PROBLEM-ORIENTED GUIDES FOR POLICE
PROBLEM-SOLVING TOOLS SERIES NO. 1

ASSESSING RESPONSES TO PROBLEMS: DID IT WORK?

AN INTRODUCTION FOR POLICE PROBLEM-SOLVERS, 2ND EDITION

John E. Eck

Center for Problem-Oriented Policing

ASU Arizona State University

PROBLEM-ORIENTED GUIDES FOR POLICE
PROBLEM-SOLVING TOOLS SERIES NO. 1

ASSESSING RESPONSES TO PROBLEMS: DID IT WORK? AN INTRODUCTION FOR POLICE PROBLEM-SOLVERS, 2ND EDITION

JOHN E. ECK

This project was supported by agreement #2013-DG-BX-K002 awarded by the Bureau of Justice Assistance, U.S. Department of Justice through CNA, a nonprofit research and analysis organization located in Arlington, VA. The opinions contained herein are those of the author(s) and do not necessarily represent the official position or policies of the U.S. Department of Justice. References to specific agencies, companies, products, or services should not be considered an endorsement of the product by the author(s) or the U.S. Department of Justice. Rather, the references are illustrations to supplement discussion of the issues.

The first edition of this publication was supported by cooperative agreement #99-CK-WX-K004 by the Office of Community Oriented Policing Services, U.S. Department of Justice.

The Internet references cited in this publication were valid as of the date of this publication. Given that URLs and websites are in constant flux, neither the author(s) nor the Bureau of Justice Assistance can vouch for their current validity.

December 2017

TABLE OF CONTENTS

TABLES

FIGURES

ABOUT THIS GUIDE

ABOUT THE PROBLEM-SOLVING TOOLS SERIES

The *Problem-Solving Tools* are one of three series of the *Problem-Oriented Guides for Police*. The other two are the *Problem-Specific Guides* and *Response Guides*.

The *Problem-Oriented Guides for Police* summarize knowledge about how police can reduce the harm caused by specific crime and disorder problems. They are guides to preventing problems and improving overall incident response, not to investigating offenses or handling specific incidents. Neither do they cover all of the technical details about how to implement specific responses. The guides are written for police—of whatever rank or assignment—who must address the specific problems the guides cover. The guides will be most useful to officers who:

- Understand basic problem-oriented policing principles and methods
- Can look at problems in depth
- Are willing to consider new ways of doing police business
- Understand the value and the limits of research knowledge
- Are willing to work with other community agencies to find effective solutions to problems

Extensive technical and scientific literature covers each technique addressed in the *Problem-Solving Tools*. The guides aim to provide only enough information about each technique to enable police and others to use it in the course of problem-solving. In most cases, the information gathered during a problem-solving project does not have to withstand rigorous scientific scrutiny. Where police need greater confidence in the data, they might need expert help in using the technique. This can often be found in local university departments of sociology, psychology, and criminal justice.

The information needs for any single project can be quite diverse, and it will often be necessary to use a variety of data collection techniques to meet those needs. Similarly, a variety of different analytic techniques may be needed to analyze the data. Police and crime analysts may be unfamiliar with some of the techniques, but the effort invested in learning to use them can make all the difference to the success of a project.

These guides have drawn on research findings and police practices in the United States, the United Kingdom, Canada, Australia, New Zealand, the Netherlands, and Scandinavia. Even though laws, customs and police practices vary from country to country, it is apparent that the police everywhere experience common problems. In a world that is becoming increasingly interconnected, it is important that police be aware of research and successful practices beyond the borders of their own countries.

The *Problem-Solving Tools* summarize knowledge about information gathering and analysis techniques that might assist police at any of the four main stages of a problem-oriented project: scanning, analysis, response, and assessment. Each guide

- Describes the kind of information produced by each technique

- Discusses how the information could be useful in problem-solving

- Gives examples of previous uses of the technique

- Provides practical guidance about adapting the technique to specific problems

- Provides templates of data collection instruments (where appropriate)

- Suggests how to analyze data gathered by using the technique

- Shows how to interpret the information correctly and present it effectively

- Warns about any ethical problems in using the technique

- Discusses the limitations of the technique when used by police in a problem-oriented project

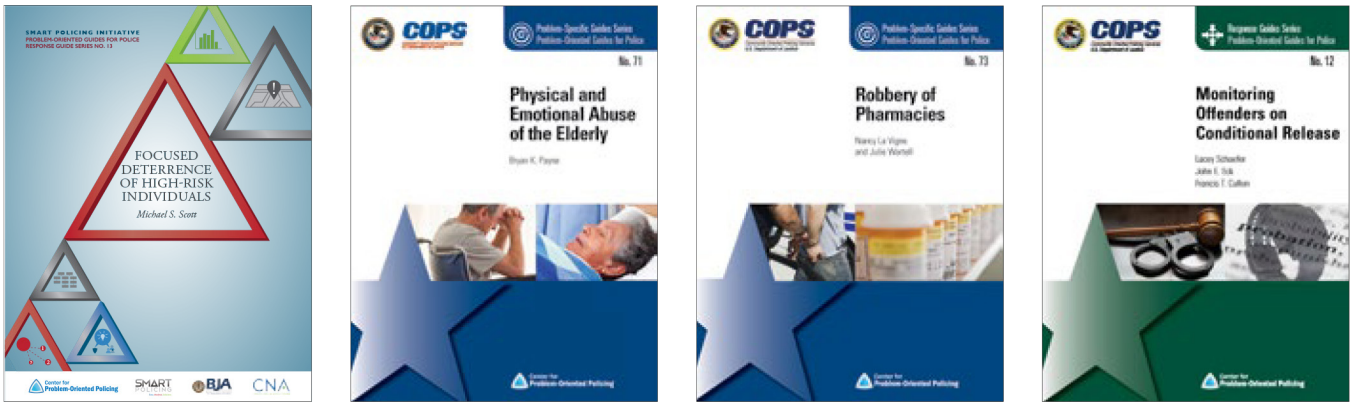
- Provides reference sources of more detailed information about the technique

- Indicates when police should seek expert help in using the technique

Each guide is informed by a thorough review of the research literature and reported police practice, and each guide is anonymously peer-reviewed by a line police officer, a police executive and a researcher prior to publication. CNA, which solicits the reviews, independently manages the process.

For more information about problem-oriented policing, visit the Center for Problem-Oriented Policing online at www.popcenter.org. This website offers free online access to:

- The *Problem-Specific Guides* series
- The companion *Response Guides and Problem-Solving Tools* series
- Special publications on crime analysis and on policing terrorism
- Instructional information about problem-oriented policing and related topics
- An interactive problem-oriented policing training exercise
- An interactive *Problem Analysis Module*
- Online access to important police research and practices
- Information about problem-oriented policing conferences and award programs



Example Problem-Oriented Policing Guides

ACKNOWLEDGMENTS

The *Problem-Oriented Guides for Police* are produced by the Center for Problem-Oriented Policing at Arizona State University under the direction of Michael S. Scott. While each guide has a primary author, other project team members, CNA and BJA staff, and anonymous peer reviewers contributed to each guide by proposing text, recommending research, and offering suggestions on format and style.

The project team that developed the guide series comprised Herman Goldstein, Ronald V. Clarke, John E. Eck, Michael S. Scott, Rana Sampson, and Deborah Lamm Weisel.

Members of the San Diego, California, National City, California, and Savannah, Georgia, police departments provided feedback on the guides' format and style in the early stages of the project.

Vivian Elliott oversaw the project for CNA. Phyllis Schultze conducted research for the guide at Rutgers University's Criminal Justice Library. Maurine Dahlberg at CNA edited this guide.

INTRODUCTION

The purpose of assessing problem-solving efforts is to help police managers make better decisions. Assessments answer two specific questions: Did the problem decline? If so, did the planned response cause this decline? Answering the first question helps decision-makers determine whether a problem-solving effort can be ended, and whether resources can be redeployed to other problems. Answering the second helps decision-makers determine whether the response should be used again to address other, similar problems.

WHAT THIS GUIDE IS ABOUT

This guide is meant to help the reader design evaluations that can answer these two questions. It was written for police officials and others who are responsible for evaluating the effectiveness of responses to problems. It assumes that the reader has a basic understanding of problem-oriented policing and the problem-solving process, including the SARA (Scanning, Analysis, Response, and Assessment) process. It is designed to be useful to readers who have no experience with evaluation and no background in evaluation and research methods.

This guide also assumes that the reader has no outside assistance. Nevertheless, the reader should seek the advice and help of researchers with training and experience in evaluation, particularly if the problem being addressed is large and complex. An independent outside evaluator can be particularly useful if there is controversy over the usefulness of the response.

Throughout, this guide refers to the importance of distinguishing between these two questions:

- **Has the problem declined following the response?**
- **Did the response cause the decline?**

It is likely that answering the first question is more critical to you than answering the second.

This guide complements the guides in the Problem-Specific Guides and Response Guides series of the Problem-Oriented Guides for Police. Each problem-specific guide describes responses to a specific problem and suggests ways of measuring the problem. Each response guide describes how and whether that response works in addressing various problem types. Though this guide is designed to work with these problem-specific and response guides, readers should be able to apply the principles of evaluation in any problem-solving project.

Because this guide is an introduction to a complex subject, it omits much that would be found in an advanced text on evaluation.[a] Readers who wish to explore the topic of evaluation in greater detail should consult the list of recommended readings at the end of this guide.

RELATED GUIDES IN THE PROBLEM-SOLVING TOOLS SERIES

This guidebook complements others in the Problem-Solving Tools series. These guidebooks address various aspects of the four phases of problem solving.[b]

Scanning phase:
- *Identifying and Defining Policing Problems* (Guide No. 13)

Analysis phase:
- *Researching a Problem* (Guide No. 2)
- *Using Offender Interviews to Inform Police Problem Solving* (Guide No. 3)
- *Analyzing Repeat Victimization* (Guide No. 4)
- *Partnering with Businesses to Address Public Safety Problems* (Guide No. 5)
- *Understanding Risky Facilities* (Guide No. 6)
- *Using Crime Prevention Through Environmental Design in Problem Solving* (Guide No. 8)
- *Enhancing the Problem-Solving Capacity of Crime Analysis Units* (Guide No. 9)
- *Analyzing and Responding to Repeat Offending* (Guide No. 11)
- *Understanding Theft of 'Hot Products'* (Guide No. 12)

Response phase:
- *Analyzing Repeat Victimization* (Guide No. 4)
- *Partnering with Businesses to Address Public Safety Problems* (Guide No. 5)
- *Understanding Risky Facilities* (Guide No. 6)
- *Implementing Responses to Problems* (Guide No. 7)
- *Using Crime Prevention Through Environmental Design in Problem Solving* (Guide No. 8)
- *Analyzing and Responding to Repeat Offending* (Guide No. 11)
- *Understanding Theft of 'Hot Products'* (Guide No. 12)

Assessment phase:
- *Analyzing Repeat Victimization* (Guide No. 4)
- *Using Crime Prevention Through Environmental Design in Problem Solving* (Guide No. 8)
- *Analyzing Crime Displacement and Diffusion* (Guide No. 10)
- Analyzing and Responding to Repeat Offending (Guide No. 11)
- *Understanding Theft of 'Hot Products'* (Guide No. 12)

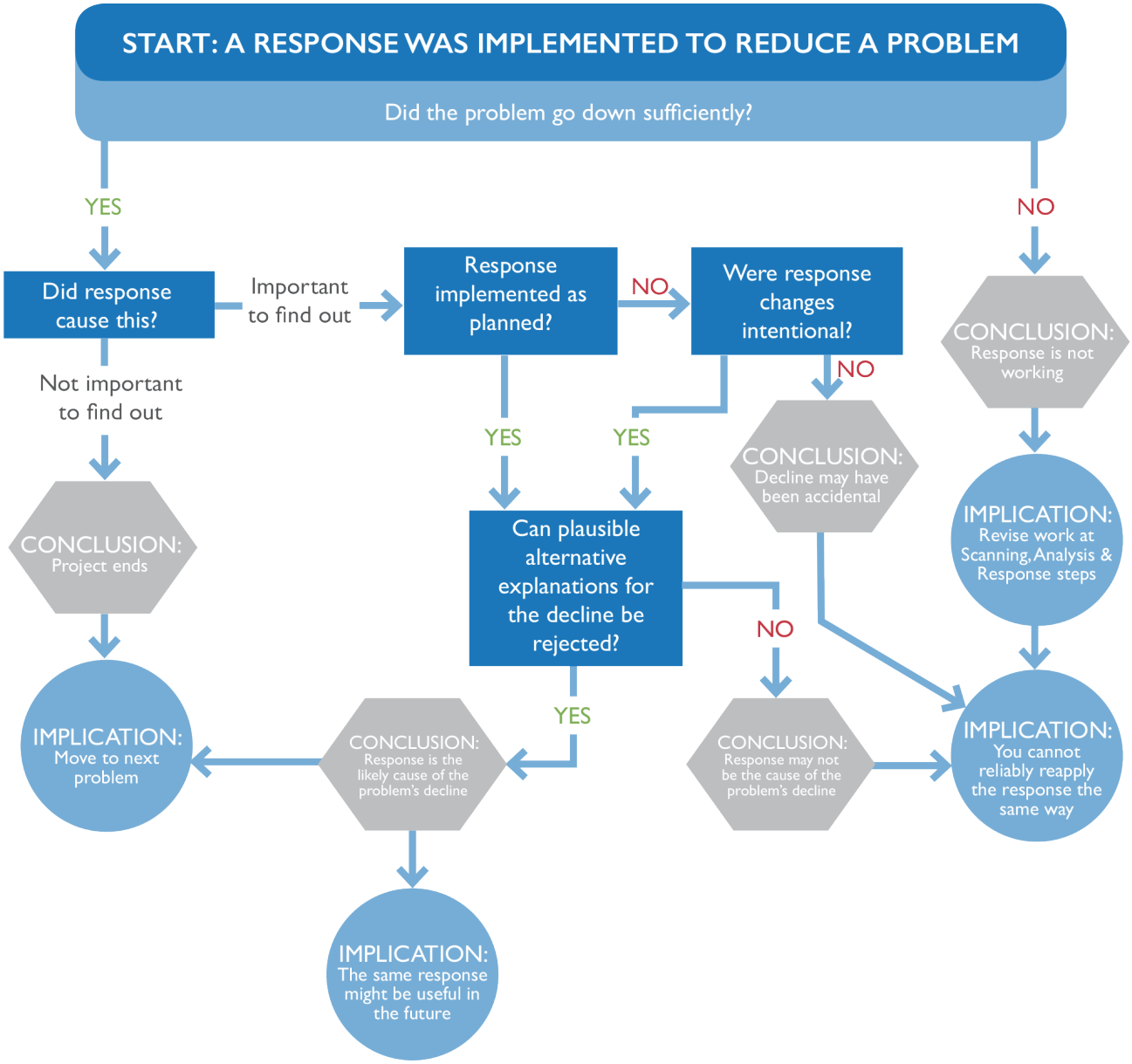
[a] Specifically excluded from this discussion are mentions of measurement theory, significance testing, and statistical estimation. A monograph of this length cannot describe those issues in enough detail for them to be useful to the reader.
[b] Some guidebooks address aspects of more than one phase of the problem-solving model.

HOW ASSESSMENT AIDS POLICE DECISION MAKING

In any problem-solving effort, two key decisions must be made. First, did the problem decrease enough that the problem-solving effort can be scaled back and the police resources applied somewhere else? If the problem did not decrease substantially, the job is not over. In such a case, the most appropriate decision may be to re-analyze the problem and develop a new response. Further, future problem solvers should be alerted so that they can develop better responses to similar problems. If the problem has declined substantially, there may be limited need to continue the problem-solving effort beyond monitoring the problem and keeping track of any response maintenance that might be required. This first decision—deciding when the problem-solving effort is done—is this guide's primary focus.

Second, if the problem did decline substantially, did the planned response cause the decline? If this decline is at least partially due to the response, it might be useful to apply a similar response to similar problems. If you cannot convincingly establish that the response caused the problem's decline, reapplying the response to similar problems may not be useful. So, future decisions about whether to apply the response are driven in part by assessment information. In this regard, assessment is an essential part of police organizational learning. Without assessments, problem solvers may repeat their or others' mistakes, or fail to benefit from their or others' successes.

Figure 1: How Assessment Aids Police Decision-Making



The process begins with the implementation of some response intended to substantially reduce the problem. The meaning of "substantially" depends on the nature of the problem and the goals of the decision-makers. The first question is whether the problem declined substantially. If the answer is no, it is clear that the response is not working well enough. Assuming that sufficient time has elapsed that you are confident the response has had time to take effect, the implication is clear: you need to go back to earlier stages of the SARA process and make revisions. Further, you now have information indicating that this response should not be recommended in the future. Let's assume that the problem has declined substantially. If you do not need to determine that the response was responsible for this decline, you can end the project and move on to the next problem. But if you take this course, you will not learn whether the response is a useful one to use in the future.

If it is important to find out whether the response caused the problem to decline, the next question is whether the response was implemented as planned. If it was not, and alterations to the response were unintentional, you do not know why the problem declined. Unintentional alterations include failure of a police unit to carry out its assigned role due to poor supervision, departures from the plan by partner agencies that are based on unforeseen circumstances (budget cuts, the appointment of a leader unsympathetic with the response, or administrative ineptitude). Success under these circumstances is welcome, but you cannot take credit for it, and you have insufficient grounds to recommend the response for similar problems.

If the response was implemented as planned, or the revisions to the response were deliberate, the next question is whether something outside of the response could have caused the problem's decline. In principle, you can never be certain. In practice, this question comes down to, "Can I credibly reject all plausible specific alternative explanations for the problem's decline?" If the answer is "no," it is possible that something other than the response is responsible for the decline. So, you cannot definitively recommend the use of the response for future problems. Any such recommendation must be cautious. If the answer is "yes," you have reasonably strong evidence that this response might work again.

Please note the qualifications in the language here. We can never be certain that a response will work the same way in the future. Instead, we can think of recommendations as being like bets: If I were to stake money on the outcome of my recommendation, am I more likely to win the bet or lose? Evidence helps you win more of these bets than you lose, but you will never win them all.

Coming to sensible conclusions requires a detailed understanding of three things: the nature of the problem; the manner by which the response is supposed to reduce the problem (see below); and the context within which the response has been implemented.[1] For this reason, the evaluation process begins as soon as the problem is first identified during the scanning stage.

This guide discusses two simple designs: pre-post, and interrupted time series. The first is only useful in the first type of decision—whether to end a problem-solving effort. The time series design can aid in both types of decisions. Designs involving comparison (or "control") groups are described in an appendix of this guide rather than in the main text. They can be difficult for a problem solver to implement successfully without receiving more advice than can be provided in this guide. Nevertheless, these designs can provide the information needed to help make the second type of decision – whether to use the response again in similar circumstances.

This guide is organized as follows.

- The body of the guide describes fundamental issues constructing simple but useful evaluations.
- The Recommended Readings list link this guide to more technical books on evaluation. Many of these clarify terminology.
- The appendices expand on material presented in the text and should be examined only after the text is read.
    - Appendix A uses an extended example to show why evaluating responses over longer periods provides a better understanding of the effectiveness of the response.
    - Appendix B describes five designs using data from a rigorous evaluation of a problem-solving project. (It includes three designs not discussed in the body of the guide: one you should never use, and two more-advanced designs.)
    - Appendix C provides a checklist designed to walk a problem solver through the evaluation process, help select the most applicable design, and draw reasonable interpretations from evaluation results.
    - Appendix D provides a summary of the strengths and weaknesses of the designs.

In summary, this guide explains in ordinary language those aspects of evaluation methods that are most important to police when addressing problems. In the next section, we will examine how evaluation fits within the SARA problem-solving process. We will then examine the two major types of evaluation—process and impact.

THE ROLE OF EVALUATION IN PROBLEM SOLVING

Assessment is the final stage in the SARA problem-solving process.[2] It is a systematic attempt to determine whether a problem declined after an effort was made to reduce it. Though assessment is the final stage of the problem-solving process, you will be making critical decisions about evaluation throughout the SARA process. The left side of Figure 2 shows the standard SARA process and some of the most basic questions asked at each stage. It also draws attention to the fact that the assessment stage may produce information requiring the problem solver to go back to earlier stages to make modifications. This is particularly the case if the response was not as successful as anticipated.

On the right side of Figure 2 are critical questions that need to be addressed in order to carry out an evaluation. During the scanning stage, you must define the problem with sufficient precision that it can be measured. Here is where you determine "what success looks like." What is the minimum necessary problem reduction that is acceptable? At this stage, you will collect baseline data on the nature and scope of the problem. At the analysis stage, you will collect data describing the details of the problem: who is impacted, when, where, and by how much? Virtually every important question to be addressed during analysis will be important in the assessment stage. This is because during assessment you want to know whether the problem has changed: information uncovered during the analysis stage becomes vital baseline information (or "pre-response" measures) for the assessment stage.

During the response stage, while developing a strategy to reduce the problem, you should also develop an accountability mechanism to be sure that various participants in the response do what they should be doing. As we will see later, one type of evaluation – process evaluation – is closely allied to accountability. Also, the type of response used will have a major influence on how the other type of evaluation – impact evaluation – will be designed.

All of these earlier decisions are brought together during the assessment stage to answer questions: Was the response implemented as planned? Did the problem change (decline)? Are there good reasons to believe that the response is the most important explanation for the changes in the problem?

In summary, you begin planning for an evaluation when you take on a problem; the evaluation builds throughout the SARA process, culminates during the assessment stage, and provides findings that help determine whether you should go back and revisit earlier stages in order to improve the response. Appendix C contains a checklist of questions that can be used as a general guide to evaluation throughout the SARA process.

Figure 2: Problem-Solving and Evaluation Planning

The SARA Process	Evaluation Questions
SCANNING What is the problem?	How should the problem be measured? What would have to decline for success to be seen?
ANALYSIS How much problem is there? Who is involved and how? Where is the problem and why?	How "much" problem is there? Who is involved and how? Where is the problem and why?
RESPONSE What should be done about the problem? Who should do it and how? Is it being done?	How will accountability be determined? How will problem reduction be measured? How will displacement and diffusion be measured? How will alternative causes for reduction be examined?
ASSESSMENT Did the response occur as planned? Is there less of the problem? What should be done next?	Was the response implemented (process evaluation)? Did the problem change? Can alternative explanations for the changes be eliminated?

TYPES OF EVALUATIONS

As we mentioned, there are two types of evaluation: process evaluation, and impact evaluation. They complement each other.

PROCESS EVALUATIONS

A *process evaluation* asks, Was the response implemented as planned? Did all of the response components work? Or, stated more bluntly, Did you do what you said you would do? This is a question of accountability.

Let's begin with a hypothetical example. Though fictitious, this example is based on an actual anti-prostitution effort in London (Matthews, 1992). We will return to this example repeatedly in this guide to illustrate numerous points.

After a careful analysis, a problem-solving team determines that to control a street prostitution problem, they will ask the city's traffic engineering department to make a major thoroughfare one-way and create several dead-end streets to thwart cruising by "johns." This will be implemented immediately after a comprehensive crackdown on the prostitutes in the target area. Arrested prostitutes, if convicted, are to be given probation under the condition that they cannot be in the target area for a year. Finally, a non-profit organization will assist women who want to leave sex work to gain the necessary skills to become legitimately employed. The vice squad, district patrol officers, prosecutor, local judges, probation office, sheriff's department, traffic engineering department, and non-profit organization have all agreed to this plan.

A process evaluation would look at whether the crackdown was implemented, how many arrests were made during the crackdown, whether the street patterns were altered as planned, how many prostitutes asked for assistance in gaining new job skills, and how many prostitutes were able to find legitimate employment. The process evaluation would also examine whether all of this occurred in the planned sequence. Here is what the process evaluation found: The crackdown did not occur until after the street alterations had been made. Only a fraction of the prostitutes operating in the area were arrested, and none of them sought job skills. Based on this, one would suspect that the plan was not fully carried out or was not carried out in the specified sequence. One might conclude that the response was a colossal failure. The fact is, however, this assessment gives us no evidence of success or failure, because a process evaluation only answers the question, "What actions were taken?" It does not answer the question, "What happened to the problem?"

IMPACT EVALUATIONS

To determine what happened to the problem, one needs an *impact evaluation*. An impact evaluation asks the questions: Did the problem decline substantially? If so, did the response cause this decline? Continuing with the example given above, let's look at how this might work. During the analysis stage of the problem-solving process, patrol officers and vice detectives conducted a census of prostitutes operating in the target area. They also asked the traffic engineering department to install traffic counters on the major thoroughfare and critical side streets to measure traffic flow. These were used to determine how customers moved through the area. The vice squad made covert video recordings of the target area to document the methods by which prostitutes interacted with potential customers. All of this was done before a response was selected, and the information gained helped the team create the response.

After the response was implemented (though not the planned response, as we have seen), the team repeated these measures. They discovered that instead of the 23 prostitutes counted in the first census, only 10 could be found. They also found that there was a slight decline in traffic on the major thoroughfare on Friday and Saturday nights, but not at other times. However, there was a substantial decline in side-street traffic on Friday and Saturday evenings. New covert video recordings showed that prostitutes in the area had altered the way they approached vehicles and that they were acting more cautiously. In short, the team had evidence that the problem had declined from what it had been before the response.

So what caused the problem to decline? This question may not be as important as it first appears. After all, if the goal was to reduce or eliminate the problem and this was achieved, what difference does it make what the cause was? It does not matter, *unless* you are interested in using the same form of response in similar situations in the future. If you have no interest in using the response again, all that matters is that the goal has been achieved. Then, the resources devoted to addressing the problem can be used on a more pressing concern. But if you believe that the response can be used again, it is very important to determine whether the response was responsible for the decline of the problem.

Let's assume that the prostitution problem-solving team believed that the response might be useful for addressing similar problems. The response, though not implemented according to plan, might have caused the decline, but it was also possible that something else caused the decline. There are two reasons that the team took this second possibility seriously. First, the actual response departed from the planned response, which had been designed to fit the problem. If the planned response had been implemented, the team would have had a plausible explanation for the decline in the problem. But the jury-rigged nature of the actual response makes it a far

less plausible explanation for the decline. Second, the impact evaluation was not particularly strong. Later, we will discuss why this was a weak evaluation and what can be done to strengthen it.

INTERPRETATION OF PROCESS AND IMPACT EVALUATIONS

Process and impact evaluations answer different questions, and their combined results are often highly informative. Table 1 summarizes the information that can be gleaned from both evaluations. As will be seen in Appendix B, the interpretation of this table depends on the type of design used for the impact evaluation.

When a response is implemented as planned (or nearly so), the conclusions are much easier to interpret (cells A and B in Table 1). When the response is not implemented as planned, we have more difficulty determining what happened and what to do next (cells C and D). Cell D is particularly troublesome because all you really know is that "we did not do it and it did not work." Should you try to implement your original plan, or should you start over from scratch?

Outcomes that fall into cell C are worth further discussion. The decline in the problem means that you could call an end to this problem-solving process and go on to something else. If the problem has declined substantially, this might be satisfactory. If, however, the problem is still large, you do not know if the response should be continued. Alternatively, you could seek a different response, on the assumption that the response is not working well and something else is needed. Additionally, you do not know whether the response will be useful for similar problems.

A process evaluation involves comparing the planned response to what actually occurred. Information about how the response was implemented usually becomes apparent while managing a problem-solving process, but only if you look for it. If the vice squad is supposed to make a series of arrests of prostitutes in the target area, one can determine this from departmental records and discussions with members of the vice squad. There will be judgment calls, nevertheless. For example, how many arrests are required? The plan may have called for the arrest of 75 percent of the prostitutes, but only 60 percent were arrested. It may be difficult to determine whether this is a serious violation of the response plan. Much of a process evaluation is descriptive: these things were done, in this order, by these people, using the following procedures. Nevertheless, numbers can help. In our example, data on traffic volume showed where street alterations had changed driving patterns, and the changes in driving patterns are consistent with what had been anticipated in the response plan.

In short, a process evaluation tells what happened in the response, when it happened, and to whom it happened. Though it does not tell whether the response made a difference in the problem, it is very useful for determining how to interpret impact evaluation results.

Table 1: Interpreting Results of Process and Impact Evaluations

		PROCESS EVALUATION RESULTS	
		Implemented nearly as planned	Not implemented or implemented in a radically different manner than planned
IMPACT EVALUATION RESULTS	Problem declined	A. Evidence that the response caused a decline in the problem	C. Suggests that other factors may have caused the decline in the problem, or the response was accidentally effective
	Problem did not decline	B. Evidence that the response was ineffective and that a different response should be attempted	D. Little was learned; perhaps better results would have been noted if the response had been implemented as planned, but this is speculative.

CONDUCTING IMPACT EVALUATIONS

An impact evaluation has two parts. The first involves the measurement of the problem: how big is it? The second involves ways of systematically comparing changes in the problem to discover if it shrank after the response or if it shrank more than other similar but untreated problems. The second part is called the evaluation design. Evaluation designs are created to provide the maximum evidence that the implemented response was the primary cause of the change in the measure. Weak designs provide little confidence that the response caused the change. Strong designs provide much greater confidence in the conclusion that the response was the cause of the problem's demise.

MEASURES

Impact evaluations require measurements of the problem *before* and *after* the response has been implemented. (Appendix B describes a commonly used bad design that does not have a before measure.) Decisions about how to measure the problem should begin at the scanning stage and be settled by the time the problem analysis has been completed. This will allow information collected during the analysis stage to be used to describe what the problem looked like before the response. During the assessment stage, measures are taken after the response has been implemented. The problem is measured the same way before and after the response.

Quantitative Measures
Measures can be qualitative or quantitative. Quantitative measures involve numbers. The number of burglaries in an apartment complex is a quantitative measure. One can count them before the response and after the response, and calculate the difference. Quantitative measures allow you to use mathematics to estimate the impact of the response. For example, burglaries went down 10 percent from before the response to after the response. In the example above, the counts of active sex workers and the traffic volume figures are both quantitative measures.

Qualitative Measures
Qualitative measures allow comparisons, but mathematics cannot be applied to them. In the example, observations of how sex workers interact with johns is a qualitative measure. Though most evaluations use quantitative measures, qualitative measures can be extremely useful. The fact that you cannot add, subtract, multiply, or divide qualitative measures does not mean they are useless. The important thing is that these measures are collected systematically and before and after the intervention so that the measures are comparable. Photos of the cleanliness of an area before and after a problem-solving effort might be useful, if they are taken at the same locations in the same lighting conditions, from the same angle and from the same distances. An arbitrary set of snapshots before and after the response is of little value in assessing the response.

Maps
Maps provide another method of qualitative measurement. Maps are very useful for showing crime and disorder patterns. Though the number of crimes is a quantitative measure, and the size and shape of the crime patterns is typically drawn using a computer algorithm, when we compare map patterns we typically use qualitative comparisons.

Measurement Validity
For both qualitative and quantitative measures, you must make sure that the measures record the problem and do not record something else. For example, counts of drug arrests are often better measures of police activity than changes in a drug problem. You should use arrest data as a measure of the problem only if you can be certain that police enforcement efforts and techniques have remained constant. On the contrary, systematic covert surveillance of a drug-dealing hotspot before and after the response could be a valid measure, if the form of surveillance was unchanged and remained undetected by the drug dealers. Measures are seldom valid or invalid; rather, they are more or less valid than alternative measures.

In short, you want to make sure that the change in the problem you measure is due to changes in the problem and not due to changes in the way you take the measures. One way of thinking about this is to compare it to physical evidence gathered at a crime scene. The reason there are strict protocols for the gathering and handling of evidence is because we do not want to confuse the activities of the offender with the activities of the evidence gatherers. The same thing is true in evaluations.

The less direct the measurement is, the less validity it has. For example, if you want to measure drug dealing, surveillance on drug-dealing sites provides direct observations of drug dealing. Arrest statistics are indirect because they involve the activities of the drug dealers and customers (the aspects of the problem you may be most interested in), as well as decisions by citizens to bring this to police attention, police decisions to intervene, and police decisions as to how they will intervene. These decisions by citizens and by the police may not always be related to the underlying reality of the problem. For example, changes in police overtime policies or the presence of special anti-drug squads can change the number of arrests, even if the drug problem remains constant. For this reason, the number of arrests of drug dealers is a less direct, and often a poor, measure of a drug problem.

Sometimes, however, it is impossible to get a direct measure of the problem and an indirect measure needs to be used. In 2004, twenty-three Chinese immigrants were drowned harvesting shellfish in the United Kingdom. A problem-solving effort was undertaken to reduce the chances of this occurring again. Evaluating the success of the response was

difficult because deaths by drowning were (fortunately) rare and multiple deaths by drowning were even less common. Therefore, counting the number of deaths by drowning before and after the effort would overestimate the success of the project, because there had been an unusually high number of such deaths in the one incident before the effort and, even if the police did nothing, there would probably be a very low number of them in the future. The police evaluators, instead, counted rescue calls to the coastal rescue service. The evidence showed that these calls declined substantially, thus providing evidence consistent with a successful response.[3]

Let's return to the prostitution problem to see another example of indirect and direct measurement. In this example, the meaning of *direct* and *indirect* depends on how one defines the problem. Men drive into a neighborhood on Friday and Saturday nights looking for prostitutes to pick up. This annoys the neighbors. They call the police to do something. You have a choice of two measures for this problem.

The first is a quantitative measure taken from automatic traffic counters strategically placed on the critical streets three months before the intervention and left there until three months after the response was completed. These devices measure traffic flow. The difference between the average Friday and Saturday night traffic volume and the average volume during the rest of the week is used as an estimate of the traffic due to prostitution.

Your second measure is based on interviews of local residents taken three months before the response and three months afterwards. Residents are asked about their perceptions of the prostitution problem using a numerical scale (0 = none, 1 = minor, 2 = moderate, 3 = heavy).

If you have defined your problem as *prostitution-related traffic*, the first measure is a more direct measure than the second. Not all of the difference between the traffic level on Friday-Saturday and the level during the rest of the week is due to prostitution, but a large part of it probably is. So, this is a reasonable approach. Asking citizens for their perceptions, however, is fraught with difficulties. Their current perceptions of prostitution may be colored by past observations. They may not see much of the prostitution traffic, particularly if they are hiding indoors to avoid the problem. They may misperceive other activities as prostitution related.

If, on the other hand, you have defined the problem as the *residents' annoyance with prostitution-related traffic*, the interviews are a more direct measure than the traffic counts. Prostitution-related traffic may have not changed, but the citizens think it has. By this measure, the response was a success. But if prostitution-related traffic (measured by the counters) has declined precipitously and the citizens are unaware of it, then, by this measure, the response has not worked.

Of course, multiple measures can be used. In this example, one could measure both the reduction in prostitution-related traffic and the perceptions of it. Only if both declined would success be unambiguous. If the traffic counters indicated a drop in traffic but the citizen surveys showed that the residents were unaware of the decline, the response could be altered to address the perceptions.

Selecting Valid Measures

How do you select specific measures for your problem? There is no one answer to this question that can be applied to any problem-solving effort. If you are working on a problem for which a problem-specific guide has been prepared, you can find some ideas for problem-specific measures listed in it. If you are working on another type of problem, the simplest approach is to use one or more of the indicators of the problem that you used to identify and analyze the problem. It is important, however, to think carefully about problem definition. As we saw in the prostitution example, seemingly minor changes in how we define the problem can have significant implications for measurement. Clearly, one needs to think about evaluation measures as soon as one begins a problem-solving process.

One way to clarify the measures to be used is to pose the question: "Why do we, the police, care about this problem?" The answers lead to outcome measures. Among other reasons, police care because: (1) citizens are annoyed (or scared); (2) people are getting hurt; (3) it's costing the city too much money; or (4) it's wasting police time. Note that these example answers are not generically valid. Prostitution activity in an industrial-warehouse area should produce different answers than the same activity in a residential area. Note also that "it is against the law" is not a valid answer. The law is a tool to help reduce problems, so compliance is not an outcome. Reduction in the problem is the outcome.

CRITERIA FOR CLAIMING CAUSE

As we discussed above, a problem-solving assessment has two goals: to determine whether the problem has changed, and to determine whether the response *caused* the change in the problem. We are particularly interested in the first goal. The second goal is only important if (1) the problem has changed, and (2) a similar response may be used to address other problems. If neither of these conditions is met, we do not need to worry about cause and the evaluation is relatively simple. If, however, the problem has changed and it is likely that the response will be used again, it is important to determine whether the response was in fact the cause of the change. If the problem decreased for reasons other than the response, then using the response again, in similar circumstances, is unlikely to produce useful results.

What if the problem has gotten worse, following the response? The response might or might not be responsible for this. If you can determine that the change in the problem was not due to the response, the response might be useful for other problems. If the response did cause the increase in the problem, you clearly do not want to use it again and should warn others not to use it.

The concept of *cause* may seem pretty straightforward, but it is not. Before you can confidently proclaim that a response caused the problem to decline, you need to meet four criteria. The first three criteria are relatively straightforward, and are often achievable. The fourth criterion cannot be achieved with absolute certainty. We discuss these below.

A Plausible Explanation of How the Response Reduces the Problem

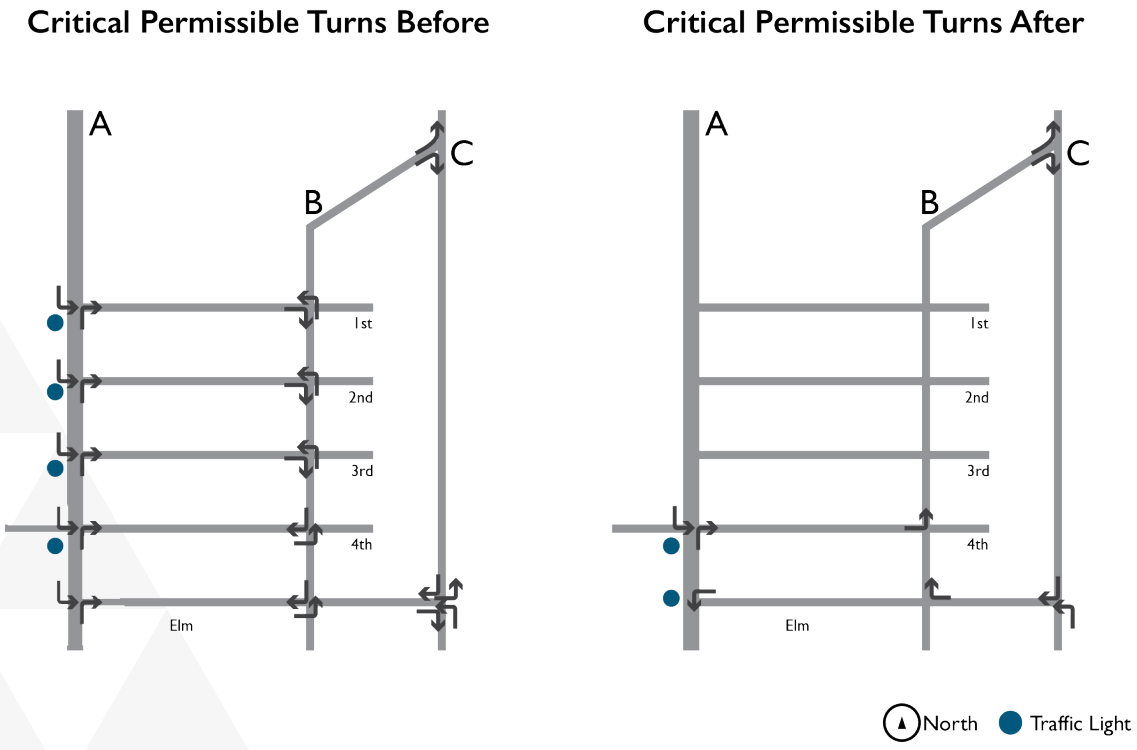
The first criterion is that there must be a convincing argument showing how the response is supposed to address the problem.[c] This explanation should be based on a detailed analysis of the problem, preferably augmented by prior research and theory. The fact that others used a similar response and were able to reduce their problem is not an explanation. Such information is useful, but there is still a need to explain how this occurred. Absent a convincing explanation, you do not know whether this prior experience was successful by accident, whether its success was unique to the situation in which it was first applied (and will not work on your particular problem), or whether it is a generally useful response.

Returning to the prostitution example, we can illustrate what is meant by a plausible explanation. We will focus on the street-pattern alterations. Police and local residents know from observations that the prostitutes congregate along a three-block stretch of roadway (between 1st and 4th streets on B Street), one block off of a very busy thoroughfare (A Street). There are traffic lights on each of the numbered streets (see Figure 3, left panel). All of the streets are two-way. The area between A and B streets largely comprises a vacant old warehouse and a light-industrial area. The prostitution activity along B Street makes use of the abandoned properties. Customers come onto B Street from A Street using the numbered streets and circle the blocks looking for women they can solicit.

Between B and C streets is an old residential neighborhood of single-family homes called the "Elms." C Street has become a thriving entertainment and arts area, and the "Elms" is being rehabilitated as older residents sell their homes to younger, more affluent couples. Residents of the Elms complain about the traffic and noise, the harassing calls of the customers and prostitutes, and the litter of drink containers, condoms, and other debris.

To address this problem, residents have proposed a series of changes to the streets. B will be made one-way north and Elm one-way west, while 4th Street will be made one-way east between A and B streets. The other numbered streets will be disconnected from A Street and their traffic lights removed. A new traffic light will be put at the corner of Elm Street and A Street but only left turns from Elm Street onto A Street will be permitted. Another traffic light will be placed at the intersection of Elm and C streets. These changes are shown in the right panel of Figure 2.

Figure 3: Street Layout Before and After a Response to Prostitution



[c] The technical term for this criterion is "mechanism."

Why do the residents think this will work? We hope their answer is a plausible explanation – it is logical and takes into account the known facts. The residents claim that this area is a hotbed of prostitution activity in large part because the streets facilitate the shopping behavior of customers and the advertising displays of the prostitutes. Customers can cruise around the block quickly, looking for prostitutes. By changing the street pattern in the manner described, circular cruising becomes more time consuming. If customers do not make a contact on the first pass, they will spend much more time on the return trip. Because the customers' convenience is reduced, fewer of them will come to the area and the problem will be reduced. In addition, once the traffic flow has been streamlined, it will be easier for the police to detect prostitution-related activities, thus increasing the risk of detection. By observing customers and prostitutes, we can verify the cruising behavior. If this explanation is logically consistent with the available information, and there is no clear and obvious contradictory information, the residents have passed the first hurdle for establishing a causal connection.

Many plausible ideas do not work when tested, so a plausible explanation by itself does not guarantee that the response will work. But it does make the response a more likely candidate for a successful solution than explanations that are not grounded in logic, fact, and experience. Prior research is important in establishing plausibility. Success of the response used in the example is made plausible by fact that previous research describes the relationship between prostitution and circular-driving patterns[4] and shows that reducing the ease of traffic movement through neighborhoods sometimes reduces crime.[5] Further, this intervention is consistent with the theory of Situational Crime Prevention, particularly the strategy of increasing the offenders' effort.[6] Too often, police, elected officials, and the public stop at the notion of plausibility and assume that if it sounds reasonable, it must be true. And just as often, evidence demonstrates this initial hunch was wrong.

In summary, the first step in demonstrating that a response has reduced the problem is a plausible explanation of (1) how the problem operates and (2) how the response is supposed to disrupt this operation. This explanation should tell how, where, when, and why the response works. If such an explanation is prepared when the response is being crafted, it can help guide the planning and implementation of the response. The more specific this explanation is, the better the response will be and the more informative the assessment will be. Ideally, this explanation would also describe the circumstances under which the response is unlikely to work. This can aid in both the process evaluation and the impact evaluation.

The Amount of the Problem and the Level of the Response Are Related

The second criterion for claiming that a response caused a decline in the problem is that there is a relationship between the presence of the response and a decline in the problem (and the absence of the response and an increase in the problem).[d]

Let's go back to the prostitution problem. How would we demonstrate a relationship here? Are there similar neighborhoods that we could compare to the Elms? Just north of the Elms, there is a neighborhood like the Elms (it is also between A and C streets with a deteriorated light-industrial area to the west and the thriving C Street development to the east), but the streets do not allow easy circular-driving patterns. Now if the ease of circular driving is associated with prostitution, we should see little or no prostitution in this other neighborhood. This would imply that changing the street pattern in the Elms might be helpful. However, if there is prostitution in this area too, there is not a strong link between prostitution and ease of circular driving and this suggests that changing the street pattern may not be effective. Either way, the evidence would not be strong, but the findings could be helpful.

We might also attempt to demonstrate a relationship by measuring the problem before and after the street changes. If we see high levels of prostitution (or high levels of resident perceptions of prostitution) before the changes but low levels on these measures after the street changes, we will have evidence of a relationship.

To clear the second hurdle in claiming causation, we must demonstrate that the situation has more of the problem in the absence of the response than when the response is in place. If so, it is tempting to declare victory at this stage; however, there are two other hurdles that must be surmounted before we can be confident that the solution was responsible for the decline in the problem. This brings us to the third criterion for demonstrating a causal connection.

The Response to the Problem Comes Before the Problem's Decline

The third criterion is that the decline in the problem comes after the response;[e] logically, a response would not have an effect before it is implemented. There is one major caveat here: by response, we include publicity —intentional or accidental—about the response. A crackdown on drunk drivers may be preceded by a widespread media campaign; if so, potential drunk drivers may alter their behavior even before the intervention. In this case, the media campaign is part of the response. A decline in drunk driving after the media campaign begins but before the crackdown, could be credited to the response.[f] However, a decline in drunk driving prior to the media campaign would be evidence that something other than the response has caused the problem to dissipate.

[d] The technical term for this criterion is "association." Typically, association is measured by the correlation between the response and the level of the problem.
[e] The technical term for this criterion is "temporal order."
[f] The technical term for this phenomenon is "anticipatory benefit."

Despite its obvious simplicity, it is surprisingly common to see violations of this criterion. Throughout the 1990s homicides declined in large cities in the United States. In the middle of the decade, a couple of years into the downward trend, several U.S. cities implemented crime-reduction strategies and gained substantial notoriety. As homicides continued to decline in these cities, proponents claimed that these reductions were due to the new strategies. In point of fact, homicides had been declining prior to the changes. Because homicides were trending downward before the changes, it is difficult to attribute the decline to changes in police strategies.[g] In short, the purported cause of the decline came after the decline began. If these same changes had been implemented in 1990, the claim that they caused the drop in homicides would be more plausible.

To demonstrate that the response preceded the problem's decline, you must know when the response began (including publicity about it) and then have measures of the problem before this time and after this time. This is called a *before-after* (or a *pre-post*) evaluation design. It is the most common evaluation design, but it is not a particularly strong design. That is, a simple pre-post design can show a decline, but it is insufficient for establishing what caused the decline.

Despite its superficial simplicity, this criterion can be difficult to demonstrate. But even if you can show that the decline in the problem came after the response, you need to achieve one more criterion before you can definitively claim that the response caused the decline: you must eliminate the alternative explanations.

Elimination of Alternative Explanations

Let's continue with the prostitution problem. You have an explanation, you have demonstrated a relationship, and you have shown that the response came before the decline in the problem. You now need to make sure that nothing else could have caused the decline in prostitution.[h] Recall that the C Street corridor and the Elms are going through a series of changes. New people are moving into the area and they are allying themselves with the remaining older residents to clean up the area. One thing they did was to call upon the police to help. Did they do anything else? Suppose the Elms' Neighborhood Association (ENA) and the C Street Corridor Business Association (CSCBA) identified the owners of the abandoned and vacant property and put pressure on them to clean up their property. This denied prostitutes access to the property. And suppose these changes got underway about the same time the street changes were being implemented. So, one could think of the ENA and the CSCBA as the cause of the street changes *and* the changes in land use. If the land-use changes were the real cause of the reduction in prostitution, and the street changes were irrelevant, you would still see a relationship between the street closures and a reduction in the prostitution, and you would still see the response before the reduction. Nevertheless, something else would be responsible for the decline in the problem.

Figure 4 diagrams the notion of an alternative explanation. The left panel shows what you believe: the response caused (shown by arrow) the decline in the problem. This belief may come from a variety of valid sources. Nevertheless, something else has caused both the response and the reduction in the problem (right panel). Here, more "something else" led to more response and, at the same time, led to a reduction in the problem.

Figure 4: Alternative Explanations

[g] There is another reason to be skeptical that the changes in policing caused the decline in homicides. Homicides declined in other large cities that had not implemented the same changes. For a more detailed examination of the police contribution to the decline in homicides through the 1990s, see Eck and Maguire (2000).

[h] The technical term for this criterion is "non-spuriousness." A spurious relationship is a false relationship: it appears that the response is causing the decline in the problem, but in reality some other factor is the cause of the decline and possibly the response, too.

The absence of an arrow between the response and the decline in the problem means that in reality the response was irrelevant to the problem. An outsider, observing more of the response and less of the problem at the same time, might wrongly conclude that the response and problem are causally connected. In situations like this, the observed relationship between the response and the decline in the problem is misleading. The possibility of a misleading relationship between a response and a problem is a threat to the validity of an evaluation's conclusions. Note that this is a possibility, not a demonstrated certainty. A threat to the validity of conclusions does not mean that the response was a failure. It means that we cannot be sure the response worked. There is substantial doubt because there is a plausible alternative explanation. Again, a jury trial is a useful example. If the prosecutor fails to eliminate others who could have committed the crime (and the defense brings this to the attention of the jury), the jury must have some doubts about the guilt of the defendant. Acquittal, in this case, does not mean that the prosecutor is wrong. It means that the prosecutor has not successfully eliminated alternative explanations.

There is a related concern: The "something else" might have occurred by coincidence at about the same time as your response. Practically speaking, it might not matter whether the "something else" accidentally occurred at the same time as your response or whether the "something else" caused both the response and the decline in the problem. In neither case did the response cause the drop in the problem.

To demonstrate a causal connection between the problem and the response, an evaluator needs to provide sound evidence that there is no "something else." To accomplish this, an evaluator needs to show evidence that there are no reasonable explanations for the decline in the problem other than the response. Eliminating all alternative explanations is difficult. You can never do so definitively, because there are many possible causes of problem fluctuations. All you can do is eliminate the most obvious known alternative explanations to the decline in the problem. We can never prove that a response caused a decline in a problem, because we cannot eliminate all possible rival explanations for the decline. We can make better or worse cases for such claims, however. And this is where the evaluation design comes in. Some designs allow for stronger statements of causality than others, just as some prosecutions are more plausible to a jury than others.

DESIGNS[i]

An evaluation design is a systematic strategy, coordinated with the response, for organizing when and where data collection will occur. If you develop the evaluation design along with the response, the evaluation will be more likely to produce useful information. If you wait until after the response has been implemented to decide how it will be evaluated, you will have more difficulty determining whether the response was effective.

There are many types of evaluation designs that can be used (see Recommended Readings). We will discuss two common practical designs: the pre-post design (which we addressed to some extent earlier) and the time series design. Neither have control, or comparison, groups. Appendix B discusses comparison-group and multiple time series designs (the bottom row of Table 2) and describes when you might want to use a control group, or control area.

Pre-post Designs

The simplest pre-post design involves a single measurement of the problem before the response and a single measurement after the response. The after measure is compared to the before measure. If there is less of the problem after than there was before, this is evidence of a decline in the problem. As we will see, this design is sometimes adequate for determining whether the problem declined, but it is insufficient for determining that the response caused the decline.

Table 2: Types of Evaluation Designs

	Single Measurement Before and After	Multiple Measurements Before and After
No Comparison (Control) Group	Pre-post design	Time series design
Comparison (Control) Group	Pre-post with a control group design	Multiple time series design

[i] Non-experimental designs are not addressed in this monograph because they often cannot demonstrate that the response came before the decline in the problem and because they are particularly poor at eliminating alternative explanations for the decline. Randomized designs are not addressed either. Though powerful for studying generic interventions for application across a class of problems, they are generally unsuited for operational problem solving where the primary interest is in the reduction of a specific problem rather than the testing of a generic solution. Information about these and other types of designs not described in this guide can be found in the Recommended Reading List.

Figure 5 shows results of a pre-post design. The Houston Police Department, working with the county's mental-health authority, implemented an approach to improve the way that mentally ill people were treated. The problem-solving project was a finalist for the 2010 Herman Goldstein Award for Excellence in Problem-Oriented Policing. As part of the evaluation of the intervention, the Houston Police Department examined whether the number of mental-health-related events had declined from the year before to the year after the intervention. The pair of bars marked "total" shows that there was a substantial reduction: about 47 percent. The police also looked at the types of events – shown in the other three pairs of bars – and found that there were notable drops in all three categories.[j]

Such a design can establish a relationship by demonstrating that there was less of a problem when the response was present than there was when no response was present. It also helps demonstrate that the response came before the decline in the problem, because the response occurs between the two measures. However, if the level of the problem normally fluctuates, what is seen as a decline in the problem may simply be a normal low before its return to higher levels.

Variations on this simple design include making sure that the measures are taken at the same time of the year, to account for seasonal fluctuations, and using two or three pre-response measures and two or three post-response measures to account for normal fluctuations.

Figure 5: Example of Impact Measurement in a Pre-post Design

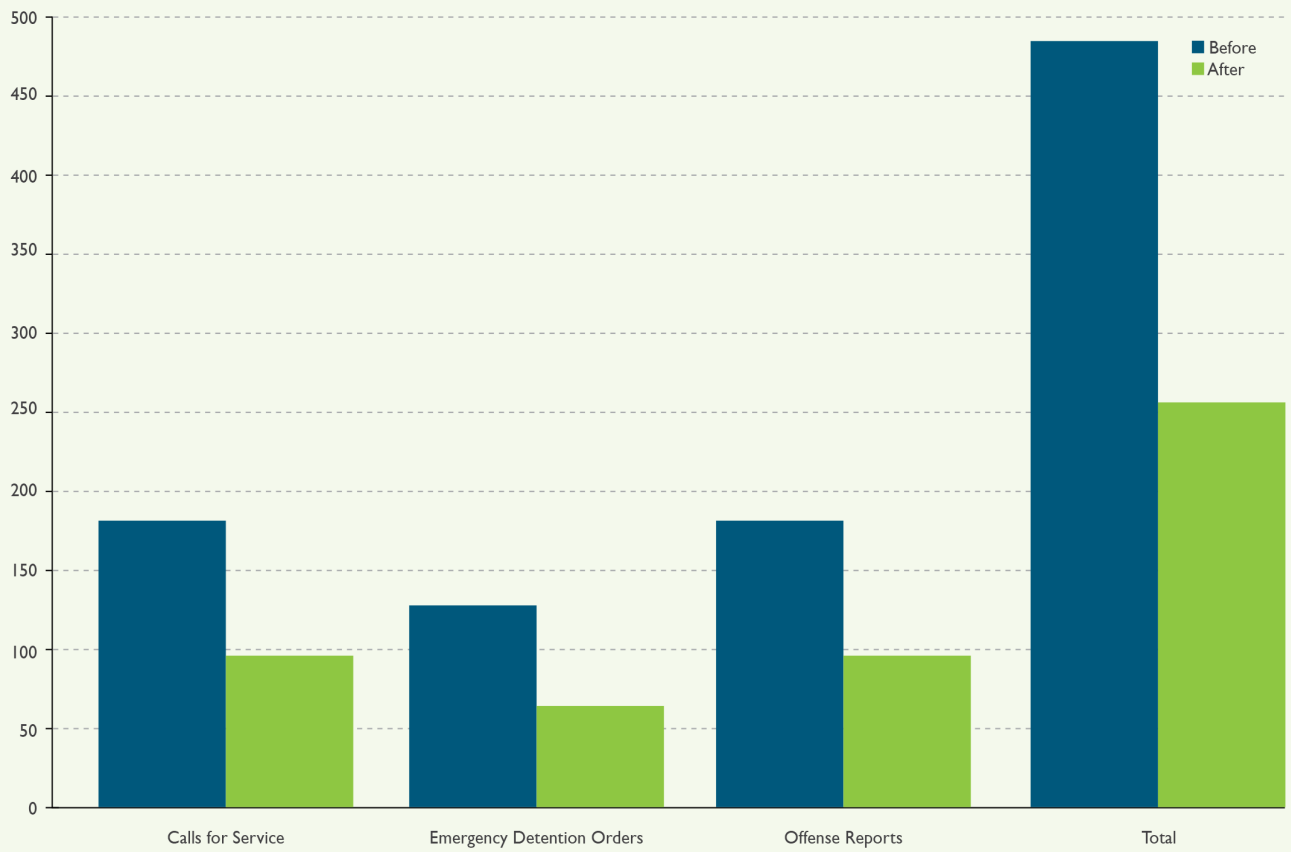


Chart derived from: Houston Police Department & Mental Health/Mental Retardation Authority of Harris County (2011). *Chronic Consumer Stabilization Initiative: A Multi-Agency Collaboration.* Presentation at the Problem-Oriented Policing Conference, 2010 Award Finalist. Miami, Florida. October. Accessible at http://www.popcenter.org/conference/conferencepapers/2011/WebLeeChronicConsumers.pdf.

[j] In most evaluation research, a test for statistical significance is used to determine whether the difference between the pre-response and the post-response is likely due to chance. In other words, one alternative explanation is that normal random fluctuations in the level of the problem caused the difference between the pre-response and post-response measures of the problem. Tests for statistical significance are most useful when the differences are small but meaningful and the number of problem events prior to the response is small. In such circumstances, normal random fluctuations in the problem are a potential cause for the change. Because of the highly technical nature of significance testing, it will not be covered in this monograph. Readers interested in significance testing can find explanations in most introductory statistics texts, in the documentation accompanying statistical software, and from statisticians and social scientists at local universities.

As we have seen, this design is weak at eliminating alternative explanations for the decline in the problem. This is because something else may have caused both the response and the decline in the problem, or because other things, occurring at the time of the response, may be responsible for the change in the problem.

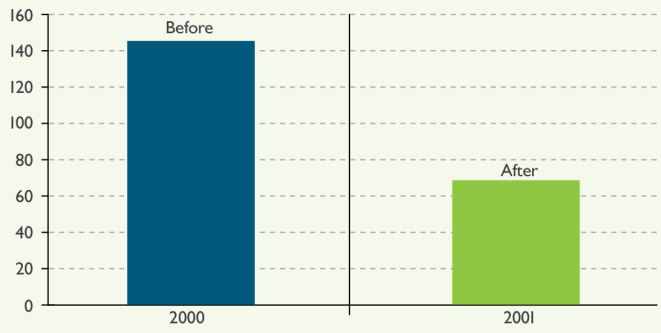
To see why a pre-post design is weak, consider the example shown in Figure 6. The data for this example come from a report on a theft-from-vehicle problem-solving effort. In the top chart of Figure 6 we see a simple pre-post comparison. The question being asked is whether the installation of CCTV in the target area caused a reduction in vehicle thefts. The answer seems to be "yes." In the lower chart we see two more years of theft data. Two things are apparent. The downward tumble in theft-from-vehicle reports begins a year before the CCTV was installed. This calls into question the validity of a conclusion that the CCTV caused the decline. Because pre-post designs do not examine long-term trends, they cannot eliminate the alternative explanation that a decline in the problem was already underway before the intervention.

Although the pre-post design is popular, the example in Figure 6 illustrates its weaknesses. A review of the four criteria of causality makes this clear. In terms of the first criterion (that we must see a plausible explanation of how the response could reduce the problem), this simple design is no better or worse than others. In regard to the second criterion (that we must see a relationship between the response and the decline of the problem), it does not fare so well. If we compare the two panels in Figure 6, our confidence that there is a relationship between the CCTV response and thefts from vehicles goes down when two more time periods are added. Although these thefts did decline after the CCTV was added, we see that the numbers of thefts were going up and down prior to the CCTV. Problems often fluctuate, even if nothing is done about them. This means that peaks are followed by troughs, followed by peaks. Consequently, any effort implemented in a peak period will virtually be guaranteed to look good because the most likely trajectory for the problem after a peak is to go down.[k] This chart raises the concern that this is what could have been going on here. We cannot be sure without more data.

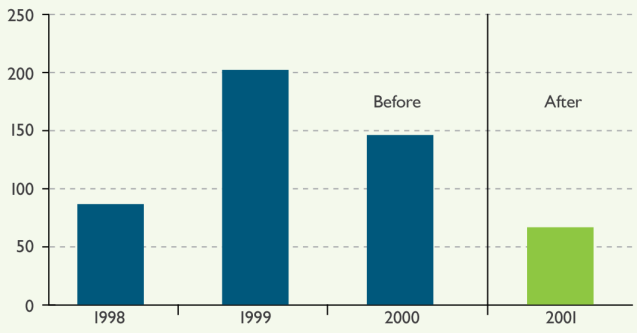
The added years of data also suggest that the third criterion (that we must be sure that the problem did not decline until after the response was applied) has been violated: thefts started going down a year before the CCTV was installed. We do not have technology that allows us to go back in time. So anytime we see a downward trend that begins before the response, we should be suspicious that the response had little or nothing to do with the decline.

Finally, the last criterion (that we need to be sure that nothing else could have caused the decline in the problem) has also not been met. Based on the data shown in Figure 6, we can imagine at least three plausible alternative explanations: (1) that thefts go up and down randomly and the CCTV was introduced while the thefts were dropping; (2) that 1999 was an unnaturally big year for thefts from vehicles, and these crimes just declined to their natural level; and (3) that some other change in the city between 1999 and 2000 created the decline (e.g., an intensive information campaign to warn drivers to remove items from the passenger compartments of their vehicles).

Figure 6: Problems With a Pre-post Design



This example comes from a description of a long-term problem-solving project to reduce theft from vehicles. The authors note that the over 50% decline in thefts from vehicles from 2000 to 2001 is possibly due to the installation of CCTV in the target area at the beginning of 2001.



Fortunately, the authors show theft-from-vehicle data for two years earlier. This illustrates one of the pitfalls of a pre-post design: it does not show the trend. Thefts from vehicles peaked in 1999, then declined 28% in 2000. So it's possible the decline from 2000 to 2001 is in part or entirely due to something that occurred long before CCTV was installed.

These charts were created from data taken from Table 5 (page 30) of Clarke, R.V., & Goldstein, H. (2003). *Theft From Cars in Center City Parking Facilities – A Case Study*. Washington D.C.: Office of Community Oriented Policing Services, U.S. Department of Justice. This table dealt with one small facet of a much larger effort to analyze a problem.

[k] The technical term for this "automatic process" is "regression to the mean."

Pre-post designs are also hard to interpret when the results indicate no change. Without knowing the long-term trend, we do not know whether the problem was trending upward before the response. If it was, and if the problem stopped getting worse following the response, then the response was successful in averting this increase. In this case, the pre-post design gives the false impression that the response was ineffective.

A final difficulty with a pre-post design is that we do not know whether the decline in the problem is sustained. Imagine that you had theft-from-vehicle data for 2002. If these data showed that there were as many thefts in 2002 as there were in 1999 or 2000, we would not be confident that the CCTV installation had made a difference. If the data showed levels of theft that were no higher than they were in 2001, we would be more confident. The longer the reduction can be maintained after the response, the more confident we are in believing that the response is working well, and that the "after" results are not some sort of fluke. It is not uncommon for programs to be successful for a short period and then the problem to bounce back after attention gets diverted to other things.

The Houston example, in Figure 5, is notable because the evaluator used multiple measures of the problem. The consistency of the drop in the problem following the response, across several different measures, gives greater validity to the conclusions. Though it is still possible that something other than the response created the declines, it is less likely that the decline is due to random fluctuations: we would not expect all measures to show the same change if randomness were the cause.

We have illustrated how the common pre-post design works, and described four concerns with interpreting the findings from such designs. All four concerns stem from not knowing the long-term trend. Next, we will examine designs that can overcome these concerns.

Time Series Designs

The time series design is far superior to the pre-post design because it can address these four concerns: there is a plausible explanation, the response is associated with a reduction in outcome, the response comes before the outcome, and the most plausible alternative explanations have been eliminated. With this design, you first take many measures of the problem prior to the response. This allows you to look at the trend in the problem before the response. It also allows you to determine the problem's normal fluctuation prior to the response. You then take many measures of the problem after the response. This allows you to determine the long-term trend in the problem after the response. You can see whether the problem bounces back or stays down. Comparing the before trend to the after trend provides an indicator of effectiveness. This is feasible using police-reported crime data or other information routinely gathered and archived by public and private organizations. It is more difficult to accomplish if you have to initiate a special data-collection effort, such as surveys of the public.

The basic approach is to use repeated measures of the problem before the response in order to forecast the likely level of the problem after the response. If the difference between the measures taken after the response and the forecast are significant and negative, this indicates that the response was effective (see Appendix A).

This type of design provides strong evidence that the response came before the problem's decline, because pre-existing trends can be identified. If the process for measuring the problem has not changed, this design eliminates most alternative explanations for the reduction in the problem.

Note that what matters is the number of measurement periods, not the length of time. So, for example, it is far less helpful to have three years of annual data before the response and three years of annual data after the response than to have 36 months of monthly measurements before and 36 months of monthly measurements after, even though the same amount of time has elapsed.

One might be tempted to take this to the extreme – if monthly data are better than annual data, why not weekly, daily, or even hourly data? The answer is that as the time interval becomes shorter, the number of crimes per time interval becomes too small to use for deriving meaningful conclusions. If the number of events is extremely large (as is sometimes the case when using calls-for-service data for large areas), very short intervals might be useful. But if the number of events is very small (like homicide, stranger-stranger rape, or vehicular-accident deaths in a modest-size city), one might have to use large time intervals.

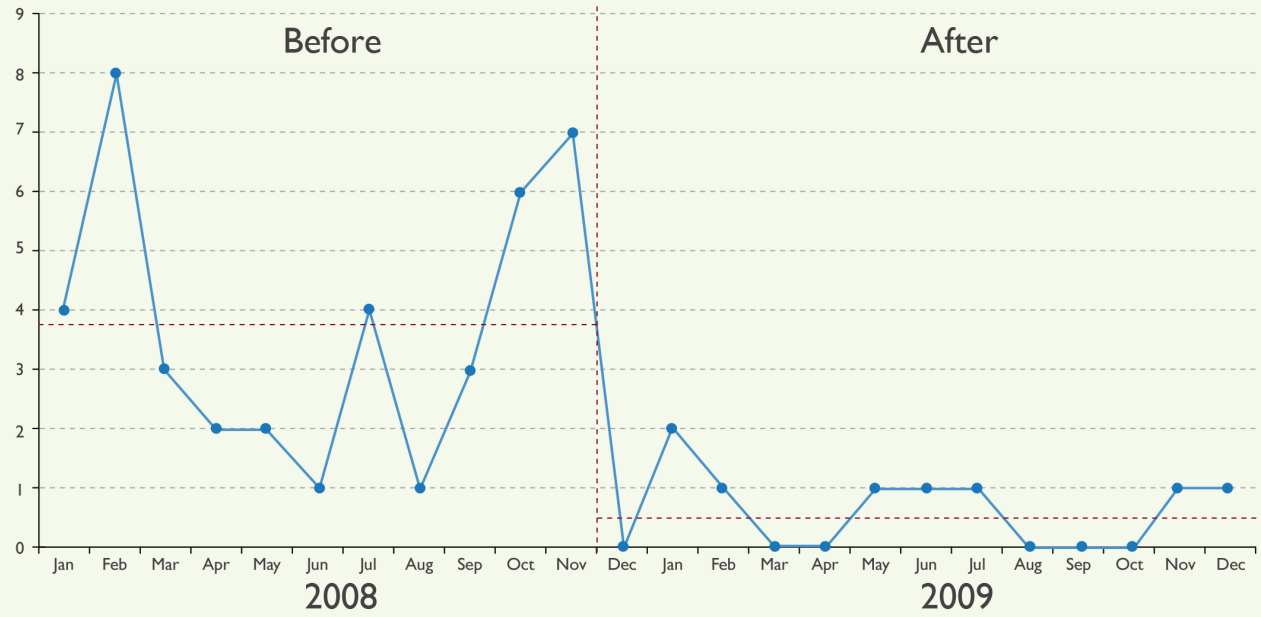
Figure 7 illustrates a simple time series design, and contrasts it to a pre-post design. This example comes from a report of the South Yorkshire Police problem-solving effort designed to combat metal theft. One form of metal theft was stealing heating boilers in residential buildings. The figure shows the frequency of such burglaries over 24 months (11 before the response and 13 after). Note the high variation in burglaries of this type prior to the intervention. A simple pre-post comparison (on the right) does not capture this, so leaves the assessment vulnerable to the problems noted earlier. It is clear from the time series chart, interrupted by the line showing when the response began, that both the number of these burglaries and the fluctuation in their numbers declined considerably, following the response. This is more convincing evidence that the trend, natural fluctuation, or lack of sustainability is unlikely to be responsible for the decline.

A comparison of the average level of the problem before and after shows a decline in the problem following the response. If the problem had been trending upward, an upward-sloping projection would have been used and the slope would have to be calculated (an example of this is illustrated in Appendix A). The more before-response time periods examined, the more confident you can be that you know the trajectory of the problem prior to the response. The more time periods examined after the response, the more confident you can be that the trajectory changed. The calculations involved in the analysis of an interrupted time series design can become quite involved; thus, if there is a great deal riding on the outcome of the evaluation, it may be worth seeking expert help.

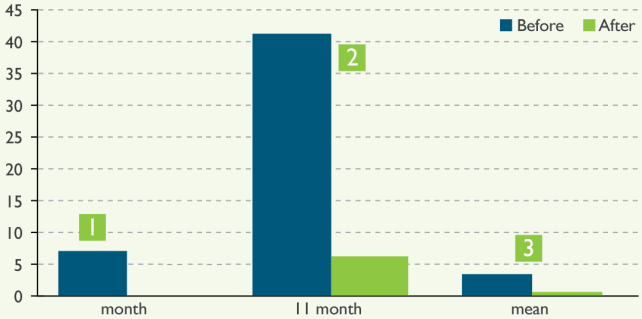
Ideally, the only difference between the time periods before the response and the time periods after the response is the presence of the response. If this can be assured, the conclusions based on this design have a high degree of validity.

The major weakness of the interrupted time series design is the possibility that something else occurred at the same time the response began and was actually what caused the observed change in the problem. To eliminate this alternative explanation, a second time series for a control group can be added (see Appendix B).

Figure 7: Impact Measurement in an Interrupted Time Series Design



The South Yorkshire Police chose to depict some of their assessment results as an interrupted time series chart (above). This chart shows 11 months of burglaries in which a heating system boiler was stolen before the intervention, and 13 months after the response. The impact of the response is evident, despite monthly fluctuations. Before the response, there were only two months with only one burglary, and most months had three or more. After the response, over half the months had zero burglaries and only one month had as many as two. The average number of burglaries, before (3.7) and after (0.5), are shown by horizontal dashed lines. Not all time series charts show such clear improvements.



If the South Yorkshire Police had used a pre-post design, they would have had at least three choices (left). 1) They could have selected the months before and after the response, showing a decline from 7 to 0. This exaggerates the impact as it implies they wiped out the problem when they did not. 2) They could have shown the total over 11 months before and 11 months after. 3) Or they could compare the average (over 11 months) number of burglaries, before and after. All three hide the monthly variations and none show that trend and natural variation are unlikely causes of the reduction.

Charts derived from: South Yorkshire Police (2010). *Shiregreen Alliance Partnership*. Goldstein Award Submission. Annual Problem-Oriented Policing Conference. Arlington, TX. September. Table 4. Accessed at http://www.popcenter.org/library/awards/goldstein.cfm?browse=abstracts&year=2010.

Even if you are only interested in determining whether the problem declined (and have little interest in establishing what caused the decline), an interrupted time series design is still superior to a pre-post design. This is because an interrupted time series design can show whether the problem declined and stayed down. As noted above, problems can fluctuate; thus, it is desirable to determine the stability of the decline. The longer the time series after the decline, the more confident you can be that the problem has been eliminated or is stable at a much lower level.

Though interrupted time series designs are superior to pre-post designs, they are not always practical. Here are five common reasons for this:

- Measurement is expensive or difficult.
- Data for many periods before the response are unavailable.
- Decision makers cannot wait for sufficient time to elapse after the response.
- Data-recording practices have changed, making inter-period comparisons invalid.
- Problem events are rare for short time intervals, forcing one to use fewer longer intervals.

Under these conditions, a pre-post design might be the most practical alternative.

Combining and Selecting Designs

Though we have examined these designs separately (here and in Appendix B), in many circumstances it is possible to use two or more designs to test the effectiveness of a response. This is particularly useful if you have several measures of the problem (for example, reported crime data and citizen-survey information) collected for different periods. A combination of designs selected to rule out particularly difficult to disprove alternative explanations can be far more useful than strict adherence to a single design.

Appendix C provides a structured checklist for selecting the design most appropriate for your problem. Appendix D summarizes the strengths and weaknesses of the designs discussed here and in Appendix B.

In considering what type of design or combination of designs to select, it is important to keep in mind that you cannot eliminate all alternative explanations for the reduction in the problem. Time, money, and evaluation expertise all argue for selecting the simplest design that eliminates the most obvious rival explanations. Consequently, it is useful to anticipate the most credible alternative explanations before you select an evaluation design. Once again, your analysis of the problem should give you some insight. It is also useful to listen to the most articulate critics of the response. Then, while planning the assessment, you can collect data and develop designs that address their concerns.

Examining How the Response Works

Many problem-solving responses comprise multiple parts, any of which might be effective, and some of which might not be. Further, as we noted when discussing process evaluations, sometimes parts of the response do not get implemented, or are implemented poorly. Gathering and examining evidence about implementation, as well as the plausibility of alternative explanations, helps determine what features of the response (if any) took a bite out of the problem, and which were toothless. This can be illustrated by going back to our hypothetical example of an effort to reduce street prostitution.

Table 3 lists a variety of explanations for *how* prostitution activity may have gone down. Such explanations are called "mechanisms." As explained earlier, mechanisms are plausible ways in which the response could reduce the problem. Mechanisms describe *how* the response works. The first five mechanisms in the table are based on the planned response. The final two are mechanisms by which alternative explanations could have reduced the problem. The second column describes what happened (or failed to happen). So, for example, in the first row we see that despite the plans, not many sex workers were arrested. And in the sixth row we see that an unplanned byproduct of the street reconstruction was the presence of road workers. The third column shows the evidence supporting or contradicting the presence of the mechanism. The last column summarizes our conclusions about the likelihood that each mechanism impacted the problem.

This table shows that most parts of the planned response probably had no impact on the problem. One part may have been responsible – the street reconfiguration. Further, the table suggests that two other alternative explanations must be considered: the presence of construction crews, and the pressure from the neighbors. It also could be a combination of these three things. The construction-crew mechanism could be refuted if the prostitution activity has stayed low long after the crews have left. If it returns, however, this might be the best explanation.

Examining the response by breaking it down into its component mechanisms and examining mechanisms from rival explanations is useful if you plan to use the response again, or even if you just want to maintain the response. Here, we might stop efforts to arrest and divert sex workers. Instead, we might monitor the impact of the street reconfiguration. If we were to use this response elsewhere, we might want to proactively mobilize neighborhood residents as part of the analysis and response stages.

Table 3 illustrates a simple procedure for examining mechanisms. The first six responses deal with the planned response and its mechanism. This helps you assess which parts of the response might have worked, and which might not have. The seventh item is an alternative to the response that might have been responsible for the change in the problem. This helps you determine whether there were reasons for the change in the problem that were not planned parts of the response. Your objectives are: (1) to eliminate the least plausible mechanisms, (2) to draw a reasonable conclusion about whether the response could have caused the change, and (3) to determine what other alternative mechanisms might also have been at work. This gives you a basic assessment of what could have caused the change in the problem. It also tells you something about whether you are likely to get the same change in the problem if you use the response again. Since you usually cannot depend on the alternative explanations operating again, the more the evidence weighs in favor of these alternatives, the less likely it is that repeated use of the response will be effective.

Displacement and Diffusion of Benefits

A common concern about problem-solving responses is that they will result in spatial displacement—the shifting of crime or disorder from the target area to nearby areas that are not being treated. This concern is probably not as great as is commonly imagined.[7] Though displacement is far from inevitable, it is a possibility that needs to be investigated.

There is increasing evidence that some responses have *positive* effects that spread beyond their target areas.[8] This is called "spatial diffusion of crime-prevention benefits." Though not all responses create benefits beyond those planned for, some do, and this possibility must also be addressed in evaluations. If we do not account for these possibilities, we could produce misleading results.

We will not examine displacement or diffusion here. See Problem-Solving Tool Guide No. 10, *Analyzing Crime Displacement and Diffusion,* for further discussion of displacement and diffusion of benefits, how to detect them, and how to measure their impact.

Table 3: Response May Have Triggered One or More of These Mechanisms to Reduce Prostitution Activity

RESPONSE USED	MECHANISM: HOW	PLANNED ACTIONS: WHAT	PROCESS EVIDENCE	CONCLUSION
Arrested sex workers	Deterred sex workers	Crackdown occurred late	Few arrests recorded	Unlikely
Banned convicted sex workers from area	Deterred sex workers	Few sex workers prosecuted	Few sex workers received banning orders	Unlikely
Provided legitimate exit from sex work	Reduced number of sex workers	Diversion plan implemented	Few requests for assistance	Unlikely
Provided sex workers legitimate employment	Reduced number of sex workers	Diversion plan implemented	Few sex workers found legitimate employment	Unlikely
Altered road network and traffic flow	Made it difficult for johns to reach sex workers	Traffic changes implemented	Traffic counts on side streets changed	Plausible
Reconstructed the street	Disrupted solicitations	Street construction work implemented	Observed that much construction activity occurred	Plausible
Applied pressure on property owners by local residents	Denied sex workers places for work	Not part of project, but occurred about the same time	Interviews with residents and property owners confirmed pressure was applied	Plausible

CONCLUSIONS

This guidebook has introduced some basic principles for assessing the effectiveness of responses to problems. All evaluations require valid measures of the problem that are systematically taken before and after the implementation of a response. There are two possible goals for any problem-solving evaluation. The first is to demonstrate that the problem declined sufficiently. This is the most basic requirement of an evaluation. For this goal, we are not concerned about whether the reduction was directly caused by the response or by something else entirely.

In many circumstances, it is also useful to determine whether the decline in the problem was due to the response. This is a second goal. If one anticipates using the response again on similar problems (or on the same problem if it returns), it is important to make this determination. This requires an evaluation design that can eliminate the most likely alternative explanations for the decline in the problem. Elimination of those explanations requires either the use of an interrupted time series design or the use of a control group (Appendix B). The control group tells you what the level of the problem is likely to have been in the absence of this problem-solving effort.

The results of an impact evaluation should be compared to the results of a process evaluation in order to form a detailed picture of whether the response was implemented as planned and what impact it had on the problem. This information helps show whether the response was the cause of the decline in the problem, and what parts of the response are the "active ingredient."

A recurring theme in this guidebook is that an evaluation design builds on knowledge gained during the analysis of the problem. Competent evaluations require the evaluators to have detailed knowledge about the problem in order to develop useful measures and to anticipate possible reasons for the decline in the problem following a response.

The evaluation of responses can be extremely complex. This guidebook is only an introduction. For small-scale problem-solving efforts, where the costs of mistaken conclusions are not serious, and weak causal inferences are tolerable, the information contained here should be sufficient. If, however, there is a great deal riding on the outcome, if it is important to show whether the response caused a drop in the problem, or if there would be serious consequences from drawing the wrong conclusion from the evaluation, you should seek professional assistance in developing a rigorous evaluation. A decision to enlist the support of an outside evaluator should be made as soon as possible once a problem has been identified so that adequate before-response measures can be made and a rigorous design can be developed.

APPENDIX A:

THE EFFECTS OF THE NUMBER OF TIME PERIODS ON THE VALIDITY OF EVALUATION CONCLUSIONS

To understand the importance of examining a large number of time periods, consider the following hypothetical example. The data here were created using a random number generator, so none of the fluctuations are systematic. This series illustrates how we can be deceived by randomness, particularly if we look at very short time intervals. All the charts that follow are from the same series.

Figure A1 shows the results of a pre-post evaluation where measures of the problem are taken just before and just after a response (time periods 19 and 20 in the series). The conclusion we would draw from this chart is that the problem experienced a moderate decline following the response.

The next chart (Figure A2) shows periods 12 through 20 of the series, so there are now eight periods before the response and one after the response. The additional time periods provide an opportunity to examine the trend in the problem leading up to the response. The straight line shows this trajectory. The extension of the trajectory to period 20 allows a comparison of what we might expect if the response were not implemented (the trajectory) to the actual level of the problem.

We can see plainly that the problem was trending downward prior to the response, so not all of the drop in the problem following the response can be attributed to the response. Nevertheless, it appears that there was a greater drop in the problem following the response than we would have expected due to the trend alone.

The periods prior to the response help establish the trajectory of the problem time series. Here we focused exclusively on the overall trend, but it is also possible to look for seasonal cycles and other recurring fluctuations.

Extending the data to periods after the response helps determine the stability of the response. Does the response continue to be effective, driving the problem further down? Or does the response wear off, allowing the problem to rebound? This is shown in Figure A3, which depicts an additional seven periods following the response. The same trend line is used based on the data prior to the response—but now projected out eight time periods after the response.

Figure A1: Two-period Pre-post Design

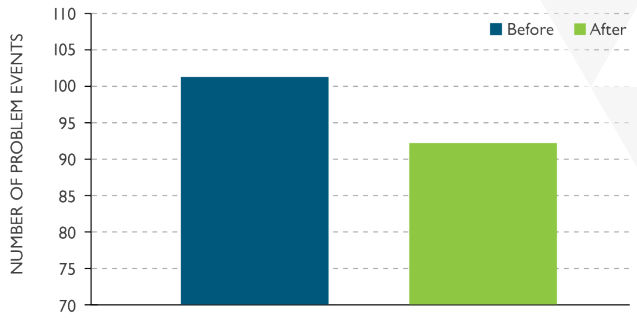
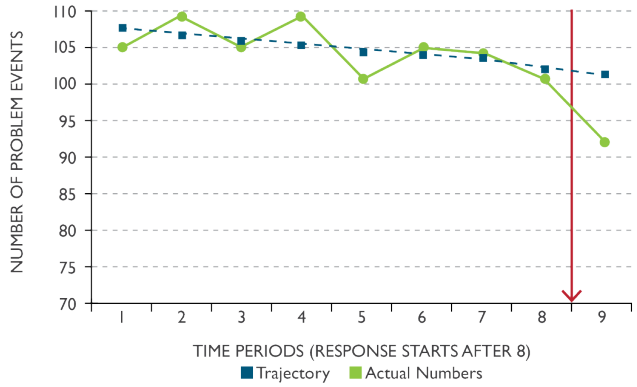


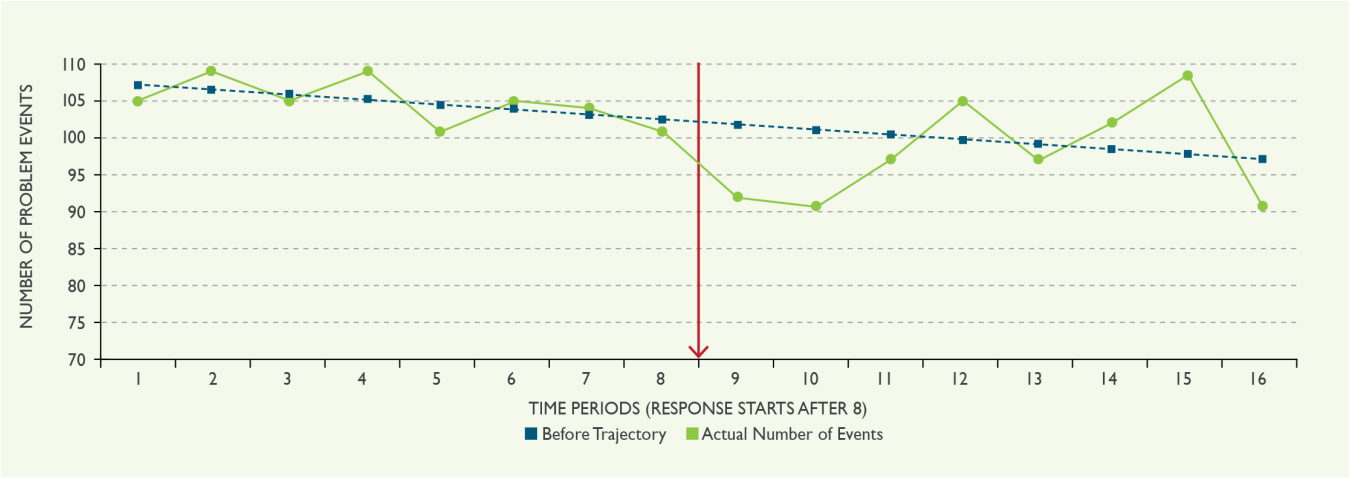
Figure A2: Nine-period Time Series Design
(with projected trajectory of problem)



We see that the problem rebounded and then seems to oscillate around the same trend line. So at best, the response was temporarily helpful.

The value of a very long time series cannot be overstated. Too often police agencies show only a few time periods even though their computer systems contain data for many more. Note how our interpretation of the trend changes when we look at the entire 40-period series from which these three charts were extracted. This is shown in Figure A4.

Figure A3: Sixteen-period Time Series Design *(with projected trajectory of problem)*

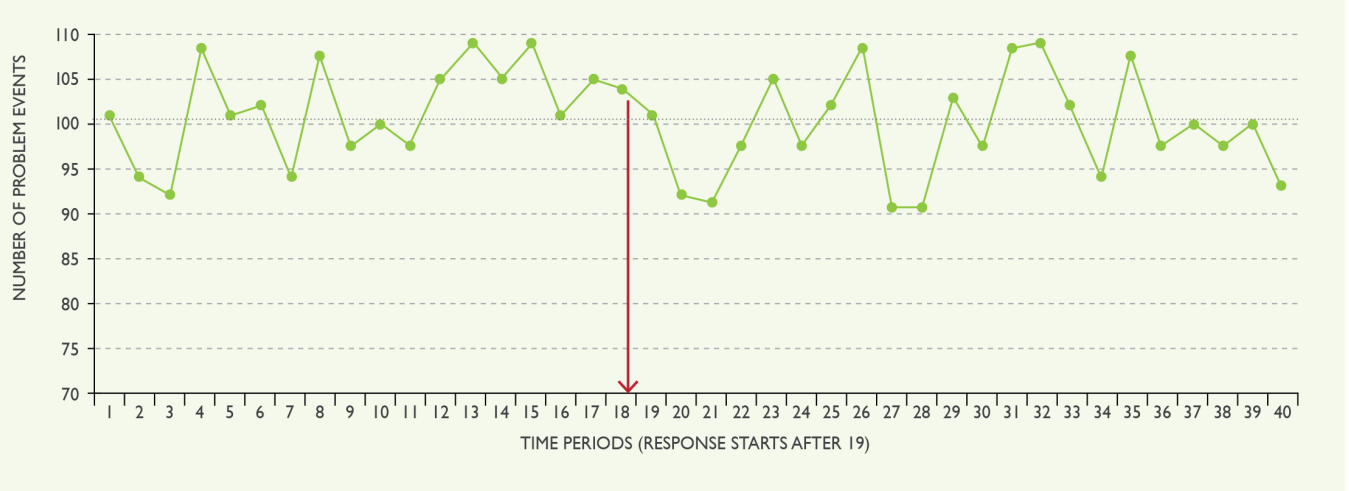


Whereas the trend in Figure A3 suggests a downward trend, Figure A4 shows that this was an illusion; in fact, the longer-term trend is flat. The underlying randomness of the data becomes much more apparent. It oscillates around 101 events per period (dotted line). Further undermining our confidence in the response, we can see that there are at least two other intervals with declines like we see after the response. So it appears that what we thought was a decline due to the response may very well be a temporary fluctuation due to normal variation in the problem.

The lesson here is that it is easy to be deceived by randomness, particularly when analyzing crime over short time periods. The police, public, news media, and elected officials are all susceptible to this deception, because comparing this month to last month, or this year to last year is so common. Using multiple measures and using a longer time series are reasonable guards against this sort of deception.

Unlike real data, where one is never quite sure of the cause, with this intentionally random data we know with absolute certainty that the variation around the 101 events per period average is random.[1] This includes the periods just before and after the response. The example illustrates the point that random fluctuations in data can be easily misinterpreted as meaningful changes. It is worth noting that a significance test to detect randomness in a pre-post design might actually suggest that the drop is *not* due to random changes. This is because the randomness affects the entire series and the pre-post design only looks at a small part of the series.

Figure A4: Forty-period Time Series Design *(with average number of events per period)*



[1] We created this data series by setting a constant level for the problem and then used a random number generator to provide the fluctuations around this level. We placed the beginning of the hypothetical response at the center of the series.

APPENDIX B:
DESIGNS WITH AND WITHOUT CONTROL GROUPS

The designs in the main body of the text focus on data for the group of people or the area receiving the response. To determine whether the response is the cause of a drop in the problem, it is helpful to use a control group. Also, control groups are critical to obtaining reasonable estimates of the amount of spatial displacement and diffusion of benefits (see Problem-Solving Tools Guide No. 10, *Analyzing Crime Displacement and Diffusion*). Control groups can be added to either the pre-post design or the time series design.

In this appendix, we will look at five designs, including the two examined in the body of this guide. We will use data from an evaluation of a problem-solving effort to reduce injurious and fatal vehicle crashes in Cincinnati. The evaluation used a multiple time series design and a very complex statistical analysis process to get a precise estimate of the number of lives saved and injuries averted by implementing a response to injury-related traffic accidents. The authors found that such accidents declined 5.7 to 10.3 percent in Cincinnati compared to the comparison areas.9 This evaluation was possible because of a long-standing partnership between the Cincinnati Police Department and the University of Cincinnati's Institute of Crime Science (based within the School of Criminal Justice).[m]

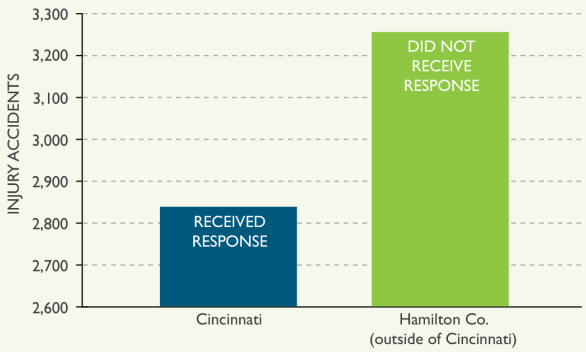
Here, we will not replicate the analysis conducted in the published paper. Instead, we will use the data to illustrate how conclusions about the effectiveness of the response can change, depending on the evaluation design used. We will start by using the data on Cincinnati traffic accidents to illustrate a design that should *not* be used: a static comparison group. This will be our baseline. We will then show why the pre-post design is an improvement. Then we will show why a control group is useful. Following this, we will return to the time series design. We will conclude by showing a time series design with a control group. This brief tutorial is an introduction to evaluation designs, meant only to illustrate their basic logic.

STATIC COMPARISON DESIGN

Let's assume that a year after the Cincinnati Police Department's Traffic Division launched its response, which was meant to decrease the number of injury-related traffic crashes, you are asked to determine whether it made a difference. A common design, which is not recommended, is to compare the number of injury accidents in Cincinnati to the numbers of those in nearby jurisdictions which did not use the response. The logic is that these nearby agencies would be exposed to the same traffic conditions and drivers, so they should have a similar level of accidents. That is, you are assuming that if the response worked, Cincinnati should have fewer accidents than the comparison, and that if Cincinnati had not used the response, its level of accidents would be similar to that of the comparison area.

Figure B1 shows the results. Cincinnati is contained within Hamilton County, so Hamilton County (without Cincinnati) is the comparison. Dividing the number of accidents over a 12-month period by the driving population of each jurisdiction (or road miles driven in the areas) would control for population differences. We do not do this here, for a simple reason: The principal problem with this design is that the comparison area is systematically different from the response area (they have different driving populations, there are more highways in one area than the other, the population is older in one area than in the other, and so on). Population is just one area in which there can be many differences.

Figure B1: Static Comparison Design



In a Static Comparison Design, you compare the problem in an area or group that received a response to a similar area or group that did not. The time period for both is after the response.

The area or group not receiving a response provides an indicator of the level of the problem in the response area or group, if the response had not been applied.

Here, the Cincinnati Police Department's response takes place only within the city. Comparing the 12 months following the response in Cincinnati to the same 12 months in the surrounding county shows fewer injury accidents.

Though it seems to show the response works, it is weak evidence. This is because Cincinnati usually has fewer accidents involving injuries.

[m] Thanks to Drs. Nick Corsaro and Robin Engel of the Institute of Crime Science, within the School of Criminal Justice, University of Cincinnati, for making these data available. The Institute of Crime Science provides scientific consulting services to police and other law enforcement agencies, including complex evaluations.

You should avoid using this type of evaluation design as it has a high risk of producing misleading results. *How* misleading can be appreciated by comparing the results in Figure B1 to the results in the next set of figures, which illustrate better designs.

PRE-POST WITHOUT A CONTROL GROUP DESIGN

We discussed this design in the main body of the guide, so will revisit it only briefly here. Figure B2 shows the results of the evaluation of the response that was designed to lower the incidence of traffic injuries in Cincinnati. The comparison is between 12 months before the response and 12 months after. We use a full-year comparison because it controls for seasonal changes in accidents. A shorter period (e.g., the September before the response to the September after it) is highly susceptible to random changes in accidents that a response cannot address. With this design, we act as if the before measure is an accurate indicator of the number of accidents Cincinnati would have had, if no response had been applied. Therefore, the difference between the before and after measures of the problem is an indicator of the reduction due to the response.

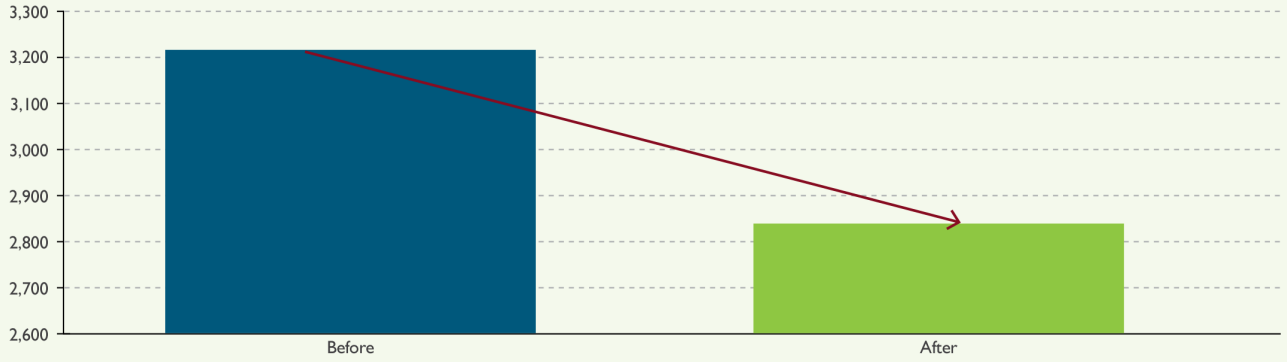
Pre-post designs are simple. They are most useful if your principal interest is in determining whether the problem did decline and you are not going to make a strong claim that the response was a major cause of the decline.

PRE-POST WITH A CONTROL GROUP DESIGN[N]

If we combine the static comparison design's use of a control with the pre-post design's use of a pre-response measure of the problem, we can improve the evaluation. The control area or group does not receive the response, even though it has a problem similar to that of the area or group that receives the response. The purpose of the control group is to demonstrate what would have occurred if no response had been taken. Knowing this can help you eliminate some alternative explanations for the decline in the problem.

This design is illustrated in Figure B3. Here we see that the county outside Cincinnati had a decline in injury-causing vehicle crashes from before to after the response inside Cincinnati. This indicates that even without a response, Cincinnati might have experienced a similar decline. However, Cincinnati's decline in vehicle-injury accidents is greater than the decline in Hamilton County (over 40% greater). This indicates that the response in Cincinnati contributed to the general decline.

Figure B2: Pre-post Design



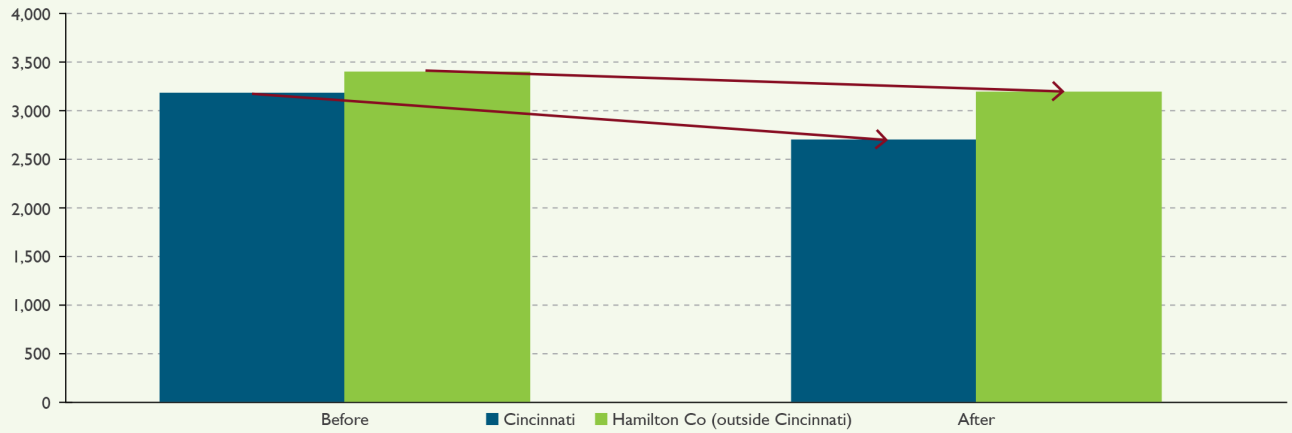
In a Pre-post Design (also called a before-after design), you compare the level of the problem after the response to the level of the problem before the response. You do not use a comparison, or control, area. You assume that the before measurement of the problem is a reasonable indicator of what the problem would be like, if you had used the response. So a meaningful decline in the problem from before to after is an indicator of success.

Here, the number of crashes with injuries is higher before the Cincinnati Police Department's response than it is after. This is consistent with an effective response.

The limitations are obvious. First, something else could have occurred at about the same time as the response that could have caused the decline in accidents. Second, it is possible that injury-related accidents were trending downward before the response was implemented.

[n] This design is usually referred to as a "non-equivalent control group design" to draw attention to the fact that members of the treatment (response) group and members of the control group may be different in important ways that could affect the outcome of the evaluation.

Figure B3: Pre-post with Control Design



This design combines features of the static comparison and pre-post designs. You compare the before-after differences of the two groups. Here, injury accidents in Cincinnati declined 11.6% from before the response to after. The Hamilton County injury accidents declined only 4.8%. We use percent decline because the two areas had different numbers of accidents before the response.

The Hamilton County decline (4.8%) is assumed to be the decline Cincinnati would have received, if it had not engaged in a problem-solving response to injury accidents. To determine the impact of the response, we subtract 4.8 from 11.6 (since the 4.8% decline would presumably have occurred whether or not a response was implemented). The Cincinnati police response may have created a 6.8% decline in injury-related vehicle crashes.

The reason this design is better than the static comparison and pre-post design is that it removes from consideration many alternative causes for the Cincinnati decline. For example, changes in state law or gas prices would impact both the county and the city. So they cannot explain the different rates of decline. Similarly, if injury accidents were trending downward throughout the state, this too would influence both of these jurisdictions, so it cannot explain the difference.

The principal limitation of this design is that there might have been something (other than the response) that occurred in Cincinnati and not the county (or vice versa) that pushed injury accidents down in the city.

Whereas in a pre-post design effectiveness is measured by calculating the percent change, when a control group is used we compare the difference between the percent declines, as illustrated for this example in Table B1. Here we see that the control area had a reduction of 164 crashes, which, when divided by the before number (3,421), is a 4.8 percent drop. Cincinnati had a reduction of 375, which, when divided by the before number (3,215) is an 11.6 percent drop. Subtracting the percent decline in the control from the percent decline in the response yields a net reduction of 6.8 percent (dividing -6.8 by -4.8 shows that the Cincinnati drop was almost 42 percent greater than the county's drop).

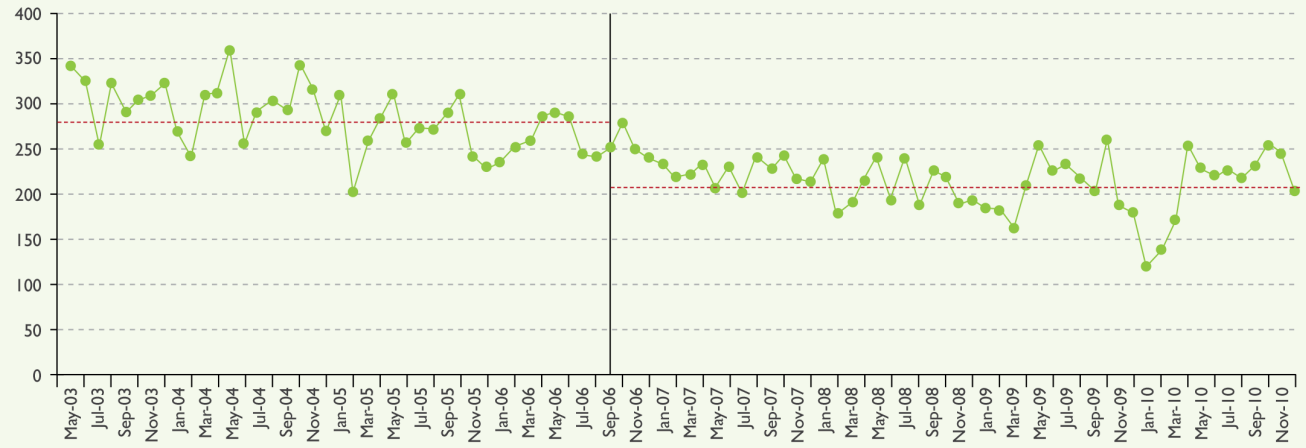
Table B1: Calculating Effectiveness with a Pre-Post with Control Design

	Before	After	Difference	% Difference
Control (County)	3,421	3,257	-164	-4.8
Response (Cincinnati)	3,215	2,840	-375	-11.6
Response % difference – Control % difference				-6.8

TIME SERIES DESIGN

This design was also discussed in the main body of the guide. Figure B4 shows the time series for the Cincinnati vehicle crashes with injuries. The horizontal dashed lines indicate the average (mean) number of such crashes per month, before and after the response. The before data are used to determine what might have occurred in Cincinnati if no response had been made. Here a simple comparison of these averages suggests an effective response. Typically, an analyst will use a more-complex statistical procedure to remove the effects of trends (here downward) and seasonal cycles. This gives a more-precise estimate of the impact. However, this type of analysis is far beyond what can be explained in this introductory guide.

Figure B4: Time Series Design



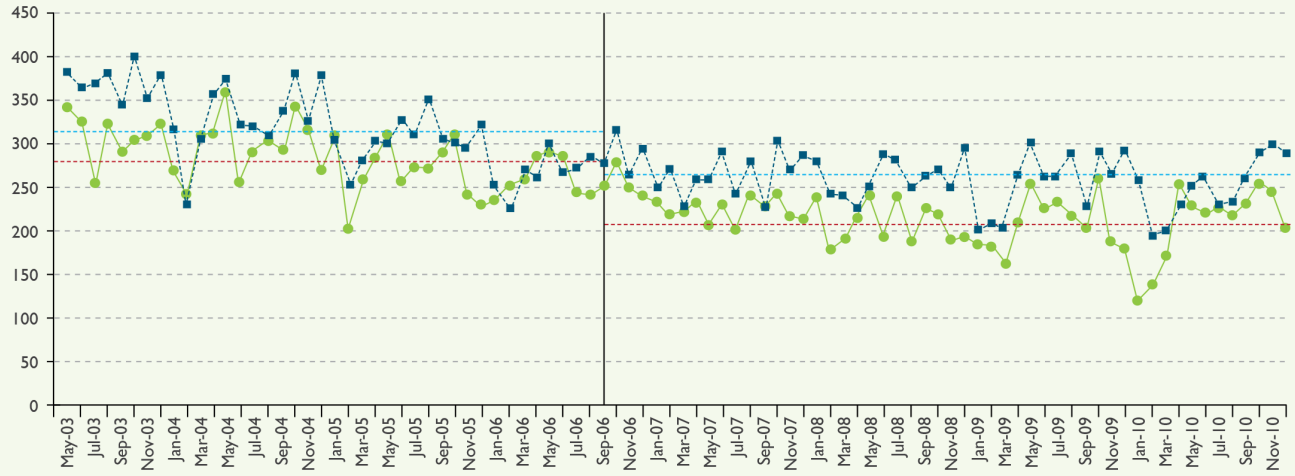
A time series design stretches the pre and post measurement. Instead of comparing two twelve-month periods, here we look at the monthly changes for over seven years. This allows us to see trends and cycles. Two things become obvious. First, the number of accidents jumps up and down a great deal. This is often the case with police problems, and is the reason we should be wary of month-to-month comparisons: random fluctuations will dominate any systematic changes. Prevention is aimed at producing a systematic change, which the random fluctuation can hide. The second thing we can see is that there is a slow downward trend from May 2003 to around May 2009. So even if the Cincinnati police did nothing new, accident injuries would have gone down.

The vertical line shows when the response began. Before the response, there were 287 crashes in an average month (dashed line). After the response, there were only 220 crashes in an average month. Typically, a complex statistical analysis of these data are used to tease out the impact of the response. Time series analysis can distinguish the real response effect from trends and seasonal cycles, as well as other measurable factors (e.g., changes in gas prices).

MULTIPLE TIME SERIES DESIGNS

When two or more times series are used the design is called a multiple time series. This design can rule out most other possible alternative explanations for the change in the problem. Figure B5 illustrates a multiple time series. This example illustrates the usefulness of adding a control time series. If we had simply looked at Figure B4, we could legitimately have assumed that much of the decline in Cincinnati's injury-related crashes could have been due to the downward trend that preceded the response. In Figure B5 we see that the trend influenced the surrounding county as well as the city. The differences in the average number of crashes per month between the city and county grew larger after the response: before, the county had an average of 35 more crashes than the city in a month; after, the county had an average of 42 more crashes than the city in a month. Like the simple time series design, analysts use highly complex statistical techniques. The study these data come from illustrates some of the complexity involved.

Figure B5: Multiple Time Series Design



A multiple time series design adds one or more control areas to the analysis. This helps eliminate possible causes unrelated to the response. Here, both the city and county had very similar trends prior to the response, though the city tended to be consistently lower than the county in injury accidents (about 287 v. 322 crashes per month). After the response, the gap between the county and the city grew (220 v. 262 crashes per month). No factor common to both jurisdictions could account for this, so the only explanation is that something different occurred in Cincinnati, relative to the county. The response is one such explanation.

Like the time series design, complex statistical analysis is needed to separate the impact of the response from the impact of other factors and random fluctuation.

The principal advantage of using a multiple time series design is that it can eliminate a large number of alternative explanations for an improvement in the problem. The only possible alternative to the claim that the response caused the decline is that something occurred in Cincinnati at about the same time as the response was implemented, and this thing did not occur in the county (or it occurred in the county but not the city). So the results of a multiple time series design, though solid, are not certain. However, for practical purposes, these results probably exceed the level of certainty we need in order to consider the response to have been successful.

APPENDIX C:
PROBLEM-SOLVING ASSESSMENT CHECKLIST

This checklist provides a summary of the issues that need to be considered when you evaluate a problem-solving effort. It is general guidance, not a set of rigid rules. It will be most useful if you use it throughout the problem-solving process, beginning in the scanning stage.

I. PRE-EVALUATION CONSIDERATIONS

The items here should be considered during the scanning, analysis, and response stages.

A. WHAT DECISION WILL THE EVALUATION HELP YOU MAKE?

- ❑ 1. Should this problem-solving effort be continued? If this is the only question the evaluation will help decide, a simple evaluation design will be sufficient (see Question III.A).
- ❑ 2. Should this response be used for similar problems (either by your agency or by other agencies)? If checked, you should consider using a control group in the impact evaluation design (see Question III.A).
- ❑ 3. No decision. If checked, an evaluation will not be helpful. Stop here.

B. DO YOU KNOW THE PROBLEM?

These questions help develop a cost-effective response and evaluate the response. If you cannot answer these questions with some precision, you need to do more to analyze the problem.

- ❑ Who is harmed by the problem? Who is not harmed?
- ❑ How can the harm be measured?
- ❑ Where does the problem occur? Where is the problem absent?
- ❑ When does the problem occur? When doesn't it occur?
- ❑ What causes the problem? What prevents or suppresses it?

C. DO YOU KNOW HOW THE RESPONSE WORKS?

These questions need to be answered in order to determine whether the response is likely to be effective, and to assure accountability during implementation. If you cannot answer them, plans for the response are inadequate and more attention needs to be put into the response stage.

- ❑ How does the response influence the causes of the problem?
- ❑ Who is responsible for carrying out the response?
- ❑ When is the response supposed to be implemented?
- ❑ Where is the response supposed to be implemented?
- ❑ How long will it be before the response is likely to have a noticeable impact on the problem?
- ❑ Who has the legal authority to carry out the response?
- ❑ What are the likely barriers to implementing the response?

II. PROCESS EVALUATION

Process evaluations begin toward the end of the response stage and continue well into the assessment stage.

A. WAS THE RESPONSE IMPLEMENTED?

The closer the actual implementation is to the planned response, the greater confidence you have that the response was the cause of the changes in the problem detected by the impact evaluation (section III, below). The more variation between what was intended and what occurred, the greater the likelihood that changes in the problem are due to things other than the response.

- ❑ Was it implemented when it was supposed to be implemented?
- ❑ Was the response implemented where it was supposed to be implemented?
- ❑ Was the response implemented for the appropriate people?
- ❑ Was the implementation carried out as planned?

B. WAS ENOUGH OF THE RESPONSE IMPLEMENTED?

The response might be implemented as planned, but lacked the resources, duration, or intensity needed to make it effective.

- ❑ Were there sufficient resources available to fully implement the response?
- ❑ Was the response carried out for a sufficient duration to have an impact?
- ❑ Was the response carried out with sufficient intensity?

III. IMPACT EVALUATION

Many of the decisions needed to conduct an impact evaluation must be considered in the analysis and response stages. This is particularly the case with measurement decisions.

A. DO YOU NEED A CONTROL GROUP?

Answering these questions helps decide on the complexity of the evaluation design.

- ❑ 1. Did you check Question I.A.1? If YES, then you do not need a control group.
- ❑ 2. Did you check Question I.A.2? If YES, then you should use a control group.

B. HOW OFTEN CAN YOU MEASURE THE PROBLEM?

Answering these questions helps you decide whether a time series design is possible.

- ❑ 1. Can you measure the problem consistently for many time periods before (at least 10 periods, though at least 30 is better) and after (at least 10, but 30 is preferred) the response was implemented? If checked, a time series design is feasible.
- ❑ 2. Can you only measure the problem a few times before the response and a few times after the response? If checked, a time series design is not feasible and some form of a pre-post design will be needed.
- ❑ 3. Do you have some measures of the problem that can be examined for many time periods before and after the response, and other measures that can only be examined for a few time periods before and after the response? If checked, then you can use a time series design *and* a form of a pre-post design.

C. WHAT TYPE OF EVALUATION DESIGN SHOULD I PICK?

Your answers to questions in A and B, immediately above, provide some basic guidance answering this question. This is shown in the following table. Obviously, precise answers depend on the particular circumstances of each problem-solving effort.

Table C1: Which Evaluation Design Makes the Most Sense?

III.B. Question Checked	III.A. Question Checked	
	1.	2.
1.	Interrupted Time Series Design	Multiple Interrupted Time Series Design
2.	Pre-post Design	Pre-post Design with a Control Group
3.	Combination of Designs Above	Combination of Designs Above

D. WHAT TYPE OF CONTROL GROUP DO I NEED?

(only applies if one of the designs from column 2 in Table D1 was selected. If the design comes from column 1, skip this section and go directly to part IV.)

- ❑ 1. Will the response be applied in an identifiable geographic area (place, neighborhood, etc.)? If YES, the control group should be a very similar geographic area that has similar problems that will not get the response, and preferably is located at some distance from the problem area (to prevent the response from leaking into the control area and contaminating it).
- ❑ 2. Will the response be applied to a set of identifiable potential victims (young males, elderly women, commuters, etc.)? If YES, the control group should be a very similar group of potential victims that will not be given the response.
- ❑ 3. Will the response be applied to a set of identifiable potential offenders? If YES, the control group should be a very similar group of potential offenders that will not be given the response.
- ❑ 4. Will the response be applied to some other identifiable set of people or things? If YES, the control group should be a very similar group of potential people or things that will not be given the response.
- ❑ 5. It is impossible to identify a control group for this evaluation. If this is checked, go back to Table D1 and pick the appropriate design from column 1. Then proceed to part IV.

If one of the questions 1 through 4 were answered YES, systematically compare the characteristics of the response group to the characteristics of the control group, and list the major differences. In the last part of this checklist (part V) you are asked to make a judgment regarding the possibility that other factors might have caused the change in the problem relative to the control. Your list of differences is a list of potential other factors that could account for the change in the problem.

IV. DRAWING CONCLUSIONS

These items fall within the assessment stage and are applicable once evaluation results have been documented. These questions are designed to help formulate conclusions that are consistent with the results from your process and impact evaluations and with your evaluation design. You will have to ask more questions than listed here to fully interpret your particular evaluation results.

A. WHAT ARE YOUR FINDINGS FROM THE PROCESS EVALUATION?

- ❑ 1. The response was not implemented.
- ❑ 2. The response was implemented in a radically different manner than was planned.
- ❑ 3. The response was implemented with inadequate resources, too limited duration, or without the intensity required.
- ❑ 4. The response was implemented nearly as planned and with adequate resources, for the necessary time, and with the required intensity.

B. WHAT ARE YOUR FINDINGS FROM THE IMPACT EVALUATION?

(Select the design(s) you applied — pre-post, pre-post with control, time series, or multiple time series. If multiple designs were used, interpret Tables 2 and 3 for each design separately.)

Pre-post design (no control) — use Table 2 to interpret your evaluation.

- ❑ 1. Problem got worse after the response.
- ❑ 2. Problem unchanged after the response.
- ❑ 3. Problem declined after the response.

Pre-post design with control — Use Table 3 to interpret your evaluation.

- ❑ 1. Problem with the response got worse, relative to control.
- ❑ 2. Problem with the response remained unchanged, relative to control.
- ❑ 3. Problem with the response declined, relative to control.

Time series design (no control) — Use Table 3 to interpret your evaluation.

- ❑ 1. Problem got worse after the response.
- ❑ 2. Problem unchanged after the response.
- ❑ 3. Problem went down after the response.

Multiple time series design (with control) — Use Table 3 to interpret your evaluation.

- ❑ 1. Problem with the response got worse, relative to control.
- ❑ 2. Problem with the response remained unchanged, relative to control.
- ❑ 3. Problem with the response declined, relative to control.

V. OVERALL IMPACT EVALUATION CONCLUSIONS

- ❑ 1. Did the people or places that received the response show less of the problem than people or places that did not receive the response?
- ❑ 2. Did the problem decline at a faster rate after the response than it did before the response? This requires a time series design (with or without a control group) to answer. Do not check if a pre-post design with or without a control group was the only design used.
- ❑ 3. Can you eliminate all other plausible explanations for the change in the problem, other than that the response caused the problem to decline? Use the list of differences (see bottom of part III) between the response and control groups to help answer these questions. Do not check if a pre-post design without a control group was the only design used.

These are judgment calls; the answers should be seen as your degree of confidence in the findings, rather than a totally objective assessment of what occurred. Other people who examine the same evidence could come to different conclusions. For this reason, these questions (and the question that follows) are best answered after several individuals with different perspectives have examined the assessment information.

Based on your answers to these three questions,
ARE YOU REASONABLY CONFIDENT THAT THE RESPONSE YOU EVALUATED WAS PRINCIPALLY RESPONSIBLE FOR THE CHANGES IN THE PROBLEM?

- ❑ YES — If a thorough examination of these questions has been conducted and *all* of them have been checked, use Table 3 to interpret your results.
- ❑ NO — If you cannot check all boxes to questions 1, 2, and 3, this is the appropriate answer. You must interpret Table 3 with extreme caution. Use Table 2 to interpret your results.

Table C2: Interpreting Results of Process and Impact Evaluations (Pre-Post Designs)

		PROCESS EVALUATION RESULTS Answers to Question IV.A.	
		4 CHECKED Implemented nearly as planned	**1, 2, OR 3 CHECKED** Not implemented or implemented in a radically different manner than planned
IMPACT EVALUATION RESULTS Answers to Question IV.B (pre-post without controls)	**3 CHECKED** Problem declined	A. **Response may or may not have caused the decline** in the problem. Nevertheless, the problem has declined.	C. Suggests that **other factors may have caused the decline** in the problem, or the response was accidentally effective. Nevertheless, the problem has declined.
	1 OR 2 CHECKED Problem did not decline	B. **Response does not seem to have worked;** though it is possible the problem would have been worse without the problem.	D. **It is unclear whether the planned response should be implemented,** or whether the problem should be reanalyzed and a different response implemented.
Regardless of the interpretation (A, B, C, or D), there is insufficient evidence to link the response to the problem. The impact-evaluation information available cannot be used to promote the use of the response for other, similar problems. Neither can it be used to rule out the use of the response in similar circumstances.			

Table C3: Interpreting Results of Process and Impact Evaluations (Other Designs)

		PROCESS EVALUATION RESULTS Answers to Question IV.A.	
		4 CHECKED Implemented nearly as planned	**1, 2, OR 3 CHECKED** Not implemented or implemented in a radically different manner than planned
IMPACT EVALUATION RESULTS Answers to Question IV.B pre-post with controls, time series, or multiple time series)	**3 CHECKED** Problem declined	A. Evidence of response causing a decline in the problem. **This response is a potentially useful option in similar circumstances.**	C. Suggests other factors may have caused the decline in the problem, or the response was accidentally effective. **You cannot recommend this response to address similar problems** since you do not know whether it would have an impact if used.
	1 OR 2 CHECKED Problem did not decline	B. Evidence that the response was ineffective or made things worse and a different response should be attempted. You should reanalyze the problem and redesign the response. **The implemented response probably should not be used in similar circumstances.**	D. Little was learned. Perhaps if the response had been implemented as planned, better results would have been noted, but this is speculative. **No recommendations, for or against the response, are valid.**

APPENDIX D:

SUMMARY OF EVALUATION DESIGNS' STRENGTHS AND WEAKNESSES

DESIGN	STRENGTHS	WEAKNESSES
STATIC COMPARISON	• Better than nothing • Can be useful for preliminary examination	• Pre-existing differences between groups are likely to be the cause of the different levels of the problem • Cannot determine if the response came before or after the pre-existing differences • Cannot eliminate the possibility that pre-existing trends created the results
PRE-POST	• Simple and quick to implement • Can easily be used with surveys • Can provide a reasonable estimate of the change in a problem following a response	• Can only show short-term changes in problems • Cannot account for pre-existing trends • Very weak at eliminating alternative explanations for the change in the problem • Cannot account for the possibility that some other factor occurred at the same time as the response, and caused the problem to change
INTERRUPTED TIME SERIES	• Very easy to use with data routinely collected over many time periods • Can eliminate pre-existing trends and many other alternative explanations	• Very hard to use if special data-collection efforts, such as surveys, are used to measure the problem • Cannot account for the possibility that some other factor occurred at the same time as the response, and caused the problem to change • Results take a long time to be established • Difficult to interpret when there are few problem events per time period *before* the response
PRE-POST WITH CONTROL	• Can easily be used with surveys • Can account for the possibility that some other factor occurred at the same time as the response, and caused the problem to change	• Can only show short-term changes in problems • Requires a problem-troubled control group that will not get the response and is similar to the response group
MULTIPLE TIME SERIES	• Easy to use with data routinely collected over many time periods • Can eliminate pre-existing trends and many other alternative explanations • Can account for the possibility that some other factor occurred at the same time as the response, and caused the problem to change	• Very hard to use if special data-collection efforts, such as surveys, are used to measure the problem • Requires a problem-troubled control group that will not get the response and is similar to the response group • Results take a long time to be established • Difficult to interpret when there are few problem events per time period before the response

REFERENCES

Clarke, Ronald V. 1992. *Situational Crime Prevention: Successful Case Studies*. Albany, New York: Harrow and Heston.

Clarke, Ronald V., and David Weisburd. 1994. "Diffusion of Crime Control Benefits: Observations on the Reverse of Displacement." In R. V. Clarke (ed.), *Crime Prevention Studies* (Vol. 2, pp. 165-184). Monsey, New York: Criminal Justice Press.

Cornish, Derek, and Ronald V. Clarke. 1986. "Situational Prevention, Displacement of Crime and Rational Choice Theory." In K. Heal & G. Laycock (eds.), *Situational Crime Prevention: From Theory into Practice*. London, U.K.: Her Majesty's Stationery Office.

Corsaro, Nick, Dan W. Gerard, Robin S. Engel, and John E. Eck. 2012. "Not By Accident: An Analytical Approach to Traffic Crash Harm Reduction." *Journal of Criminal Justice* 40 (6), 502–514.

Eck, John E. 1993. "The Threat of Crime Displacement." *Criminal Justice Abstracts* 25, 527-546.

Eck, John E. 1997. "Preventing Crime at Places." In Lawrence W. Sherman, Denise Gottfredson, Doris MacKenzie, John Eck, Peter Reuter and Shawn Bushway eds. *Preventing Crime: What Works, What Doesn't, What's Promising – A Report to the Attorney General of the United States.* Washington, DC: United States Department of Justice, Office of Justice Programs.

Eck, John E., and Edward R. Maguire. 2000. "Have Changes in Policing Reduced Violent Crime? An Assessment of the Evidence." In A. Blumstein & J. Wallman (eds.), *The Crime Drop in America* (pp. 207-265). New York: Cambridge University Press.

Eck, John E., and William Spelman. 1987. *Problem Solving: Problem Oriented Policing in Newport News*. Washington, D.C.: Police Executive Research Forum.

Hesseling, René B. P. 1995. "Displacement: A Review of the Empirical Literature." In R. V. Clarke (ed.), *Crime Prevention Studies* 3, pp. 197-230. Monsey, New York: Criminal Justice Press.

Lancashire Constabulary. 2006. *Operation SeaQuest.* Submission to the Tilley Award. Available at www.popcenter.org/library/awards/tilley/2006/06-52(W).pdf.

Matthews, Roger. 1992. "Developing More Effective Strategies for Curbing Prostitution." In R. V. Clarke (ed.), *Situational Crime Prevention: Successful Case Studies* (1st ed., pp. 89-98). New York: Harrow and Heston.

Office of Community Oriented Policing Services. 1998. *Problem-Solving Tips: A Guide to Reducing Crime and Disorder through Problem-solving Partnerships*. Washington, D.C.: U.S. Department of Justice, Office of Community Oriented Policing Services.

Pawson, Ray, and Nick Tilley. 1997. *Realistic Evaluation*. London: Sage.

ABOUT THE AUTHOR

JOHN E. ECK

John E. Eck is a professor at the School of Criminal Justice at the University of Cincinnati, where he teaches graduate courses in police effectiveness and crime prevention. Dr. Eck received his Ph.D. from the University of Maryland in 1994 and his masters of public policy from the University of Michigan in 1977. From 1977 to 1994 he directed research at the Police Executive Research Forum, in Washington D.C., where he conducted studies of criminal-investigations management and drug markets, and helped test and implement problem-oriented policing in agencies throughout the United States. From 1995 to 1998, Dr. Eck was the evaluation coordinator for the Washington/Baltimore High Intensity Drug Trafficking Area, where he developed methods for assessing the effectiveness of drug-trafficking enforcement. Dr. Eck has written extensively on problem-oriented policing, crime mapping, drug markets, and crime prevention at places. Dr. Eck's research interests focus on crime places: how they arise and what can be done to reduce crime at these locations. Among his many publications, Dr. Eck is the co-author, with Ronald V. Clarke, of *Crime Analysis for Problem Solvers: In 60 Small Steps*. In the summers, Dr. Eck helps his wife restore old cemeteries and repair broken tombstones, which has fueled an interest in preventing vandalism of cemeteries.

RECOMMENDED READING LIST

There are numerous books and articles describing how to conduct evaluations. Listed here are some that I have found helpful to graduate students, are essential standard readings, or were written specifically for police and are available on the web.

If you want to develop an expertise in the area of evaluation, then you will need to take masters- and doctoral-level courses. But much can be learned and usefully applied by reading beyond this introductory guide.

Bachman, Ronet, and Russell K. Schutt. 2013. *The Practice of Research in Criminology and Criminal Justice*. Thousand Oaks, California: Sage. This college level text provides a well written description of the theory and practice of data collection, measurement, and research design as applied to criminal justice research and evaluation.

Campbell, Donald T., and Julian C. Stanley. 1963. *Experimental and Quasi-Experimental Designs for Research*. New York: Houghton Mifflin. Though over 50 years old, this is still the "bible" of evaluation designs. Virtually every methods text adapts material from this source. It is still indispensable and though short and to the point, this book is not a fast read.

Clarke, Ronald V., and John E. Eck. 2005. Crime Analysis for Problem Solvers: In 60 Small Steps. Washington, DC: Office of Community Oriented Policing. Available at: http://www.popcenter.org/learning/60steps/. Contains several brief sections describing the evaluation of problem-solving efforts, as well as a brief summary of displacement and diffusion of benefits.

Converse, Jean M., and Stanley Presser. 1986. *Survey Questions: Handcrafting the Standardized Questionnaire.* Thousand Oaks, California: Sage. This book is a standard reference in survey research.

Blair, Johnny, Ronald F. Czaja, and Edward Blair. 2013. *Designing Surveys: A Guide to Decisions and Procedures.* Los Angeles: Sage. A good introduction to designing surveys.

Eck, John E., and Nancy LaVigne. 1993. *Police Guide to Surveying Citizens and Their Environment*. Washington, D.C.: Bureau of Justice Assistance. NCJ Number: 143711. Available at http://www.popcenter.org/library/reading/PDFs/Surveying_Citizens.pdf. This monograph describes the basics of conducting surveys of the public and surveys of physical environment. It contains a number of examples and survey instruments.

Eck, John E. 2005. "Evaluations for Lesson Learning." In N. Tilley, ed. *A Handbook for Crime Prevention: Theory, Policy and Practice*. Pp. 699-733. Cullompton, Devon: Willan. An academic treatment of the role of evaluation in improving crime prevention.

Guerette, Rob T. 2009. *Analyzing Crime Displacement and Diffusion.* Problem-Solving Tools Guide No. 10. Washington, D.C.: Office of Community Policing Services. Available at: http://www.popcenter.org/tools/displacement/. A thorough readable guide to displacement written for police practitioners.

Kosslyn, Stephen M. 1993. *Elements of Graph Design*. New York: W.H. Freeman. This well-organized graphical-design book offers practical and straightforward advice on how to create effective charts, graphs and figures with data. It is filled with good and bad examples.

Shadish, William R., Thomas D. Cook, and Donald T. Campbell. 2001. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference.* Belmont, California: Wadsworth. This is the much longer follow-up to the Campbell and Stanley monograph. It is dense, but thorough. If you seek to master the topic of evaluations, you must consult this standard text.

Trochim, William, and James P. Donnelly. 2007. *The Research Methods Knowledge Base.* Boston, Massachusetts: Cengage Learning. This college text was designed for use on-line, but is available in a paperback version. It is very practical and shows how to create complex evaluation designs out of simpler designs in order to address peculiar situations. It also contains an excellent discussion of measurement and sampling.

Weisburd, David, and Chester Britt. 2014. *Statistics in Criminal Justice*. New York: Springer. This is a very well-written introductory college text in statistics, taking the reader from the very basics to an intermediate level.

Weisel, Deborah. 1999. Conducting Community Surveys: A Practical Guide for Law Enforcement Agencies. Washington, D.C.: Bureau of Justice Statistics and Office of Community Oriented Policing. (October) NCJ #:178246. Available at http://www.popcenter.org/library/reading/PDFs/Conducting_Surveys.pdf. This practical guide for law enforcement agencies accompanies the Crime Victimization Survey Software developed by the Office of Community Oriented Policing Services and the Bureau of Justice Statistics. It describes how surveys have been used to improve policing services, ways to identify survey goals and procedures for survey administration and analysis.

ENDNOTES

[1] Pawson & Tilley (1997).

[2] Eck and Spelman (1987); Office of Community Oriented Policing Services (1998).

[3] Lancashire Constabulary (2006).

[4] Matthews (1992).

[5] Eck (1997).

[6] Clarke (1992).

[7] Cornish and Clarke (1986); Eck (1993); Hesseling (1995).

[8] Clarke and Weisburd (1994).

[9] Corsaro et al. (2012).