Chapter 10

Evaluation Research and Problem Analysis

In this chapter, our attention centers on applied criminal justice research. Evaluation studies are conducted to learn whether (and why) programs have succeeded or failed. Problem analysis helps officials plan their actions and anticipate the possible effects of new programs.

Introduction 255

Topics Appropriate for Evaluation Research and Problem Analysis 255

The Policy Process 256

Linking the Process to Evaluation 257

Getting Started 260

Evaluability Assessment 260

Problem Formulation 261

Measurement 263

Designs for Program Evaluation 266

Randomized Evaluation Designs 266

Home Detention: Two Randomized Studies 269

Quasi-Experimental Designs 271

Other Types of Evaluation Studies 273

Problem Analysis and Scientific Realism 273

Problem-Oriented Policing 274

Auto Theft in Chula Vista 275

Other Applications of Problem Analysis 276

Space- and Time-Based Analysis 276

Scientific Realism and Applied Research 280

The Political Context of Applied Research 282

Evaluation and Stakeholders 282

When Politics Accommodates Facts 283

Politics and Objectivity 284

Introduction

Evaluation research and problem analysis are increasingly important activities for researchers and public officials alike.

Evaluation research—sometimes called *program evaluation*—refers to a research purpose rather than a specific research method. Its special purpose is to evaluate the effects of policies such as mandatory arrest for domestic violence, innovations in probation, and new sentencing laws. Another type of evaluation study, **problem analysis**, helps public officials plan and select alternative actions. Virtually all types of designs, measures, and data collection techniques can be used in evaluation research and problem analysis.

Growth of evaluation research over the last several years no doubt reflects desire on the part of criminal justice researchers to actually make a difference in the world. At the same time, we cannot discount the influence of two additional factors: (1) increased federal requirements for program evaluations to accompany the implementation of new programs and (2) the availability of research funds to meet that requirement.

By the same token, increased interest in program evaluation and problem analysis has followed heightened concern for the accountability of public officials and public policy. Criminal justice agencies are expected to justify the effectiveness and cost of their actions. If traditional approaches to probation supervision, for example, do not deter future lawbreaking, new approaches should be developed and their effectiveness assessed. Or if using temporary detention facilities fabricated from recycled semitrailers is less costly than constructing new jails, public officials should consider whether the lower-cost alternative will meet their needs for pretrial detention and short-term incarceration.

Justice agencies have come to rely more on **evidence-based policy**, in which the actions of justice agencies are linked to evidence used for planning and evaluation. Traditional practices are being reevaluated against evidence provided by social science research. The *Problem-Oriented Guides* series summarizes evidence concerning responses by police and others to problems ranging from *Acquaintance Rape of College Students* (Sampson 2002) to *Witness Intimidation* (Johnson 2006). CompStat and its variations base police actions on evidence about the location and circumstances of crime problems. Corrections policies are increasingly evaluated to sort out those that do in fact reduce reoffending (Cullen and Sundt 2003). This trend represents an expansion of applied research that moves beyond collaborations between justice professionals and professional researchers.

Topics Appropriate for Evaluation Research and Problem Analysis

Problem analysis and evaluation are used to develop justice policy and determine its impact.

Evaluation research is appropriate whenever some policy intervention occurs or is planned. A policy intervention is an action taken for the purpose of producing some intended result. In its simplest sense, evaluation research is a

process of determining whether the intended result was produced. Problem analysis focuses more on deciding what intervention should be pursued. Given alternative courses of action, which is likely to be least costly, most effective, or least difficult to implement? Our focus, of course, is on the analysis and evaluation of criminal justice policy and criminal justice agencies. However, it will be useful to first consider a simple general model of the policy-making process in order to understand various topics appropriate to evaluation and problem analysis.

The Policy Process

Figure 10.1 presents our model, adapted from Robert Lineberry's (1977, 42–43) classic summary of a policy system. A similar type of input–output model is described in a National Institute of Justice publication on evaluation guidelines (McDonald and Smith 1989). Although we will examine each step in turn, recognize that the policy process, like the research process generally (see Chapter 3), is fluid and does not always start at the beginning and conclude at the end.

The policy process begins with a demand that normally appears as support for a new course of action or opposition to existing policy. Such demands can emerge from within a public organization or from outside sources. Newspaper stories alleging racial discrimination in drug sentencing can generate demand for revised sentencing policies or a prosecutor may independently decide to review all sentence recommendations made by deputies who prosecute drug cases. Before any action can be taken, demands must find a place on the policy agenda.

The next step, as shown in Figure 10.1, actually encompasses several steps. Policy makers consider ultimate goals they wish to accomplish and different means of achieving those goals. Does our prosecutor seek absolute equality in sentences recommended for all white and African American drug defendants, or should there be ranges of permissible variation based on criminal history, severity of charges, and so on? Resources must be allocated from available inputs, including personnel, equipment, supplies, and even time. Who will review sentence recommendations? How much time will that take, and will additional staff be required? Because the word *policy* implies some standard course of action about how to respond to a recurring problem or issue, routine practices and decision rules must be formulated. Will sentence recommendations for each case be reviewed as they are prepared, or is it sufficient to review all cases on a weekly basis?

Policy outputs refer to what is actually produced, in much the same manner that a manufacturer of office supplies produces paper clips and staples. In our hypothetical example, the prosecutor's policy produces the routine review of sentence recommendations in drug cases. Or, to consider a different example, a selective traffic enforcement program intended to reduce auto accidents on a particular roadway may produce a visible police presence, together with traffic citations for speeding.

In the final stage, we consider the impact of policy outputs. Does the prosecutor's review process actually eliminate disparities in sentences? Are auto accidents reduced in the targeted enforcement area?

The distinction between policy outputs and their impacts is important for understanding applications of evaluation to different stages of the policy process. Unfortunately, this difference is often confusing to both public officials and researchers. *Impacts* are fundamentally related to policy goals; they refer to the basic question of what a policy action is trying to achieve. *Outputs* embody the means to achieve desired policy goals. The prosecutor seeks to achieve equality in sentence recommendations (impact), so a review process is produced as a means to achieve that goal (output). Or a police executive allocates officers, patrol cars, and overtime pay to produce traffic citations (outputs) in the expectation that citations will achieve the goal of reducing auto accidents (impact).

Now consider the left side of Figure 10.1. Our policy model can be expressed as a simple cause-and-effect process such as we considered

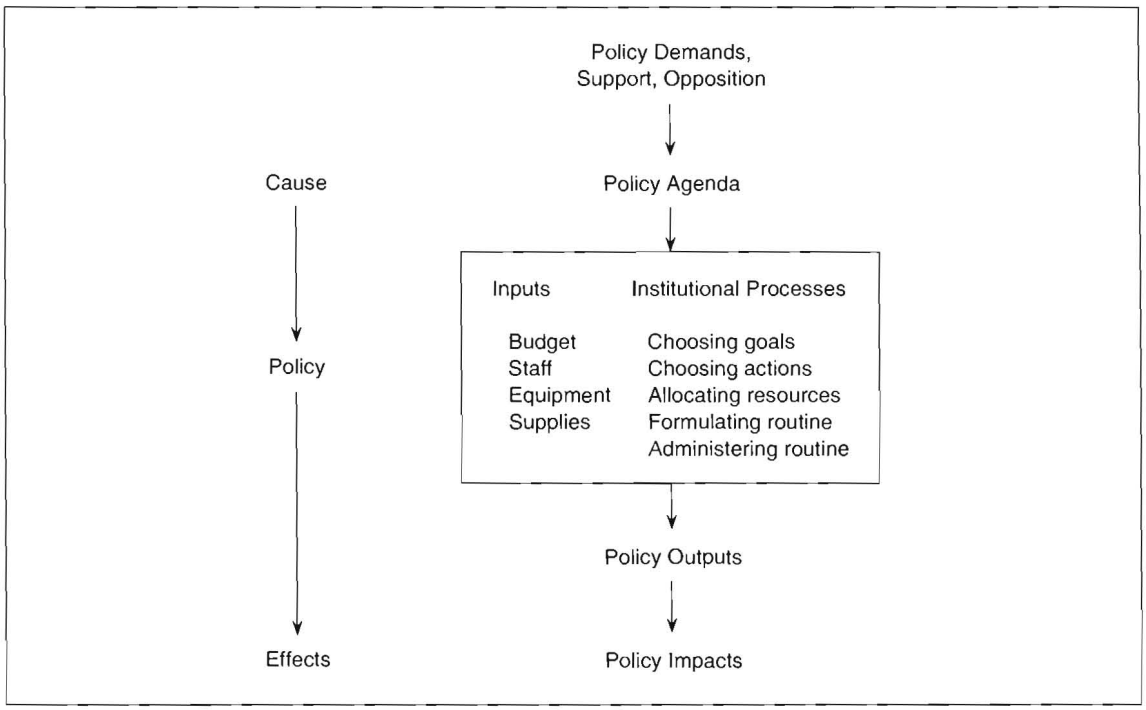


Figure 10.1 The Policy Process
Source: Adapted from Lineberry (1977, 42–43).

in earlier chapters. A cause has produced the variation in sentences for African American and white defendants, or a cause has produced a concentration of auto accidents. Policies are formulated to produce an effect, or impact. In this sense, a policy can be viewed as a hypothesis in which an independent variable is expected to produce change in a dependent variable. Sentence review procedures are expected to produce a reduction in sentence disparities; targeted enforcement is expected to produce a reduction in auto accidents. Goal-directed public policies may therefore be viewed as if-then statements: if some policy action is taken, then we expect some result to be produced.

Linking the Process to Evaluation

By comparing this simple model with a general definition of program evaluation given in one of the most widely used texts on the subject (Rossi, Freeman, and Lipsey 1999), the topics appropriate to applied research will become clearer. Peter Rossi and associates (1999, 4; emphasis in original) define program evaluation as

> *the use of social science research procedures to systematically assess the effectiveness of social intervention programs.* More specifically, evaluation researchers (evaluators) use social research methods to study, appraise, and help improve social programs in all their aspects, including the diagnosis of the social problems they address, their conceptualization and design, their implementation and administration, their outcomes, and their efficiency.

We have been discussing systematic social scientific research procedures throughout this book. Now let's substitute *criminal justice* for *social programs* and see how this definition and Figure 10.1 help us understand program evaluation applications.

Problem Analysis Activities listed under "Institutional Processes" in Figure 10.1 refer to conceptualization and design. For example, faced with a court order to maintain prison populations within established capacity, corrections officials might begin by conceiving and designing different ways to achieve this demand. Problem analysis is an example of a social scientific research procedure that can help corrections officials evaluate alternative actions, choose among them, and formulate routine practices for implementing policy to comply with a court order.

One approach might be to increase rated capacity through new construction or conversion of existing facilities. Another might be to devise a program to immediately reduce the existing population. Still another might be to cut back on the admission of newly sentenced offenders. A more general goal that would certainly be considered is the need to protect public safety. Each goal implies different types of actions, together with different types and levels of resources, that would be considered within constraints implied by the need to protect public safety. If officials from other organizations—prosecutors, judges, or state legislators—were involved in conceptualization and design, then additional goals, constraints, and policies might be considered.

Increasing capacity by building more prisons would be the most costly approach in financial terms, but it might also be viewed as the most certain way to protect public safety. Early release of current inmates would be cheaper and faster than building new facilities, but this goal implies other decisions, such as how persons would be selected and whether they would be released to parole or to halfway houses. Each of these alternatives requires some organizational capacity to choose inmates for release, place them in halfway houses, or supervise compliance with parole. Refusing new admissions would be least costly. Political support must be considered for each possible approach. Each alternative—spending money on new construction, accepting responsibility for early release, or tacitly passing the problem on to jails that must house inmates refused admission to state facilities—requires different types of political influence or courage.

Many other topics in criminal justice research are appropriate for problem analysis. Police departments use such techniques to help determine the boundaries of patrol beats and the allocation of other resources. In most large cities, analysts examine the concentration of calls for service in terms of space and time and consider how street layout and obstacles might facilitate or impede patrol car mobility.

A growing number of law enforcement agencies are using computerized crime maps to detect emerging patterns in crime and develop appropriate responses. Producing computer-generated maps that display reported crimes within days after they have occurred is one of the most important policy planning tools for the New York City Police Department (Silverman 1999). Other departments have taken advantage of funding and technical assistance made available by federal funding to enhance mapping and other crime analysis capabilities (Boba 2005).

Program Evaluation Problem analysis takes place in the earlier stages of the policy process. In contrast, program evaluation studies are conducted in later stages and seek answers to two types of questions: (1) Are policies being implemented as planned? (2) Are policies achieving their intended goals? Evaluation, therefore, seeks to link the intended actions and goals of criminal justice policy to empirical evidence that policies are being carried out as planned and are having the desired effects. These two types of questions correspond to two related types of program evaluations: process evaluation and impact assessment. Returning to our example of policies to reduce prison population, we will first consider impact assessment and then process evaluation.

Let's assume that corrections department policy analysts select an early-release program to reduce the population of one large institution.

Inmates who have less than 120 days remaining on their sentence and who were committed for nonviolent offenses will be considered for early release. Further assume that of those inmates selected for early release, some will be assigned to parole officers, and some will serve their remaining sentence in halfway houses—working at jobs during the week but spending evenings and weekends in a community-based facility.

The program has two general goals: (1) to reduce prison population to the court-imposed ceiling and (2) to protect public safety. Whereas the first goal is fairly straightforward, the second is uncomfortably vague. What do we mean by "protecting public safety"? For now, let's say we will conclude that the program is successful in this regard if, after six months, persons in the two early-release conditions have aggregate rates of arrest for new offenses equal to or less than a comparison group of inmates released after completing their sentences.

Our **impact assessment** would examine data on the prison population before and after the new program was implemented, together with arrest records for the two types of early releases and a comparison group. We might obtain something like the hypothetical results shown in Table 10.1.

Did the program meet its two goals? Your initial reaction might be that it did not, but Table 10.1 presents some interesting findings. The prison population certainly was reduced, but it did not reach the court-imposed cap of 1,350. Those released to halfway houses had lower arrest rates than others, but persons placed on early parole had higher arrest rates. Averaging arrest rates for all three groups shows that the total figure is about the same as that for persons released early. Notice also that almost twice as many people were released to early parole as were placed in halfway houses.

The impact assessment results in Table 10.1 would have been easier to interpret if we had conducted a **process evaluation**. A process evaluation focuses on program outputs, as represented in Figure 10.1, seeking answers to the question of whether the program was implemented as intended. If we had conducted a process evaluation of this early-release program, we might have discovered that something was amiss in the selection process. Two pieces of evidence in Table 10.1 suggest that one of the selection biases we considered in Chapter 5, "creaming," might be at work in this program. Recall that creaming is the natural tendency of public officials to choose experimental subjects least likely to fail. In this case, selectivity is indicated by the failure of the early-release program to meet its target number, the relatively small number of persons placed in halfway houses, and the lower re-arrest rates for these persons. A process evaluation would have monitored selection procedures and probably revealed evidence of excessive caution on the part of corrections officials in releasing offenders to halfway houses.

Ideally, impact assessments and process evaluations are conducted together. Our example illustrates the important general point that process evaluations make impact assessments more interpretable. In other cases, process evaluations may be conducted when an impact assessment is not possible. To better understand how process evaluations and impact assessments complemenr each other, let's now look more closely at how evaluations are conducted.

Table 10.1 Hypothetical Results of Early-Prison-Release Impact Assessment

	Percent New Arrests After 6 Months	Number of Persons
Normal release	26%	142
Early release to halfway houses	17	25
Early parole	33	46
Subtotal early release	27	71
Total	26	213

Note: Preprogram population = 1,578; actual population after implementation = 1,402; court-imposed population cap = 1,350.

Getting Started

Learning policy goals is a key first step in doing evaluation research.

Several steps are involved in planning any type of research project. This is especially true in applied studies, for which even more planning may be required. In evaluating a prison early release program, we need to think about design, measurement, sampling, data collection procedures, analysis, and so on. We also have to address such practical problems as obtaining access to people, information, and data needed in an evaluation.

In one sense, however, evaluation research differs slightly in the way research questions are developed and specified. Recall that we equated program evaluation with hypothesis testing; policies are equivalent to if-then statements postulating that an intervention will have the desired impact. Preliminary versions of research questions, therefore, will already have been formulated for many types of evaluations. Problem analysis usually considers a limited range of alternative choices; process evaluations focus on whether programs are carried out according to plans; and impact assessments evaluate whether specified goals are attained.

This is not to say that evaluation research is a straightforward business of using social scientific methods to answer specific questions that are clearly stated by criminal justice officials. It is often difficult to express policy goals in the form of if-then statements that are empirically testable. Another problem is the presence of conflicting goals. Many issues in criminal justice are complex, involving different organizations and people. And different organizations and people may have different goals that make it difficult to define specific evaluation questions. Perhaps most common and problematic are vague goals. Language describing criminal justice programs may optimistically state goals of enhancing public safety by reducing recidivism without clearly specifying what is meant by that objective.

In most cases, researchers have to help criminal justice officials formulate testable goals, something that is not always possible. Other obstacles may interfere with researchers' access to important information. Because of these and similar problems, evaluation researchers must first address the question of whether to evaluate at all.

Evaluability Assessment

An evaluability assessment is described by Rossi and associates (1999, 157) as sort of a "preevaluation," in which a researcher determines whether conditions necessary for conducting an evaluation are present. One obvious condition is support for the study from organizations delivering program components that will be evaluated. The word *evaluation* may be threatening to public officials, who fear that their own job performance is being rated. Even if officials do not feel personally threatened by an impact assessment or other applied study, evaluation research can disrupt routine agency operations. Ensuring agency cooperation and support is therefore an important part of evaluability assessment. Even if no overt opposition exists, officials may be ambivalent about evaluation.

This might be the case, for example, if an evaluation is required as a condition of launching a new program. This and other steps in evaluability assessment may be accomplished by scouting a program and interviewing key personnel (Rossi, Freeman, and Lipsey 1999, 135). The focus in scouting and interviewing should be on obtaining preliminary answers to questions that eventually will have to be answered in more detail as part of an evaluation. What are general program goals and more specific objectives? How are these goals translated into program components? What kinds of records and data are readily available? Who will be the primary consumers of evaluation results? Do other persons or organizations have a direct or indirect stake in the program? Figure 10.2 presents a partial menu of questions that can guide information gathering for the evaluability assessment and later stages.

1. Goals
   a. What is the program intended to accomplish?
   b. How do staff determine how well they have attained their goals?
   c. What formal goals and objectives have been identified?
   d. Which goals or objectives are most important?
   e. What measures of performance are currently used?
   f. Are adequate measures available, or must they be developed as part of the evaluation?
2. Clients
   a. Who is served by the program?
   b. How do they come to participate?
   c. Do they differ in systematic ways from nonparticipants?
3. Organization and Operation
   a. Where are the services provided?
   b. Are there important differences among sites?
   c. Who provides the services?
   d. What individuals or groups oppose the program or have been critical of it in the past?
4. History
   a. How long has the program been operating?
   b. How did the program come about?
   c. Has the program grown or diminished in size and influence?
   d. Have any significant changes occurred in the program recently?

Figure 10.2 Evaluation Questions
Source: Adapted from Stecher and Davis (1987, 58–59).

The answers to these and similar questions should be used to prepare a program description. Although official program descriptions may be available, evaluation researchers should always prepare their own description, one that reflects their own understanding of program goals, elements, and operations. Official documents may present incomplete descriptions or ones intended for use by program staff, not evaluators. Even more importantly, official program documents often do not contain usable statements about program goals. As we will see, formulating goal statements that are empirically testable is one of the most important components of evaluation research.

Douglas McDonald and Christine Smith (1989, 1) describe slightly different types of questions to be addressed by criminal justice officials and evaluators in deciding whether to evaluate state-level drug control programs:

> How central is the project to the state's strategy?
> How costly is it relative to others?
> Are the project's objectives such that progress toward meeting them is difficult to estimate accurately with existing monitoring procedures?

Such questions are related to setting both program and evaluation priorities. On the one hand, if a project is not central to drug control strategies or if existing information can help determine project effectiveness, then an evaluation should probably not be conducted. On the other hand, costly projects that are key elements in antidrug efforts should be evaluated so that resources can be devoted to new programs if existing approaches are found to be ineffective.

Problem Formulation

We mentioned that evaluation research questions may be defined for you. This is true in a general sense, but formulating applied research problems that can be empirically evaluated is an important and often difficult step. Evaluation research is a matter of finding out whether something is or is not there, whether something did or did not happen. To conduct evaluation research, we must be able to operationalize, observe, and recognize the presence or absence of what is under study.

This process normally begins by identifying and specifying program goals. The difficulty of this task, according to Rossi and associates (1999, 167), revolves around the fact that formal statements of goals are often abstract statements about ideal outcomes. Here are some examples of goal statements paraphrased from actual program descriptions:

- Equip individuals with life skills to succeed (a state-level shock incarceration program; MacKenzie, Shaw, and Gowdy 1993).
- Provide a safe school environment conducive to learning (a school resource officer program; Johnson 1999).

- Encourage participants to accept the philosophy and principles of drug-free living (an urban drug court; Finn and Newlyn 1993).
- Provide a mechanism that engages local citizens and community resources in the problem-solving process (a probation-police community corrections program; Wooten and Hoelter 1998).

Each statement expresses a general program objective that must be clarified before we can formulate research questions to be tested empirically. We can get some idea of what the first example means, but this goal statement raises several questions. The objective is for individuals to succeed, but succeed at what? What is meant by "life skills"—literacy, job training, time management, self-discipline? We might also ask whether the program focuses on outputs (equipping people with skills) or on impacts (promoting success among people who are equipped with the skills). On the one hand, an evaluation of program outputs might assess individual learning of skills, without considering whether the skills enhance chances for success. On the other hand, an evaluation of program impacts might obtain measures of success such as stable employment or not being arrested within some specified time period.

In all fairness, these goal statements are taken somewhat out of context; source documents expand on program goals in more detail. They are, however, typical of stated goals or initial responses we might get to the question, "What are the goals of this program?" Researchers require more specific statements of program objectives.

Wesley Skogan (1985) cautions that official goal statements frequently oversell what a program might realistically be expected to accomplish. It's natural for public officials to be positive or optimistic in stating goals, and overselling may be evident in goal statements. Another reason officials and researchers embrace overly optimistic goals is that they fail to develop a micromodel of the program production process (Weiss 1995). That is, they do not adequately consider just how some specified intervention will work. Referring to Figure 10.1, we can see that developing a micromodel can be an important tool for laying out program goals and understanding how institutional processes are structured to achieve those goals. Skogan (1985, 38; emphasis in original) describes a micromodel as

> part of what is meant by a "theory-driven" evaluation. Researchers and program personnel should together consider just how each element of a program should affect its targets. If there is not a good reason why "X" *should* cause "Y" the evaluation is probably not going to find that it did! Micromodeling is another good reason for monitoring the actual implementation of programs.

A micromodel can also reveal another problem that sometimes emerges in applied studies: inconsistent goals. Michael Maxfield and Terry Baumer (1992) evaluated a pretrial home detention program in which persons awaiting trial for certain types of offenses were released from jail and placed on home detention with electronic monitoring. Five different criminal justice organizations played roles in implementation or had stakes in the program. The county sheriff's department (1) faced pressure to reduce its jail population. Under encouragement from the county prosecutor (2), the pretrial release program was established. Criminal court judges (3) had the ultimate authority to release defendants to home detention, following recommendations by bail commissioners in a county criminal justice services agency (4). Finally, a community corrections department (5) was responsible for actually monitoring persons released to home detention.

Maxfield and Baumer interviewed persons in each of these organizations and discovered that different agencies had different goals. The sheriff's department was eager to release as many people as possible to free up jail space for convicted offenders and pretrial defendants

who faced more serious charges. Community corrections staff, charged with the task of monitoring pretrial clients, were more cautious and sought only persons who presented a lower risk of absconding or committing more offenses while on home detention. The county prosecutor viewed home detention as a way to exercise more control over some individuals who would otherwise be released under less restrictive conditions. Some judges refused to release people on home detention, whereas others followed prosecutors' recommendations. Finally, bail commissioners viewed pretrial home detention as a form of jail resource management, adding to the menu of existing pretrial dispositions (jail, bail, or release on recognizance).

The different organizations involved in the pretrial release program comprised multiple **stakeholders**—persons and organizations with a direct interest in the program. Each stakeholder had different goals for and different views on how the program should actually operate—who should be considered for pretrial home detention, how they should be monitored, and what actions should be taken against those who violated various program rules. After laying out these goals and considering different measures of program performance, Maxfield and Baumer (1992, 331) developed a micromodel of home detention indicating that electronic monitoring is suitable for only a small fraction of defendants awaiting trial.

Clearly specifying program goals, then, is a fundamental first step in conducting evaluation studies. If officials are not certain about what a program is expected to achieve, it is not possible to determine whether goals are reached. If multiple stakeholders embrace different goals, evaluators must specify different ways to assess those goals. Maxfield (2001) describes a number of different approaches to specifying clear goals, a crucial first step in the evaluation process.

Measurement

After we identify program goals, our attention turns to measurement, considering first how to measure a program's success in meeting goals. Obtaining evaluable statements of program goals is conceptually similar to the measurement process, in which program objectives represent conceptual definitions of what a program is trying to accomplish.

Specifying Outcomes If a criminal justice program is intended to accomplish something, we must be able to measure that something. If we want to reduce fear of crime, we need to be able to measure fear of crime. If we want to increase consistency in sentences for drug offenses, we need to be able to measure that. Notice, however, that although outcome measures are derived from goals, they are not the same as goals. Program goals represent desired outcomes, whereas outcome measures are empirical indicators of whether those desired outcomes are achieved. Furthermore, if a program pursues multiple goals, then researchers may have to either devise multiple outcome measures or select a subset of possible measures to correspond with a subset of goals.

Keeping in mind our program-as-hypothesis simile, outcome measures correspond to dependent variables—the Y in a simple $X \rightarrow Y$ causal hypothesis. Because we have already considered what's involved in developing measures for dependent variables, we can describe how to formulate outcome measures. Pinning down program goals and objectives results in a conceptual definition. We then specify an operational definition by describing empirical indicators of program outcomes.

In our earlier example, Maxfield and Baumer (1992) translated the disparate interests of organizations involved in pretrial home detention into three more specific objectives: (1) ensure appearance at trial, (2) protect public safety, and (3) relieve jail crowding. These objectives led to corresponding outcome measures: (1) failure-to-appear rates for persons released to pretrial home detention, (2) arrests while on home detention, and (3) estimates of the number of jail beds made available, computed by multiplying

the number of persons on pretrial home detention by the number of days each person served on the program. Table 10.2 summarizes the goals, objectives, and measures defined by Maxfield and Baumer.

Measuring Program Contexts Measuring the dependent variables directly involved in an impact assessment is only a beginning. As Ray Pawson and Nick Tilley (1997, 69) point out, it is usually necessary to measure the context within which the program is conducted. These variables may appear to be external to the experiment itself, yet they still affect it.

Consider, for example, an evaluation of a job-skills training program coupled with early prison release to a halfway house. The primary outcome measure might be participants' success at gaining employment after completing the program. We will, of course, observe and calculate the subjects' employment rates. We should also be attentive to what has happened to the employment and unemployment rates of the community and state where the program is located. A general slump in the job market should be taken into account in assessing what might otherwise seem to be a low employment rate for subjects. Or if all the experimental subjects get jobs following the program, that might result more from a general increase in available jobs than from the value of the program itself.

There is no magic formula or set of guidelines for selecting measures of program context, any more than there is for choosing control variables in some other type of research. Just as we read what other researchers have found with respect to some topic we are interested in—say, explanatory research—we should also learn about the production process for some criminal justice program before conducting an evaluation.

Measuring Program Delivery In addition to making measurements relevant to the outcomes of a program, it is necessary to measure the program intervention—the experimental stimulus or independent variable. In some cases, this measurement will be handled by assigning subjects to experimental and control groups, if that's the research design. Assigning a person to the experimental group is the same as scoring that person "yes" on the intervention, and assignment to the control group represents a score of "no." In practice, however, it's seldom that simple.

Table 10.2 Pretrial Home Detention with Electronic Monitoring: Goals, Objectives, and Measures

Actor/Organization	Goals
Sheriff	Release jail inmates
Prosecutor	Increase supervision of pretrial defendants
Judges	Protect public safety
Bail commission	Provide better jail resource management
Community corrections	Monitor defendant compliance
	Return violators to jail
Objectives	**Measures**
Ensure court appearance	Failure-to-appear counts
Protect public safety	Arrests while on program
Relieve jail crowding	*N* defendants × days served

Source: Adapted from Farrington and associates (1993, 108).

Let's continue with the job-training example. Some inmates will participate in the program through early release; others will not. But imagine for a moment what job-training programs are actually like. Some subjects will participate fully; others might miss sessions or fool around when they are present. So we may need measures of the extent or quality of participation in the program. And if the program is effective, we should find that those who participated fully have higher employment rates than those who participated less.

Other factors may further confound the administration of the experimental stimulus. Suppose we are evaluating a new form of counseling

designed to cure drug addiction. Several counselors administer it to subjects composing an experimental group. We can compare the recovery rate of the experimental group with that of a control group (a group that received some other type of counseling or none at all). It might be useful to include the names of the counselors who treat specific subjects in the experimental group, because some may be more effective than others. If that turns out to be the case, we must find out why the treatment works better for some counselors than for others. What we learn will further elaborate our understanding of the therapy itself.

Obtaining measures that reflect actual delivery of the experimental intervention is very important for many types of evaluation designs. Variation in the levels of treatment delivered by a program can be a major threat to the validity of even randomized evaluation studies. Put another way, uncontrolled variation in treatment is equivalent to unreliable measurement of the independent variable.

Specifying Other Variables It is usually necessary to measure the population of subjects involved in the program being evaluated. In particular, it is important to define those for whom the program is appropriate. In evaluation studies, such persons are referred to as the program's *target population*. If we are evaluating a program that combines more intensive probation supervision with periodic urine testing for drug use, it's probably appropriate for convicted persons who are chronic users of illegal drugs, but how should we define and measure chronic drug use more specifically? The job-skills training program mentioned previously is probably appropriate for inmates who have poor employment histories, but a more specific definition of employment history is needed.

This process of definition and measurement has two aspects. First, the program target population must be specified. This is usually done in a manner similar to the process of defining program goals. Drawing on questions like those in Figure 10.2, evaluators consult program officials to identify the intended targets or beneficiaries of a particular program. Because the hypothetical urine-testing program is combined with probation, its target population will include persons who might receive suspended sentences with probation. However, offenders convicted of crimes that carry nonsuspendable sentences will not be in the target population. Prosecutors and other participants may specify additional limits to the target population—employment or no previous record of probation violations, for example.

Most evaluation studies that use individual people as units of analysis also measure such background variables as age, gender, educational attainment, employment history, and prior criminal record. Such measures are made to determine whether experimental programs work best for males, those older than 25, high school graduates, persons with fewer prior arrests, and so forth.

Second, in providing for the measurement of these different kinds of variables, we need to choose whether to create new measures or use ones already collected in the course of normal program operation. If our study addresses something that's not routinely measured, the choice is easy. More commonly, at least some of the measures we are interested in will be represented in agency records in some form or other. We then have to decide whether agency measures are adequate for our evaluation purposes.

Because we are talking about measurement here, our decision to use our own measures or those produced by agencies should, of course, be based on an assessment of measurement reliability and validity. If we are evaluating the program that combined intensive probation with urinalysis, we will have more confidence in the reliability and validity of basic demographic information recorded by court personnel than in court records of drug use. In this case, we might want to obtain self-report measures of drug use and crime commission from subjects themselves, rather than relying on official records.

By now, it should be abundantly clear that measurement must be taken very seriously in evaluation research. Evaluation researchers must carefully determine all the variables to be measured and obtain appropriate measures for each. However, such decisions are typically not purely scientific ones. Evaluation researchers often must work out their measurement strategy with the people responsible for the program being evaluated.

Designs for Program Evaluation

Designs used in basic research are readily adapted for use in evaluation research.

Chapter 5 introduced a variety of experimental and other designs that researchers use in studying criminal justice. Recall that randomly assigning research subjects to experimental or control groups controls for many threats to internal validity. Here our attention turns specifically to the use of different designs in program evaluation.

Randomized Evaluation Designs

To illustrate the advantages of random assignment, consider this dialogue from Lawrence Sherman's book *Policing Domestic Violence: Experiments and Dilemmas* (1992b, 67):

> When the Minneapolis domestic violence experiment was in its final planning stage, some police officers asked: "Why does it have to be a randomized experiment? Why can't you just follow up the people we arrest anyway, and compare their future violence risks to the people we don't arrest?"
>
> Since this question reveals the heart of the logic of controlled experiments, I said, "I'm glad you asked. What kind of people do you arrest now?"
>
> "Assholes," they replied. "People who commit aggravated POPO."
>
> "What is aggravated POPO?" I asked.
>
> "Pissing off a police officer," they answered. "Contempt of cop. But we also arrest people who look like they're going to be violent, or who have caused more serious injuries."
>
> "What kind of people do you not arrest for misdemeanor domestic assault?" I continued.
>
> "People who act calm and polite, who lost their temper but managed to get control of themselves," came the answer.
>
> "And which kinds of people do you think would have higher risks of repeat violence in the future?" I returned.
>
> "The ones we arrest," they said, the light dawning.
>
> "But does that mean arrest caused them to become more violent?" I pressed.
>
> "Of course not—we arrested them because they were more trouble in the first place," they agreed.
>
> "So just following up the ones you arrest anyway wouldn't tell us anything about the effects of arrest, would it?" was my final question.
>
> "Guess not," they agreed. And they went on to perform the experiment.

Sherman's dialogue portrays the obvious problems of selection bias in routine police procedures for handling domestic violence. In fact, one of the most important benefits of randomization is to avoid the selectivity that is such a fundamental part of criminal justice decision making. Police selectively arrest people, prosecutors selectively file charges, judges and juries selectively convict defendants, and offenders are selectively punished. In a more general sense, randomization is the great equalizer: through probability theory, we can assume that groups created by random assignment will be statistically equivalent.

Randomized designs are not suitable for evaluating all experimental criminal justice programs. Certain requirements of randomized studies mean that this design cannot be used

in many situations. A review of those requirements illustrates many of the limits of randomized designs for applied studies.

Program and Agency Acceptance Random assignment of people to receive some especially desirable or punitive treatment may not be possible for legal, ethical, and practical reasons. We discussed ethics and legal issues in Chapter 2. Sometimes practical obstacles may also be traced to a misunderstanding of the meaning of random assignment. It is crucial that public officials understand why randomization is desirable and that they fully endorse the procedure.

Richard Berk and associates (2003) describe how researchers obtained cooperation for an evaluation of a new inmate classification system in the California Department of Corrections (CDC) by appealing to the needs of agency managers. Preliminary research suggested that the experimental classification system would increase inmate and staff safety at lower cost than classification procedures then in use. In addition:

> Plans for the study were thoroughly reviewed by stakeholders, including CDC administrators, representatives of prison employee bargaining unions, . . . California State legislative offices, and a wide variety of other interested parties. There was widespread agreement that the study was worth doing. (Berk et al. 2003, 211)

At the same time, justice agencies have expanding needs for evaluations of smaller-scale programs. John Eck (2002) explains how designs that are less elaborate are more likely to be accepted by public agencies.

Minimization of Exceptions to Random Assignment In any real-world delivery of alternative programs or treatments to victims, offenders, or criminal justice agency staff, exceptions to random assignment are all but inevitable. In a series of experiments on police responses to domestic violence, officers responded to incidents in one of three ways, according to a random assignment procedure (Sherman et al. 1992). The experimental treatment was arrest; control treatments included simply separating parties to the dispute or attempting to advise and mediate. Although patrol officers and police administrators accepted the random procedure, exceptions were made as warranted in individual cases, subject to an officer's discretionary judgment.

As the number of exceptions to random assignment increases, however, the statistical equivalence of experimental and control groups is threatened. When police (or others) make exceptions to random assignment, they are introducing bias into the selection of experimental and control groups. Randomized experiments are best suited for programs in which such exceptions can be minimized. The prison classification study by Berk and associates offers a good example. Random assignment was automatic—inmates having odd identification numbers at intake were assigned to the treatment group, while those having even numbers were in the control group. This procedure produced treatment and control groups that were virtually identical in size: 9,662 in treatment and 9,656 controls (2003, 224–225).

Adequate Case Flow for Sample Size In Chapter 6, we examined the relationship between sample size and accuracy in estimating population characteristics. As sample size increases (up to a point), estimates of population means and standard errors become more precise. By the same token, the number of subjects in groups created through random assignment is related to the researcher's ability to detect significant differences in outcome measures between groups. If each group has only a small number of subjects, statistical tests can detect only very large program effects or differences in outcome measures between the two groups. This is a problem with statistical conclusion validity and sample size, as we discussed in Chapters 5 and 6.

Case flow represents the process through which subjects are accumulated in experimental and control groups. In Sherman's domestic violence evaluations, cases flowed into experimental and control groups as domestic violence incidents were reported to police. Evaluations of other types of programs will generate cases through other processes—for example, offenders sentenced by a court or inmates released from a correctional facility.

If relatively few cases flow through some process and thereby become eligible for random assignment, it will take longer to obtain sufficient numbers of cases. The longer it takes to accumulate cases, the longer it will take to conduct an experiment and the longer experimental conditions must be maintained. Imagine filling the gas tank of your car with a small cup: it would take a long time, it would test your patience, and you would probably tarnish the paint with spilled gasoline as the ordeal dragged on. In a similar fashion, an inadequate flow of cases into experimental groups risks contaminating the experiment through other problems. Getting information about case flow in the planning stages of an evaluation is a good way to diagnose possible problems with numbers of subjects.

Maintaining Treatment Integrity Treatment integrity refers to whether an experimental intervention is delivered as intended. Sometimes called *treatment consistency*, treatment integrity is therefore roughly equivalent to measurement reliability. Experimental designs in applied studies often suffer from problems related to treatment inconsistencies. If serving time in jail is the experimental treatment in a program designed to test different approaches to sentencing drunk drivers, treatment integrity will be threatened if some defendants are sentenced to a weekend in jail while others serve 30 days or longer.

Criminal justice programs can vary considerably in the amount of treatment applied to different subjects in experimental groups. For example, Gottfredson and associates (2003) acknowledge that the drug-court treatment in Baltimore County was unevenly implemented. Only about half of those assigned to the experimental group received certified drug treatment. In contrast, the classification system tested by Berk and associates was a relatively simple treatment that was readily standardized. There was no danger of treatment dilution as was the case in the drug-court experiment.

Midstream changes in experimental programs can also threaten treatment integrity. Rossi and associates (1999, 297) point out that the possibility of midstream changes means that randomized designs are usually not appropriate for evaluating programs in early stages of development, when such changes are more likely. For example, assume we are evaluating an intensive supervision probation program with randomized experimental and control groups. Midway through the experiment, program staff decide to require weekly urinalysis for everyone in the experimental group (those assigned to intensive supervision). If we detect differences in outcome measures between the experimental and control groups (say, arrests within a year after release), we will not know how much of the difference is due to intensive supervision and how much might be due to the midstream change of adding urine tests.

Summing Up the Limits of Randomized Designs Randomized experiments therefore require that certain conditions be met. Staff responsible for program delivery must accept random assignment and further agree to minimize exceptions to randomization. Case flow must be adequate to produce enough subjects in each group so that statistical tests will be able to detect significant differences in outcome measures. Finally, experimental interventions must be consistently applied to treatment groups and withheld from control groups.

These conditions, and the problems that may result if they are not met, can be summarized as two overriding concerns in field experiments: (1) equivalence between experimental and control groups before an intervention, and

(2) the ability to detect differences in outcome measures after an intervention is introduced. If there are too many exceptions to random assignment, experimental and control groups may not be equivalent. If there are too few cases, or inconsistencies in administering a treatment, or treatment spillovers to control subjects, outcome measures may be affected in such a way that researchers cannot detect the effects of an intervention.

Let's now look at an example that illustrates both the strengths of random experiments and constraints on their use in criminal justice program evaluations.

Home Detention: Two Randomized Studies

Terry Baumer and Robert Mendelsohn conducted two random experiments to evaluate programs that combine home detention with electronic monitoring (ELMO). In earlier chapters, we examined how different features of these studies illustrated measurement principles; here our focus is on the mechanics of random assignment and program delivery.

In their first study, Baumer and Mendelsohn evaluated a program that targeted adult offenders convicted of nonviolent misdemeanor and minor felony offenses (Baumer and Mendelsohn 1990; also summarized in Baumer, Maxfield, and Mendelsohn 1993). The goal of the program was to provide supervision of offenders that was more enhanced than traditional probation but less restrictive and less costly than incarceration. Several measures of outcomes and program delivery were examined, as we have described in earlier chapters.

Baumer and Mendelsohn selected a randomized posttest-only design, in which the target population was offenders sentenced to probation. Subjects were randomly assigned to an experimental group in which the treatment was electronically monitored home detention or to a control group sentenced to home detention without electronic monitoring. Figure 10.3 summarizes case flow into the evaluation experiment. After a guilty plea or trial conviction, probation office staff reviewed offenders' backgrounds and criminal records for the purpose of recommending an appropriate sentence. The next step was a hearing, at which sentences were imposed by a criminal court judge.

Persons sentenced to probation were eligible for inclusion in the experiment. Their case files were forwarded to staff in the community corrections agency responsible for administering the home detention programs. On receiving an eligible case file, community corrections staff telephoned the evaluation researchers, who, having prepared a random list of case numbers, assigned subjects to either the treatment or control group. Subject to two constraints, this process produced 78 treatment subjects and 76 control subjects.

Thinking back on our consideration of ethics in Chapter 2, you should be able to think of one constraint: informed consent. Researchers and program staff explained the evaluation project to subjects and obtained their consent to participate in the experiment. Those who declined to participate in the evaluation study could nevertheless be assigned to home detention as a condition of their probation. The second constraint was made necessary by the technology of ELMO: subjects could not be kept in the treatment group if they did not have a telephone that could be connected to the ELMO equipment.

Notice that random assignment was made after sentencing. Baumer and Mendelsohn began their evaluation by randomizing subjects between stages 2 and 3 in Figure 10.3. This produced problems because judges occasionally overruled presentence investigation recommendations to probation, thus overriding random assignment. After detecting this problem, Baumer and Mendelsohn (1990, 27–29) moved randomization downstream, so that judicial decisions could not contaminate the selection process.

Baumer and Mendelsohn (1990, 26) obtained agreement from community corrections staff, prosecutors, and judges to use random

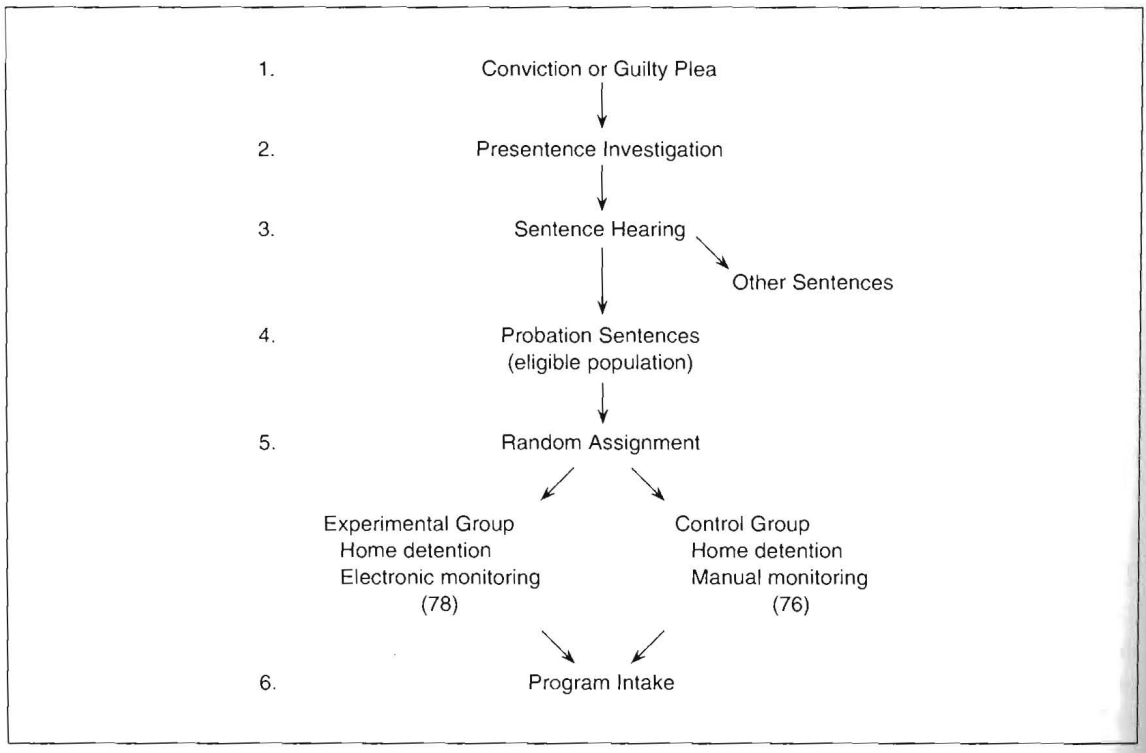


Figure 10.3 Home Detention for Convicted Adults: Case Flow and Random Assignment

assignment by getting all parties to accept an assumption of "no difference":

> That is, in the absence of convincing evidence to the contrary, they were willing to assume that there was no difference between the ... methods of monitoring. This allowed the prosecutor to negotiate and judges to assign home detention as a condition of probation only, while permitting the community corrections agency to make the monitoring decision.

Convinced of the importance of random assignment, the community corrections agency delegated to researchers the responsibility for making the monitoring decision, a "decision" that was randomized.

In this example, the experimental condition—ELMO—was readily distinguished from the control condition, home detention without ELMO. There was no possibility of treatment spillover; control subjects could not unintentionally receive ELMO because they had neither the bracelet nor the home-base unit that embodied the treatment. ELMO could therefore be readily delivered to subjects in the experimental group and withheld from control subjects.

The second ELMO evaluation conducted by Baumer and Mendelsohn reveals how program delivery problems can undermine the strengths of random assignment (Baumer, Maxfield, and Mendelsohn 1993). In their study of juvenile burglars, they used similar procedures for randomization, but eligible subjects were placed in one of four groups, as illustrated in this table:

	Electronic Monitoring?	
Police Visits?	**No**	**Yes**
No	C	E1
Yes	E2	E3

Juvenile burglars could be randomly assigned to three possible treatments: ELMO only (E1), police visits to their home after school only (E2), or ELMO and police visits (E3). Subjects in the control group (C) were sentenced to home detention only. As in the adult study, outcome measures included arrests after release.

Although there were no problems with random assignment, inconsistencies in the delivery of each of the two experimental treatments produced uninterpretable results (Maxfield and Baumer 1991, 5):

> Observations of day-to-day program operations revealed that, compared with the adult program, the juvenile court and cooperating agencies paid less attention to delivering program elements and using information from . . . the electronic monitoring equipment. Staff were less well trained in operating the electronic monitoring equipment, and police visits were inconsistent.

The box titled "Home Detention" in Chapter 1 elaborates on differences in the operation of these two programs and a third ELMO program for pretrial defendants. However, the lesson from these studies bears repeating here: *randomization does not control for variation in treatment integrity and program delivery.*

Randomized experiments can be powerful tools in criminal justice program evaluations. However, it is often impossible to maintain the desired level of control over experimental conditions. This is especially true for complex interventions that may change while an evaluation is underway. Experimental conditions are also difficult to maintain when different organizations work together in delivering some service—a community-based drug treatment provider coupled with intensive probation, for example.

Largely because of such problems, evaluation researchers are increasingly turning to other types of designs that are less fragile—less subject to problems if rigorous experimental conditions cannot be maintained.

Quasi-Experimental Designs

Quasi-experiments are distinguished from true experiments by the lack of random assignment of subjects to an experimental and a control group. Random assignment of subjects is often impossible in criminal justice evaluations. Rather than forgo evaluation altogether in such instances, it is usually possible to create and execute research designs that will permit evaluation of the program in question.

Quasi-experiments may also be nested into experimental designs as backups should one or more of the requisites for a true experiment break down. For example, Shaddish, Cook, and Campbell (2002) describe how time-series designs can be nested into a series of random experiments. In the event that case flow is inadequate or random assignment to enhanced or standard counseling regimens breaks down, the nested time-series design will salvage a quasi-experiment.

We considered different classes of quasi-experimental designs—nonequivalent groups, cohorts, and time series—in Chapter 5, together with examples of each type. Each of these designs has been used extensively in criminal justice evaluation research.

Nonequivalent-Groups Designs As we saw in Chapter 5, quasi-experimental designs lack the built-in controls for selection bias and other threats to internal validity. Nonequivalent-groups designs, by definition, cannot be assumed to include treatment and comparison subjects who are statistically equivalent. For this reason, quasi-experimental program evaluations must be carefully designed and analyzed to rule out possible validity problems.

For evaluation designs that use nonequivalent groups, attention should be devoted to constructing experimental and comparison groups that are as similar as possible on important variables that might account for differences in outcome measures. Rossi and associates (1999) caution that procedures for constructing such groups should be grounded in a theoretical

understanding of what individual and group characteristics might confound evaluation results. In a study of recidivism by participants in shock incarceration programs, we certainly want to ensure that equal numbers of men and women are included in groups assigned to shock incarceration and groups that received another sentence. Alternatively, we can restrict our analysis of program effects to only men or only women.

One common reason for using nonequivalent-groups designs is that some experimental interventions are intended to affect all persons in a larger unit—a neighborhood crime prevention program, for example. It may not be possible to randomly assign some neighborhoods to receive the intervention while withholding it from others. And we are usually unable to control which individuals in a neighborhood are exposed to the intervention.

Different types of quasi-experimental designs can be used in such cases. In a Kansas City program to reduce gun violence, police targeted extra patrols at gun crime hot spots (Sherman, Shaw, and Rogan 1995). Some beats were assigned to receive the extra patrols, while comparison beats—selected for their similar frequency of gun crimes—did not get the special patrols. Several outcome measures were compared for the two types of areas. After 29 weeks, gun seizures in the target area increased by more than 65 percent and gun crimes dropped by 49 percent. There were no significant changes in either gun crimes or gun seizures in the comparison beat. Drive-by shootings dropped from 7 to 1 in the target area and increased from 6 to 12 in the comparison area. Homicides declined in the target area but not in the comparison area. Citizen surveys showed less fear of crime and more positive feelings about the neighborhood in the target area than in the comparison area.

Time-Series Designs Interrupted time-series designs require attention to certain validity threats because researchers cannot normally control how reliably the experimental treatment is implemented. Foremost among these issues are instrumentation, history, and construct validity. In many interrupted time-series designs, conclusions about whether an intervention produced change in an outcome measure rely on simple indicators that represent complex causal processes.

In their evaluation of legislation to provide for mandatory minimum sentences in Oregon, Nancy Merritt and associates (Merritt, Fain, and Turner 2006) examined changes in sentences for different types of offenses. They found that sentences for offenses clearly covered by the law did in fact increase in the first five years after its passage. However, they also found declines in the number of cases filed that were clearly included in the mandatory provisions. Meanwhile, more charges were filed for offenses covered by discretionary provisions. Of course, criminal case prosecution and sentencing are complex processes. The authors could not directly control for different circumstances surrounding cases processed before and after the law took effect. However, their time series analysis does clearly show changes in case filings, suggesting that prosecutors exercised discretion to evade the mandatory provisions of Oregon's legislation.

Understanding the causal process that produces measures used in time-series analysis is crucial for interpreting results. Such understanding can come in two related ways. First, we should have a sound conceptual grasp of the underlying causal forces at work in the process we are interested in. Second, we should understand how the indicators used in any time-series analysis are produced.

Patricia Mayhew and associates (Mayhew, Clarke, and Elliott 1989) concluded that laws requiring motorcycle riders to wear helmets produced a reduction in motorcycle theft. This might seem puzzling until we consider the causal constructs involved in stealing motorcycles. Assuming that most motorcycle thefts are crimes of opportunity, Mayhew and associates argue that few impulsive thieves stroll about carrying helmets. Even thieves are sufficiently

rational to recognize that a helmetless motorcycle rider will be unacceptably conspicuous—an insight that deters them from stealing motorcycles. Mayhew and colleagues considered displacement as an alternative explanation for the decline in motorcycle theft, but they found no evidence that declines in motorcycle theft were accompanied by increases in either stolen cars or bicycles. By systematically thinking through the causal process of motorcycle theft, Mayhew and associates were able to conclude that helmet laws were unintentionally effective in reducing theft.

Other Types of Evaluation Studies

Earlier in this chapter, we noted how process evaluations are distinct from impact assessments. Whereas the latter seek answers to questions about program effects, process evaluations monitor program implementation, asking whether programs are being delivered as intended.

Process evaluations can be invaluable aids in interpreting results from an impact assessment. We described how Baumer and Mendelsohn were better able to understand outcome measures in their evaluation of ELMO for juvenile burglars because they had monitored program delivery. Similarly, process evaluations were key elements of CCTV evaluations reported by Gill and Spriggs (2005). They were able to describe whether cameras were placed and monitored as intended. In many cases camera placement was modified, something the authors suggest was related to the relative success of different CCTV installations. Without a process evaluation, information about program implementation cannot be linked to outcome measures.

Process evaluations can also be useful for criminal justice officials whose responsibility centers more on the performance of particular tasks than on the overall success of some program. For example, police patrol officers are collectively responsible for public safety in their beat, but their routine actions focus more on performing specific tasks such as responding to a call for service or, in community policing, diagnosing the concerns of neighborhood residents. Police supervisors are attentive to traffic tickets written, arrests made, and complaints against individual officers. Probation and parole officers are, of course, interested in the ultimate performance of their clients, but they are also task-oriented in their use of records to keep track of client contacts, attendance at substance abuse sessions, or job performance. Process evaluations center on measures of task performance—on the assumption that tasks are linked to program outcomes. So process evaluations can be valuable in their own right, as well as important for diagnosing measures of program effects.

Problem Analysis and Scientific Realism

Problem analysis, coupled with scientific realism, helps public officials use research to select and assess alternative courses of action.

Program evaluation differs from problem analysis with respect to the time dimension and where each activity takes place in the policy process. Problem analysis is used to help design alternative courses of action and choose among them.

In reality, there is not much of a difference between these two types of applied research. Similar types of research methods are used to address problem analysis questions (What would happen? What should we do?) as are brought to bear on program evaluation questions (What did happen? What have we done?). Consider, for example, a definition of a similar approach, policy analysis, from a prominent text: "Attempting to bring modern science and technology to bear on society's problems, policy analysis searches for feasible courses of action, generating information and marshalling evidence of the benefits and other consequences that would follow their adoption and implementation" (Quade 1989, 4). Except for

the form of the verb ("would follow"), this is not too different from the way we defined program evaluation.

Results from program evaluations are frequently considered in choosing among future courses of action. Problem analysis and policy analysis depend just as much on clearly specifying goals and objectives as does program evaluation. And the achievement of goals and objectives worked out through problem analysis can be tested through program evaluation. Measurement is also a fundamental concern in both types of applied studies.

Problem-Oriented Policing

More than an alternative approach to law enforcement, the core of problem-oriented policing is applying problem analysis methods to public safety problems. Problem-oriented policing depends on identifying problems, planning and taking appropriate action, then assessing whether those actions achieved intended results.

This approach centers on problems, not individual incidents. For example, traditional policing responds to reports of auto thefts, writing up details about the incident to support an insurance claim, then moving on to the next incident. Let's consider this *incident-oriented policing*. In contrast, problem-oriented policing would begin by analyzing a number of auto theft reports. Reports would be examined for similarities, such as where and when they occurred, types of autos stolen, whether stolen cars were eventually recovered, and if so in what condition. Such analysis would define a more general problem of auto theft. Subsequent steps would consider what kinds of actions might be taken to address the problem.

Problem solving is a fundamental tool in problem-oriented policing. As initially defined by Ronald Clarke and John Eck, problem solving involves four analytic steps:

> (1) carefully defined specific problems ... ; (2) conduct in-depth analysis to understand their causes; (3) undertake broad searches for solutions to remove these causes and bring about lasting reductions in problems; (4) evaluate how successful these activities have been. (2005, step 7-1)

You can easily see how problem-solving merges the application of problem analysis and evaluation (assessment) of the effects of interventions.

Problem-oriented policing is an especially useful example of applied research because a large number of resources are available. We'll briefly describe three types of such resources: how-to-do-it guides, problem and response guides, and case studies. Most of the first two categories have been prepared with support from the Community Oriented Policing Services (COPS) office in the U.S. Department of Justice. Resources are available at the Center for Problem-Oriented Policing website: www.popcenter.org (accessed May 16, 2008).

How-to-Do-It Guides Ronald Clarke and John Eck (2005) have prepared a general guide to crime analysis to support problem-oriented policing. Adapted from a document originally prepared for the Jill Dando Institute of Crime Science in London, this publication offers succinct guidance on analysis and reporting results. The COPS office has also sponsored guides that provide more detail on different problem analysis tools: assessment and evaluation (Eck 2003a); understanding the process of repeat victimization (Weisel 2005); conducting background research on problems (Clarke and Schultze 2005); interviewing offenders (Decker 2005); and collaborating with private-sector interests to solve problems (Chamard 2006).

Crime mapping and other methods of space-based analysis are important tools in problem-oriented policing. *GIS and Crime Mapping* by Spencer Chainey and Jerry Ratcliffe (2005) is an excellent general guide. John Eck and associates (2005) focus on the use of mapping to identify crime hot spots.

Problem and Response Guides In an earlier chapter, we mentioned that justice agencies frequently adopt programs that appear to have

been successful in other jurisdictions. While this can sometimes be advisable, a key principle of problem-oriented policing is to base local actions on an understanding of local problems. Instead of trying an off-the-shelf program or so-called "best practice," appropriate interventions should be considered only after analyzing the data.

This principle is evident in two series of guides that describe what is known about effective responses based on past experience. *Problem guides* describe how to analyze very specific types of problems (for example, "Financial Crimes Against the Elderly"), and what are known to be effective or ineffective responses. Response guides describe very general kinds of actions that might be undertaken to address different types of problems (for example, "Video Surveillance of Public Places").

Case Studies and Other Research One of the hallmarks of applied research is to use research to change practice. The two groups of guides discussed so far were prepared for use by criminal justice professionals, but they were developed following many years of research. Many examples of research that contributed to changes in justice policy have been published in the series *Crime Prevention Studies*. We now turn to an example that illustrates the application of problem analysis, as well as other research principles presented in this and earlier chapters.

Auto Theft in Chula Vista

Chula Vista is a medium-sized city of just under 200,000 residents, bordered by the Pacific Ocean on the west, and sandwiched by San Diego on the north and southwest; the city is about seven miles north of the U.S.-Mexico border. Nanci Plouffe and Rana Sampson (2004) began their analysis of vehicle theft by comparing Chula Vista to other southern California cities. After noting that theft rates tended to increase for cities closer to the border, they began to disaggregate the problem by searching for areas where vehicle thefts and break-ins were concentrated. Deborah Weisel (2003) refers to this as "parsing," or breaking down a large-area measure to examine smaller areas.

Plouffe and Sampson first determined that 10 parking lots accounted for 25 percent of thefts and 20 percent of break-ins in the city. Furthermore, 6 of those 10 lots were also among the top 10 calls-for-service locations in Chula Vista. This meant that auto-theft hot spots also tended to be hot spots for other kinds of incidents. Continuing their analysis, the analysts found some notable patterns:

- Recovery rates for stolen cars and trucks were lower in Chula Vista than in areas to the north.
- Recovery rates in 4 of the 10 hot parking lots were especially low, under 40 percent.
- Smaller pick-up trucks and older Toyota Camrys had even lower recovery rates.
- High-risk lots were close to roads that led to the Mexico border.

Together these findings suggested many cars stolen from the high-risk areas were being driven into Mexico.

Plouffe and Sampson next moved beyond using existing data from police records. This is again consistent with the methods of problem analysis: use existing data to identify problems and their general features, then collect additional data to better understand the mechanisms of problems. For Plouffe and Sampson that meant conducting environmental surveys of high-risk parking lots, observing operations and interviewing officials at U.S.-Mexico border crossings, and interviewing a small number of individuals arrested for auto theft from target lots. They sought to understand why particular lots were targeted and whether stolen cars could be easily driven into Mexico.

We described environmental surveys in Chapter 8. In conducting theirs, Plouffe and Sampson discovered that the highest-risk lot was a two-minute drive from vehicle entry points into Mexico. The lot served a midrange general shopping mall with typical open parking. Access was easy and thieves could expect that vehicles would be unguarded for some time.

Information gathered from the border crossing confirmed that few cars entering Mexico were stopped, and vehicle identification documents were rarely requested.

In-person interviews with auto thieves used a 93-item questionnaire, asking about target selection, techniques, and other routines. Thieves preferred older cars because they could be easily stolen—steering column locks wear out and can be broken with simple tools. They watched people entering stores, judged that their vehicle would be unguarded for a time, then drove the few minutes into Mexico. Cars were rarely stolen from parking garages because thieves would have to produce a ticket in order to exit.

With this and other information, Plouffe and Sampson discussed strategies with Chula Vista police and security staff at parking lots and shopping malls. More diligent screening at the border was rejected, largely because most vehicles had been driven into Mexico before the theft was even discovered. They recommended that high-risk shopping malls install gates at entrance and exit points for parking lots. Drivers would take a ticket upon entering and would have to produce it when leaving. This, it was argued, would substantially increase the effort required to steal vehicles from parking lots near the border.

Other Applications of Problem Analysis

Partly because it has proved helpful in law enforcement applications, problem analysis is being adopted by other criminal justice agencies. Veronica Coleman and associates describe how local and federal prosecutors in several U.S. cities have formed planning teams to identify crime problems and develop appropriate interventions. Teams include U.S. attorneys, researchers, and other criminal justice professionals who pursue a form of problem analysis labeled Strategic Approaches to Community Safety Initiatives (SACSI). SACSI involves five steps, four of which should look familiar (Coleman et al., 1999, 18):

1. Form an interagency working group.
2. Gather information and data about a local crime problem.
3. Design a strategic intervention to tackle the problem.
4. Implement the intervention.
5. Assess and modify the strategy as the data reveal effects.

We have only scratched the surface of problem analysis applications in criminal justice. This is an area of applied research that is growing daily. Other examples draw on methods of systems analysis, operations research, and economics for such purposes as cost-benefit studies, police patrol allocation, and decisions about hiring probation officers. Cost-benefit analysis, in particular, is used to assess the relative value and expense of alternative policies. Although the mathematical tools that form the basis of problem analysis can be sophisticated, the underlying logic is relatively simple. For example, police departments traditionally used pin maps to represent the spatial and temporal concentrations of reported crime.

Space- and Time-Based Analysis

Pin maps are examples of "low-tech" problem analysis that are nonetheless conceptually identical to computer models of hot spots used in many departments to plan police deployment. Growing numbers of justice agencies, especially police and sheriff's departments, have taken advantage of rapid advances in computing and telecommunications. Computerized mapping systems now permit police to monitor changes in crime patterns on a daily or hourly basis and to develop responses accordingly. Furthermore, simultaneous advances in computing power and declines in the cost of that power make it possible for even small agencies to use mapping tools (Harries 1999). The ongoing technological advances in mapping have fueled the application of statistical models to geographic

clusters of crime problems. Thomas Rich (1999) describes this as *analytic mapping*, whereby statistical tools supplement the "eyeballing" approach to locating concentrations of crime.

Crime maps usually represent at least four different things: (1) one or more crime types, (2) a space or area, (3) some time period, and (4) some dimension of land use, usually streets. The most useful crime maps will show patterns that can help analysts and police decide what sort of action to take. That's part of applied research. An example will illustrate some basic features of crime maps.

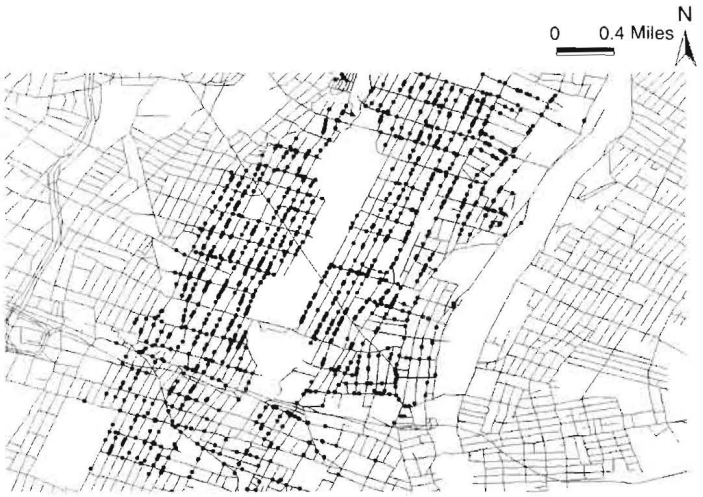
Figure 10.4 shows four crime maps prepared by Shuryo Fujita, a graduate student at the Rutgers University School of Criminal Justice, for a midsized city in the northeast United States. All four maps show completed auto theft, but for different areas and time periods. The map in panel A shows auto thefts for the year 2005 in one of four police precincts in the city. About 1,750 completed thefts are represented, about 33 percent of all thefts in the city. You will probably notice two things about panel A. First, car theft seems to be everywhere in this area, except for blank spots in the center and on the right side of the map—a large park and a river, respectively. Second, because car theft seems to be everywhere, the map is not especially useful. Much of the district appears to be a hot spot. Panel B changes the time reference, showing the 30 car thefts that occurred in the first week of August 2005. You might think this is somewhat more useful, showing more theft in the southern part of the district. But while panel A shows too much, panel B shows smaller numbers that don't seem to cluster very much.

Panel C shifts the geographic focus to one sector within the district, to the left of the park. This sector happens to have the highest volume of car theft, 464 completed thefts in 2005; it's the hottest sector in the hottest precinct in the city. Again, car theft seems to be all over the sector. A closer look shows more dots on the longer north-south streets than on cross streets. This is more clear in panel D, which shows a crime density map of the sector. Crime density is a numerical value showing how close some dots are to each other, and how distant those clusters are from outlying dots. These values are mapped, showing patterns much more clearly than simple dots. The darker areas of panel D represent more dense concentrations of car theft. There seem to be two corridors of car theft, running north-south below the diagonal street that bisects the map. These corridors are sort of connected in the middle, showing a rough H-shape. This shape happens to correspond with some major thoroughfares in the area. You might be able to imagine cruising up, across, and down, looking for cars to steal. That's useful information a crime analyst can provide for police managers. During the summer months of 2006, police in this city deployed special patrols on the streets within the H-shaped area depicted in panel D.

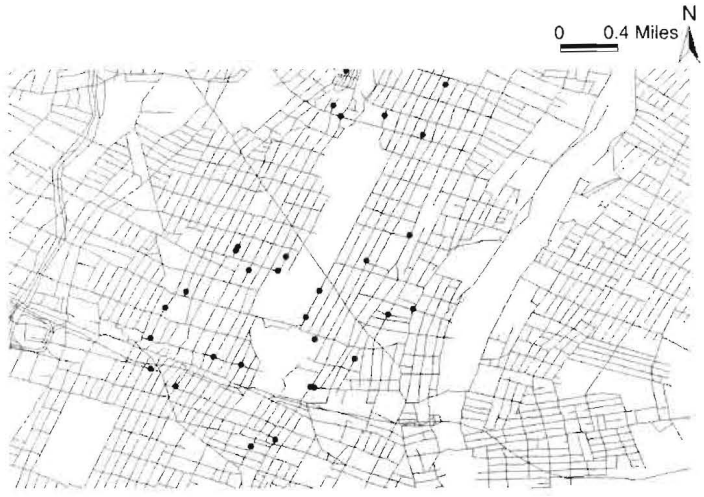
Tools for mapping crime and other problems are similar to the tools of statistical analysis, a topic we consider in the final chapter. Maps and statistics are most useful when we seek to understand patterns in a large number of observations. Very small police departments that report very few incidents need neither statistical nor geographic analysis to understand crime problems. But departments serving cities like the one in Figure 10.4 can really benefit from space-based analytic tools like crime mapping and density analysis.

Computerized crime mapping has been used for many years in a small number of departments and is spreading to many large and midsized cities. Software is more powerful, and web-based mapping programs have been used to make crime maps generally available. More published guides are appearing that describe how to combine maps with other analysis programs and sources of data. Jerry Ratcliffe (2004) describes how to classify crime concentrations across space and time dimensions to produce a hot spot matrix.

Crime mapping and other types of problem analysis illustrate another advantage of



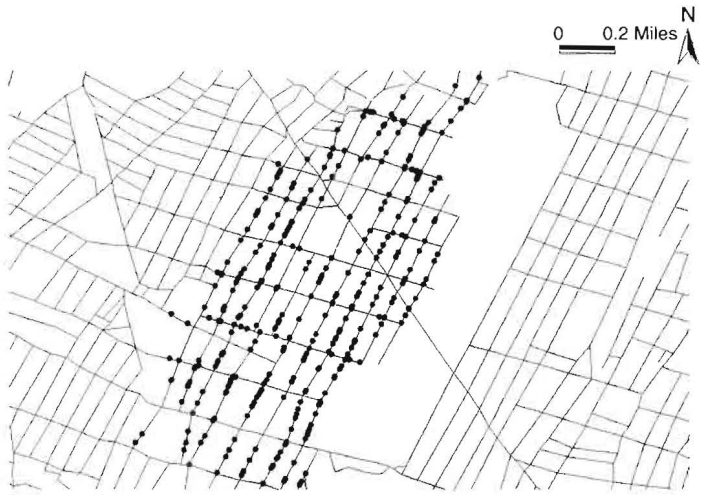
A. District 2, 2005



B. District 2, August 1–7, 2005

Figure 10.4 Mapping Auto Theft

Source: Maps prepared by Shuryo Fujita.



C. Sector 212, 2005

D. Sector 212, 2005 Density

Figure 10.4 (*continued*)

incident-based data—the potential for use in the kind of problem analysis we have described. Most crime mapping and similar tools are developed and used by individual departments, reflecting the fact that crime analysis is based on locally generated data. With incident-based reporting, crime analysis can be conducted on larger units. For example, Donald Faggiani and Colleen McLaughlin (1999) describe how National Incident Based Reporting System (NIBRS) data can show state or regional patterns in drug arrests and offenses. Using NIBRS data for Virginia, the authors demonstrate differences in types of arrests and drugs for different areas of the state.

Scientific Realism and Applied Research

Traditional research and evaluation are based on the model of cause and effect we considered in Chapters 3 and 5. An independent variable (cause) produces some change in a dependent variable (effect). Experimental and quasi-experimental designs seek to isolate this causal process from the possible effects of intervening variables. Designs thus try to control for the possible effects of intervening variables.

Problem analysis as we have described it represents a bridge between traditional research approaches and applied research that is the foundation of scientific realism. Ray Pawson and Nick Tilley (1997) propose that, instead of trying to explain cause in a traditional sense, evaluators should search for mechanisms acting in context to explain outcomes. As we have seen, experiments do this by producing pretest statistical equivalence between groups of subjects who receive an intervention and groups of subjects who do not. Quasi-experiments using nonequivalent groups might seek to control intervening variables by holding possible intervening variables constant. So, for example, if we believe that employment status might be an intervening variable in the relationship between arrest and subsequent domestic violence, we will try to structure an evaluation to hold employment status constant between treatment and control groups.

Scientific realism treats employment status as the *context* in which an arrest *mechanism* operates on the *outcome* of repeat domestic violence. Rather than try to control for employment status, a scientific realist will study the mechanism in context and conclude, for example, that arrest is effective in reducing subsequent violence in situations in which an offender is employed but is not effective when the offender is unemployed. This finding will be no different from what Sherman and associates (1992) conclude in their assessment of a series of randomized experiments.

What is different is that the scientific realist approach is rooted in the principle that similar interventions can naturally be expected to have different outcomes in different contexts. Most notably, this approach is more compatible with the realities of evaluation than is the experimental approach. Pawson and Tilley (1997, 81) put it this way: "Ultimately, realist evaluation would be *mechanism-* and *context-driven* rather than program-led" (emphasis in original). This means that interventions should be designed not so much as comprehensive programs that apply equally in all situations. Instead, interventions should be developed for specific contexts, and evaluations of those interventions must consider context as a key factor in whether the intervention achieves the desired outcome.

Situational crime prevention (Clarke 1997b) is an example of the scientific realist approach that bridges problem analysis and evaluation because it focuses on what mechanisms operate for highly specific types of crime in specific situations. So, for example, rather than develop and evaluate large-scale programs intended to reduce auto theft generally, situational crime prevention seeks specific interventions that will be effective in reducing particular types of auto theft. Ronald Clarke and

Patricia Harris (1992) distinguish several types of auto theft by their purposes: joyriding, temporary transportation, resale or stripping, or insurance fraud. Theft of certain models for joyriding may be reduced by modest increases in security, while theft of expensive cars for resale or export requires different approaches. Many types of auto theft can be reduced by placing attendants at the exits of parking garages, but car break-ins may not be affected by that intervention.

As we mentioned in Chapter 5, the realist approach resembles a case-study approach. Both are *variable-oriented* strategies for research—they depend on measures of many variables to understand and assess a small number of cases. Detailed data and information are gathered about specific interventions, often in very small areas. Whereas an experimental evaluation uses probability theory to control for intervening variables, the case-study approach depends on detailed knowledge to understand the context in which mechanisms operate.

In his discussion of applied research tools for problem solving, Eck (2002, 2003b) makes the case even more strongly. Public officials, he argues, are more interested in solving local problems than in identifying robust cause-and effect relationships. Both problem solving and evaluation are concerned with answering the question "Did the problem decline?" But eliminating alternative explanations for a decline, which is the central concern of internal validity and the rationale for stronger evaluation designs, is important only if officials wish to use the same intervention elsewhere.

In what Eck terms "small-claim, small-area problem solving," analysts develop appropriate interventions for problems in context. This is the essence of the problem-solving process. Like Eck, we emphasize *process*—systematically studying a problem, developing appropriate interventions, and seeing if those interventions have the intended effect. This is quite different from what Eck terms "large-claim interventions"—such as Drug Abuse Resistance Education (D.A.R.E.) or corrections boot camps—that are developed to apply in a wide variety of settings. Because small-claim, small-scale interventions are tailored to highly specific settings, they cannot easily be transferred intact to different settings. However, the *process* of diagnosing local problems, selecting appropriate interventions, and then assessing the effects of those interventions can be generally applied. Anthony Braga (2002) offers more examples of this reasoning. Gloria Laycock presents an even stronger case for scientific realism in applied criminal justice research generally (2002) and in making specific plans for crime prevention (Tilley and Laycock 2002).

Randomized or quasi-experimental evaluations should be conducted when such designs are appropriate. But it is important to recognize the formidable requirements for deploying these designs. The scientific realist approach to evaluation is flexible and may be appropriate in many situations. A scientific realist evaluation or case study can be especially useful in smaller-scale evaluations in which interest centers on solving some particular problem in a specific context more than on finding generalizable scientific truths. In any case, a variety of approaches can satisfy the definition of program evaluation we discussed early this chapter, by systematically applying social science research procedures to an individual program or agency.

Our general advice in this regard is simple: do the best you can. This requires two things: (1) understanding the strengths and limits of social science research procedures, and (2) carefully diagnosing what is needed and what is possible in a particular application. Only by understanding possible methods and program constraints can we properly judge whether any kind of evaluation study is worth undertaking with an experimental, quasi-experimental, or nonexperimental design, or whether an evaluation should not be undertaken at all.

The Political Context of Applied Research

Public policy involves making choices, and that involves politics.

Applied researchers bridge the gap between the body of research knowledge about crime and the practical needs of criminal justice professionals—a process that has potential political, ideological, and ethical problems. In the final section of this chapter, we turn our attention to the context of applied research, describing some of the special problems that can emerge in such studies.

Some similarities are evident between this material and our discussion of ethics in Chapter 2. Although ethics and politics are often closely intertwined, the ethics of criminal justice research focuses on the methods used, whereas political issues are more concerned with the substance and use of research findings. Ethical and political aspects of applied research also differ in that there are no formal codes of accepted political conduct comparable to the codes of ethical conduct we examined earlier. Although some ethical norms have political aspects—for example, not harming subjects relates to protection of civil liberties—no one has developed a set of political norms that can be agreed on by all criminal justice researchers.

Evaluation and Stakeholders

Most applied studies involve multiple stakeholders—people who have a direct or indirect interest in the program or evaluation results (Rossi, Freeman, and Lipsey 1999, 204–205). Some stakeholders may be enthusiastic supporters of an experimental program, others may oppose it, and still others may be neutral. Different stakeholder interests in programs can produce conflicting perspectives on evaluations of those programs.

Emil Posavec and Raymond Carey (2002) describe such problems as dysfunctional attitudes toward program evaluation. Program supporters may have unrealistic expectations that evaluation results will document dramatic success. Conversely, they may worry that negative results will lead to program termination. Agency staff may feel that day-to-day experience in delivering a program imparts a qualitative understanding of its success that cannot be documented by a controlled experiment. Staff and other stakeholders may object that an evaluation consumes scarce resources better spent on actually delivering a program.

We have two bits of advice in dealing with such problems. First, identify program stakeholders, their perspectives on the program, and their likely perspectives on the evaluation. In addition to agency decision makers and staff, stakeholders include program beneficiaries and competitors. For example, store owners in a downtown shopping district might benefit from an experimental program to deploy additional police on foot patrol, whereas people who live in a nearby residential area might argue that additional police should be assigned to their neighborhood.

Second, educate stakeholders about why an evaluation should be conducted. This is best done by explaining that applied research is conducted to determine what works and what does not. The National Institute of Justice (NIJ) and the Office of Community Oriented Policing Services (COPS) have issued brief documents that describe how evaluation can benefit criminal justice agencies by rationalizing their actions (Maxfield 2001; Eck 2003a). Such publications, together with examples of completed evaluations, can be valuable tools for winning the support of stakeholders.

Also keep in mind that applied research is very much a cooperative venture. Accordingly, researchers and program staff are mutual stakeholders in designing and executing evaluations. Evaluators' interest in a strong design that will meet scientific standards must be balanced against the main concern of program sponsors—obtaining information that is useful for developing public policy.

When Politics Accommodates Facts

Tony Fabelo

The 1994 federal anticrime bill, and related politics emanating from this initiative, put pressure on the states to adopt certain sentencing policies as a condition for receiving federal funds. Among these policies is the adoption of a "three strikes and you're out" provision establishing a no-parole sentence for repeat violent offenders. Facts have prevented a criminal justice operational gridlock in Texas by delineating to policy makers the operational and fiscal impact of broadly drafted policies in this area. Facts established through policy analysis by the Criminal Justice Policy Council (CJPC) have clearly stated that a broad application of the "three strikes and you're out" policy will have a tremendous fiscal impact.

Therefore, state policy makers have carefully drafted policies in this area. For example, during the last legislative session, the adoption of life with no parole for repeat sex offenders was considered. State policy makers, after considering facts presented by the CJPC, adopted a policy that narrowly defined the group of offenders for whom the law is to apply. They also adopted a 35-year minimum sentence that must be served before parole eligibility, rather than a life sentence with no parole. The careful drafting of this policy limited its fiscal impact while still accomplishing the goal of severely punishing the selected group of sex offenders.

Unlike Texas, politics did not accommodate facts in California, where lawmakers adopted a fiscally unsustainable "three strikes and you're out" policy.

For my part, I need to maintain personal integrity and the integrity of the CJPC in defining the facts for policy makers. I have to be judged not only by "objectivity," which is an elusive concept, but by my judgment in synthesizing complex information for policy makers. To do this, I follow and ask my staff to follow these rules:

1. Consider as many perspectives as possible in synthesizing the meaning of information, including the perspectives of those stakeholders who will be affected.
2. State the limits of the facts and identify areas where drawing conclusions is clearly not possible.
3. Consult with your peers to verify methodological assumptions and meet accepted criteria to pass the scrutiny of the scientific community.
4. Provide potential alternative assumptions behind the facts.
5. Set clear expectations for reviewing reports and releasing information so that facts are not perceived as giving advantage to any particular interest group.
6. Judge the bottom-line meaning of the information for policy action based on a frame of reference broader than that of any particular party or constituency.
7. Finally, if the above are followed, never succumb to political pressure to change your judgment. Integrity cannot be compromised even once. In the modern crowded marketplace of information, your audience will judge you first for your motives and then for your technical expertise.

Source: Adapted from Fabelo (1997, 2, 4).

The flip side of being cautious about getting caught in stakeholder conflict is the benefit of applied research in influencing public policy. Evaluation studies can provide support for continuing or expanding successful criminal justice programs, or evidence that ineffective programs should be modified or terminated. And problem analysis results can sometimes be used to influence actions by public officials. For an example, see the box titled "When Politics Accommodates Facts," in which Tony Fabelo describes how problem analysis dis-

suaded Texas legislators from costly, ineffective lawmaking.

Politics and Objectivity

Politics and ideology can color research in ways even more subtle than those described by Fabelo. You may consider yourself an open-minded and unbiased person who aspires to be an objective criminal justice researcher. However, you may have strong views about different sentencing policies, believing that probation and restitution are to be preferred over long prison sentences. Because there is no conclusive evidence to favor one approach over the other, your beliefs would be perfectly reasonable.

Now, assume that one of the requirements for the course you are taking is to write a proposal for an evaluation project on corrections policy. In all likelihood, you will prepare a proposal to study a probation program rather than, say, a program on the use of portable jails to provide increased detention capacity. That is natural, and certainly legitimate, but your own policy preferences will affect the topic you choose.

Ronald Clarke (1997b, 28) describes political objections to applied studies of situational crime prevention: "Conservative politicians regard it as an irrelevant response to the breakdown in morality that has fueled the postwar rise in crime. Those on the left criticize it for neglecting issues of social justice and for being too accepting of the definitions of crime of those in power." By the same token, ELMO is distrusted for being simultaneously too lenient by allowing offenders to do time at home and too close to a technological nightmare by enabling the government to spy on individuals. Evaluations of situational crime prevention or ELMO programs may be criticized for tacitly supporting either soft-on-crime or heavy-handed police state ideologies (Lilly 2006; Nellis 2006).

It is difficult to claim that criminal justice research, either applied or basic, is value-free. Our own beliefs and preferences affect the topics we choose to investigate. Political preferences and ideology may also influence criminal justice research agendas by making funds available for some projects but not others. For example, in 2004, the National Institute of Justice awarded money for projects to study these topics: "Chinese Connection: Changing Patterns of Drug Trafficking in the Golden Triangle" and "Assessment of Risk Factors Associated with Sexual Violence in the Texas Prison System." No funds were awarded, however, for such projects as "The Scope of Institutionalized Racism in the War on Drugs" or "Exploratory Research on Torture in Federal Detention Camps." It is, of course, possible for researchers—consciously or unconsciously—to become instruments for achieving political or policy objectives in applied research.

It may sometimes seem difficult to maintain an acceptable level of objectivity about or distance from evaluation results in criminal justice research. This task can be further complicated if you have strong views one way or another about a particular program or policy. Researchers who evaluate, say, an experimental program to prevent offenders from repeating probably sincerely hope that the program will work. However, substantially less consensus exists about other criminal justice problems and policies. For example, how do you feel about a project to test the effects of restrictive handgun laws or mandatory jail sentences for abortion protesters? We conclude this chapter with one final example that we expect will make you think about some of the political issues involved in applied research.

In 1990, the elected prosecutor of Marion County, Indiana—in which Indianapolis is located—was sharply criticized in a series of newspaper stories that claimed to present evidence of racial disparity in drug sentences handed down in the county. Convicted minority offenders, it was asserted, received longer prison terms than white offenders. The prosecutor immediately responded, criticizing the data collected and methods used by the investigative reporter. He also contacted Maxfield and

asked him to conduct an independent analysis of drug cases accepted for prosecution.

In the first place, the prosecutor claimed, he had had previous feuds with the author of the newspaper stories. Second, he categorically denied any discriminatory policies in making sentence requests in drug cases. Third, he said he knew that the data and methods reported in the newspaper stories were deficient even though the reporter would not reveal details about his sources and information. Finally, if any pattern of racial disparity existed, it was certainly inadvertent, and the prosecutor wanted to know about it so that the problem could be fixed. Maxfield accepted the project and was paid to produce a report.

How do you feel about this example? Did Maxfield sell out? How would you feel if Maxfield turned up clear evidence of disparity in sentences? Or no evidence of disparity? What about political party affiliation—would it make a difference if the prosecutor and Maxfield identified with the same party? With different parties?

✪ *Main Points*

- Evaluation research and problem analysis are examples of applied research in criminal justice.
- Different types of evaluation activities correspond to different stages in the policy process—policy planning, process evaluation, and impact evaluation.
- An evaluability assessment may be undertaken as a scouting operation or a preevaluation to determine whether it is possible to evaluate a particular program.
- A careful formulation of the problem, including relevant measurements and criteria of success or failure, is essential in evaluation research.
- Organizations may not have clear statements or ideas about program goals. In such cases, researchers must work with agency staff to formulate mutually acceptable statements of goals before proceeding.
- Evaluation research may use experimental, quasi-experimental, or nonexperimental designs. As in studies with other research purposes, designs that offer the greatest control over experimental conditions are usually preferred.
- The use of randomized field experiments requires careful attention to random assignment, case flow, and treatment integrity.
- Randomized designs cannot be used for evaluations that begin after a new program has been implemented or for full-coverage programs in which it is not possible to withhold an experimental treatment from a control group.
- Process evaluations can be undertaken independently or in connection with an impact assessment. Process evaluations are all but essential for interpreting results from an impact assessment.
- Problem analysis is more of a planning technique. However, problem analysis draws on the same social science research methods used in program evaluation. Many variations on problem analysis are used in applied criminal justice research.
- The scientific realist approach to applied research focuses on mechanisms in context, rather than generalizable causal processes.
- Criminal justice agencies are increasingly using problem analysis tools for tactical and strategic planning. Crime mapping and other space-based procedures are especially useful applied techniques.
- Problem solving, evaluation, and scientific realism have many common elements.
- Evaluation research entails special logistical, ethical, and political problems because it is embedded in the day-to-day events of public policy and real life.

✪ *Key Terms*

evaluation research, p. 255
evidence-based policy, p. 255
impact assessment, p. 259
problem analysis, p. 255
problem solving, p. 274
process evaluation, p. 259
stakeholders, p. 263

✪ *Review Questions and Exercises*

1. In presentations to justice practitioners, Maxfield describes evaluation as answering two questions: "Did you get what you expected?" and "Compared to what?" Discuss how particular sections of this chapter relate to those two questions.

2. When programs do not achieve their expected results, it's due to one of two things: the program was not a good idea to begin with, or it was a good idea but was not implemented properly. Discuss why it is necessary to conduct both a process and an impact evaluation to learn why a program failed.
3. What are the principal advantages and disadvantages of randomized designs for field experiments? Are such designs used in problem analysis? Explain your answer.

✪ *Additional Readings*

Clarke, Ronald V., and John Eck, *Crime Analysis for Problem Solvers in 60 Small Steps* (Washington, DC: U.S. Department of Justice, Office of Community Oriented Policing, 2005; www.popcenter.org/learning/60steps/; accessed May 16, 2008). This guide assumes some knowledge of crime mapping and some experience in doing crime analysis. But it is still a source of countless (well, maybe just 60) tips about doing applied research in crime prevention.

Eck, John E., *Assessing Responses to Problems: An Introductory Guide for Police Problem-Solvers* (Washington, DC: U.S. Department of Justice, Office of Community Oriented Policing Services, 2003; www.popcenter.org/tools/assessing_responses/; accessed May 16, 2008). This highly recommended guide was written to accompany a series of problem-solving guides that aid police in addressing a wide range of problems. Eck summarizes key elements of user-oriented evaluation.

Pawson, Ray, and Nick Tilley, *Realistic Evaluation* (Thousand Oaks, CA: Sage, 1997). The authors describe scientific realism and apply it to the evaluation of crime prevention and other criminal justice policy. This book also presents an interesting critique on the inappropriate use of experimental and quasi-experimental designs in criminal justice evaluation.

Rossi, Peter H., Howard E. Freeman, and Mark W. Lipsey, *Evaluation: A Systematic Approach,* 6th ed. (Thousand Oaks, CA: Sage, 1999). Of the many available "handbooks" on evaluation methods, this is the most widely read. Although the book is uneven in its coverage of recent developments, the authors provide a good general foundation in evaluation methods.

Tilley, Nick (ed.), *Analysis for Crime Prevention: Crime Prevention Studies*, vol. 13 (Monsey, NY: Criminal Justice Press, 2002); *Evaluation for Crime Prevention: Crime Prevention Studies*, vol. 14 (Monsey, NY: Criminal Justice Press, 2002). These companion volumes present innovative thinking about how problem analysis and program evaluation can be used by public officials in preventing crime. Some of the articles will be controversial. All are interesting and mostly fun to read.