

Practical Evaluation for Public Managers

Getting The
Information
You Need



Department of Health & Human Services
Office of Inspector General

FOREWORD

Trying to make a decision without information is like shooting at a target blindfolded. You probably won't hit the bull's eye. You may not hit the target at all. And there is a chance you may injure an innocent bystander. As the Inspector General for the Department of Health and Human Services, I get to see first-hand many of the effects and side-effects of off-target decision-making.

It is my job to protect HHS programs against fraud, waste, and abuse, and to promote program effectiveness and efficiency. While I have been given a special trust to protect and improve these programs on behalf of the public, this is really the trust given to all government employees. Without the cooperation of the employees of the programs we oversee, my office would accomplish very little. Further, no matter how well my staff and I, or any other oversight agency, do our jobs, most of the effectiveness of government programs rests day-to-day with program managers and their staffs, and is based on their decisions.

Before coming to HHS, I served as Inspector General for the Department of the Interior, for NASA, for the Department of Defense, and for the Navy's Pacific Fleet at Pearl Harbor, Hawaii. So I know the inspector general job pretty well. Earlier, I held management positions at the Navy Finance Center in Cleveland and at Interior. If there is one thing I have learned in these public service capacities, it's that *there is nothing more powerful than meaningful information in the hands of someone who can use it to make government work better*. It's always been one of my foremost goals as Inspector General to give decision-makers solid, timely, and useful information.

While most Inspectors General would share my philosophy, many rely on their auditors and investigators for program information. However, my office also depends on a small, distinct group of analysts for a great deal of the information we collect on HHS programs. The Office of Evaluation and Inspections (OEI) has adapted techniques from evaluation, policy analysis, survey

research, and other disciplines to develop a practical system for providing rapid feedback to decision-makers on program and policy issues. We have found this system works well. In fact, both Vice President Gore, as noted in the report on the National Performance Review, and Paul Light, as shown in his book Monitoring Government, see an evaluation role in the future of federal IGs.

Four individuals, Mike Mangano, my principal deputy, Penny Thompson, the head of OEI's Health Care Branch, and Jack Molnar and Brenda Stup, OEI senior analysts, have written this volume, which describes our evaluation system and their experiences. The authors make a strong case for program managers themselves to go beyond routine data collection to analyze and evaluate their programs, and the book presents low-cost, accessible strategies for accomplishing this.

Many thanks to Mike, Penny, Jack, and Brenda for keeping us "on target." I see this book as a valuable contribution to the ongoing endeavor to make our public programs more efficient, more effective, and more responsive.

June Gibbs Brown
 Inspector General
 November 1994

Table of Contents

SETTING THE STAGE
 Page i

CHAPTER 1
 Making the Commitment 1

CHAPTER 2
 What's It All About? 7

CHAPTER 3
 Client Satisfaction Survey? 21

CHAPTER 4
 Performance Indicator Studies 33

CHAPTER 5
 Compliance Reviews 43

CHAPTER 6
 Effective Practice Reviews 51

CHAPTER 7
 Early Implementation Reviews 59

CHAPTER 8
 Emerging Issue Reviews 67

CHAPTER 9
 What's Next? 75

SELECTED BIBLIOGRAPHY
 Page 85

SETTING THE STAGE

How many times have you discovered you were missing a critical piece of information and...

- ▶ Couldn't figure out why service-delivery was inefficient?
- ▶ Couldn't understand why clients seemed dissatisfied?
- ▶ Couldn't answer an inquiry about your program's accomplishments or weak spots?
- ▶ Couldn't make a strong case for additional budget dollars?
- ▶ Couldn't assess the extent of program fraud or abuse?

Couldn't be sure you made the right decision about which action would make the most significant program improvement?

You may have missed an opportunity because you lacked timely and reliable information. Armed with the right information, you could have made more informed decisions, eliminated a bottleneck, understood your clients' problems, documented your program's strengths and weaknesses, effectively argued for resources, protected the program from profiteers, or known which changes would produce the biggest payoff.

If you are a program manager or analyst, you are probably inundated with data and statistics. You may not even recognize an information deficit. Yet you may often find you don't have data-based answers to policy and operational questions when you need them, despite the huge investment being made in data collection and reporting in public programs.

This book is about getting the information you need in a timely manner and at a relatively modest cost. By "information," we mean something more than the data typically produced by management information systems. Rather, we use information to mean a collection of facts and

logical conclusions which answer the types of questions we posed above. By learning and using a variety of strategies for obtaining information, you can better address specific problems, gain insight into what's happening in your program, and determine what directions you should be taking. These strategies can help you reduce complexity and uncertainty in your day-to-day decision-making and see beyond the many immediate problems you face.

Sharing Techniques and Approaches. This book describes the methods used by a group of professional analysts dedicated to putting practical program information in the hands of the people who need it in time to make a difference. The strategies we present in the next chapters are drawn from our experience as analysts in the Inspector General's office for the U.S. Department of Health and Human Services (HHS). A short description of our operation, mission, and philosophy might help explain why we are writing this book.

Our organization, the Office of Evaluation and Inspections, comprises about 100 analysts who come from a variety of educational and occupational backgrounds. Stationed across the country, we work in teams to conduct short turnaround evaluations of HHS programs. These generally take six to nine months from the time we settle on a project design to the time we present preliminary findings to our audience. We can move even faster when the need arises. Our studies usually fall into one or more of the following categories: client satisfaction surveys; performance indicator studies; compliance reviews; effective practice reviews; early implementation reviews; and emerging issue reviews. We have organized this book around these categories.

We tend to be very pragmatic in our outlook and approach. We search out the methods and data sources that will produce the best information in time to be useful. In fact, we measure our success in terms of program improvements and dollars saved or put to better use. We strive to anticipate critical issues. We define our studies using a relatively sharp focus. Our techniques allow us to reach logical conclusions rather than definitive answers. We believe timing is the key

to our success and that, no matter how accurate, information provided too late is of little value.

Because of this, some people might call our work "quick and dirty" or question the accuracy of calling what we do evaluation at all. Indeed, you may tend to associate the field of evaluation with drug testing, for example, more readily than with program management. The most scientific evaluations involve control and test groups to assess the effects of a particular intervention (e.g., a drug). In its purest form, such an evaluation involves randomly assigning individuals to either the control (no treatment or intervention) or the test group, so that any measurable differences in the groups over time can be attributed to the one difference between the groups: the presence or absence of the treatment or intervention. On the strength of evidence amassed in such "pure" evaluations, our society makes critical decisions, not only about which medical treatments work best, but about how individuals respond to all sorts of interventions.

Yet, as part of an agency charged with overseeing some of the largest federal programs in existence, we often must assess and evaluate programs or other interventions without the luxury of conducting a scientific experiment. To do this, we use methods consistent with principles of evaluation but quite different from the exacting work of scientists who supply us with the kind of definitive information described above. We see our approach as far more attainable and our goals as very much like those of most program managers. We examine and report on issues of efficiency, fraud and abuse, and program effectiveness. Similar responsibilities--to use public dollars wisely, to maintain public confidence, and to achieve intended results--fall on program managers. Like many program managers, we move quickly from one issue to the next, as we anticipate and respond to the interests and needs of decision-makers in the executive and legislative branches of government.

We are interested in making our commonsense approach to evaluation--*providing outcome-oriented, practical information for immediate use*--available to program managers and others who may not consider themselves evaluators. We realize that establishing a distinct evaluation

component like ours may be an unreachable or even inappropriate goal for many programs. It requires the help of experienced evaluators, a long-term commitment, and more than a little risk-taking. It's hard work. However, with an understanding of the way evaluation techniques can apply to program management, non-evaluators can effectively use the services of evaluators to obtain needed information. Similar to any relationship based on a contract for services, the more you know about the concepts and applications of evaluation, the stronger your partnership with professional evaluators will be, and the more likely you will be to end up with information that meets your needs. With this objective, we will share our strategies with you in the chapters that follow.

Since our goal is to make evaluation more accessible, we've limited our use of professional jargon. We've avoided technical terms that may be unfamiliar to non-evaluators. We haven't filled our discussion with theory or references, although we offer suggestions for further reading if you are interested in learning more. We concentrate on the techniques with which we've had the most success over a number of years. We hope this information will start you thinking about how to use evaluation to improve the program you manage or to expand your repertoire as an analyst.

We certainly don't claim to have all the answers or to have originated a new evaluation system. We've had enough successes as well as failures to know what works and what doesn't. We've borrowed extensively from the fields of journalism, auditing, investigation, and management analysis, as well as evaluation, in developing our techniques. We will describe our strategies, along with examples of how we have used them, and explain why they work.

We know that many of you have long been convinced of the value of producing timely, relevant information to monitor the effects of your program and that you have developed your own techniques to do this. For those of you who fall into this category, we hope to offer some new ideas for collecting and analyzing program information. Others of you may be wary of committing any of your dwindling resources to evaluation efforts. We hope to convince you that evaluation is critical

to your program's success, and possibly its very survival, and to give you some simple ways to get started.

Organization of the Book. Our discussion starts with some compelling reasons for conducting program evaluations, then examines a variety of effective evaluation techniques, with examples of their use, and ends with a review of some of the practical considerations involved in developing an evaluation capacity. The first chapter answers the question: Why should program managers conduct evaluations? It describes the legislative, private-sector, and public forces pushing managers in this direction. Chapter 2 discusses what's involved in program evaluation. Chapters 3 through 8 cover six types of information-gathering strategies you may find useful. In Chapter 9, we talk about what to do after you've made the decision to add program evaluation to your management system.

About the Authors. Michael Mangano is Principal Deputy Inspector General for the U.S. Department of Health and Human Services. Penny Thompson is Chief of the Health Care Branch for the Office of Evaluation and Inspections, where she devotes her attention to evaluations of the Medicare and Medicaid programs. Jack Molnar and Brenda Stup are senior analysts in the Office of Evaluation and Inspections and are both currently involved in evaluations involving Social Security issues. All four have published articles on program evaluation and related topics in professional journals and books. Together, the authors have over 50 years of hands-on experience in conducting studies of over 250 programs within HHS.

CHAPTER 1

MAKING THE COMMITMENT

The American public is demanding better government. As the private sector devotes more attention to meeting customers' needs, people want and expect the same responsiveness from government. We look to the federal government to provide for a common defense, to collect taxes, to assist the needy, to regulate health care, to monitor air quality, to operate parks, to keep defective products off the market,...the list goes on and on. However, the staffing levels and budgets that previously sustained these programs and services are shrinking.

Is it really possible to "do more with less," and "work smarter, not harder"? The evidence from the private sector seems to say it is. Over the last decade or more, innovative managers in a number of industries have shown us a virtual quality explosion coupled with significant cost reductions. They did it by investing in employee training and modern equipment, reducing the ranks of middle management and empowering all staff, redesigning products to stress quality and "delight" customers, rethinking basic processes, and focusing on the long term. This is how the U. S. automotive industry, particularly Ford and Chrysler, turned itself around and became competitive again.

People have begun to realize that if such dramatic improvement can be achieved in the private sector, it can be accomplished in the public sector as well. We are seeing the results of this awareness in several initiatives to improve accountability in the management of federal programs.

Reporting Out the Results. In November 1990, the Congress passed and the President signed the Chief Financial Officers Act. One of the objectives of the Act is to "provide for the production of complete, reliable, timely and consistent financial information for use by the executive branch of the government and the Congress in the financing, management and evaluation of federal programs." The Act also requires agencies to collect information needed for "the systematic measurement of performance" and to report the information in their financial statements. Managers of federal agencies covered by the Act

are required to establish indicators for assessing their programs and to report the results of their assessments to the Congress. The measurements can include inputs, outputs, outcomes, and impacts. Managers will now be held publicly accountable for program accomplishments.

The Government Performance and Results Act of 1993 further strengthens federal government accountability by mandating quantifiable program results. In passing this law, Congress set out to "improve the confidence of the American people in the capability of the federal government, by systematically holding federal agencies accountable for achieving program results." The law requires each agency to "establish performance indicators to be used in measuring or assessing the relevant outputs, service levels, and outcomes of each program activity."

Managing for Quality. These new requirements for accountability to the Congress, and, more importantly, to the taxpayer, demand new information-gathering strategies for managing federal programs. As we noted in the introduction, current management information systems often provide a profusion of data without answering basic questions about program effectiveness. A change in philosophy within government highlights this "information deficit" perhaps more than any new statute. Learning the lessons of the private sector, many public managers have adopted Total Quality Management (TQM) as a guide to improving performance. While there are differences among approaches to TQM, all have in common some basic tenets: continuous improvement, client focus, employee empowerment, investment in training, strategic planning, and quality measurement.

TQM requires good information. To continuously improve your program, you have to know which processes and systems need improving and where the biggest impact on your operations can be gained by new strategies and ways of doing business. You will need to evaluate the effects of the changes you make, to determine when new problems crop up, and to deal with unforeseen obstacles that prevent you from achieving the good results you intend. You will need to set performance standards appropriate to the fundamental purpose of your

organization. By showing progress toward those standards, you can convince others of the value of your improvements. Documented progress will help garner support from your staff and, eventually, the public. It will enable you to build on your successful initiatives. Further, it will assist the executive and legislative branches in resource allocation, to your benefit.

However, good measurement systems need "care and feeding"--ongoing analysis and refinement. This requires some investment in establishing and maintaining those systems. TQM teaches us to consult regularly with our customers, in our case, the public, the Congress, and other organizations impacted by our products, in instituting performance standards and measurement systems.

Reinventing Government. These new philosophies about government's role and about how government should respond to changing expectations are described in Osborne and Gaebler's book *Reinventing Government*. By challenging many of the traditional roles and methods of government, Osborne and Gaebler demand that public managers reassess what they do. One of the most notable responses to this challenge has been Vice President Gore's 1993 National Performance Review (NPR). Modeled after the Texas Performance Review, the NPR entailed a major assessment of all federal programs. On a smaller scale, such efforts now are taking place at all levels of government, and not as one-time reviews. We are entering a period where programs need to "reinvent" themselves as times and challenges change. You as a program manager will need to evaluate your agency's performance and to assess the external environment on an ongoing basis to ensure your success.

Do it First or Others Will Do it for You. The forces of change are evident. New laws call for the reform and use of financial and other performance information systems to better plan and assess program effectiveness. Management philosophies like TQM have gained legitimacy and achieved success by focusing on quality, clients, and performance measurement. The Vice President personally, through the NPR, has led an effort to assess every federal program with an eye

toward giving citizens their money's worth, and this movement is evident at all levels of government.

The direction is clear. To be successful, government managers will have to develop the skills and tools to evaluate program performance. Your ability to respond to public pressure for effective and efficient service delivery will depend on your ability to get the information you need when you need it. You can plan an active evaluation program to meet your own information needs and set your own performance standards...or you can wait for others to tell you what to do.

First Choices. Evaluation, as we defined it earlier, means obtaining information for use in decision-making. So the first step in deciding what kinds of evaluations you should undertake is to decide what your questions are. Some examples of key questions that may not be easily answered from data you now collect are: What are the unintended harmful effects of your program or bottlenecks impeding efficient service? How do beneficiaries, interest groups, or your own workers view the program? How vulnerable is your program to criminal abuse?

The next chapter contains basic information about how to evaluate these types of issues. The chapters that follow describe a number of specific types of evaluation techniques we employ that you may want to use to help answer your key questions.

- ◆ Client satisfaction surveys determine how the beneficiaries of your program or others affected by it feel about it. This technique can be used to monitor performance over time and to help pinpoint problems, successes, and areas for improvement.
- ◆ Performance indicator studies help you forge ahead in the movement to develop performance indicators and measures by identifying and testing these yardsticks for program success.
- ◆ Compliance reviews assess how well your program is doing at meeting its legal or regulatory requirements.

- ◆ Effective practice reviews identify and describe methods that work better than others and why.
- ◆ Early implementation reviews provide an immediate reading on how well a new program is operating. They emphasize identifying problem areas so you can take corrective action quickly.
- ◆ Emerging issue reviews gather information on emerging problems or opportunities that require your attention.

The last chapter deals with other early decisions concerning how you might go about building an evaluation capacity. The possibilities include contracting the work out, either by individual study or on a long-term basis; hiring professional evaluators to become part of your staff; or developing the evaluation skills of your current employees.

CHAPTER 2

WHAT'S IT ALL ABOUT?

Evaluation is a process for getting information. Like most processes, an evaluation has definable phases and technical components. Fortunately, evaluation has a lot to do with common sense. It's useful to think of an evaluation as having four distinct phases:

- ◆ *Design.* You'll need a plan of attack. What is the purpose of your evaluation? What are the specific questions you hope to answer? What information do you need and how will you collect it? How will you use this information to accomplish your original purpose?
- ◆ *Data Collection.* You'll need to collect the right information and make sure it's accurate. What information is available and how do you go about getting it? Who will be responsible for interviewing key people, reviewing policies and procedures, and obtaining data from records? How do you make sure the information you get is reliable?
- ◆ *Data Analysis.* Once you have all the information, you'll need to make good use of it. How does it answer your study questions? How do you learn all you can from it? How do you separate the wheat from the chaff? This is the stage in which you distill key findings from your data and develop recommendations for action.
- ◆ *Presentation.* After you develop findings and recommendations, you'll probably want to share your results with others, perhaps a variety of people with different backgrounds and interests. How do you present the information you've collected in a compelling way that makes sense to all your audiences?

Understanding these phases, and some technical concepts that are important for each one, will give you a frame of reference for our later discussion of evaluation strategies.

Designing the Evaluation. We strongly believe a good design is the most important element of a successful evaluation. However, since this phase is less visible than others, it may be the one where you are most tempted to cut corners. PLEASE DON'T! A comprehensive design process ensures that the evaluation is targeted to critical issues (and that your report will be used, not just added to a pile on someone's desk). A well-crafted design lays out a coherent strategy for the implementation phases--collecting, analyzing, and presenting the data. It helps everyone working on the project stay focused and "on task." Consequently, we urge you to spend the time and resources to do it well, and we will devote some effort here to helping you do this.

Purpose. The first and maybe the hardest part of the design phase is formulating the purpose. It might seem like a simple and straightforward task to write a purpose statement and a list of questions you want the study to answer. However, to truly focus an evaluation, the purpose statement should be *complete, concise, and concrete*. A complete yet concise statement sets out exactly what you intend to do--no more and no less. "Concrete" means your goal is real and measurable. Avoid vague purposes that begin "To review," "To explore," or "To study." Instead, the statement should call for a specific action, such as to determine, to identify, or to measure, for example, "To describe the enforcement of State laws prohibiting the sale of cigarettes to minors."

As you develop the purpose statement, be sure to identify and respond to the needs of the study's sponsor, whether that is you, the program manager; a Congressional committee; your boss; or the budget office. Ask yourself if fulfilling the purpose as you've stated it will meet those needs. If not, go back to the drawing board.

Issues. The purpose statement should be supported by issues. These are the key questions that are directly relevant to your purpose and which your evaluation will attempt to answer. We cannot

overemphasize the need for "doing your homework" as you develop your study issues. Learn the history of the program and review previous studies related to your purpose. Talk with knowledgeable people. Developing the issues is generally a very unstructured, informal, and iterative process, and the insights to be gained are absolutely critical to a successful evaluation.

While the issue questions need to address the purpose of the evaluation fully, their connection should be solid. A tightly focused study represents the best use of your resources. If you find your list of key issues expanding beyond three or four, it may be a good idea to revisit the purpose statement to see if the scope of your review is too broad or if you're getting sidetracked. Each issue you choose to examine may have several sub-issues associated with it, but you should be able to explain the focus of your review to your sponsor or audience in a few minutes. Limiting the number of issues forces you to concentrate on those that are the most important.

Your sub-issues, or study questions, should point directly to the data you will need to collect. As you develop the sub-issues, avoid the tendency to pose questions that would be easy to answer but fail to yield useful information. One way to do this is to ask yourself, "What action could I take with this piece of information?" Questions that are too vague can mean different things to different people. Our advice here is to be specific. For example, you may have identified an issue such as, "What do our clients think of our new application form?" Go on to questions like "What do our clients like or dislike about our new application form?" "Do our clients find the instructions on the new application form helpful or confusing?" "Is the new form easier or harder to understand than the old form?" Specific questions will be more likely to give you information you can act on to improve your program. Well developed issues and sub-issues provide a solid foundation for the entire study.

Methodology. The methodology part of the design describes the data sources to be used and how you will select and access them. Many data sources and methodological approaches are possible. We'd

like to limit this part of our discussion to those that you may find helpful in conducting the kinds of evaluations we present later on.

Data sources include individuals (e.g., program clients, program officials, or interest group officials), case files, policy and procedure manuals, or even the evidence of your own eyes. In our work, we often use several data sources for a single evaluation because we recognize that each has its own bias. That is, each data source may be undependable in some way. Respondents may not be objective. Case files may be missing important pieces of information. Procedure manuals may tell us how things are supposed to work, but not how they actually work. Even our own eyes may deceive us if we fail to recognize or understand what we see.

Of course, the most important factor in deciding which data sources to use is how well they will answer your questions. Some other practical considerations include: how easy it will be to access the data; how consistent, complete and reliable the data are; and how easy the data will be to analyze. A final consideration, and an extremely important one, is whether the information you get from these data sources will be convincing. It's often helpful to discuss your decisions about methodology with key audiences before proceeding with the evaluation. Even statistically valid data may not convince people if that isn't what they believe is needed to prove a point. This doesn't mean you have to give them exactly what they want, no matter what. But you should be sensitive to which arguments and data will convince the skeptical audience, regardless of the outcome. Find out whether the sources and levels of evidence you've chosen will convince them to accept your results.

Once you've decided which data sources to use, the next question is how to select and access them. If you're going to interview clients, which clients (of maybe millions) should you interview? If you're going to look at case files, which ones should you examine? The answer is, "It depends." It depends on your purpose, your audience, your resources and your capabilities.

Sometimes it is important to find and examine all instances of a phenomenon. A good illustration of this is a review we did of U. S. hospitals that closed. The Secretary of HHS asked us to examine every closure of certain hospitals over the prior year to determine whether Medicare payment rates were responsible. We looked at all 69 of these hospitals that closed in 1987 and found that few Medicare patients had been using them, so Medicare payments had had little effect. We didn't have to convince anyone that our findings applied to all the closed hospitals, because we had original data on every one of them (OEI, 1992c).

If resource constraints had made it impossible to access all the closed hospitals, we could have selected a sample of them to examine. When this is your situation, normally you'll want to use a sample that is representative of the larger group, so you can talk not only about the cases you examined, but also about the cases you *didn't* examine. How can you do this and be confident in your results?

The most straightforward technique is a simple random sample. This method assures that each and every case has an equal chance of being selected for your sample. If you select a sufficient number of cases using a random sample, you can confidently project your results to the entire group of cases. Of course, any sample may be biased in some way. However, a random sample maximizes your chances of selecting cases that represent the group. This method also lets you calculate the likelihood that you might be wrong and the extent of how wrong you could be. As a result, random sampling gives you and your audience confidence in your statements. This can help you convince others that the issue you identified is an important one. Moreover, it can be extremely valuable when your issue already is being hotly debated and you need to defend your findings from possible attack.

A common misconception is that the best studies use the largest samples. On the contrary, far better results may come from using a sound statistical technique to select a smaller but representative sample, and then using the savings to fine-tune your questionnaire or to achieve a higher response rate. In selecting a sample of any kind and performing calculations of confidence and precision, you'll need

statistical expertise, your own or from someone on your staff or an outside consultant. If not available in-house, your parent organization probably has a component that specializes in providing this kind of assistance. If you don't have a background in statistics, you might find it helpful to take an introductory course, such as one offered by the Office of Personnel Management, "Statistics for Management." This will give you an understanding of basic statistical concepts and will enrich your partnership with statistical consultants.

As you plan a particular evaluation, you may find that using a simple random sample would be too difficult or time consuming, or would be inappropriate. This could happen because cases are spread over too broad a geographic area or because the incidence of the factor you want to study is so low that the required sample size would be too large to manage. For such situations, which are not uncommon, there are other selection options. These include other forms of random sampling, such as staged or hierarchical sampling, and non-random sampling, known as purposeful or judgmental sampling. Another selection approach is the case study which uses one of the sampling techniques above to select cases in a confined area.

For example, we used a two-stage random sample to answer a congressional inquiry about physicians who referred Medicare patients to clinical laboratories in which the doctor had an ownership interest (OEI, 1989b). In the first stage, we randomly selected 10 areas of the country. In the second stage, we randomly selected Medicare beneficiaries *in those areas* who had received laboratory services. We sent questionnaires to the laboratories that had provided these services to determine if they were physician-owned; we also randomly selected physicians who billed Medicare and asked them if they had ownership in facilities to which they referred their patients. Thus we were able to estimate how many Medicare physicians owned facilities to which they referred their patients and how many facilities billed Medicare for laboratory services that were physician-owned. Because using a pure random sample would have involved data collection all over the country, we compromised by selecting a limited number of geographic locations and then selecting beneficiaries and physicians in those

locations. However, because we randomly chose the locations, we retained the ability to project our results nationally.

Purposeful sampling involves selecting sites or individuals with a specific intent, rather than leaving the selection to chance. Some form of random sample is preferable to a purposeful sample if your goal is to estimate or describe the characteristics of a larger group. On the other hand, you may not need to project the results to make a case for your findings if, for example, you selected the eight States with the highest rates for a particular event and this represents 90 percent of the events for the country. You may be satisfied to talk about the 90 percent. However, you do need to be careful not to assume or imply that your results are representative of the country as a whole.

Another selection technique to consider is the case study. While case study results cannot be projected to a larger group, this option provides for in-depth examination of issues, processes or outcomes at a particular site. Case studies attempt to describe the range and nature of the event under study in a confined environment and can provide interesting insights and pinpoint areas needing further study. They can answer questions of how much or how frequently with precision, but only for the site under review. When you've chosen a case study approach to examine a phenomenon or issue, be careful not to extrapolate your findings. Also be wary of a tendency of your audience to "project" your results for you, regardless of how scrupulous you are in reporting them.

Years ago we studied the practice of patient "dumping"--hospitals turning away or transferring emergency room patients simply because they aren't insured and can't pay the bill. We tried to determine how often this happens and failed miserably because the available hospital records couldn't tell us what we needed to know (OEI, 1988a). Around the same time, some researchers in Chicago did an in-depth case study of emergency room admissions to Cook County hospital and concluded that the problem was pervasive there. When they reported out their findings, they suggested that *if Cook County hospital were representative of all hospitals* (a very conditional statement), the number of patient dumping cases across the country could be as high as

250,000 a year. With no reliable estimate of the actual incidence nationwide, that number took on a life of its own. Even today we see it quoted as an "estimate" of how often patient dumping takes place.

It's important not to see random sampling as the "best" method and all other selection methods as "second best." Many times, you won't be interested in what is happening in the system as a whole or in generalizing your results, and a random sample would not be appropriate. For example, Tom Peters examined the cultures and methods of successful organizations for his book *In Pursuit of Excellence*. He didn't randomly select corporations. He purposefully selected organizations that seemed to exemplify excellence and then explored them in depth. When you're interested in the exception rather than the rule, or in understanding the dynamics of a changing environment, purposive sampling and non-random case studies can be your best choices.

Finishing the Design. Once you have decided on your methodology, other topics you need to cover in the design for your evaluation include an analysis plan to show how you will use the data to answer your issue questions, and the type of report or presentation you will produce as a result of your work. We will address these topics later in this chapter. To be most useful, your plan should also provide a schedule and a list of needed resources.

In our organization, all the staff involved in or responsible for the evaluation meet to review and discuss the design and to reach consensus on any needed revisions before proceeding with the data collection. We also share the design with our audience and any sponsor, to be sure it will meet their needs. We think these are good practices.

Data Collection. Having planned your evaluation, you're ready to do actual data collection.

Frequently, data collection is based on an instrument or guide that lists questions to be asked of respondents, data elements to be gleaned from records, or details to be noted through direct observation. It's

often the most labor-intensive part of the evaluation. Data collection instruments may be mailed to respondents to fill out or may be completed by you or your staff based on telephone interviews, in-person interviews, record reviews, etc.

Questions contained in data collection instruments can take many forms. A useful and common distinction is between open-ended and closed-ended questions. A closed-ended question asks the respondent to select from a list of options: yes or no; good, fair or poor; less than 15 minutes, 15 minutes to an hour, or more than an hour. An open-ended question lets the respondent create his or her own answer. The open-ended question "Why were you satisfied with the service you received?" could be answered with the response, "I find the people there courteous and caring. They helped my daughter when she had a problem with her check." While the open-ended response can give you richer data, it is also far more difficult to compile, categorize and analyze.

The types of questions, the setting in which they are asked, how they're worded, who asks them, and in what order, all affect the degree to which the resulting data is objective, valid, reliable and sufficient. We discuss these issues further in Chapter 3, "Client Satisfaction Surveys."

Finally, you must face the challenge of actually obtaining the answers to your questions from those who have them. A well-chosen sample of records could give you great information--but what if the owners of the records won't provide access to them? A group of clients could tell you a lot about their satisfaction with a program--but what if they don't answer your questions? One of the big challenges for evaluators is finding ways to increase response rates and minimize non-respondent bias. Non-respondent bias can occur when the people who don't respond would have given you information different from those who did. Not surprisingly, the higher the proportion of sampled individuals who answer your questions, the less the effect of those who don't. You also can compare known characteristics of the respondents and non-respondents to determine whether the two groups differ. The

results will help you judge the possible effect of non-respondent bias (Krushat & Molnar, 1993).

Data Analysis. The analysis phase is as cerebrally intensive as data collection is labor-intensive. Now you attempt to answer the questions, "So what? What does all this mean?" The goal of analysis is to link the data back to the original study questions and purpose. The results of your analysis will determine your findings--the answers to the issues you explored--and whether your findings have policy implications. Stay open, too, to the possibility of "serendipitous" results, which could point the way to future studies.

There are many ways to analyze data. For example, you can examine data in simple form, each variable alone (univariate analysis), or in increasingly complex forms in relation to other data (bivariate or multivariate analysis). Univariate analysis involves one variable, such as the number of clients served. How many clients were served in all? What was the average number of clients served in each program site? Frequency distributions, and other statistical concepts such as mean, mode, and median are useful here. Bivariate analysis involves relating one variable to another, such as number of clients served to staffing levels. Did the sites with more staff serve more clients? Statistical techniques such as chi-square or t-tests are useful here. Multivariate analysis involves examining relationships among a number of variables, such as the number of clients served, staffing levels, and types of outreach used to attract clients. Did staffing levels have more to do with the number of clients served than type of outreach used at each site? Regression analysis is helpful here. Many good references on these statistical analysis techniques are available and also will explain the important concept of *statistical significance*, the likelihood that differences you see are real and not simply due to chance.

Some data cannot be easily manipulated or understood using the statistical techniques of frequency distribution, chi-square, t-tests, or regression analysis. This type of data would be difficult to quantify; it isn't easily represented by a number for use in a mathematical equation. For instance, how do you analyze regulations? How do you assess procedure manuals? Probably not by statistical tests. Analysis of this

kind of data calls for comparisons against norms or rules and reviews for internal logic and consistency. You may have reviewed procedures, interviewed clients, and collected organizational material in a number of different program sites. You will need to sift through this information, much of it qualitative, to find common elements. Analysis of qualitative data is just as important as analysis of quantitative data. Although there is sometimes a bias toward "hard" (quantitative) data as more meaningful than "soft" (qualitative) data, experts in evaluation have come to agree that both types lend themselves to rigorous examination and analysis and can yield important information for policy-makers.

As you go about analyzing the various data you collected, follow a process similar to that for developing issues and sub-issues. Force yourself to distill your data and analysis down to a few key findings. The findings should directly address the purpose of your review and answer the questions you set out in the issues. The discipline imposed by the need to condense all the data will help keep you on track and provide a solid basis for your presentation and any recommendations.

Whether or not to develop recommendations depends on the purpose of the evaluation, the significance of the issues, the quality and significance of the findings, and your audience. As we mentioned earlier, your original purpose should be action-oriented; answers to your issue questions should naturally lead to ideas for actions that would improve program effectiveness. If you make recommendations, they should be feasible, supported by the findings (which are in turn supported by the data), and stated unambiguously. We have found that providing a list of options for achieving a recommended improvement can increase the likelihood that it will be implemented. Another critical, although sometimes subtle, consideration in developing recommendations is what we call the "political climate." It's a fact of life that some recommendations, no matter how well you support them, will not be accepted by those in authority due to factors beyond your control. Just be aware of these factors so that you can develop an alternative recommendation or recognize that your recommendation may not be implemented until the climate changes or until others have helped "tip the scale."

Presentation. Once the data have been analyzed and findings and recommendations developed, the study team must summarize the information in a presentation, whether a written report or series of reports, a set of charts, a briefing, a videotape, or some combination of techniques. Again, many helpful texts are available on how to present information, so we'll limit the time we spend on it here. However, based on our experience, we'd like to offer the following advice:

Before developing a presentation, the study team should prepare a one-page outline of the major findings and recommendations that address the original purpose of the review. The outline forms the basis for a presentation that is focused and well-organized. Don't try to write a research paper. You want to present a logical flow of information that clearly sets out and supports your significant findings and any recommendations. If a piece of information doesn't meet this objective, it doesn't belong in your presentation. (As noted before, this requires some discipline, since you will probably feel that *all* your findings are meaningful and important.)

Stakeholders--people with an interest in the evaluation's outcome--should have input on the final content of the findings and recommendations as well as the presentation. Ideally, you will have communicated frequently with these individuals during the course of the study, to minimize misunderstandings and surprises.

At least one product should be geared toward the senior-level official whose time is at a premium. This product should be crisp, clear, and concise. Stick to significant, action-oriented statements, presented logically, and supported by concrete, reliable information. Leave technical details about your methodology and analysis for presentations to other staff, or include them in an appendix. (We routinely prepare a 1- to 2-page "executive summary" to preface our reports.)

There is a great deal of literature available on the subjects of design, data collection and analysis, and presentation. We recommend that you read more to become familiar with the basic principles. Some of the books we use frequently are contained in the bibliography.

Additionally, OEI has developed Technical Assistance Guides for Conducting Program Evaluations and Inspections (OEI, 1990a). These guides outline the basic methods we use for conducting our program reviews, called inspections, and provide details of the key steps in moving through the evaluation process. They include three guides on design: *Focusing the Inspection*, *Targeting the Information Needed*, and *Specific Sources of Information*. These are followed by *Gathering Information* and *Analyzing the Information Gathered*. The last three are on presentation: *Specific Steps for Writing an Inspection Report*, *Presenting an Effective Inspection Briefing*, and *Using Graphics Effectively*. We use these guides not only to train new OEI staff and to refresh experienced staff, but also as a way to share our approach with other evaluators, program analysts, and policy analysts.

CHAPTER 3

CLIENT SATISFACTION SURVEYS

Traditionally, government managers have relied on their own expertise and operating reports to make decisions on how to run their programs. Program assessments have centered around inputs and outputs, based on criteria internal to the organization's operations. If you are a manager, more than likely you could tell us at the drop of a hat your processing time and accuracy for a particular workload, whether this is acceptable performance based on the size of your staff, and whether this meets regulatory standards. What you may not know is how important or acceptable this performance is to your clients. Total quality management (TQM) quickly teaches us to look to the client to define quality. In this new operating environment, program quality is often measured in terms of success in meeting clients' needs and expectations.

TQM also broadens our concept of "client." In the conventional view, the clients of a program are the end-users of its products or services, such as Social Security or Medicare beneficiaries. However, when we define clients as the people or organizations who affect or are affected by the program, as prescribed under TQM, we begin to identify many other kinds of clients. These may include:

- ◆ program staff;
- ◆ service providers;
- ◆ other government agencies; or
- ◆ private sector groups that are subject to program regulation or that rely on program data.

As you and other government managers go about "reinventing" your programs to meet clients' needs and expectations, you will have to develop client-based performance goals and indicators to assess your progress. The basic way to do this is to get input directly from your clients.

What Are They? Client satisfaction surveys use public opinion survey techniques to determine the nature and extent of clients' experiences with a program and to assess their satisfaction with the services they received. These surveys can determine the extent to which the entire population of clients is satisfied with service, the satisfaction levels of certain groups of clients, and what kinds of experiences or services are related to high or low rates of satisfaction.

We conduct two ongoing evaluations which assess beneficiary satisfaction. Not surprisingly, they concern our Department's two largest programs: Social Security and Medicare.

The Social Security client satisfaction survey was initiated in the mid-1980s at a time when the Social Security Administration began reducing staff and automating the way it served the public (OEI, 1994). The impetus for the survey was congressional and public concern that the quality of services to beneficiaries would decline as result of these changes.

In 1989, we began a periodic Medicare client survey to assess overall client satisfaction with the Medicare program and to determine how well Medicare contractors are serving beneficiaries (OEI, 1993A). (The Medicare program contracts with local insurance companies to process claims and provide program information. These contractors are beneficiaries' primary points of contact with Medicare.) We've also done more focused beneficiary surveys in response to concerns about the performance of certain Medicare contractors. For these evaluations, we compared the local findings to national findings to determine if any of the perceived problems were reflected in measures of service quality. In effect, the national results provided benchmark data against which we could evaluate the performance of an individual State or contractor.

We also have conducted a number of client satisfaction surveys which were not directed to program beneficiaries. These include surveys of program staff, service providers, other government agencies, and private sector groups:

- ◆ *Program Staff.* As Social Security began its program to automate claims processing, we conducted a survey of staff to discover what they knew about the new process, their views on the process and what was needed to succeed, and early operational problems. The results helped Social Security improve the phase-in process.
- ◆ *Service Providers.* When the Head Start program was mid-way through a major expansion of services, they asked us to survey local grantee agencies to find out how well the federal program staff were assisting them. The review identified problems with the quality and types of technical assistance they provided, responses to policy questions, and the availability of staff. By responding to each problem, the Head Start program improved the success of their expansion (OEI, 1991).
- ◆ *Other Government Agencies.* The Assistant Secretary for Planning and Evaluation at HHS asked us to survey the Department agencies and offices that use the data produced by the National Medical Expenditure Survey (NMES). Conducted every 10 years, the survey produces a vast array of information on health care costs, utilization, sources of payment, and insurance coverage. The Public Health Service, the Health Care Financing Administration, the Office of the Secretary, and others need the data to understand levels of specific health care spending so they can plan future policy in areas from primary to long term care. The review assessed agencies' views on the priority, timeliness, and usefulness of NMES data. The results were used to improve the next data collection process (OEI, 1993b).
- ◆ *Private Sector Groups.* We surveyed drug manufacturers concerning their experience in getting drug applications reviewed and approved by the Food and Drug Administration (FDA). We asked about the adequacy of FDA's guidance to manufacturers; their timeliness in reviewing applications; their communication with manufacturers; the capabilities of staff; and the uniformity and fairness of reviews. The findings showed substantial

problems at FDA that were all corrected through a comprehensive action plan (OEI, 1990g).

In fact, some form of client satisfaction survey is part of nearly every evaluation we do. For example, when we assess States' compliance with federal rules, we also ask State respondents about the adequacy of federal guidance and technical assistance. As the customer for this kind of information, it's important to know if States see the federal government as helping or hindering their efforts to comply with program rules.

How Are They Useful? The client satisfaction survey is an ideal means of validating your operating reports, but it can be more than a report card. It is an opportunity to learn clients' views on how well your program is meeting their needs, the criteria they use to gauge your program's effectiveness, and their ideas for ways to improve it. Client-based data can be seen as the ultimate barometer of quality (Molnar & Stup, 1994).

Surveys are one of the most effective ways of getting feedback from clients about their experiences with and expectations for your program. As stated earlier, client surveys are focused on outcomes, the effects of your program on its clients. Knowing, and being able to document, their opinions is a valuable asset in budget negotiations and in countering external criticism. For example, Congressional hearings often are based on complaints about a particular issue or program. While these negative indicators are useful, they only describe service delivery failures. A carefully designed client survey, on the other hand, can measure the full range of service; as such, it is more accurate and valid in describing performance than a haphazard collection of complaints. Several years ago, constituent complaints resulted in congressional hearings on problems with Social Security's 800-number service. We discovered from responses to the annual client survey that clients who called the 800 number were highly satisfied with service and in fact rated the quality of service higher than did callers to local offices.

Detailed data on clients' experiences with your organization, along with their opinions about those experiences, can provide you with guidance on how to increase satisfaction levels. You can ask clients what makes them satisfied. You can identify the types of clients or transactions associated with good or poor service. Further, using the more sophisticated statistical technique of regression analysis, you can determine what factors have the most influence on satisfaction.

For example, by analyzing responses to the Social Security client survey, we've been able to document that service delays--waiting lines, busy signals, and slow responses to requests--contribute to dissatisfaction. Not surprisingly, we've also found that competent, courteous staff contribute to higher satisfaction. Further, the survey has helped us pinpoint which clients--those involved with the disability program and those who frequently contact Social Security--are dissatisfied and to identify the reasons for frequent contacts. Findings such as these lead naturally to other, more focused evaluations. When our analysis indicated that disability program clients were less satisfied with services than others, we began work on a detailed review of these clients' perspectives on the disability claims process, also using the client survey approach.

Periodic surveys, using a common methodology and instrument, can be used to track the results of organizational initiatives. Likewise, changes over time in the responses to key questions on your survey will give you an invaluable "early warning system" for emerging service problems, or early confirmation of success. A decline in clients' ratings of the clarity of Social Security's mail led us to conduct a technical review of several widely-used notices. This review also included a client survey. For the first time in 1993, we reported a statistically significant decline in overall satisfaction with Social Security's services.

For many years, Social Security has used results from the annual client survey to monitor overall satisfaction, the clarity of agency notices, and the courtesy clients receive from agency representatives. We recently expanded on this use of the survey for performance measurement. Two years ago we worked with Social Security officials to tailor our questionnaire to the service objectives in Social Security's

strategic plan. We fine-tuned questions to solicit information about response times to public inquiries, explanations of Social Security programs, referrals to related programs, and other service indicators (Molnar & Stup, 1994).

How Are They Done? A client satisfaction survey uses public opinion survey techniques to determine the nature and extent of clients' experiences with a program and assess their satisfaction with the services received. Most typically, we use a structured format to ask a sample of clients a series of questions about the program. The survey can be administered by telephone, in mail, or in person.

There are several basic steps involved in conducting a client survey, along the lines of the evaluation phases we discussed in Chapter 2. They include: defining the client population and selecting a sample; determining how you will collect the data and developing a survey instrument; administering the survey and recording the data; and analyzing and reporting the results. Some special considerations involving a typical client satisfaction survey will be covered here. Primarily in this discussion we focus on beneficiary satisfaction surveys, where a random sample is drawn in order to provide generalizable information on issues of satisfaction. Different considerations may apply with small samples of technically proficient clients, such as regulated companies, program staff, or officials at other levels of government or in sister agencies.

The Universe and Sample. After you determine the purpose of your survey, you must decide what questions you need to answer to accomplish that purpose; the universe consists of the group of people who can provide the information to answer those questions. In defining this population, you should specify the nature of their program experience, as well as how you will identify them. The sample consists of the subgroup of the universe you will actually contact. As we noted in Chapter 2, there are many ways to select a sample. When you do the selection, also collect any administrative data that's available. This could include basic demographic data, such as age and sex, or programmatic information. Using this data to compare the survey

respondents to the universe and to non-respondents can help you detect possible bias in your results.

The Survey Instrument. How you define your universe may determine how you will survey the people in your sample. For example, if your universe consists of office visitors, and if only some of those visits are recorded, the only way to identify and contact a representative sample of this population may be through an "exit poll" at the site. Similarly, if the only identifier you have is a telephone number, then you will almost certainly have to conduct the survey by phone.

Whatever mode of contact you choose, allow plenty of time and resources for developing a survey questionnaire. This process involves several cycles of writing, testing (using actual population members), and rewriting. Strive for questions that (1) are concise, objective, and non-threatening, (2) use an appropriate language level, with no agency jargon, and (3) follow a logical order. As we discussed earlier, there are two basic types of questions--open-ended, where the respondent provides an answer, and closed-ended, where you ask the respondent to choose from a list of options. The primary advantage of using closed-ended questions is the relative ease of recording and analyzing the responses. However, in using closed-ended questions, you risk "missing the boat." To illustrate, suppose you ask a closed-ended question, such as "What was the main reason for your visit?," giving several possible answers, and 30 percent of your respondents mark "other." You would find it hard to draw any valid conclusions about the reasons for clients' visits. If you decide to use closed-ended questions, then one goal of pretesting should be to scope out all the likeliest responses to your questions.

When asking respondents to rate the quality of service, include a neutral option unless you wish to "force" a positive or negative answer. We usually find a five-point scale ideal for such questions. In asking questions about a past event, consider giving a "don't remember" option. Keep the survey to a reasonable length by asking only the questions you need to address the issues for your study; leave out the "nice to know" questions.

If you are planning an ongoing survey, you should first identify a few key program goals that are unlikely to change very soon. Then develop questions, in terms of clients' perceptions, that will indicate how well the goals are being met. These key questions need not be elaborate or profound, but should be very basic to your program. The key question in the Social Security client survey is "Overall, how would you rate the service that Social Security has given you?" We present five options ranging from "very good" to "very poor." To effectively compare results over time, avoid making any major changes to key questions, whether in wording, scaling or placement, so be sure to ask the right questions from the beginning.

Data Collection. There are a number of steps you can take to maximize your response rate and ensure the validity of your results. In administering the questionnaire, assure respondents that participating in the survey will not affect their connection to the program and that their answers will be held confidential. Be sure to give clear instructions, especially for mailed surveys. Note if respondents should "check all that apply" or "check one." Mark clearly the places where certain respondents should skip questions, depending on their previous answers. (To minimize analysis problems later, take care to identify and mark all of these skip patterns.) If the question refers to a specific event, name it (e.g., "During your last visit.."). If the person had a proxy for the program contact, who should respond to the survey?

We have taken a number of steps to keep our response rate between 75 and 80 percent. The questionnaires have brightly colored covers and are desktop-published. Since many of our respondents are elderly or disabled, we use a type size larger than average to permit easy reading. We send a pre-survey "you have been selected" postcard to each client prior to mailing the questionnaire. When the questionnaire arrives, it includes a signed cover letter explaining the importance of the survey and assuring the respondent of confidentiality. We include a postage-paid return envelope of the appropriate size. We send a reminder postcard, a second questionnaire, and a "last notice" postcard at three week intervals to encourage response. We also include our own 800 number for questions about the survey and for the Social Security survey the agency's 800 number for questions about services.

In the case of surveys conducted by telephone or in person, the interviewer should consistently ask the questions as written and in the same order, should answer any questions the respondent may ask in a non-leading manner, and should use a neutral probing technique when necessary. With multiple interviewers, a training session to assure this consistency is advisable. It should go without saying that common courtesy will go a long way toward gaining the cooperation of potential respondents. After all, you want them to do you the favor of participating in the survey.

Survey responses can be entered into a data base either directly, as when using computer-aided telephone interviewing, or by transcribing the written answers. Particularly with mailed surveys, this, along with human nature, can create the problem of inconsistency: skipped questions or pages, too many options marked, or an answer which conflicts with an earlier answer. You will need some editing method to check for these occurrences. We suggest a review of each completed questionnaire, along with an automated analysis of the responses using test questions to detect problems. For instance, if someone answered "none" to the question "How many times have you visited Social Security in the past 12 months?," then they should not have answered the series of questions about recent visits. Your test should check for the presence of such inappropriately answered (or skipped) questions. You should decide ahead of time what rule you will use to handle this situation, such as assuming the first answer is the correct one.

Limitations of Client Surveys. While client-based information has unique value, it must be used with caution. A primary weakness is that it comprises reported behavior and perceptions, not actual behavior or events. Consequently, its accuracy is affected by the time elapsed since the event in question and the ability of the respondent to recall or evaluate it. For example, of clients who had visited a Social Security office during the past year, 38 percent who had appointments reported that they were served within 15 minutes. In fact, we don't know how long they actually waited. Similarly, Medicare clients may be highly satisfied with a medical provider's substandard service simply because they have no real basis for comparison. For this reason, some program

officials, advocates, or other stakeholders may dismiss client input as peripheral or even misleading.

This is less of a concern if your primary audience or sponsor recognizes the limitations of such qualitative data but highly values the views of the program's customers. If client-based information doesn't have strong support in your program, one way to demonstrate its value is to corroborate survey results using other data sources. For example, when we surveyed patients about their satisfaction with outpatient cataract surgery, we also contracted for a review of their medical records to determine whether the procedure was medically necessary and whether outpatient surgery was appropriate. The medical record review validated the patients' reports about their medical condition and the results of the surgery.

A second weakness is that client-based data generally require special collection techniques. If the data collection is conducted by program staff or by others at the service site, there is a potential for bias. Awareness of the monitoring activity will almost certainly influence the behavior of your public contact staff toward better service. Further, in selecting the sample respondents, interviewers could favor cooperative clients who are "articulate" or "knowledgeable." Clients may also be influenced to respond positively simply because they want to be nice to the interviewer or because they expect to receive assistance at that site. While much of this may occur subconsciously, the biasing effect can be pronounced. In an on-site survey of clients visiting Social Security field offices, overall satisfaction rates were, in fact, slightly higher than for the annual mail-out survey. One possible explanation for this difference is the unintentional biasing of clients' responses.

To reduce this potential bias, whenever possible use non-operations staff to conduct client surveys, and use sites other than the service site. Of course, this raises the problem and expense of identifying and contacting the client. We have found this is sometimes difficult. For example, callers to an 800 number often can be identified only by the telephone numbers from which they called. However, many clients call from business numbers, pay telephones, or other shared numbers such as those for rooming houses or hospitals. In an ongoing survey

involving its 800-number callers, Social Security is successful less than half the time in reaching the person who placed the original call.

In-person, off-site interviews of a sufficient number of clients often are prohibitively expensive. Both the Social Security and Medicare client satisfaction surveys rely on questionnaires mailed to randomly selected samples of clients. However, while inexpensive and generally non-biasing, mailing to potential respondents raises the issue of achieving adequate response rates and therefore confidence levels. As we mentioned before, there are many texts available to help you with formulating a mail survey to achieve high response rates.

We believe the value of client-based data far outweighs the limitations and problems. As noted earlier, under TQM, client perceptions provide a barometer of service quality. For example, it can be argued that waiting on a client within an arbitrary amount of time, such as 15 minutes, is irrelevant. What is relevant is whether the client believes he or she waited too long or was served more quickly than expected. That kind of information, and reasons for a client's satisfaction or dissatisfaction, can come only from the client. Client-based data also can be used to establish performance goals, for example, by letting the client tell you whether 5, 15, or 20 minutes is an acceptable waiting time.

We don't mean to suggest that a client satisfaction survey is the only true measure of your program's performance. It is, however, a critical component in assessing quality and determining whether the program is meeting your clients' needs and expectations.

CHAPTER 4 PERFORMANCE INDICATOR STUDIES

As we discussed earlier, performance measurement is becoming very important to public managers, who are being held increasingly accountable for the success of government programs. Both the Chief Financial Officers Act and the Government Performance and Results Act require federal managers to measure performance (Duquette & Stowe, 1992). The management concepts that form the basis of TQM and reinventing government also depend on the use of performance measurement.

We're all familiar with performance measures. Teachers grading students are measuring performance. Supervisors rating employees also are measuring performance. Other measures are far less systematic or formal, such as the number of requests for a recipe or the number of complaints about a barking dog. While the process is generally imperfect and often criticized, performance measurement, in one form or another, is a routine part of our lives.

Yet the science, or perhaps art, of developing and using performance indicators and measures is still somewhat in its infancy in government. In the past, public programs have captured a great deal of data on inputs (budgets, staff) and outputs (numbers of clients served, number of claims processed). For the most part, these things are easily measured. However, the public sector hasn't collected much information on outcomes or impact, the ways in which a program fulfills its fundamental purpose. Output measures provide feedback on your program's achievements but may show you very little about whether those achievements led to meaningful results. Outcome measures give you feedback on whether the results you were looking for actually occurred but less information about whether your program was responsible for them. In such an imperfect world, program managers need to develop and test a variety of measures and obtain advice from a wide range of stakeholders.

What Are They? Performance indicator studies identify and test possible gauges of inputs, outputs, and outcomes that indicate how

effective and efficient a program is at achieving its intended purpose. While performance measurement terminology is still somewhat variable, **indicators** can be described as representing "what," or a quantity of inputs, outputs, or results. **Measures**, on the other hand, express "how well" as a ratio of input to output, of output to results, or of output or results over time. Thus, the relationships between and among indicators can give you useful information about the success of your program. Since indicators form the basis for performance measurement, they will be our focus in this chapter. (Although for simplicity we discuss the "program" as a whole, performance indicator studies also can be used in managing individual components of a program, such as customer service or quality assurance.)

These studies use a variety of evaluation techniques to: determine the intended purpose of the program under review; identify potential indicators of success; and test those indicators for validity. Effective indicators then become standards against which to measure program performance. This chapter deals almost exclusively with the **development** of indicators, since the program evaluation techniques presented in other chapters can be used to assess performance on established indicators. For example, as we mentioned earlier, Social Security tracks the ratings for courtesy, clarity of mail, and overall satisfaction from our annual survey of Social Security clients and uses these to measure agency performance over time.

How Are They Useful? Sound performance indicators can tell you a lot about your organization. First, they can tell you how well your own program is doing. Second, if other organizations have the same or similar objectives, you can compare your performance to theirs. Further, after you've collected data for more than one period of time, you can tell if your performance got better or worse. However, all this is useful only if your performance indicators accurately reflect your program's basic purpose. Ideally, performance measurement shouldn't be just a report card; it should give you enough information to pinpoint trouble spots and identify obstacles to achieving your goals. Otherwise, the information can't help you decide on your next steps for program improvement or issues for further investigation.

Besides giving you valuable information to better manage your program, performance measurement can be used to convince others of the importance of your program's goals and objectives. Arguably the most significant benefit of developing and adopting meaningful performance indicators relates to the old saying "What gets measured gets done." Good performance indicators can galvanize and empower your entire organization toward achieving lasting impact. Finally, performance measurement can help you prove your program's success and impact to those who oversee your operations as well as those who control your funding.

How Are They Done? Performance indicator studies involve a number of interrelated steps: (1) purpose review; (2) stakeholder analysis; (3) feasibility assessment; and (4) validity testing. Existing indicators can be critiqued using the same process, with the added advantage that you can evaluate their actual effectiveness in monitoring performance.

Purpose Review. The primary objective of the purpose review is to research and document the intended results of the program. Your review also may reveal potential measures of success in achieving those results. (Our suggestions for conducting literature reviews, shown later in Chapter 6, may help with this research.) First, you should identify any statutorily mandated goals or standards and learn whether statutory language, or legislative debate or other action, indicates specific interest in certain program results. Reviewing past related studies, especially those involving control and test groups, can yield important information about expected program outcomes as well as what factors seem most closely related to success. Also look at any recent program directives or planning documents issued by top agency officials, since they may set out performance goals that need only be described in measurable terms to become indicators. You may find disagreement among sources as to the importance of some program goals, or you may find some expected outcomes never have been clearly articulated, much less measured. You can address these issues in the stakeholder analysis.

Stakeholder Analysis. Managers often make the mistake of assuming they know what's most important to others. Indeed, if you stay in close contact with your stakeholders--clients, staff, and others with a stake in your program--you probably do have a pretty good idea of their priorities. However, you may not be aware of the nature or relative importance of some of the factors they use to judge the program's success, so you need to ask them. (You may even be unaware of who some of your stakeholders are. Analyzing each of your program's processes--whom it touches and in what way--can help you identify your stakeholders and their relationship to your program.)

Your contacts with stakeholders should give you a good idea of what factors they relate to program success and what kinds of measures would have **face validity** as indicators of how well the program is doing in those areas. By face validity we mean acceptance by reasonable people that the measure is valid, not necessarily based on in-depth study but simply on its face, relying on their experience, knowledge, and common sense. You also should gain an understanding of how your stakeholders rank these factors in importance.

For example, we assessed the performance of Head Start grantees by working with national program managers to establish a set of 18 performance measures, including the child-to-staff ratio and the percentage of children receiving medical screening and treatment (OEI, 1993c). In another study, we asked State and local welfare workers to rate a series of proposed indicators in terms of their value and relationship to intended outcomes.

A number of techniques can be used in various combinations to solicit information from stakeholders about their priorities and assessments:

- ◆ Individual interviews. One-on-one interviews, by telephone or in person, can provide a rich understanding of key stakeholders' personal views. In your discussion, you can explore the degree of flexibility in their views and probe for the reasons behind strongly held opinions.

- ◆ Focus groups. Here, a trained moderator guides a round-table discussion among a small group of stakeholders. The discussion should allow for a free exchange of ideas among participants--for participants to disagree with each other, to build on each other's responses, to try out their ideas (and possibly to change their minds), and to create new ideas through synergy. These forums are particularly useful for developing ways to measure program success as well as identifying what aspects of the program are important.
- ◆ Mail surveys. These are an economical way to solicit the views of a large number of stakeholders and are useful for validating the information obtained from individual interviews and focus groups. A common format is to ask respondents to rank the importance of a list of factors of program success and then to suggest, select, or rate possible indicators for those judged most important.

Analyzing these results can be complex and time-consuming. You shouldn't be surprised to find that different stakeholders will identify different factors as more or less important. For example, program participants might value customer service more highly than do budget staff, who might rank cost-efficiency as most important. The following approaches may help you use the information you've gained from your research and stakeholder contacts to make some preliminary decisions on performance indicators:

- ◆ List the factors respondents identified with program success. Array responses on a scale showing the percentage of respondents who mentioned the factor as important, from 100 percent at one end to 0 at the other. Where there is consensus or near consensus on a factor (perhaps 75-100 percent), that factor should probably be included as part of your performance measurement plan. You will need to explore further the areas with mixed (or almost no) support: Are the stakeholders who value this factor especially important for some reason? Is this factor related to a legal mandate or noted as important in the

literature you reviewed? Are there other compelling arguments for including it?

- ◆ After you have developed a preliminary list of factors to be measured, analyze it by stakeholder type. To the extent possible, your list should reflect the priorities of all key groups (otherwise you're at risk of alienating a group of your stakeholders).
- ◆ Share the list with your key stakeholders (we suggest a mix of your original respondents and some "cold readers"), and ask for their reactions both to individual items and to the entire list. Find out if measuring this combination of factors will fairly portray the success of the program as whole.
- ◆ Once you've developed this list of factors, go back to the data you've collected to look for potential indicators for those factors. You may find it helpful again to array your responses, develop a preliminary list, and consult your key stakeholders.

Feasibility and Resource Assessment. You should now have a list of potential program indicators--quantifiable inputs, outputs, and outcomes related to program success--and you need to explore ways to measure them. It's important to find out whether relevant data already exist somewhere in useable form, whether existing data systems can be accessed to provide you with the needed information, or whether you will need to modify or create a system to obtain the data you need. Before you decide to place additional demands on an existing system or develop a new one, consider the possible disruption or resource costs that may result and explore the alternatives. To the extent you can translate performance measurement into dollar savings, you can use a cost-benefit analysis to defend requests for additional funds for this purpose. For example, you may be able to project the savings that would result from monitoring the readability of program correspondence and thereby reducing the number of clients who visit or call the office for explanations.

When you consider using an existing information system for collecting performance data, examine the system and its current use

carefully. If quality control is limited or if the information has limited operational importance, the people who submit the information may have little concern for its accuracy. Also consider whether those who provide or collect the data could have a vested interest in the results. These factors all affect the reliability of the data and should enter into your decision about whether or not to use that data system for performance measurement.

Testing. The final step in developing performance indicators is assessing how well they work. You should actually collect data and then check the data's validity using case files or other source evidence. Analyze the data and develop preliminary findings based on your indicators. Preferably, you should conduct the data collection more than once so you can assess the effects of time. Then go back and compare your results to the information you gathered earlier about the purpose of your program, desired outcomes and their relative importance, and your processes. Also examine the data collection costs you expect to incur if you adopt these indicators for your program.

As you do this analysis, ask yourself, as well as significant stakeholders, these questions:

- ◆ Does the indicator appear to be valid and reliable? In other words, does it measure what it's intended to measure, and is it essentially free of error? Does it demonstrate these characteristics consistently over time?

Wide and unexpected variations from one program site to another, or strange results across the board, might suggest validity or reliability problems. In one study on the delivery of preventive health services to children in the Medicaid program, we were surprised to note some States reporting they had provided services to more than 100 percent of the eligible children! The formula on which the performance measure was based was in error (OEI, 1992a).

- ◆ Do the results seem reasonable or do they suggest potential problems? Does there appear to be virtually no variation at all?

That could be a sign that your performance measure lacks the ability to distinguish true differences and doesn't really measure anything.

- ◆ For existing performance indicators, how is the information being used? Is it feeding into budget deliberations, legislative agendas, or new program initiatives? Is it being totally ignored? If the latter, does this reveal a lack of trust in the numbers, in the methodology or data sources, or in the measure itself? Or does it reflect a programmatic problem so frustrating that people don't know what to do with the information, no matter how credible.

After you eliminate unproductive indicators and make needed adjustments and corrections, the performance measurement system you've designed can be a vital tool for managing and improving your program. As you use this system, you'll become aware of needs and opportunities to fine-tune it; you can use the same techniques presented in this chapter to refine existing indicators. Also keep in mind that as the environment changes--new legislative interests, judicial decisions, demographic shifts, etc.--so too will your performance indicators and measures need to change.

Limitations of Performance Indicator Studies. Performance indicator studies may be the most difficult of the techniques we use in program evaluation. The methodology often depends on institutional and individual intuition and judgment about what matters enough to measure. Outcome measures are particularly difficult to develop. A particular outcome may be absolutely critical to the success of the program and absolutely impossible to quantify, so a measurable proxy has to be found to serve as the indicator. Causal connections are frequently difficult to establish, so these studies rely heavily on the factors that Congress, advocacy groups, and other entities involved in formulating public policy believe contribute to desired outcomes. To counteract this element of subjectivity, try to select indicators that are supported by multiple sources and to obtain consensus among a variety of stakeholders that those indicators are the right ones.

CHAPTER 5 COMPLIANCE REVIEWS

As a public sector manager, more than likely nearly everything you do is governed by law and regulations. Your program--its purpose and authority, its funding, and possibly even your job--is established by law. Details about the processes you use to administer your program may be legislated. If your organization funds other public agencies or service providers, the nature of their involvement in your program may be mandated also. These laws, and the regulations that implement them, serve as standards against which your program's performance can be measured.

You probably already produce or receive information about program compliance--audits and legislatively-mandated reports. However, at any point in time, your chief executive may want to know if you, your contractors, or grantees are meeting legal standards that aren't routinely monitored. Legislative oversight bodies may, too. Shouldn't you be the first to ask *and answer* these questions? Using evaluation techniques to investigate compliance issues can prepare you for such challenges as well as give you a deeper understanding of what's happening in your field operations.

What Are They? Like performance measures, compliance reviews involve assessing an activity or output and comparing it to a standard. As covered in the last chapter, performance measures assess how well the program is meeting its basic purpose. This purpose may or may not be stated explicitly in statute or other legally binding form. On the other hand, compliance reviews are designed to determine if specific activities are occurring in accordance with legal, regulatory, and administrative requirements.

Compliance reviews determine if you, or your grantees or contractors, are doing what is prescribed by your program's enabling law and regulations. They also can be used to determine compliance with other administrative laws or regulations dealing with personnel, fiscal or similar requirements. Some examples of administrative requirements are internal controls and computer security standards,

hiring and procurement procedures, and fiscal reporting and accountability standards.

How Are They Useful? Compliance reviews are most frequently used to fulfill an oversight function. For example, they could be used by a funding agency to assure that a grantee or contractor is performing in compliance with the terms and conditions of the grant or contract. Accordingly, State and federal agencies often conduct compliance reviews to ensure that their grantees are following applicable laws and regulations.

Agencies also can use compliance reviews to monitor their own operations. For instance, an organization with a number of field offices could use a compliance review to assure that these offices were operating according to established procedures. Our office uses the same program evaluation techniques that we use for oversight activities to monitor our own regional and branch offices' compliance with internal requirements for maintaining work paper files.

Oversight organizations performing compliance reviews usually are looking only to document the degree of compliance or noncompliance. However, as a program manager, you can use compliance reviews to learn a great deal more. Compliance reviews conducted periodically can provide valuable information on performance trends. Where you find evidence of noncompliance, you will want to know why and take corrective action. Even where compliance exists, it could be extremely useful to find out the costs and benefits of that compliance. Sometimes compliance reviews are seen as just a "paper exercise." Using an evaluation approach can add meaningful information that will help you get to the heart of any compliance problems.

How Are They Done? Having decided to conduct a compliance review, you need to determine what the specific standards are. In many cases, the wording of the applicable law or regulation will stipulate the standard. It may state how you must operate or what your grantee "must" or "shall" do as a condition of funding. For example, many federal grant programs require States to conduct a needs assessment of the target population. In conducting a compliance review for this

provision, the evaluator only needs to verify the existence of the needs assessment. However, if the program requires a **comprehensive** needs assessment, the evaluator must go beyond verifying the existence of the needs assessment and evaluate it. While the standards usually are found in law or regulation, they may not be stated in a measurable form. Using the same approach discussed in the last chapter, you need to develop a measurable standard.

Closely related to measurability is the quality of the data you need to collect. The validity of your results is more likely to be questioned for this type of study than any other because your respondents have a vested interest in them. To what extent do you need to validate statements by officials and staff of the organization under review, attesting to their compliance? Should you verify their testimony by asking for documentation? Should you then conduct an independent and objective verification, collecting raw data to determine if the documentation provided is accurate?

We adopted the last approach when we tested for compliance with the guidelines of the National Cholesterol Education Program by groups performing cholesterol screening in local drugstores, shopping malls, and other public places (OEI, 1990b). The guidelines prescribe the practices that health personnel must follow to safely and accurately screen for cholesterol in the blood. We selected 70 local cholesterol screening sites across the country, and then our analysts covertly participated in the actual screening process. (We never promised you program evaluation would be painless!) Using the guidelines and direct observation, we assessed the methods used in the screening process as we and others were tested. We found a surprising lack of compliance with the guidelines, noting unsanitary conditions, poor handling of blood, lack of staff training, and incorrect advice to participants regarding their cholesterol levels. The publication of our results caused a panic in the industry and a rush to bring screening sites into compliance.

In designing a compliance review, you must decide on a number of methodological questions. Do you want to evaluate a sample of agencies or all of them? Whom will you interview--the senior official

or a specialist? Or both? Do you want to interview them together or separately? Should you interview a group of all responsible staff? Will you review or copy records? There may be no right or wrong answer. It will depend on your audience, available resources (staff time and travel funds), and on the level of confidence you want to have in your findings.

When you develop a data collection instrument for a compliance review, consider whether you plan to use it periodically to measure performance over time. In a review covering a number of issues, it might be useful to have three-quarters of the instrument remain stable over time, while one-quarter could be devoted to questions about new initiatives or special items of interest to your stakeholders. Remember that, even for a comprehensive review, you normally should not try to measure compliance with all details of all applicable laws and rules. Concentrate on substantive requirements that are universally viewed as having an important relationship to successful program outcomes.

The data collection instrument you develop serves a number of functions. First, it serves as the agenda for the review. Through the instructions and the content and ordering of questions, it specifies what issues will be covered, in what order, and by whom, and whether respondent-provided documents, or raw data will be required. With a few exceptions, we have found it beneficial to share the compliance review instrument with the respondents well in advance of the data collection. This facilitates the interview process by giving respondents an opportunity to pull together information needed to answer your questions and to locate and duplicate, if necessary, key documents you will need. Second, development of a data collection instrument goes hand-in-hand with setting priorities for you and your staff. This exercise will help you focus on what matters most to you by defining what you believe to be the important elements of compliance for the subject agency or respondent, which in turn sends an important message to that entity.

An interesting example of a compliance review is one in which we assisted the Administration on Aging (AoA) in designing and conducting an assessment of their State agencies (OEI, 1993d).

AoA's major program is a \$750 million formula grant to States to fund community-based, long-term care services authorized by the Older Americans Act. As a condition of the formula grant, States must implement some specific planning requirements, and inform and monitor their local grantees regarding a broad array of programmatic requirements. The States' implementation activities, in turn, are to be monitored by AoA. As Congress went through its periodic reauthorization of the Older Americans Act, it asked AoA for information on States' implementation of the formula grant requirements.

Unable to fully answer Congress' questions, AoA requested our help. Due to significant reductions in staff and travel funds, AoA had not made routine visits to monitor States in almost 10 years. We recommended a compliance review, both to provide AoA with information to answer Congress' questions and to reestablish AoA's oversight role.

The first step was to identify the standards that had to be met as a condition of the grant. We began by locating the various sources of these program requirements--laws, regulations, etc. Next, we actually started listing the requirements, which included activities such as defining client eligibility, conducting needs assessments, and holding public hearings. As we attempted to list all the requirements, we found too many to include in one data collection. Accordingly, we limited our review to five critical components of the program. Even then, we found over 100 requirements, which we called "review criteria." For each one, we decided to look at three aspects of States' implementation of the requirement: 1) was the State aware of the requirement; 2) had the State made the local grantee aware of it; and 3) did the State monitor local implementation of it?

The next step was to develop the data collection instruments. We developed five, one for each of the major programmatic areas. We designed the instruments to meet a number of needs. They needed to address all three aspects of implementation for each of the review criteria. We wanted them to be the single place for recording all the

review information and data. They also needed to include all the instructions for completing the review.

Developing the instruments was complicated by several factors. First, the lack of data collection experience of the AoA staff called for providing detailed instructions. Second, the instruments were to be shared with each State prior to the visit, so some of the instructions had to be addressed to the State. Third, the instrument had to accommodate two methods of data collection, personal interviews and document reviews.

Each State review was conducted in two phases: document review and site visit. In the first phase, we sent the review criteria to the State and asked for the documents used to make grantees aware of the criteria and to monitor grantees' implementation of them. The AoA reviewed the State's response for completeness. In the second phase, two staff visited the State to conduct interviews and to review sample grantee plans, assessments, and financial audits.

Actual data collection took place in a sample of States. We had estimated that it would take two AoA staff four days on site to complete the five instruments for one State. However, AoA's limited travel funds wouldn't cover visits to all 57 States and territories. We decided to use a random sample of States that would allow us to project findings to the universe of all States and territories. Consequently, at the end of the study, we had compliance reviews for 20 States and a database containing information representative of the country as a whole.

There were two products of the entire project. The first consisted of a set of reports on States' compliance in each of the five programmatic areas. These reports together presented a national picture of implementation of the Older Americans Act. We also used the results to prepare an action plan for AoA to strengthen its guidance to the States.

Limitations of Compliance Reviews. Compliance reviews are generally input- and output-focused. This is because, as a rule, laws and regulations spell out "what" and "how" agencies should perform, but not "why" or "to what point." (It will be interesting to see if Congress moves toward *legislating* more outcome-based performance standards as a result of current trends.) Program officials, grantees, and contractors usually are not held directly accountable by law for the intended ultimate outcome of their program. The danger of compliance reviews, then, is that they may show that everything is in order while failing to show that nothing is being accomplished. Nonetheless, compliance reviews can be instrumental in identifying bad actors, avoiding scandal, redirecting misspent funds, and alerting you to fundamental vulnerabilities in your program.

CHAPTER 6 EFFECTIVE PRACTICE REVIEWS

How can you improve your program in times of scarce resources? One way is with effective practice reviews. These studies search for successful methods, techniques, and strategies for providing services more efficiently or effectively. This chapter describes effective practice reviews: what they are, how they are used and conducted, and their limitations.

What Are They? Simply put, effective practice reviews are a way to learn from the successes of others--to discover methods, processes, projects, or practices that have the potential to work well for similar programs (Mangano, 1991). Often, program managers rely on their own personal experience and judgment to assess and decide among alternatives. Those wanting definitive answers can invest in evaluations that assess impact over long time periods. While personal experience is often too limiting, such impact evaluations are often too costly and time-consuming. Effective practice reviews provide a middle-ground approach. They look for lessons learned by a number of professionals or at a number of service-delivery sites. They sacrifice some of the precision of full-scale impact evaluations in exchange for faster (and less costly) information.

Effective practice reviews draw on a number of analytical techniques including:

- ◆ reviewing prior studies to see if certain innovations have already been proven effective,
- ◆ collecting opinions from the people most familiar with the program or issue about whether or not a practice works well or holds promise,
- ◆ analyzing available data to determine if the implementation of a practice has resulted in improved program outcomes over time, and

- ◆ making on-site visits to directly observe the effectiveness of program activities.

The findings from effective practice reviews provide program managers with a basis for saying "This idea might work for us."

How Are They Useful? We have been conducting effective practice reviews for a number of years because program managers like you have told us you want to know what others are doing and what works best for them. Our findings have been used to improve program effectiveness, stretch tight budgets, and address problems.

Two examples illustrate the usefulness of these reviews. In one instance, the HHS official responsible for child welfare services expressed concern about a growing problem for the health care and social service communities: children born to mothers addicted to crack cocaine (OEI, 1990c). We helped her with an effective practice review. We made on-site visits to hospitals and other service-providers in 12 metropolitan areas to observe their practices for dealing with the problem. As a result, we identified and reported on a number of constructive ways for communities to tailor services to crack-addicted mothers and their babies, such as community outreach, early intervention, and co-located comprehensive services.

In another case, we used an effective practice review to help save money as well as improve service. The Omnibus Budget Reconciliation Act of 1987 banned the use of physical and chemical restraints for nursing home patients who don't need them for treatment of a medical condition (OEI, 1992b). Many facilities had relied heavily on restraints to manage their patients. Using an effective practice review, we sought out the lessons learned by a group of nursing homes that successfully reduced their reliance on restraints without incurring special expenses. Our report, presented in question-and-answer format, catalogued their inexpensive yet effective approaches to maintaining a restraint-free home.

How Are They Done? Identifying effective practices is a dynamic process. You are looking for practices with a "track record"--ideas for

handling problems more quickly, reducing costs, or providing better service. To find them, search for program sites that appear to perform better than others, the *positive exceptions*. How do you do this? The best way we have found is an interactive process combining a program review and literature search, discussions with key people, on-site reviews, and data analysis. Here's what these steps involve.

Program Review. An effective practice review begins with a detailed program review. Even though you may be highly knowledgeable about your program, there is value in re-examining its legislation, regulations and important policy documents. These provide a reminder of the purpose the legislature and administration intended the program to fulfill and of how important policy issues have been resolved. Be sure to review documentation of current program objectives to get a sense of how they may have changed over time. A few brief visits to local sites for discussions with managers and staff can give you a fresh perspective on program operations and on the current role of your program in the community. This program review sets the stage for the rest of the review by providing the context--the purpose, objectives, and current operational climate of your program.

Literature Review. Typically, your next step will be examining prior studies of your program or issue, looking for examples of effective practices that already have been documented. The literature also may suggest useful approaches to conducting your review, such as ways to analyze programs, identify experts in the field, describe successful programs, or present comparative data. Possible sources of prior studies include:

- ◆ public and private agency-sponsored evaluations;
- ◆ oversight reviews by the General Accounting Office, Inspectors General, State legislative audit offices, and the Congressional Research Service;
- ◆ university and foundation-sponsored research;

- ◆ studies by professional associations like the American Medical Association, the American Public Welfare Association, and the American Bar Association;
- ◆ publications by "think tanks" like the Urban Institute, American Enterprise Institute, and Brookings Institution;
- ◆ reviews by policy analysis groups like the Office of Technology Assessment; and
- ◆ articles and reports by interest and advocacy groups like the National Governors Association, American Association of Retired Persons, and Children's Defense Fund.

It's important, however, to evaluate the quality of these reports. You need to understand any shortcomings or limitations of a previous study before deciding on the weight it should carry in your review. With a critical eye, examine how the research was conducted (a statistician's advice could prove helpful in this). Some important questions to consider include: did the evaluators' methods introduce potential bias; did they or their funding source have a vested interest in achieving the reported outcome; was the data obtained independently or from the subject of the review; have the findings been confirmed by others; does the sample size support the extent to which the findings were generalized; and was the research recent enough that the results are still valid? Be suspicious of any report that doesn't provide enough information for you to reach a reasonable conclusion as to its reliability.

Some effective practice reviews may rely almost exclusively on a literature review, especially if there have been numerous studies on the subject. In one effective practice review on problems of the elderly with medication, we found that many effective practices were already cited in prior research. However, no one had analyzed and presented the collective findings from this body of work. In the reports for over 70 studies, we found good ideas for ensuring that the elderly take needed medication in the prescribed amounts, and we used our findings to prepare a user-friendly collection of these best practices (OEI, 1989c).

Expert Opinions. Consult the key people in your program; they usually know what works. A good place to start is with nationally recognized experts, who typically have a broad perspective on the program and an understanding of numerous approaches to service delivery and who are respected by others in their field. Many have conducted their own research and published reports. Some are instrumental in running or assessing national programs. You can identify them through the literature review or by asking national associations or professional groups. In the study of medication problems of the elderly, researchers, academics, and national experts in geriatrics and pharmacology were some of our best sources of good practices. However, since some experts hold strong biases, be careful to validate their views through other sources.

Local practitioners are another excellent source of effective practices. While conducting on-site reviews, you can ask program officials and related service providers to identify local experts. These providers usually have a good grasp of what is working in the community and will note which local programs they believe are effective. Local experts helped us identify many good practices we later described in our report on babies prenatally exposed to crack cocaine. We then confirmed the effectiveness of these practices during site visits, interviews, and data analysis.

On-site Reviews. You should be reluctant to rely exclusively on what others tell you, so most effective practice reviews should include site visits. Local conditions can affect which practices are most effective, and there is no substitute for first-hand observation. We recommend that your study team personally inspect the service-delivery sites whenever possible. The on-site review should include:

- ◆ talking with local program officials, front-line workers, and recipients;
- ◆ observing the program in action,
- ◆ collecting local program performance data (e.g., number of people served, cost of services, quality of service, etc.), and

- ◆ gaining an in-depth understanding and appreciation of program dynamics.

Since an effective practice must have the potential for use in a variety of settings, site visits also are a good way to assess their transferability.

Selecting the local sites can be tricky. You have to decide which criteria to use. Do you want to go only to sites reputed to be the best? If so, how will you identify them? Would you rather get a mix, so you can observe their differences and compare their results? Do you want diversity (in geographic settings, population groups, income levels, etc.)? The answers are important and will have a significant effect on what you find.

A good example of the use of on-site reviews was a study we did of the clinical role of the community pharmacist (OEI, 1990d). Clinical pharmacy services are patient-oriented activities that identify, resolve, and prevent drug-related problems. The study team selected 6 local pharmacies to visit out of 35 suggested by members of professional pharmaceutical groups. Final sites were selected based on the nature of each pharmacist's practice, the population served, the pharmacist's therapeutic specialty, and geographic diversity. In each case, the study team observed the pharmacists' daily activities, interviewed their customers, and spoke with the physicians with whom they collaborated. The study compared the practices of these six practitioners against the average level of clinical services offered in community pharmacy settings and resulted in a report describing the exceptional range of clinical care provided by these model pharmacists.

Data Analysis. At some point, all effective practice reviews include data analysis; some begin with it. By analyzing available program data, you can discover trends, problems, and possible solutions. Management information, census data, financial reports, as well as educational and health statistics, are all examples of useful data. Your study team needs to determine what data are available, how reliable they are, which to use, and how to access them. These data bases are often the quickest way to identify or confirm effective practice sites by

pointing you to those which provide the most efficient and effective service to their clients (i.e., at less cost and with better results).

Our review of effective methods for increasing child support payments, for example, began with data analysis. To identify successful local programs, we analyzed national statistics on out-of-wedlock births and the number of paternities established (OEI, 1990e). We also sought recommendations from prominent special interest groups, federal, congressional and regional child support staff, attorneys, and researchers. Trend data from the sample sites were important in determining which methods were most effective. In this case, the study team defined effective practices as procedures which improve the number of paternities established, case decision accuracy, or case management efficiency. On-site reviews and interviews with key people validated the results of the data analysis.

In summary, the techniques used to conduct effective practice reviews are highly interactive, with information gained through each method--literature review, key interviews, on-site reviews, and data analysis--guiding other review activities. Ideally, several sources will validate each other in identifying the best practices. The stronger the corroboration, the more credible the findings.

Limitations of Effective Practice Reviews. We believe effective practice reviews are a useful and practical tool for finding ways to improve program services. However, we acknowledge some potential shortcomings that may limit their usefulness:

- ◆ The success of an effective practice may not continue over time.
- ◆ What works well in one setting may not be effective in others.
- ◆ It is difficult to isolate the specific practices that led to success.
- ◆ Reliable data may not be available to substantiate (or refute) expert opinions.

- ◆ Effectiveness may be more a result of characteristics of the practitioners than of their practices.

Despite these limitations, we believe effective practice reviews are a sound investment. In a budget-saving and timely way, they help you discover what works as you seek out, analyze, and describe successful practices that can improve your program's performance and help you achieve success in times of limited resources.

CHAPTER 7 EARLY IMPLEMENTATION REVIEWS

New mandates and requirements provide one of your best opportunities to test the usefulness of developing an evaluation capacity for your program. Any venture into uncharted territory comes with risks. Your role in implementing new responsibilities may involve taking on new and different functions: providing service to new clients with unfamiliar needs or characteristics; interacting with new provider groups or industries; or creating new policies, procedures, or data systems. Regardless of how well you've planned the implementation of these new requirements, some unanticipated problems will almost certainly arise, and they can have serious consequences.

What Are They? Early implementation reviews are studies that assess the vulnerabilities, problems, and successes of a new initiative or program during the start-up period.

How Are They Useful? These reviews allow you to identify and correct trouble spots before too much damage is done and before faulty processes or assumptions become ingrained. These reviews can even demonstrate where adjustments are needed in your assigned mandate. Since things are still new and evolving, you're less likely to encounter resistance to change now than you will be later on when program officials, staff and the public become vested in the current way of doing things. On the positive side, early identification of the successful aspects of implementation lets you capitalize on them.

We used an early implementation review to evaluate the Comprehensive Perinatal Care Program in its initial stages of operation (OEI, 1990f). This program provides services to low-income mothers and their children, at or near the time of birth, in communities with the highest infant mortality rates. We found a flaw in the design of the program's grant application process that prevented it from reaching some of the most needy areas. The Public Health Service immediately instituted changes to accept applications from these excluded communities. By alerting officials to the problem early in the life of the

program, we minimized the period during which many poor families were disadvantaged.

A well-designed early implementation review can provide the basis for ongoing monitoring. You may want to repeat the review at regular intervals to know if you're getting better at fulfilling your new responsibilities as time goes on. You may want to issue periodic "report cards" to provide information on your successes, to discuss problem areas and your corrective action plans, and to demonstrate your ongoing commitment to monitoring the implementation of the new initiative.

How Are They Done? Needless to say, the best time to plan for an early implementation review is early, preferably when you've just received your new mandate and are planning the implementation. Your planning group can make the review an integral part of its implementation plans. When you announce your implementation plan and schedule, you also can announce your intention to conduct an early implementation review and describe how the results of that evaluation will help you meet everyone's expectations.

Because the initiative is new, you'll have the luxury of a fresh perspective on the originators' intentions. Using this information, the planning staff can identify, in practical and measurable terms, how they will determine if the program is achieving its intended goals. In addition, the staff may want early feedback on how well certain processes are working, particularly if these activities represent a new role for the agency or are somehow problematic.

For example, suppose you've been required to expand the services of your program to a new group of clients. Advocacy groups, legislative staff, and policy analysts may voice the concern that this population won't have adequate access to your services. Perhaps your intake staff (who have the job of enrolling clients) or your caseworkers (the folks who actually provide the services) aren't physically located near the new client base, or maybe these clients won't be aware they can use your services. To address these concerns, you've developed an outreach and education initiative. Moreover, you decide you'll measure

how well you're doing in bringing in the new clients six months to a year later. In other words, you commit to testing your theory about how to reach out to this population, early enough to minimize the impact of any problems and to take maximum advantage of the activities that are working well. Perhaps you'll find you're failing miserably at bringing in your new clients. This information could indicate a need for a legislative change such as requiring that established service providers for these clients provide enrollment forms or information about your services.

Don't worry too much about being too early with your early implementation review. We recently examined how much progress States had made in implementing a new set of federal rules for inspecting nursing homes. These rules were a sharp departure from the old way of doing things, and we fully expected States to have trouble adjusting to them. At that point, only a few inspections had taken place under the new system, and, in many people's eyes, it was too soon to conduct any sort of evaluation of these first, tentative efforts. Yet we believed this was exactly the right time to review the States' experiences and progress in carrying out these reforms. Through conversations with a variety of State inspectors, as well as a review of data on the inspections and their outcomes, we identified stumbling blocks which could be remedied with more technical assistance and federal direction. We also alerted policy-makers to problems inherent in the legislation itself that we feared would prove troublesome.

In conducting an early implementation review, it is important to go beyond your automated management information system. While computerized reports of program statistics can give you important information, they cannot, by themselves, constitute an early implementation review. This is because a critical factor in an initiative's implementation phase is **uncertainty**. Remember, much of the purpose of an early implementation review is to discover unanticipated problems and the things that are working best. It involves examining the effectiveness of the process you've established with an eye out for things you've overlooked. That means you and your team have got to get out of the office and away from spreadsheets to talk with knowledgeable individuals, examine the mechanisms you've put in

place, and observe real life. It is essential to talk with front-line staff, program recipients, and other stakeholders, and it may be necessary to collect and analyze data on an ad hoc basis, in order to obtain a full picture of a new program's performance.

Do Your Homework. The first step in an early implementation review is to make sure you and your project team fully understand what's "new" about this initiative--not just the words on a piece of paper, but its meaning and spirit. What does the initiative represent--a change in philosophy, a shake-up of the old order, or simply an extension of the current way of doing business? Does this program represent a response to a previously unrecognized or unacknowledged need? If so, is it similar to existing programs? Just how does this program differ from the status quo?

An understanding of the nature of the change can come from many sources which may include accounts of the legislative history, newspaper or journal articles, and interviews with key staff. Try to become conversant about the deliberations that preceded the decision to adopt the new program or initiative by answering these questions:

- ◆ Why was this program needed? What were people concerned about? What are they trying to accomplish?
- ◆ If the initiative represents a different way of achieving a goal, what was wrong with the old way of doing things?
- ◆ What groups have new responsibilities? Why them?
- ◆ What obstacles did knowledgeable individuals anticipate?

The answers will help you focus your review on major issues. Further, if the new program involves a significant expansion of your agency's current responsibilities, you may want to examine structural capacity. Do you have the personnel, budget, and data systems to do the job you've been given? Does the new initiative build on existing relationships and capacity or require the development of new ones? Other issues to examine are whether the change introduces new

problems or vulnerabilities and whether the new procedures are being followed consistently, sporadically, or not at all. The answers to these questions will suggest the appropriate next steps for your review.

Consider the purpose and nature of the change in deciding on your methodology and data sources. If the new program represents a significant shift in priorities or philosophy for your agency, you might want to focus on the attitudes of front-line employees and whether they buy into the change or are resisting it. If the change is less intrinsic but involves new policies or procedures, you might want to focus on communication issues and whether staff understand their new responsibilities. Here's a case where we did both. Following the passage of the Family Support Act of 1988, we conducted a review focused on local case managers responsible for processing applications for Aid to Families with Dependent Children (OEI, 1988b). We were interested in testing their knowledge of the new legislation and how that knowledge affected their performance. Analysis showed that those staff most knowledgeable about the Act were also the most enthusiastic and supportive of it. We advised that training front-line workers on the Act's provisions appeared to be a good investment.

Find Out How Things Are Working. As mentioned above, it is critical that your project team review the process that's been put in place to implement the new initiative and then go to the front lines to observe it.

Ask the people closest to the program about their views on how the new process is working. Are the directions clear? Does it take too much time? Does it cover all contingencies? These folks usually know almost immediately if a new process is workable or has fundamental flaws. To save time in taking applications for disability benefits, Social Security began asking the claimants to complete the medical summary. We went to claimants and front-line workers to see how this process was working. We discovered that, while there were some savings, many applicants lacked the confidence, knowledge, and ability to accurately complete such a complex form. Many completed it so poorly that a Social Security representative had to fill out a new one. Based on this information, the agency decided to simplify the form and

institute a screening process to identify applicants who are able to complete it (OEI, 1993e).

Review documentation. Is the appropriate information being collected and entered in the files? Does it indicate that the new procedures are being followed? If not, where does the process break down? Are there enough data for you to make these judgments, or do you need to collect more?

Watch the action. Do the people involved in the process seem comfortable or are they on edge? Are participants getting mixed signals? Why? Is there confusion about processes or goals?

You also will want to record or collect information on production. How many cases have been processed under the new system? How much time did they take? How does that compare with the number of cases and processing time for the same period last year under the old system? How many clients have been served by the new program? How many enforcement actions have been taken? How do all these outputs differ from one area or site to the next?

Limitations of Early Implementation Reviews. For most programs, impact takes time. Consequently, early implementation reviews have one principal limitation--they generally can't measure impact. They are basically process reviews. Only in very rare situations will you be able to assess the impact of a new program--if there is a clear relationship between outputs and outcomes and if the scope and goals are straightforward and fairly modest. You can, however, identify barriers or problems that could prevent the program from having its intended effect over the long term.

A more subtle limitation is related to a very human trait: resistance to change. Much of the information for an early implementation review comes from people who are adjusting to their new role in a new process or program. Their perceptions and attitudes about the change may be more intense and less stable than they would regarding a longstanding, and more familiar, program. You'll want to be sensitive

to this and temper your observations with this understanding of human nature.

CHAPTER 8 EMERGING ISSUE REVIEWS

It is paradoxical that sometimes the most significant changes in a program's mission or responsibilities hit program managers almost completely without warning. Managers should be the first to know about emerging issues and trends likely to impact their program. Yet we often are so busy managing the responsibilities we already have, we don't watch for new problems on the horizon, problems that we could constructively address. Instead we wait for legislatures, advocacy groups, political appointees, and the media to make new policies and, in effect, tell us what to do. In this chapter we argue that such a wait-and-see approach is short-sighted and results in missed opportunities to contribute our talents, knowledge and creativity to finding solutions to emerging problems.

What Are They? Emerging issue reviews are evaluations that help you scan the environment to identify new trends that represent potential problems, challenges, or opportunities for your program. Think about how you would use evaluation if you worked in the private sector. In addition to assessing your clients' satisfaction with your product or service, you would test the marketplace to see how you could make your clients even happier. You would explore ways to expand your client base, your services, or your product line. You would continuously monitor the environment and its effects on your clients' willingness to use your products or services.

A strong analogy can be drawn between market research in the private sector and environmental scanning in public programs. Program managers can pursue policy and issue development purposely and reflectively, in the same way as managers in private enterprise. While the bottom-line motivation for the private sector is beating out the competition, the objectives of environmental scanning for both private and public managers are surprisingly similar. For example, you might "test the marketplace" to determine:

- ◆ If your customers' needs are changing. You may be able to offer a new product or service to increase your clients' support for your program.

In a review of the Head Start program, for example, we found that the children's families were increasingly suffering from social and medical problems (OEI, 1991). These problems resulted in higher incidence of child abuse and neglect, with predictable effects on the children. Local Head Start grantees had to expand their referrals and adjust their program content and structure to deal successfully with these children and their families.

- ◆ If your customer profile is changing, e.g., if your customers are older or younger, or more urban, suburban, or rural than they used to be. You could use this information to determine the effect on demand for your product or service and to anticipate future changes.
- ◆ If your customers, despite liking your product or service, are having concerns about its effect on their health or the environment. You might use this information to change your product or service to address customers' concerns.

Fear of this effect led to rapid improvement in quality control for public cholesterol screening (as we mentioned earlier in the chapter on compliance reviews). While customers liked the convenience of cholesterol screening in neighborhood settings, our evaluation of the screening found serious health risks. Realizing a loss of public confidence would translate into lost business, the industry went to great lengths to improve the safety of their services.

- ◆ If your advertisements communicate your message clearly to potential new customers. You should use this information to know if you are targeting your information to the right groups and communicating effectively with those audiences.

You may first become aware of emerging issues through news articles and professional publications related to your program; through contacts with clients, staff, or other people involved in your field; or through findings from other evaluations of your program, especially those conducted periodically. You may find it helpful to occasionally convene a small group of experts, policy analysts, program managers, or beneficiaries in an informal environment to talk about their problems and concerns. Keep your antennae up for any changes with the potential to impact your program. Seek out enough information about such issues to decide if you should evaluate the situation right away or if you can afford to just keep an eye on things for awhile.

How Are They Useful? The more you understand about the changes occurring in your operational and policy environments, the more influence you can have on the inevitable changes in your program. Emerging issue reviews are effective in three fundamental ways: they help you anticipate the need to expand your program in new areas or extend services to new groups, allowing you to take part in formulating new, related policies; they give you information you'll need to defend your program against popular but ill-conceived changes; and they give you an early warning of trends or events that could spell trouble for your current operations.

Depending on the scope of your program's mission, this type of evaluation may be essential on a daily basis or it may be relevant only once in awhile. However, very few programs would not benefit from developing the capacity to conduct emerging issue reviews. For example, constant environmental scanning is critical for the Public Health Service, which has widespread responsibility for improving the nation's health. On the other hand, such reviews, while highly useful, might be done more sporadically in connection with a county waste disposal program. Emerging issues for the waste disposal program could involve people's acceptance of recycling programs, problems associated with traditional methods of waste disposal, and changes in the public's consumption patterns.

How Are They Done? Emerging issue reviews involve charting how frequently a new event is happening or how quickly a change is

occurring, describing the characteristics of the trend, and understanding the environment in which the trend is evolving. Here's an example of some study questions:

- ◆ How frequently does an event happen? (*How often do households recycle?*)
- ◆ What does it look like when it does happen? (*What kinds of items are being recycled? What are the characteristics of households that recycle? How well do households prepare their items for recycling?*)
- ◆ In what ways are our policies affecting the event? (*Does providing curbside pick-up increase participation in recycling compared to providing central collection sites? Does door-to-door distribution of informational pamphlets increase the number of households that recycle?*)

Due to the great variation in types of emerging issues, decisions about the appropriate evaluation techniques to use depend to a very large extent on the nature of the issue. For example, in our own work, we examined several emerging issues related to adolescent and child health at the request of the Surgeon General. Her questions included:

- ◆ Is the use of spit tobacco among youth growing? What contributes to its popularity? Is there a need for stronger laws?
- ◆ Do children have access to cigarettes despite State and local laws against the sale of cigarettes to minors? If so, what changes in laws or enforcement techniques might be warranted?
- ◆ Is the number of children being born addicted to crack cocaine creating a burden for the country's child welfare system? If so, how can we minimize the strain on the system and still give these children the care they need?

We used different techniques to address each of these issues. In the study on spit tobacco, we advertised a toll-free line to receive

information from anonymous informants. In the study on cigarettes, we conducted a large survey of law enforcement personnel, students, and parents. In the study on babies addicted to cocaine, we purposely selected inner city hospitals and social service agencies to visit, where we assessed the situation in depth.

Assessing Incidence. In all three studies mentioned above, policy-makers were interested in knowing about incidence, or how widespread the problem was. However, we did not include a statistically projectable random sample in any of them. The techniques we did use proved successful for describing the characteristics of the problem and gave us enough information to establish that the problems were significant and growing. Had it been necessary, we could have followed up our work with surveys using valid samples to produce scientifically rigorous estimates of incidence. However, you may not need precise estimates of incidence to prove to yourself and others that you have a problem that needs to be addressed.

Using statistically valid surveys of a large population for these reviews can be challenging, or even dangerous, for two reasons. First, if you require information from people who are not familiar with your program, you may need to include a certain amount of education as part of your contact with them. For example, a State arts council with a goal of expanding its customer base may want to determine the cultural needs of a demographically changing community. This may require surveying a representative sample of residents, many of whom know nothing about the arts program, rather than those who in the past have supported council-sponsored events. Second, new trends commonly begin with particular areas or populations; the AIDS crisis is a notorious example. Where this appears to be the case, relying on incidence in the population as a whole might seriously underestimate the significance of the problem, and purposive sampling should be used. Also, be aware that the lower the expected incidence of an event, the larger a random sample needs to be to find it.

Describing Characteristics. Often in emerging issue reviews, you will want to know more about **how** an event occurs than about **how often**. This is especially true for trends that represent opportunities for

program improvement. Recently, we examined the trend toward school-based health services, where a growing number of public schools are providing students with primary health and counseling services. Many of these students have no other source of health care, so about 20 percent of the cost of the schools' services are covered under our department's programs. We wanted to understand the phenomenon of school-based health services--issues such as pros and cons, coordination with other service providers, and potential for filling gaps in service.

In cases like this, a variation of the case study approach often is effective. You might solicit from the field examples of the event under study and analyze the cases in some detail to understand and document how the phenomenon operates. You will want to determine its potential for use in other sites or if the circumstances surrounding this event are somehow unique. The concepts and techniques we discussed in the chapter on effective practice reviews also might be useful.

Understanding the Environment. To truly understand an emerging trend and predict its potential significance, you will need to understand the context in which it is occurring. Analyzing the environment can tell you a great deal about the trend: whether your program's law, regulations, or policies encourage or discourage the activity; how conducive the current environment is to the continuation of the trend; and whether your monitoring systems are able to detect a change in the level of the activity.

We've used this technique successfully in a number of studies and recommend it highly as an important component of any evaluation involving an emerging trend. A good example is our study of Social Security's policies relating to drug addicts and alcoholics receiving Supplemental Security Income (SSI). The number of substance-addicted SSI recipients has increased four-fold over the past few years. In connection with our review of Social Security's compliance with legal restrictions on these benefits, we analyzed the environment in an effort to find reasons for the increase in this recipient population.

An environmental analysis usually begins with a review of legal and other requirements that could have a bearing on the trend--federal and State legislation, local ordinances, regulations and administrative requirements. It assesses the extent to which there is confusion about, or flexibility in, the meaning or application of these rules. A knowledgeable attorney, perhaps in your office of general counsel, often can do this legal analysis quickly, or an experienced analyst can do it. In fact, we strongly suggest that you develop this capacity, since, as you've probably noted already, it's an important element of most of the evaluation strategies we've discussed.

An environmental analysis goes on to assess other relevant influences: budgets; mission statements; goals and objectives; social contracts; values; and institutional sympathies and alliances. All of these factors and many more can influence a given event. There are no hard and fast rules for which ones to explore, so your perceptiveness, your ability to ask the right questions, and your intuition will be your best guides. We will, however, caution you not to limit yourself to obvious and easily measurable influences, such as budgets, when more intangible factors, such as values, may be having a more pronounced effect on the trend you are studying.

Environmental influences form the basis for the debate over the shockingly low immunization rates for young children in the United States. Some researchers have pointed to the rising cost of vaccines as the primary cause. Other research has suggested that the principal factor is the failure of parents to understand the risks associated with not having their children immunized. It may be that budgetary influences are far less important in addressing this problem than the more abstract element of parental responsibility.

Limitations of Emerging Issue Reviews. Of all the types of studies we have discussed, emerging issue reviews may be the most subject to personal or organizational bias. By definition, these studies reveal new information, which we might not be ready to accept. Our perception of "the way things ought to be" can limit our ability to properly analyze a new phenomenon and may even hinder our inquiries and observations.

Since the issue is still emerging, the incidence of the phenomenon is likely to be small and difficult to detect in the population at large. It also may be limited to select groups, making the task of estimating how widespread it is quite difficult. On the other hand, knowing that a problem is significant or growing, along with understanding its nature and context, will be invaluable to you in meeting future challenges to your program, even in the absence of statistically precise data on incidence.

CHAPTER 9 WHAT'S NEXT?

By now, we hope we've convinced you of the benefits of developing or improving the evaluation capacity of your program. We'd like to remind you of the forces leading you in this direction: the Chief Financial Officers Act, the Government Performance and Results Act, the National Performance Review, and total quality management, all of which are designed to make government programs more accountable to the public they serve. We also hope that our discussion of different evaluation strategies has given you some practical ideas (and enthusiasm!) for putting evaluation to good use in your program.

Once you've made program evaluation an integral part of your management system, we believe you'll quickly see significant uses for the results, whether you are testifying before a legislative committee, meeting with your Chief Financial Officer, vying for program resources, striving to meet clients' needs and expectations, or planning for the future. At times, the payoff from your evaluation activities may seem small and you'll be tempted to put your money or staff elsewhere. Yet the benefits of evaluation are more than the findings of a single study. Since program evaluation is basically a process of asking and answering the right questions and using the results to achieve improvements, it builds on itself. It can energize your approach to the problems and opportunities you encounter. Program evaluation ultimately can change the entire culture of your organization as it frees people from rigid, short-sighted patterns of thinking and as it focuses everyone's attention on the real purpose for the program's existence.

We realize your next question probably is "What does it take to do this kind of evaluation?" No doubt you'd like to know what it costs, how it actually gets done, what kind of expertise is needed, and whether it's harder (or easier) than it sounds in this book. We can't give you a precise answer for your situation, since it depends on your own needs and preferences as well as the type of evaluation program you choose. We will give you answers from our own experience, and we encourage you to solicit the answers of other people who use or conduct evaluations. One way to get a good idea of what it would take

to incorporate evaluation into your program is to start with a single study. For example, you could include an early implementation review as part of your plans for a new initiative. Take note of the resources you use--contractor costs (if any), staff time, and any telephone, travel or printing expenses. In the past, we have completed evaluations that cost little more than six weeks' salary for one analyst.

Options. You have a number of options for acquiring an evaluation capacity. One approach is to draw on the expertise of professional evaluators in federal, State, or local government, possibly even within your own agency. You can buy the services of private-sector consultants on an ongoing or an as-needed basis. You can hire trained evaluators to become part of your staff. You can develop the talents of your own staff. In all likelihood, some combination will prove the most effective, and you'll need to find the best mix of program staff and professional evaluators. As you and your staff become more experienced in using evaluation, that mix probably will change. Even though our organization was created for the express purpose of conducting evaluations and most of our projects are done in-house, on occasion we seek help from public or private outside organizations.

Your involvement in evaluation can range from simply using the information from previous studies to doing it all yourself. The level that works best for you may fall somewhere on this list:

- ◆ Using the results of evaluations done by others;
- ◆ Contracting with experienced evaluators to address issues in your program;
- ◆ Planning your own projects and using outside evaluators to collect and analyze the data;
- ◆ Tackling evaluation projects in-house under the guidance of experienced evaluators; or
- ◆ Training and/or hiring staff to conduct evaluations so you can handle the entire process in house.

Regardless of your level of involvement, becoming a more informed "consumer" of evaluation will enhance your partnership with professional evaluators. Not only will your ability to contribute to the evaluation process help guarantee a worthwhile product that meets your needs, it will be appreciated by the evaluators with whom you work.

An outside evaluator might serve your particular program well, especially if the work costs you virtually nothing. Consider calling on your agency's or department's planning and research group, Inspector General's office, or quality assurance component. Many reputable for-profit organizations are available to work with you on evaluation projects if you decide to look outside government for this expertise. Although they are generally more costly than the alternatives, using their services can minimize disruption, may give you a greater measure of control than using other government groups, and could represent the best approach for some programs. Managers often find it useful to develop ongoing relationships with specific consultants, whose staff then become familiar with the program's culture, needs, and issues.

If you ultimately decide to develop an in-house evaluation capacity, be ready for a long-term commitment and a lot of work. You'll need to balance program demands against the benefits to be gained from evaluation activities. In the short term, you may lose productivity as you move people out of their current assignments (writing regulations, responding to inquiries, or serving clients) to learn evaluation skills. Another short-term sacrifice is the cost of continuous improvement: the discomfort that comes with disrupting current processes, upsetting the balance of power, and questioning the status quo. It takes time for people to adjust to an environment of change, even when that change is for the better. On the other hand, these sacrifices can be seen as investments. We believe that monitoring the performance of a program is so vital to its success, you may find the long-term benefits of building an evaluation capacity from within are well worth the costs.

There is another argument against in-house evaluation that's been a topic of discussion among professional evaluators for a long time. No one likes bad news. Sometimes this is especially true of organizations

hearing bad news from their own staff. Unfortunately some groups, to avoid taking responsibility for problems, may keep their evaluation staff busy on trivial issues or find some reason to dismiss troubling findings. You should carefully consider how your agency officials are likely to react to problems you bring to their attention. If you believe they are inclined to avoid confronting problems, you should probably use an outside source for evaluations. Findings and recommendations by outside evaluators are more likely to be viewed as objective and independent.

Despite the challenges, we believe the "grow-your-own" approach offers a number of advantages. First, your staff already know the program--the operation, the data systems, and the names of people to contact for more information. While outside evaluation organizations might bring a fresh perspective and less chance for bias or error, they may also bring higher costs in time and money while they learn how your program works and the background for the evaluation. In contrast, your own employees have the knowledge, insight, and experience to get to the heart of program issues and problems quickly. Second, working with your own staff puts you in an ideal position to ensure that the evaluation meets your needs by addressing the right issues, using reliable data sources, meeting your timetable, and arriving at valid conclusions. Your direct involvement increases the likelihood of getting worthwhile results you can use, so in a sense you can guarantee the project's success. Finally, performing evaluations fosters your staff's analytical skill and creativity, and you reward those skills every time you implement one of their recommendations for improvement. Empowering your employees in this way leads to the "culture change" and continuous improvement we noted earlier. Moreover, staff expertise will grow as their evaluation experience grows.

The next discussion describes our organization's process and structure in some detail. We include this information primarily to help you predict the relative costs and benefits of doing evaluation in your own program. You may want to incorporate some of these elements into your evaluation activities.

Personnel. Almost all our evaluations are conducted by a team of analysts and a project leader, with the assistance of other analysts who serve as program and technical specialists. These individuals come from a variety of backgrounds; few have earned degrees in program evaluation. Most hold liberal arts degrees and are skilled in critical thinking. Many also have held positions in the programs we review. After joining our organization, analysts receive training in evaluation planning, instrument design, data collection, statistical theory and application, report writing, and oral presentations. They move from entry-level to mid-level positions as they acquire this training and demonstrate key skills and abilities in their work. Eventually they take on the role of project leader or senior specialist. With each evaluation, team members learn new skills or refine existing skills, develop or build on relationships with key stakeholders, and deepen their understanding of important program issues.

We believe we can train a good analyst to become an effective evaluator in our organization if the individual is motivated and if we fulfill two basic conditions: we assign an experienced evaluator to work closely with a new analyst; and we provide the time and funds for training in evaluation concepts and techniques, both formally and on the job.

Most of our evaluators are generalists stationed across the country. Our headquarters includes branches with specialists in each of the Department's major program areas, plus a staff of technical support specialists in statistics and computer processing. The interaction between headquarters and regional staff gives us the advantage of multiple perspectives on the issues we study. This is an important factor in our success, since all our reviews are national in scope. Our regional project teams are quite small; we have two-, three-, and four-person teams. For some evaluations, we have used "mega-teams" comprising several teams. Some teams are convened for the purpose of conducting a specific evaluation; others are standing teams which move from one evaluation to the next (Molnar & Nudelman, 1993).

Costs. We believe in giving our staff the necessary tools to do their jobs well. As noted above, we value training. We allot 80 hours of

training every two years for each employee, and many people exceed that target, especially early in their careers. We also support the work of our employees by providing access to the latest technologies (fully-supported computer hardware and software, fax machines, etc.). Despite these substantial investments, the overall cost for an average evaluation is quite reasonable relative to the nature and scope of our work, due to the small size of our teams and the short turnaround times of most of our studies. For each evaluation we conduct, we spend about \$100,000, which includes salaries and benefits. (Our total payroll in 1993 was \$7.9 million in a total budget of \$8.8 million; other costs--for rent, equipment, travel, training, printing, etc.--made up the difference.) With this investment, we gave policy-makers in our department over 80 final reports carrying recommendations for program improvements and savings. On average, we return about \$180 to the taxpayer for every dollar we spend.

Process. Our evaluations are based on the steps we described in Chapter 2. This process and our structure allow the project team a great deal of independence, while engaging others inside and outside the organization in advisory roles.

At several milestones in an evaluation, the project team prepares a product for review by a program specialist, a statistical specialist, and managers in the region and headquarters. These products include the project design, the initial outline of findings and recommendations, and the draft and final reports. The team and all the reviewers meet, usually by telephone conference, for the purpose of reaching consensus on the team's product. After the meeting, the team revises its document to reflect the agreed-upon changes.

Our process also provides for collaboration with stakeholders outside our organization. We consult key staff in the area under review on an informal basis throughout the project. These contacts help us gain insight into program issues and stay informed on new developments. We also hold formal entrance and exit conferences where we inform staff in the program about our plans and findings. At the entrance conference, we invite input on the scope of the review, critical issues we should explore, available data sources, and any related

studies completed or in progress. The exit conference gives program staff a chance to review a preliminary version of the report and to share their concerns and observations about our findings and recommendations. In some cases, we brief senior staff. In almost all cases, we ask agency officials to comment on a formal draft report; their response is considered, and included verbatim, in producing the final report. The final report is available to any member of the public upon request.

To be effective, an evaluation process should specify avenues for collaboration and decision-making. Here's a list of questions you should answer as you go about designing your own evaluation process. We've included our answers for your information.

- ◆ How will potential study issues be identified? How will decisions be made about the issues to be examined? Who will be involved and who will make the final decision?

Study ideas can be suggested by anyone in our organization, and we encourage policy-makers to request specific studies in areas where they lack adequate information. Program specialists are responsible for preliminary research on study ideas and for developing work plan proposals, which are submitted for final decisions to a work-planning committee composed of managers.

- ◆ How will priorities be set?

We use a strategic planning process to identify areas where serious program vulnerabilities exist or where significant policies are being developed. Work plan proposals submitted in those areas get top consideration. We also ask our project teams to identify their preferred projects, and we use their level of interest as a further indication of a proposal's priority.

- ◆ How will external stakeholders be involved, either in the planning or the performance of evaluations?

We publish our work plan and circulate it in draft for our stakeholders to review in order to learn their perspectives and priorities. As noted earlier, we use entrance and exit conferences to discuss our study designs and results with stakeholders. We sometimes ask stakeholders to comment on a data collection instrument or analysis plan.

- ◆ At what stages will the project team generate a product? Who will review it?

We use key milestones in the project to involve a wide range of individuals within and outside the organization. Within the organization, a team of managers and technical specialists reviews a written product with the project team at each milestone.

- ◆ Are formal comments solicited at any point? From whom?

We solicit formal comments to the draft report from responsible components both within and outside the Department before producing the final report. Those comments are included in the final report.

- ◆ What kind of product results from the evaluation?

We publish final, public reports on our study results. We sometimes deliver oral briefings to Department officials, speak at conferences and at other forums involving outside organizations, and publish our results in professional journals.

- ◆ Will the project team make recommendations? To whom?

When our evaluation results indicate the need for specific improvements, the team makes recommendations to the responsible component.

- ◆ Is there a mechanism to ensure the information is used?

Yes. Our reports are signed out by the Inspector General to the head of the responsible component, who by law must respond to our recommendations and develop a corrective action plan. A formal conflict resolution process is available for handling disagreements.

Now that you have some ideas for evaluation projects and you've made some initial decisions about staffing, funding, and how your process will work, the next step is to make it happen. We sincerely hope we've given you the information you need to make constructive plans, gain the support of your colleagues and staff, and get started in practical program evaluation. We all wish you great success!

Bibliography

- Duquette, D., Stowe, A. (1992). *Entering the era of performance measurement reporting*. *Government Accounts Journal*. Summer 1992, 19-28.
- Fink, A. (1993). *Evaluation Fundamentals*. Newbury, CA:Sage.
- Guba, E.G., Lincoln, Y.S. (1982). *Effective Evaluation*. San Francisco. CA:Jossey-Bass.
- Hatry, H. (1981). *Practical Program Evaluation for State and Local Governments*. Washington, D.C.:The Urban Institute Press.
- Herman, J.L.(Ed.). (1987). *Program Evaluation Kit*. Newbury, CA:Sage.
- Krushat, W., Molnar, J. (1993). *Nonrespondent analysis*. *Evaluation Practice*. 14(1), 25- 34.
- Majchrzak, A. (1981). *Methods for Policy Research*. Newbury, CA:Sage.
- Mangano, M. (1990). *The inspections program within the U.S. Department of Health and Human Services Office of Inspector General. New Directions for Program Evaluation--Inspectors General: A New Force in Evaluation*. Number 48.
- Mangano, M. (1991). *Improving services through effective practice reviews*. *Evaluation Practice*. 12(3).
- Mangano, M. (1992). *A marketing approach to evaluation: four lessons for evaluators*

from the Honda motor company. *Evaluation and Program Planning*. 15(3).

Mangano, M. (1993). *Rapid, decision-oriented evaluation. The Public Manager*. 21(4).

Molnar, J., Nudelman, J. (1993). *How OEI uses teams. Federal Quality Forum*. 1(1), 9-13.

Molnar, J., Stup, B. (1994). *Using clients to monitor performance. Evaluation Practice*. 15(1), 29-35.

Office of Evaluation and Inspections. (1988a). *Patient Dumping After COBRA: U.S. Department of Health and Human Services Response to Complaints*. Washington, D.C.:Office of Inspector General, Department of Health and Human Services.

Office of Evaluation and Inspections. (1988b). *Impact of New Requirements: AFDC Standard Filing Unit*. Washington, D.C.:Office of Inspector General, Department of Health and Human Services.

Office of Evaluation and Inspections. (1989a). *Write to the Point*. Washington, D.C.:Office of Inspector General, Department of Health and Human Services.

Office of Evaluation and Inspections. (1989b). *Financial Arrangements Between Physicians and Health Care Businesses*. Washington, D.C.:Office of Inspector General, Department of Health and Human Services.

Office of Evaluation and Inspections. (1989c). *Medicare Drug Utilization Review*. Washington, D.C.:Office of Inspector General, Department of Health and Human Services.

Office of Evaluation and Inspections. (1990a). *Technical Assistance Guides for Conducting Program Evaluations and Inspections*. Washington, D.C.:Office of Inspector General, Department of Health and Human Services.

Office of Evaluation and Inspections. (1990b). *Public Cholesterol Screening*. Washington, D.C.:Office of Inspector General, Department of Health and Human Services.

Office of Evaluation and Inspections. (1990c). *Crack Babies*. Washington, D.C.:Office of Inspector General, Department of Health and Human Services.

Office of Evaluation and Inspections. (1990d). *Clinical Role of the Community Pharmacist*. Washington, D.C.:Office of Inspector General, Department of Health and Human Services.

Office of Evaluation and Inspections. (1990e). *Effective Paternity Establishment Practices*. Washington, D.C.:Office of Inspector General, Department of Health and Human Services.

Office of Evaluation and Inspections. (1990f). *Comprehensive Perinatal Care Programs*. Washington, D.C.:Office of Inspector General, Department of Health and Human Services.

Office of Evaluation and Inspections. (1990g). *Perspectives of Drug Manufacturers: Abbreviated New Drug Applications*. Washington, D.C.:Office of Inspector General, Department of Health and Human Services.

Office of Evaluation and Inspections. (1991). *Readiness to Expand Head Start*.

- Washington, D.C.:Office of Inspector General, Department of Health and Human Services.
- Office of Evaluation and Inspections. (1992a). *Early and Periodic Screening, Diagnosis and Treatment: Performance Measures*. Washington, D.C.:Office of Inspector General, Department of Health and Human Services.
- Office of Evaluation and Inspections. (1992b). *Minimizing Restraints in Nursing Homes*. Washington, D.C.:Office of Inspector General, Department of Health and Human Services.
- Office of Evaluation and Inspections. (1992c). *Hospital Closure: 1990*. Washington, D.C.:Office of Inspector General, Department of Health and Human Services.
- Office of Evaluation and Inspections. (1993a). *Medicare Beneficiary Satisfaction 1993*. Washington, D.C.:Office of Inspector General, Department of Health and Human Services.
- Office of Evaluation and Inspections. (1993b). *The National Medical Expenditure Survey*. Washington, D.C.:Office of Inspector General, Department of Health and Human Services.
- Office of Evaluation and Inspections. (1993c). *Evaluation of Head Start Expansion through Performance Indicators*. Washington, D.C.:Office of Inspector General, Department of Health and Human Services.
- Office of Evaluation and Inspections. (1993d). *Action Plan to Strengthen the Implementation of Title III of the Older Americans Act*. Washington, D.C.:Office of Inspector General, Department of Health and Human Services.
- Office of Evaluation and Inspections. (1993e). *Claimant Provision of Disability Information*. Washington, D.C.:Office of Inspector General, Department of Health and Human Services.
- Office of Evaluation and Inspections. (1994). *Social Security Client Satisfaction: Overview*. Washington, D.C.:Office of Inspector General, Department of Health and Human Services.
- Patton, M.Q. (1984). *Practical Evaluation*. Newbury, CA:Sage.
- Patton, M.Q. (1986). *Utilization-Focused Evaluation*. Newbury, CA:Sage.
- President's Council on Integrity and Efficiency. (1993). *Quality Standards for Inspections*. Washington, D.C.
- Thompson, P. (1989). *The (evaluation) role of the inspector general. Evaluation Practice*. 10(3).
- Thompson, P., Yessian, M. (1992). Policy analysis in the Office of Inspector General. In Carol Weiss(Ed.) *Organizations for policy analysis: helping government think*. Newbury, CA:Sage.
- U.S. General Accounting Office. (1982). *Content Analysis*. Washington, D.C.
- U.S. General Accounting Office. (1984). *Designing Evaluations*. Washington, D.C.
- U.S. General Accounting Office. (1985). *Using Structured Interview Techniques*. Washington, D.C.

U.S. General Accounting Office. (1986). *Developing and Using Questionnaires*. Washington, D.C.

Rossi, P.H., Freeman, H.E. (1993). *Evaluation: A Systematic Approach*. Newbury, CA:Sage.

Wholey, J.S. (1983). *Evaluation And Effective Public Management*. Boston, MA:Little Brown.

OFFICE OF INSPECTOR GENERAL
June Gibbs Brown, Inspector General
Michael F. Mangano, Principal Deputy Inspector General

The mission of the Office of Inspector General (OIG), as mandated by Public Law 95-452, as amended, is to protect the integrity of the Department of Health and Human Services' (HHS) programs as well as the health and welfare of beneficiaries served by those programs. This statutory mission is carried out through a nationwide network of audits, investigations, and evaluations. The OIG also informs the Secretary of HHS of program and management problems and recommends courses to correct them.

OFFICE OF EVALUATION AND INSPECTIONS
George F. Grob, Deputy Inspector General for
Evaluation and Inspections

The OIG's Office of Evaluation and Inspections (OEI) conducts short-term management and program evaluations (called inspections) that focus on issues of concern to the Department, the Congress, and the public. The findings and recommendations contained in these inspection reports generate rapid, accurate, and up-to-date information on the efficiency, vulnerability, and effectiveness of departmental programs.

Numerous individuals reviewed earlier versions of this volume and provided comments. The authors thank them all.

For additional copies of this volume, or to request copies of OEI reports and documents referenced in this volume, call (202) 691-0480, or write:

Office of Inspector General
Office of Evaluation and Inspections
330 Independence Avenue, SW
Room 5660
Washington, DC 20201
<http://www.hhs.gov/progorg/oei/>