will be particularly helpful for the users of this guide:
 - o Worksheets for Planning, Conducting, and Managing an Evaluation
 - o Evaluation and Research Designs (describes a variety of non-experiment, quasi-experimental, and true experimental designs that can be used in program evaluations)
 - o Report Formatting and Presentation Guidelines
- Evaluation glossary (www.epa.gov/evaluate/glossary.htm)
- ESD training materials (www.intranet.epa.gov/evaluate/training/index.htm): The training slides present a detailed and interactive guide to evaluation concepts.

Other Online Evaluation Resources

- Logic Modeling:
 - o Clegg Logic Model Game (<http://cleggasociates.com/html/modules.php?name=Content&pa=showpage&pid=38&cid=3>): Interactive game designed to teach the concepts of logic modeling
 - o University of Wisconsin Extension (www.uwex.edu/ces/pdande/progdev/index.html)

- Program Evaluation:
 - o W.K. Kellogg Foundation's Evaluation Toolkit (www.wkkf.org/default.aspx?tabid=75&CID=281&NID=61&LanguageID=0): Contains resources on developing evaluation questions, plans, budgeting for evaluation, managing a contractor, and checklists. Includes the Evaluation Handbook and Logic Model Development Guide.
 - o U.S. Government Accountability Office (www.gao.gov/policy/guidance.htm): Policy and guidance materials on evaluations, evaluation design, case study evaluation, and prospective evaluation methods.
 - o The Evaluation Center at Western Michigan University (www.wmich.edu/evalctr/): Excellent resource for evaluation checklists, instructional materials, publications, and reports.
 - o Online Evaluation Resource Library (<http://oerl.sri.com/>): Contains evaluation instruments, plans, reports, and instructional materials on project evaluation design and methods of collecting data.
 - o Collaborative & Empowerment Evaluation Web site (<http://homepage.mac.com/profdavidf/empowermentevaluation.htm>)
 - o Centers for Disease Control and Prevention Evaluation Resources (www.cdc.gov/healthyyouth/evaluation/resources.htm)
 - o Web Center for Social Research Methods (www.socialresearchmethods.net/): Site provides resources and links to other locations on the Web that deal in applied program evaluation methods, including an online hypertext textbook on applied methods, an online statistical advisor,

and a collection of manual and computer simulation exercises of common evaluation designs for evaluators to learn how to do simple simulations.

Helpful Program Evaluation Publications:

- Logic Modeling
 - o *Logic Model Workbook* (http://www.innonet.org/index.php?section_id=64&content_id=185): Innovation Network Inc. 2005.
 - o *Guide for Developing and Using a Logic Model* (www.cdc.gov/dhdsp/CDCynergy_training/Content/activeinformation/resources/Evaluation_Guide-Developing_and_Using_a_Logic_Model.pdf): Centers for Disease Control and Prevention
- Program Evaluation:
 - o *Program Evaluation & Performance Measurement: An Introduction to Practice*. McDavid, J. and Hawthorn, L. 2006. Thousand Oaks, CA: SAGE Publications.
 - o *Handbook of Practical Program Evaluation*. Woley, J., Hatry P., and Newcomer, K. 1994. San Francisco: Jossey-Bass Publishers.
 - o *The Manager's Guide to Program Evaluation: Planning, Contracting, and Managing for Useful Results*. Mattessich, P. 2003. Saint Paul, MN: Wilder Publishing Center.
 - o *Real World Evaluation: Working Under Budget, Time, Data, and Political Constraints*. Bamberger, M., Rugh, J. and Mabry, L. 2006. Thousand Oaks, CA: Sage Publications.
 - o *Utilization-Focused Evaluation: The New Century Text*. 3rd ed. Patton, M. 1997. Thousand Oaks, CA: Sage Publications.

Useful Tools:

- OPEI's Program Evaluation Competition (http://intranet.epa.gov/evaluate/capacity_building/competition.htm): Provides a source of financial and technical support open to all headquarters and regional offices.
 - Information Collection Request Center (www.epa.gov/opperid1): An EPA-wide site that provides a basic guide to the ICR process.
 - SurveyMonkey (www.surveymonkey.com): Free online survey package.
 - Survey Suite (http://intercom.virginia.edu/cgi-bin/cgiwrap/intercom/SurveySuite/ss_intex.pl): An internet tool to help design surveys.
- Outside Evaluation Opportunities:
- The Evaluators' Institute (www.evaluatorsinstitute.com): Offers short-term professional development courses for practitioners.
 - American Evaluation Association (<http://eval.org>): Professional society for evaluators with links to evaluation Web sites.

Appendix C: Case Study

Hospitals for a Healthy Environment (H2E) is an EPA Partnership Program launched in 1998 with the goal of advancing waste reduction and pollution prevention efforts in the nation's hospitals. Specifically, H2E directed its efforts towards 1) virtually eliminating mercury-containing waste, 2) reducing the overall volume of regulated and non-regulated waste, and 3) identifying hazardous substances for pollution prevention and waste reduction opportunities by providing a variety of tools and resources to its partners.

In 2004, H2E was spurred to undertake a program evaluation because of an upcoming PART assessment. Program managers and staff realized that the questions included in the PART assessment were not sufficient, however, to answer questions about H2E's internal processes, customer satisfaction, the varying roles of their diverse partners, however, or the identification of potential program improvements that were most needed by the program. Managers and staff understood that a program evaluation was the appropriate performance management tool to provide them with the information that they needed to make important decisions about the program's future; they decided that an impact evaluation would provide the most benefit.

H2E realized early on that the resources and expertise needed to conduct an impact evaluation exceeded the program's internal capacity, so the staff submitted a proposal to the Office of Policy, Economics, and Innovation's (OPEI's) annual Program Evaluation Competition to access additional funding and program evaluation expertise. The competition provided H2E with partial funding, a contractor with evaluation ex-

pertise, and an EPA staff person with evaluation expertise to manage the contract. The contractor advised H2E that an impact evaluation might not be the best choice for the program because in order to make causal claims, the study would need to control for a wide variety of factors that influence hospitals' green behavior, and the data available were not of adequate quality to do so. After consulting with the contractor and stakeholders, H2E decided to focus on measuring short-term and intermediate outcomes and customer satisfaction, which would provide useful information to the program and could be achieved with the data available and within a reasonable budget.

When H2E began the evaluation process, the program looked to involve stakeholders that would represent the diversity of its stakeholders. The evaluation team identified program managers, team leaders, program staff, and partners as the key stakeholders they needed to consult with at key stages in the evaluation process (such as logic model development, finalization of evaluation questions and evaluation design, and the development of conclusions and recommendations).

Additionally, a core evaluation team was involved in the day-to-day management of the evaluation. This team included the program manager, the internal evaluation expert provided to them through the competition, and the contractor. This team worked to ensure that the evaluation was carried out with methodological soundness and with intelligent program insight so that it would provide the program with the most useful results possible. On the team, the contractor served as the "skeptic," asking those closely involved with the program to think

critically about their assumptions. The collaborative nature of the evaluation and diversity of stakeholders involved allowed H2E to address a broader set of questions critical to program improvement than the program originally intended and, in the opinion of program staff, served to strengthen the ultimate quality of the evaluation and maximized the return on the resources expended during the process.

Because logic models were not in wide use at the Agency until the mid-2000s, H2E did not have a logic model of the program when it decided to conduct a program evaluation. At the time of the evaluation, H2E had been in existence for seven years, and revisiting the goals expressed in its original charter and reflecting on if and why those goals had changed provided valuable insight. H2E began its logic model as soon as the program was selected for funding through the Program Evaluation Competition. Managers and staff found the process of developing the logic model to be very useful in its own right, as it allowed the program and its stakeholders to reflect on how each group conceptualized the program's goals, activities, outputs, and customers. Once they had access to expertise, they were able to finalize a logic model that clarified their expectations for the evaluation and helped to build consensus among stakeholders about which questions were of highest priority. Participating in the logic modeling process was also beneficial for the evaluation experts who were working on the evaluation as a means to familiarize themselves with H2E.

After developing the logic model, H2E decided to answer four evaluation questions that can be traced along the logic model: 1) What types of environmental activities are H2E partner

hospitals engaged in? 2) How can H2E be improved in terms of the services it offers? 3) How satisfied are H2E partners with the key elements of the program? and 4) What measurable environmental outcomes can H2E partner hospitals show? When deciding what questions to answer, practical constraints—especially data availability and quality—were balanced against programmatic priorities. H2E used the logic modeling process to help make these decisions about tradeoffs. By developing a set of carefully focused evaluation questions, the program felt it had enhanced the manageability of conducting a program evaluation. The question of environmental outcomes was the central focus of the evaluation; however, the other three questions supported this question by illuminating the logical links between program activities and outcomes.

After developing the evaluation questions, H2E combined the evaluation expertise of the contractor and the program staff to identify the best evaluation design. A collaborative approach to designing the evaluation, guided by its contractor and the EPA evaluation advisor, led H2E to a design that would compare participants with non-participants on self-reported waste behavior (a quasi-experimental design). The evaluation used surveys to collect primary and secondary data that yielded both qualitative and quantitative data.

To collect these data, the program used 1) a survey of hospitals, administered by the American Hospital Association, involving a sample of partner and non-partner hospitals, 2) data from the H2E Facility Assessment and Goal Summary report forms submitted by partners to EPA, and 3) a customer satisfaction survey of the

program, administered by EPA. H2E was able to avoid the ICR process by accessing a generic customer service ICR that had already been approved and using data collected by an outside entity. Although the expert evaluator designed the evaluation to minimize some of the limitations associated with surveys, including self-selection bias, these factors did influence how the program qualified its findings in its final report.

H2E staff and the evaluation team were very active in the early stages of the evaluation; however, they took a more hands-off approach at the implementation stage. Their primary role during implementation involved establishing contact between the contractor and the partner hospitals that would provide data for the evaluation. This role was instrumental in providing the necessary data to the contractor so that that data could be analyzed.

Because the relationships that form the core of H2E are voluntary, data collection proved difficult, as the burden placed on partners had to remain reasonable. Although H2E served as facilitator and “data police,” the contractor conducted the data analysis, and H2E assumed a less involved role, limited to monthly check-ins with the contractor. During these check-ins, the contractor would ask program staff for their reaction to preliminary results and to clarify any anomalies that appeared. During this stage, H2E also considered how it could facilitate future evaluation efforts by developing innovative and efficient ways to collect and store data.

H2E organized internal briefings so that the contractor could begin communicating the evaluation results, including the data analysis process and initial findings of the evaluation. Stakehold-

ers then worked with the contractor to draw conclusions from these findings. Based on these conclusions, the team developed a series of recommendations.

The principal audience for the evaluation was internal, and the contractor tailored the final communication of the evaluation results to meet the needs of this audience. The evaluation process concluded with a technical report that outlined the results of the evaluation and presented some of the limitations in terms of data and measurement that H2E faced. Summary tables organized around each evaluation question helped with interpretation. By presenting a detailed description of methodology and limitations, the report presented a credible response to H2E’s initial questions and earned partial credit on the evaluation questions included in the PART assessment that followed.

At the end of the evaluation process, H2E managers and staff were pleased with their experience. In addition to programmatic recommendations outlined in the report, team members identified several management improvements they could undertake to ready themselves for more complex evaluations in the future, such as enhancing recordkeeping, identifying baseline data, identifying new sources of measurement; and developing ways to control for other factors that influence the behavior of H2E partners. In 2006, H2E became an independent nonprofit organization and expanded its waste reduction goals. The final evaluation report is published on the Evaluation Support Division Web site (www.intranet.epa.gov/evaluate).



United States
Environmental Protection Agency
National Center for Environmental Innovation
(1807T)
Washington, D.C.
March 2009