The *Art* of Appropriate Evaluation

A Guide for Highway Safety Program Managers

<table>
<tr><td>1. Report No.
DOT HS 811 061</td><td colspan="2">2. Government Accession No.</td><td>3. Recipient's Catalog No.</td></tr>
<tr><td colspan="3">4. Title and Subtitle
The Art of Appropriate Evaluation: A Guide for Highway Safety Program Managers</td><td>5. Report Date
December 2008</td></tr>
<tr><td>6. Performing Organization Code</td><td colspan="3">7. Editor(s)
Nancy C. Pullen-Seufert and William L. Hall</td></tr>
<tr><td>8. Performing Organization Report No.</td><td colspan="3">9. Performing Organization Name and Address
University of North Carolina Highway Safety Research Center
730 Martin Luther King Jr, Blvd
Suite 300, CB 3430
Chapel Hill, NC 27599-3430</td></tr>
<tr><td>10. Work Unit No. (TRAIS)</td><td colspan="3">11. Contract or Grant No.
DTNH22-05-D-05043
Task Order No. 006</td></tr>
<tr><td colspan="3">12. Sponsoring Agency Name and Address
U.S. Department of Transportation
National Highway Traffic Safety Administration
1200 New Jersey Avenue SE.
Washington, DC 20590</td><td>13. Type of Report and Period Covered</td></tr>
<tr><td>14. Sponsoring Agency Code</td><td colspan="3">15. Supplementary Notes
Ian Reagan served as the NHTSA Contracting Officer's Technical Representative for the project.</td></tr>
<tr><td colspan="4">16. Abstract
The guide, updated from its original release in 1999, is intended for project managers who will oversee the evaluation of traffic safety programs. It describes the benefits of evaluation and provides an overview of the steps involved. The guide includes case examples to illustrate the described concepts and discusses what to look for when hiring a professional evaluator for large or complex evaluations.</td></tr>
<tr><td>17. Key Words
Evaluation</td><td colspan="3">18. Distribution Statement</td></tr>
<tr><td>19. Security Classif. (of this report)</td><td>20. Security Classif. (of this page)</td><td>21. No. of Pages
80</td><td>22. Price</td></tr>
</table>

Table of Contents

CHAPTER 1

Introduction

Once upon a time there was a manager who was responsible for starting up a new pedestrian safety program. Because it was new, her boss asked her to evaluate the program to find out how well it worked. Alarm bells rang in her head; she had never done an evaluation and it seemed way beyond her ability. When she discussed this assignment in her regular staff meeting, one of the staff volunteered to take on the responsibility. Greatly relieved, she gave him free rein.

The staff member immediately busied himself designing data collection forms and survey instruments. He wrote instruction manuals for filling out the forms and distributed them to the folks who were involved in publicizing the program. His research designs called for dividing the city into four regions that would each receive different combinations of the program's components. His weekly project reports were filled with detailed accounts of new forms, focus group protocols, new data collection and analytical procedures, and statistical tests. It seemed that everything was under control.

As the program reached its peak of activity, things took a turn for the worse. Data collectors weren't filling out the forms correctly, and no one could get a handle on the mountains of data the survey produced. The evaluator spent most of his time analyzing the change in public perception of the program. The difference was statistically significant, but so small as to be practically negligible. The progress reports started documenting why it was impossible to conduct a valid evaluation, with terms like "changes in data definitions," and "confounding variables" leading the list of excuses.

The net result was that more than 20 percent of the project's resources were spent on evaluation and no one could answer the simple question "did it work?" The project manager vowed "Never again!"

The term *evaluation* sometimes evokes similar nightmares for real-life managers. We have all heard stories about expensive evaluation efforts that yield reams of complex data that end up confusing people. None of us wants an evaluation like that. We want to document the good parts of our program and find the things that need to be changed.

Who the Guide is for

Before we go any further, it's time to share the assumptions we have made about who you are. This Guide will be useful to you if you are a state or local traffic safety project director, or if you oversee these kinds of projects and need to become acquainted with the reasons for evaluation and the basics of how it is supposed to be done.

If you are starting a new project, ask your project officer or funder about any specific reporting requirements. For example, if you're conducting a high visibility enforcement campaign to reduce impaired driving, the funder will likely want to know information such as the number of checkpoints held, the number of citations issued and the number of newspaper articles printed.

Our assumption is that you do not have a background in experimental design or statistics, but you do need to understand a few basic topics:

- What type of evaluation is reasonable for the type of program you are implementing
- What you can do to maximize the success of a program evaluation
- Where you can get help

If that is what you are looking for, this Guide is for you!

What is Evaluation?

Evaluation is a term that refers to the process by which someone determines the value of something.

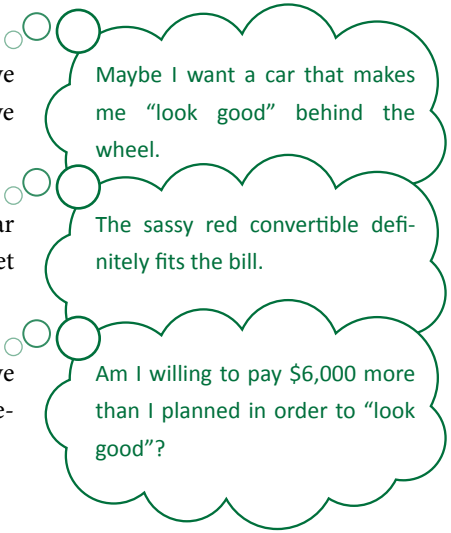
Value doesn't only mean monetary value; so evaluation doesn't necessarily involve converting something into a dollar and cents issue. It is simply examining, appraising, or judging th worth of a particular item or program.

We all conduct evaluations whenever we are contemplating a major purchase. If we are thinking about buying a new car, we must decide if a vehicle is worth the price being asked fo We go through three distinct evaluation processes to make that det mination.

STEP 1 We first determine what we need in a car and what we would like to have.

STEP 2 We then determine if the car we are looking at will meet these needs and wants.

STEP 3 If it does, we must decide if we are willing to pay the price being asked.



Once we have purchased the car, we probably continue to evaluate, but we may call it "having second thoughts." After the purchase is made, we try to determine if we made a good choice. Did the car deliver on the advertising promises? Did it meet our personal needs and wants? Did it actually cost what we planned, or did it require a lot of expensive maintenance to keep it running? If I had it to do over, would I buy the same car? Would I recommend it to a friend?

When you are implementing a traffic safety program, you should be making the same types of judgments. You build evaluation into your program so that you can determine:

☑ The exact nature of the traffic safety problem you are trying to address	Last year, 10 percent of the 50 traffic-related deaths were child bicyclists. None of the children were wearing bike helmets.
☑ Reasonable goals and objectives for reducing this problem	To decrease the number of bicyclist fatalities by increasing bike helmet usage to 80 percent among child bicyclists.
☑ How well the program you implemented accomplished your objectives	Bike helmet usage increased from 45 percent before the program to 85 percent after the program.

What do you notice about these three statements? They are all specific, focused, and practical.

First, the evaluator identified a specific problem (the kids who died were not wearing bicycle helmets). Next there is one focused, promising program approach to address this problem (increase bicycle helmet use). Note that there is no mention of *how* this will be done—free helmets, school programs, bike safety events—those are specifics that are decided later in the process. Finally, there is a practical measure of the progress your program made (document the change in bicycle helmet use).

Why You Want to Read this Guide

A lot has been said over the years about the importance of program evaluation in traffic safety. At various times, program managers have been required to allocate a specified percentage of their program budgets to program evaluation. Training programs have been developed on how to evaluate traffic safety programs using complex statistical tools such as time series analysis and multiple regression analyses. And despite all of this attention, criticism continues to pour in about the fact that most traffic safety programs are never actually evaluated. And it is no wonder. Some program managers view

The Guide provides an overview of the steps that are involved in program evaluations and gets you thinking about how these steps fit into your implementation plans.

evaluation as too complex to undertake and a waste of resources. This Guide will convince you otherwise!

This Guide is designed to alleviate your fears about program evaluation and convince you that conducting an appropriate evaluation actually makes your job easier rather than harder. The focus is on what evaluation can do for you, not the other way around.

The Guide provides an overview of the steps that are involved in program evaluations and gets you thinking about how these steps fit into your implementation plans. It also will provide you with some handy suggestions on how to find and work with an evaluation consultant. Finally, it will provide you with descriptions of two traffic safety evaluations and a glossary of evaluation terms and concepts so you speak with more confidence when the topic turns to "proving results." (When you encounter an underlined term such as Before and After Design, you can refer to the Glossary for its definition.)

It is equally important that you know what this Guide is not. It will not give you detailed, step-by-step instructions on how to evaluate a traffic safety program. For that

Small projects may not have the funding to hire an evaluator. However, this does not mean evaluation is unnecessary. In fact, having information about what the project did and any changes in behavior that occurred might increase the likelihood that a larger effort could be funded in the future. Three options for small projects are:

1. If funded by a State Highway Safety Office, ask if there is assistance available for evaluation.
2. Seek help from a local university to see if the evaluation can be conducted as part of a student or class project. While it may take more time, it may have the benefit of oversight and advice from a faculty advisor.
3. Explore how you can do it on your own by referring to "Detailed Evaluation Guides" in the Resources section.

level of information, consider attending a US DOT Transportation Safety Institute data and analysis course (see www.tsi.dot.gov) or similar training, or refer to "Detailed Evaluation Guides" in the Resources section. Our assumption is that you are already too busy to take on a new career as an evaluation specialist but that a basic understanding of evaluation will help you get the results you want. There are talented individuals in your own community who can help you design and conduct an appropriate evaluation. This Guide will tell you how to find and work with them.

The focus of this Guide is on using limited resources to their maximum and practical advantage for you. This means conducting an evaluation that is appropriate to the size and scope of the program you are implementing.

How the Rest of the Guide is Organized

The remainder of this Guide includes:

Chapter Two: The Importance of Program Evaluation
This is where we try to convince you that program evaluation is always a good idea.

Chapter Three: What a Program Evaluation Can Accomplish
A discussion of what you can reasonably expect a state or community program evaluation to do.

Chapter Four: Evaluation Step-By-Step
An overview of the steps involved in program evaluation, from defining your problem to reporting results.

Chapter Five: Getting Help
What you should expect from an evaluator, where to find them, and how to work with them.

Chapter Six: Closing Comments
A wrap-up of the arguments in support of always evaluating your program efforts.

Glossary of Terms
Some basic evaluation terms defined to increase your comfort level around evaluators.

Case Examples
Example evaluations from two traffic safety programs.

Resources
Where to find evaluation information that is beyond the scope of this Guide.

CHAPTER 2

The Importance of Program Evaluation

Evaluation is like regular exercise. We all know that exercise is good for us. And we all try it from time to time. But the majority of Americans fall far short of the recommended exercise requirements. Why is that?

In many cases, we don't exercise regularly because we have convinced ourselves that exercise requires too much effort, that we're too busy and that it probably won't benefit us much anyway.

Why do many people shy away from regularly conducting program evaluations?

It's the same way for evaluation. Most managers and funders agree that it is a good idea to evaluate a program, particularly one that uses taxpayer dollars. But when it comes time to build an evaluation into a program plan, dozens of reasons are offered as to why it just can't or shouldn't be done. Why do many people shy away from regularly conducting program evaluations?

Most excuses for not doing an evaluation are variations on the following themes.

- "Evaluation is too complicated. Program evaluations require complex research designs for sophisticated experiments. I don't fully understand what it is involved but it sounds pretty intimidating."
- "If I conduct an evaluation, I may discover that my great idea was a total flop. I want to avoid that embarrassment."
- "I have a very limited budget. Evaluations are expensive and time-consuming. I just can't spare the resources required."
- "Evaluation is a lot of work, and I don't have the time. Besides, I never understand the technical jargon that I read in evaluation reports. So why bother?"

Do some of these sound familiar? Have you found yourself thinking these thoughts? Let's go through them one by one to show you why they aren't true and may be standing in the way of your success in traffic safety.

"Evaluation is too complicated."

Many are intimidated by the whole concept of evaluation. A mystique has built up that program evaluation is very complicated with a hundred ways to do it wrong and only one, very difficult, way to do it right. It's true that in some cases conducting a valid evaluation can be complicated and difficult. When you are talking about establishing a direct cause and effect relationship between a specific traffic safety countermeasure and a reduction in traffic deaths, you need a solid research design with random assignment to comparison and treatment groups, and sophisticated statistical analyses. More importantly, you need large numbers of cases in order to detect any real change in traffic deaths.

In reality, however, traffic safety evaluation is an applied science that works within the constraints of state and local program implementation. Most local communities simply do not have the volume of traffic deaths and injuries to conduct countermeasure effectiveness evaluations measuring changes in deaths and injuries. Instead, these communities can focus their evaluation efforts on determining if the particular program they implemented achieved its specific objectives and whether behavior (like seat belt use) changed.

If you are implementing an occupant protection program, your evaluation dollars would be better spent demonstrating an increase in seat belt use rather than trying to prove that seat belts save lives. As will be discussed in Chapter Three, documenting an increase in seat belt use over baseline levels involves a much simpler evaluation and should not deplete your program resources.

"I may not like the answer so I better not ask the question.

When you implement a program, you put a lot of yourself into the effort. You believe that it is a good program and you do everything you can to make it work. But you never know what might happen.

Some people shy away from evaluations because they don't want their good idea to be proven wrong. The mistake they are making is viewing evaluation as the last step in a process, like a final exam. If an evaluation is tacked on at the end of the project, you may very well come up with answers that you don't like.

The key to successful evaluation is to build evaluation in from the start so it can help you frame the questions you are asking, and even clarify the problem that you are trying to solve. A well-planned evaluation should not yield unpleasant, last-minute surprises. Instead, it will provide useful information that helps fine-tune the program every step of the way. It can also tell you what's working and what can be improved.

"I have a limited budget; I prefer to spend my dollars on implementation.

These days, everyone is expected to do more with less. Project directors struggle to stretch every program dollar to the limit. When asked to choose between delivering more high school presentations or conducting a program evaluation, many choose more presentations because they believe that their impact will be greater if they can reach more kids. There are two errors in this logic:

1. They believe that an evaluation will cut into their program implementa-

When it is time to ask your funding source to extend your program, your proposal will be much more persuasive if it includes solid data demonstrating that you accomplished your objectives.

tion resources, when in fact, there are many ways to conduct an evaluation inexpensively.

2. Reaching more teens with an effective traffic safety message is always a good idea. But how do you know if your message is being well-received and achieves the desired effects unless you do some sort of evaluation? Collecting and monitoring feedback throughout a project provides the opportunity to fine-tune your message while the project is underway. At the conclusion of the program you will also have the data you need to improve the entire program before you offer it again.

Another point to keep in mind when thinking about conserving program dollars is satisfying your funding source. When it is time to ask your funding source to extend your program for another year, your proposal will be much more persuasive if it includes solid data demonstrating that you accomplished your objectives.

"Evaluation is too much work."

Evaluation *can* be labor-intensive (note that we didn't say expensive) and very tedious. This is why evaluation is one of those areas that, as a project director, you should delegate to someone else. This could be an employee from another department in your agency, a faculty member or graduate student at a local university, or a private evaluation consultant. In addition to doing the time-consuming work involved in data collection, you should expect an evaluation specialist to be able to explain the results in a language you and everyone else can understand. (Chapter Five talks about what to look for in evaluators and how best to work with them.)

Now that we have discussed some commonly heard reasons for not evaluating a program, we can concentrate on reasons why evaluations are truly necessary.

Benefits of Evaluation

Let's go back to our exercise analogy for a moment. We all know some people who are fully committed to getting regular exercise. They don't have to think about if they will exercise or how they will fit it into their schedule. Exercise is an essential part of their daily existence, just like eating and sleeping. These individuals report that exercise improves every aspect of their life, not just their physical conditioning. They have more energy, they are more productive at work, and they are less prone to depression. What separates them from the rest of us is their exercise mentality.

An evaluation mentality can't ensure that each project you implement will be a resounding success, but it can ensure that you fully understand what you tried to do and why things turned out as they did.

To get the full range of benefits from evaluation, you need to have an evaluation mentality. This means that you never even consider implementing a program without first thinking like an evaluator. You approach each new problem with the same set of questions:

- What do I know about the problem?
- If I tried to fix this problem, what could I accomplish?
- How could I measure my results?
- How can I collect the data I will need?
- What are my criteria for success?

With the answers to these questions in hand, you are prepared to convince any funding source that you know what needs to be done and that your ideas have a high probability of success. Your evaluation mentality will also ensure that at the end of this project, you can report back to these same funding sources with solid information on what you accomplished.

An evaluation mentality can't ensure that each project you implement will be a resounding success, but it can ensure that you fully understand what you tried to do and why things turned out as they did.

If you have an evaluation mentality, you design your program and your evaluation at the same time. The benefits of this approach are substantial. An evaluation mentality will enable you to:

Accurately identify the problem you are trying to solve
All too often people jump into implementing a program without really understanding the underlying cause of their problem. Is underage DWI a problem because the liquor stores are selling to teenagers or because the police are not targeting the locations where kids are drinking? It is not enough to suspect that seat belt use is low in your community. You need to look at belt use rates and what groups make up the non-belt users up front. An evaluation mentality frees you from having to take a "shot-gun" approach to solving every problem.

Uncover problems you didn't know you had
You might assume that your pedestrian safety problem involves the very young and very old until you discover that a significant percentage of your pedestrian fatalities are working age adults who had been drinking prior to being hit by a car. This problem would require a totally different set of countermeasures than would a child pedestrian problem.

Establish reasonable, practical objectives for dealing with these problems
Global objectives are the hardest to accomplish. With good problem identification data, you can focus your objectives on the specific problem you are trying to solve. Instead of trying to "reduce unsafe driving behaviors" you might want to reduce red-light running, which is the unsafe behavior that is causing the most concern in your community.

Determine if you have accomplished your program objectives

A major purpose of any evaluation is to determine if your program accomplished its objectives. Well-thought-out objectives are an important first step, but an evaluation mentality will also help you identify what you truly need to measure. Too often, project directors waste time and money collecting data that they can never use, either because they can't compare it to any baseline or because it does not relate to their program objectives.

Recognition of the importance of evaluation has become integrated into highway safety programs. In fact, State Highway Safety Plans and Federal transportation legislation (such as SAFETEA-LU) require evaluation of highway safety programs and specify highway safety goals to be met. This reinforces the importance of evaluating programs so resources can be directed to sustaining and disseminating programs that have been proven to be effective.

Provide information to funding sources, the media and the public to continue support for program

Support for a good idea can evaporate if there is no evidence that the idea really works. With an evaluation mentality, you create consensus from the beginning on the criteria for success. This allows you to stay focused on the data you will need to measure your performance against those criteria. By providing updates to funding sources, media and the public, you have a better chance of continuing support for your program. Keep in mind, though, that a statistically significant but barely discernible change may not constitute success in everyone's eyes.

Determine if and how a program should be revised to increase its effectiveness

With an evaluation mentality, you don't wait until the end to find out how your idea worked—very few projects work perfectly. There are always aspects that could be tweaked to make them more effective. With an evaluation mentality, you monitor performance throughout the project so that you can institute any necessary mid-course corrections, and you'll be better prepared

to recommend any revisions for the future. Your evaluation might reveal that the project did not attain the expected results. For example, impaired driving rates did not decrease as had been anticipated. Keep in mind that discovering a project does not result in the desired behavior change is valuable information so that future time and resources can be appropriately allocated in response to the evaluation findings.

You do not need to be an evaluation specialist to have an evaluation mentality, just as you don't have to be a marathon runner to have an exercise mentality. You just have to recognize that evaluation, when built in from the beginning, provides benefits throughout the life of your program.

CHAPTER 3

What a Program Evaluation Can Accomplish

The program evaluation in the introductory scenario didn't demonstrate anything other than a lot of good intentions and confused activity. The person running that evaluation clearly did not have an evaluation mentality and didn't design an evaluation that was appropriate to the size of the project and to the data that was available.

One of the most critical elements in a successful evaluation (that is, one that actually proves something) is deciding what should be demonstrated. This decision should be based on the type of project you are implementing and the types of data that are collectable (or available). Your goal is to set up an evaluation that is appropriate for a program's individual circumstances.

What Makes an Evaluation Appropriate?

When the National Highway Traffic Safety Administration (NHTSA) promotes a traffic safety countermeasure as an effective tool in reducing traffic deaths and injuries, NHTSA does so only after thoroughly evaluating it in realistic conditions to make sure it works. This requires conducting several full-scale evaluation research projects that verify the effectiveness of the countermeasure. NHTSA can also study large volumes of national and state level crash data with enough records to confirm, with a high degree of confidence, that changes can be attributed to the countermeasure. A full-scale countermeasure effectiveness evaluation project is the only type of evaluation that would be appropriate in these circumstances.

Think about a case where a program manager in a city of 75,000 reads about a countermeasure in a NHTSA publication and decides that it might be just what is needed to solve a troubling traffic safety problem in his community. This program manager has a solid evaluation mentality so he immediately considers what type of evaluation would be appropriate for his circumstances. He does not need

to conduct the same type of evaluation that NHTSA conducted for two reasons:

1. He is not trying to prove to the nation that it works; this has already been proven by NHTSA's evaluation results.
2. His community experiences only a few crashes of the type affected by this countermeasure (but he still would like to reduce that number even further).

He needs to determine what an appropriate evaluation would be for these circumstances. There are two basic evaluation questions that are appropriate for most local, and even some State, programs:

1. Did you implement the program as planned?
2. Did you accomplish your objectives?

Did You Implement the Program as Planned?

At the most fundamental level, you can do an evaluation to determine if you implemented the program as planned. This may sound pretty obvious, but in fact many projects take a wrong turn at the start. This approach, which is called a process evaluation, does not require any elaborate data collection efforts or even a research design. All that it requires is an understanding of what is supposed to happen during a program and a systematic approach to tracking what actually happens.

A process evaluation does not require any elaborate data collection efforts or even a research design.

Let's go back to the bicycle helmet program in Chapter One. Suppose you decide you're going to have two safety fairs over the summer and you're going to give away free helmets donated by a community sponsor. A process evaluation would keep track of the number of helmets you obtained and the number you gave away. It should also document such things as the age, gender and neighborhoods of the children who received the helmets, the number of people who participated in the safety fairs, and the amount and sources of publicity you received about the fairs.

If you monitor your program from the beginning, you will be able to spot any implementation problems early and determine if the problem can be fixed or if the whole idea should be scratched. There is no sense wasting dollars and effort going through the motions of implementing a program with fatal flaws.

Some managers might dismiss this type of process evaluation as simple "bean counting" that doesn't demonstrate anything worthwhile. You will be surprised by what you can learn merely by checking to see whether everything is going as planned.

An important element of documenting how the program was implemented is tracking the resources as they are being spent. Every project should have a detailed budget for items including staffing and supplies. A good evaluation should document whether the project was completed within budget or over budget. The rate at which resources are being spent can sometimes give a good indication if the project is being implemented as planned. For example, if local police are not putting in the budgeted amount of overtime, maybe the sobriety check-points are not being conducted as frequently as planned.

For example, a community decided to implement an occupant protection traffic enforcement blitz, complete with highly visible public information and media coverage. The evaluator kept track of the number of police officer hours spent and the number and type of citations issued. The program staff were surprised to find that although lots of seat belt citations were issued during the first week, there were no citations issued for child safety seat violations. The police officers did not seem to fully understand the requirements of the State law. This discovery led to a police roll call training session on the child safety seat law and on the importance of enforcing it. During the second week of the blitz, forty-seven citations and warnings were issued for child safety seat violations.

Did You Accomplish Your Objectives?

You don't need an evaluation mentality to realize that you conduct an evaluation to determine if you accomplished your objectives. But it does help you understand what objectives to measure.

People usually write goals and objectives to impress a funding source. They are frequently written in grandiose terms that sound impressive but lack a clear focus.

- To reduce traffic deaths (Do you want to promise that in your small town?)
- To increase support for traffic safety (How will you measure this?)
- To improve safe driving behaviors (What behaviors do you care about?)

When challenged, the individuals who wrote these objectives were able to revise them to focus on what an individual project was specifically designed to do, not what sounded good on paper.

"Reducing traffic deaths" was changed to "increase seat belt use"—that was what they were really aiming for. "Increase support for traffic safety" was changed to "get 1,500 signatures on a petition for passage of a bicycle helmet ordinance," and "improve safe driving behavior" was changed to "reduce the incidence of red-light running."

We cannot emphasize enough the importance of carefully defined objectives. They make the difference between a successful evaluation and a frustrating one. Read Chapter Four for more detailed suggestions on writing SMART Objectives.

What May Not Be Appropriate to Demonstrate?

It is very difficult to link a countermeasure program to a reduction in deaths and injuries at the local level (and sometimes even at the State level). There are several reasons for this.

- Traffic crashes are a serious national problem, but traffic deaths in any community are relatively rare events. Most communities will experience fewer than a dozen traffic-related fatalities a year resulting from all causes. Furthermore, the number of deaths might fluctuate considerably from year to year, for no apparent reason.

We cannot emphasize enough the importance of carefully defined objectives. They make the difference between a successful and a frustrating evaluation.

- Traffic deaths are influenced by a variety of factors, all of which can influence whether fatalities climb or drop. These factors, called variables, include:
  - The amount of driving in the community (an increase in gasoline costs could reduce the amount of miles traveled, or a new shopping mall on the outskirts of town could increase vehicle travel).
  - The weather conditions (a very bad winter could lead to an increase in fender bender type collisions, but major injuries might go down because people drive less and at slower speeds in bad weather).
  - A change in the population (a downward trend in population growth could reduce the number of drivers on the road).
  - Previous extremes (a shift back to "normal" levels after reaching an extreme value, either high or low, also called "regression to the mean" by statisticians).

If you are trying to establish a connection between a particular countermeasure and a reduction in deaths and injuries, you have to ensure that none of these variables, or any others you might think of, contributed to that change.

If you were committed to demonstrating a reduction in fatalities, you would need to aggregate your data over several years in order to have enough deaths to show a real decrease because the number of fatalities that occurs in most communities is so

Given that the number of deaths might go up or down regardless of what new program you implemented, you might not want to raise expectations that your program will save lives. It would be far better, for example, to demonstrate that your program resulted in an increase in seat belt use.

small. This approach, however, creates an entirely different problem related to existing data: it is very difficult to compare data that were collected in widely separated time periods whether you are looking for fatalities or some other measure, such as citations issued. Over time, changes in data collection procedures, data definitions, and enforcement thresholds can change significantly. For example, a community may change its policy concerning the collection of blood alcohol concentration data on traffic fatalities, making it difficult to compare the number of alcohol-related deaths over a five-year period. Or a Traffic Records Department may change its definition of a "reportable" crash from $250 or more in damages to $2,000 or more in damages. Obviously, this would decrease the number of reported crashes.

It is very difficult to compare data that were collected in widely separated time periods whether you are looking for fatalities or some other measure, such as citations issued.

These problems with linking countermeasures directly to bottom line changes in fatality levels are not insurmountable. However, they do require a significant increase in the complexity and cost of an evaluation. You should consider undertaking this extra effort only when it really is necessary, like when you are trying a countermeasure that has never been tried anyplace else. If you are trying something that has not been tried before, it is especially important that you have an evaluator.

If your program involves countermeasures that have been proven to work, such as those that follow below, you can concentrate your evaluation time on documenting that you met your objectives (see "Develop Reasonable Objectives" in Chapter Four) and changed behaviors rather than trying to demonstrate that the countermeasure saved lives:

- High-visibility enforcement to address seatbelt use or impaired driving
- Publicize, enforce, and adjudicate laws prohibiting alcohol-impaired driving

- Enforcement of graduated driver licensing and zero tolerance laws for young drivers
- Increase use of bicycle helmets or motorcycle helmets

For more information on proven countermeasures, refer to the most recent edition of the NHTSA publication "Countermeasures That Work" on the NHTSA Web site (www.nhtsa.dot.gov).

In order to prove that program objectives were met, you will still have to collect data and document your accomplishments. In the case of high-visibility enforcement to increase seat belt use or DWI enforcement, you will probably need to observe seat belt use before and after you implement you strategy, or collect enforcement data for a comparable period before you instituted your "blitz." If seat belt usage or DWI enforcement increased, your program was a success.

Learning that something did not work does not make your evaluation a failure.

If it did not increase, then you should look at the strategies you used. Perhaps these techniques were not as effective as other options (e.g., a public information campaign alone will not be yield the same change in behavior as an enforcement campaign coupled with continuing media coverage). Learning that something did not work does not make your evaluation a failure. It simply provides an opportunity to learn more about your problem and to revise your approach in the future.

In summary, a program evaluation can demonstrate that you implemented the program as planned, what resources were spent, and whether your program accomplished its objectives.

That level of detail is appropriate for most local and State level evaluations. The next chapter provides an overview of what will be involved in program evaluation.

CHAPTER 4

Evaluation Step-by-Step

Evaluation is an integral part of program implementation and needs to be included from the start. Therefore, the seven primary steps involved in evaluation mirror the steps followed to conduct a program.

In this chapter, every step includes a description and an example of how a real-life program conducted that step. The example follows a high visibility seat belt enforcement campaign conducted in Chemung County, New York. For more information, see "Achieving a High Seatbelt Use Rate: A Guide for Selective Traffic Enforcement Programs" in the Resources section.

1. **Identify the problem**
2. **Develop reasonable objectives**
3. **Develop a plan for measuring results**
4. **Gather baseline data**
5. **Implement your program**
6. **Gather data and analyze results**
7. **Report results**

Identify the Problem

STEP 1

It may sound obvious but you need to understand the problem you're facing before you can expect to solve it. All too often, decisions are made to implement a program based on a reaction to a single, tragic fatal crash. Problem identification serves two important functions:

- It provides the information necessary for selecting an appropriate countermeasure and target audience for your program. You will be looking for information on the magnitude of the problem, the underlying causes, and the target groups most affected. This information should enable you to select the most effective countermeasure.
- It may provide some of the baseline data needed to determine if the program meets its objectives. You may start your problem identification with crash data, but you will also need to

collect other types of data in order to understand the problem you have and to select the most effective strategy for dealing with it. This might include baseline observations of seat belt use, measures of enforcement levels, public opinion and awareness surveys or speed counts. At this stage, it is also helpful to gather any trend data that may have been collected over the prior few years so that you will be able to show a trend before and after your program.

During the problem identification step, you also lay the foundation for your data collection efforts throughout the program evaluation. As you collect your baseline data, it is critical that you carefully document the procedures you follow, so that data collected later in the project can be compared with your baseline. In order for the data to be compared, it has to be collected at the same locations and times of day, using the same collection forms, and ideally the same observers. Failure to follow the same data collection procedures can make it difficult to document your accomplishments.

Seat belt use rates in the US continued to remain at unacceptably low levels into the late 1990s even after many years of public information and education as well as enforcement campaigns. The seat belt use rate in Chemung County, NY, on October 1, 1999, was only 63 percent.

In the early 1990s, there was interest in evaluating the effectiveness of Selective Traffic Enforcement Programs (sTEPs). Modeled after Canadian efforts, these programs combined high levels of enforcement with high levels of publicity about the enforcement, and were subsequently found to produce large gains in seat belt use. The National Highway Traffic Safety Administration and the Insurance Institute for Highway Safety chose to implement a modified sTEP to increase seat belt use in Chemung County, NY.[1,2]

Develop Reasonable Objectives

STEP 2

Once you have identified the problem and selected a strategy for addressing it, you need to define what you expect to accomplish. Many would argue that this is the most critical step in the evaluation process because it determines what success will be and how it will be measured. To increase your chances of success, find out what programs have already been done to effectively address your problem. (See the Resources section.)

Volumes have been written on how to write program objectives, each with its own set of do's and don'ts. These rules are all similar and it is not important which set you follow. The one advantage to the list shown at right is that it is easy to remember.

Program objectives should be SMART:

- ☑ **Specific**
- ☑ **Measurable**
- ☑ **Action-oriented**
- ☑ **Reasonable**
- ☑ **Time-specific**

Program objectives should be SMART (Specific, Measurable, Action-oriented, Reasonable and Time-specific).

Objectives should be SPECIFIC

Avoid using generalities like "improving traffic safety" or "increasing awareness." If you identify exactly what you want to happen, then you can document your success. Sometimes you can be specific about the amount of change you anticipate, expressed either in absolute (increase seat belt use to 75 percent) or relative (increase citations by 15 percent over the baseline) terms. At other times, you can simply observe and record the change in behavior.

Objectives should be MEASURABLE

For an objective to be measurable, there must be something you can quantify, like DWI citations, and you must be able to detect a change over time. When possible, isolate the targets of the countermeasure. For example, you want to increase by 10 percent the number of DWI citations issued to young drivers. Make sure to use data that you can obtain or that you can collect if it does not exist in another source.

You will learn more about this in the next section.

Objectives should be ACTION-ORIENTED

Action is good. You usually can see an action and count the number of times it happens. It is much easier to document that seat belt laws were enforced by counting the number of traffic stops and citations, than it is to document if public support for seat belt law enforcement increased. When resources allow for it, it is ideal to measure behavior(s) directly related to your objectives.

Objectives should be REASONABLE

A small community implemented a public information campaign on the value of traffic safety enforcement. The published objective of this public service campaign was to reduce traffic deaths community-wide. While this would be a desirable outcome, it is not a reasonable one. Public information and education programs can change knowledge, attitudes and awareness, but they have not been shown to change behavior in traffic safety, unless linked to highly visible enforcement. This community should take another look at the problem they are trying to solve, select a specific countermeasure that will address that problem, and then establish a reasonable target for success.

Objectives should be TIME-SPECIFIC

Projects don't last forever and objectives should have deadlines. Deadlines make it clear to everyone when results can be expected. They also keep people focused on what needs to be accomplished by when so it will be very obvious if you meet them or not. They challenge you to accomplish what you set out to do and serve as a constant reminder of your criteria for success. This is all the more reason to be honest and practical when you write them.

SAMPLE OBJECTIVES

Not So Smart	S.M.A.R.T
To encourage increased seat belt enforcement	To increase seat belt citations by 15 percent in 6 months
To reduce underage drinking	To reduce the number of liquor establishments that serve minors by 40 percent in 12 months
To get tough on speeders	To decrease average vehicle speed on Smith Road from 55 mph to 45 mph in 6 months

Once you have drafted the program's objectives, circulate them to those decision-makers who hold the fate of your program in their hands. You need to get buy-in at the outset as to what you are trying to accomplish. If they're expecting dramatic bottom line results (i.e., a reduction in fatalities), now is the time to explain to them why that would be difficult, if not impossible, to demonstrate, especially in the short term. If you wait until the program is over, they will likely come to the conclusion that the program failed because it didn't meet *their* objectives even if the program accomplished *your* objectives!

This does not mean that your community may not experience a reduction in injuries and deaths over time. If you continue to implement effective countermeasures targeting specific traffic safety problems, you should begin to observe a downward trend in crashes and, ultimately, injuries or deaths. Your decision-makers need to understand, however, that this improvement will not occur overnight.

EXAMPLE STEP 2: DEVELOP REASONABLE OBJECTIVES

In October 1999, National Highway Traffic Safety Administration and the Insurance Institute for Highway Safety conducted a sTEP called "Buckle Up NOW" in Chemung County, NY, with the objective to demonstrate that seat belt use can be increased to 90 percent or more in a period of three weeks.

STEP 3

Develop a Plan for Measuring Results

Before you can begin implementing a program, you have to plan how you will conduct your evaluation. This plan will address the questions:

- What will you measure?
- How will you measure it?
- How will you analyze your results?

While all of these questions are important, the first, what you will measure, is critical to the success of your evaluation.

Remember that you need to understand any reporting requirements from the funder and incorporate the necessary information collection into your plan.

What will you measure?

What you will measure must be tied directly to the objectives you have established for your program. If your objective is to reduce speeding on a given roadway, the most logical thing to measure would be average speeds on that given road. Since your objective is tied to speeding, you don't need to spend time or money trying to measure a reduction in crashes.

The problem that you will face again and again is that everyone else will be urging you to tie program success to saving lives. Rather than getting backed into that corner—remember what we said in Chapter Three about raising expectations that your program will save lives—you should point out that the traffic safety literature indicates that excessive speed contributes to serious crashes. Since you have documented that there is a problem with excessive speeds on specific highways in your community, you are going to implement a countermeasure whose objective is to reduce speeding. You will measure program success by monitoring speeds on the selected road segments before, during and after your program is in effect.

Wherever possible you should try to measure observable phenomena, things you can see and quantify, and that occur with a high degree of frequency. The phenomena can include behaviors, knowledge, opinions, attitudes or program activities such as media coverage or enforcement actions.

Measuring success becomes harder to do, and more difficult to credit to your program, as the measurement moves from *what activities occurred* (e.g., number of traffic citations) towards *outcomes*. However, outcomes are what hold the greatest importance, such as whether there's been a decrease in traffic crashes or crash-related injuries and deaths. On the following page is a listing of some measures that can be used to evaluate traffic safety programs. They are listed in descending order of importance for measuring outcomes, but in ascending order of convenience and ease of data collection; in other words, as you read down the list, the data are easier to obtain, but have less of a direct link to measuring reductions in crashes, injury, and deaths.

Changes in these observable phenomena can be caused by your program or by some other confounding factors such as engineering improvements along a roadway. It will be important to understand what these confounding variables might be and how you can control them. This is an area in which an evaluation specialist can be extremely valuable.

How will you measure it (and when)?

Once you have decided what you will measure to determine if your program achieved its objectives, you will need to decide how you will gather the information needed to make the measurement. To collect valid data you, or your evaluation specialist, will need to determine where and when to collect the data, how much data will be needed, and what procedures will be followed to collect and record the data.

There are five basic ways that you can measure program effects:

1. Observational Surveys
2. Knowledge/Attitudinal Surveys
3. Activity Records
4. Data Records
5. Media Coverage

Important vs. Convenient Possible Evaluation Measures[3]

Primary Outcomes	
Changes in crashes (the number or severity)	
Reductions in fatalities and the severity of injuries	
Secondary Outcomes (also called Proxy Measures)	**Examples**
Changes in observed behavior	• Observing the use of seat belts and child safety seats • Observing the use of bicycle and motor-cycle helmets • Measuring the speed of vehicles • Observing red-light running • Counting the number of pedestrians who jaywalk
Changes in reported behavior (what people say they do when asked)	• How often do you wear your seat belt? • Have you ever driven after having too much to drink?
Changes in attitude (what people believe)	• Support for legislative initiatives • Knowledge of a seat belt law • Teen attitudes about drinking and driving
Changes in awareness (what messages people have heard)	• Awareness of high visibility enforcement campaigns • Perceived risk of getting a traffic ticket
Changes in activities conducted (new program implemented)	• Citations issued by the police • Special police patrols and check-points • Presentations • Training programs • Media coverage • Legislation changes

Source: Devon County Council, 1999.

Observational Surveys

Observational surveys are used to measure changes in safety behaviors. They can detect the presence or absence of a behavior (wearing or not wearing a helmet), or record some measurement of a condition (a vehicle's speed, or the size of a traffic gap that a person accepts before pulling into traffic).

Knowledge/Attitudinal Surveys

These surveys are used to collect attitude, knowledge and opinion information about individuals. They can be administered in person or by telephone, mail, e-mail or Internet. Each of these approaches has its own strengths and weaknesses which an evaluation specialist can describe for you. Surveys can provide a wealth of information but the survey instrument must be designed very carefully and tested thoroughly and the procedures for selecting individuals to survey (the sampling plan) must be well thought-out or else bias will be introduced into the data. Questions should directly relate to the behavior you are trying to change with your program. Getting a result that reports "parents really enjoyed the teen driving safety program," does not tell you anything about their attitudes towards their teen's supervised driving needs under their State's graduated driver licensing (GDL) requirements or how the teens responded to it.

Surveys conducted at State driver license offices are a common way to obtain information from the general public. Usually the survey includes questions to determine whether the program's primary messages have been heard and whether changes in attitudes or behavior occurred that might be attributed to the program. See page 46 for an example survey form used for a high visibility seat belt use campaign.[4]

Activity Records

Activity records are used to track process data such as the number of presentations made (and where and when), the number of patrol hours and the number of visits made to liquor establishments. These forms should be tailored to the specific data you need to capture and should be designed in coordination with the people who will

be using them. Consideration also has to be given to the conditions under which the forms will be completed. These forms should be tested with your specific target audience prior to using them to be sure that they are understandable and that they give you the kind of information you intended to collect.

Data Records

These records document a variety of issues. They are powerful because their use allows you to consider trends such as how a behavior, like speeding, has changed over time. Sources include:

- Police crash records
- Department of Motor Vehicle driver records
- Traffic citations logs
- EMS transport records
- Emergency room records
- Traffic court files
- Hospital disposition records

Beware, however, that unless your program is conducted over a long period of time and reaches a large area (like a county or state), records are not likely to reveal a change in behaviors.

The biggest challenge you will face with data records is getting access to them. Any organization that maintains databases with any personal information will have very strict guidelines for who can access the information and what can be done with it. Make sure that your evaluation specialist understands these data sources and has experience accessing them. Keep in mind that records may change over time as improvements are made in data collection. You will need to check each source to see if it is consistent over the time period you want to use.

If you are conducting a State-level evaluation that focuses on fatalities and injuries, your State may have Crash Outcome Data Evaluation System (CODES) data that can be useful. CODES data links

crash information together with hospital injury outcome and cost data and other available traffic records for more detailed evaluations. All States can use the Fatality Analysis Reporting System (FARS), the General Estimates System (GES), and the National Automotive Sampling System (NASS), all of which are maintained by NHTSA. Information about these sources, as well as the availability of CODES data for your State, can be obtained from:

The National Center for Statistics and Analysis, US DOT, NHTSA
Phone: 800-934-8517 or 202-366-4198
NHTSA Hotline: 888-327-4238
Web site: www.nhtsa.dot.gov

Media Coverage

High visibility enforcement campaigns have been demonstrated to change behavior. They work on a general deterrence model which means that the general public is more likely to obey the law when they perceive that consequences of breaking the law are quick and certain. As a result, campaigns should reach the public with messages that convey that a driver who breaks the law (speeding, not wearing a seat belt, driving while impaired) will be caught and punished. High levels of enforcement, coupled with extensive media coverage that provides information about the enforcement activities has been shown to be effective in reducing alcohol-related crashes and increasing seat belt use. As a result, these programs are becoming more widely implemented. Campaigns often include earned media coverage, such as newspaper articles, and occasionally paid media, such as television ads. In order to understand the outcome of your media campaign, there are two components to measure:

1. The amount and types of media coverage, such as which television channels mentioned the enforcement campaign and the length of the segment.
2. Audience awareness of the enforcement campaign. While it's good to know about the media coverage obtained, it's even bet-

ter to find out if the target audience received the message: if they heard it, believed they might get a citation and hopefully changed their behavior. Surveys conducted at driver license offices are one way to collect this information.

Conducting an effective, high-visibility enforcement campaign requires some careful planning. Take advantage of the NHTSA communications campaigns tools and information available at: www.trafficsafetymarketing.gov and www.stopimpaireddriving.org.

Develop data collection procedures

Once you have determined the type of data you will be collecting and its source, you will need to develop systematic procedures for data collection. You cannot leave this important step to chance. You will likely have multiple people collecting data and you want to minimize any variations in how they interpret what they are seeing. You accomplish this by designing data collection forms that can be used by everyone, and by providing training on how to make observations, read police forms, and other related tasks. You want each person to collect data in exactly the same way. If you collected observational data as part of your problem identification activity in Step 1, use the same procedures you used then so that you can make a valid comparison. Your evaluation specialist will be responsible for ensuring systematic data collection procedures.

You want each person to collect data in exactly the same way. If you collected observational data as part of your problem identification activity in Step 1, use the same procedures you used then so that you can make valid comparisons.

There is one last consideration when planning for measurement: the timing of your data collection efforts. We have discussed how your pre-and post data should be collected under similar conditions, including time of year and time of day. You must also consider when the post data should be collected in relation to the implementation schedule. For exam-

ple, you will want to collect seat belt use data immediately after a major enforcement blitz to determine if belt use changed. Traditionally, each increase after an enforcement blitz will drop off slightly over time. So you also need to know what the long-term effects of that enforcement blitz may be. You will therefore need to plan for follow-up data collection at scheduled intervals after implementation is complete. Your schedule for data collection should be determined before implementation begins so that it will not be influenced by the implementation itself.

A plan was developed to collect data on observed seat belt use, enforcement activity, paid and earned media, self-reported seat belt use, attitudes related to the seat belt law, and recall of activities undertaken to encourage compliance with the law. Observational seat belt surveys measured seat belt use rate in the community shortly before the campaign began and daily for three weeks.

Gather Baseline Data

STEP 4

During problem identification, you gathered preliminary data on such factors as seat belt use, and documented how you collected this information so that you can repeat these procedures after implementation. Now that you have refined your program objectives and developed a plan for measuring results, you may collect some additional data about other aspects of your program. You may need to conduct an awareness survey to document what the public thinks about DWI enforcement, before you implement a campaign to conduct sobriety checkpoints on weekends. This information should all be gathered before you actually start implementing anything, so that you can easily isolate any effect your program may have.

Before the program began, a seat belt observation form was used to measure the belt use rate. During the program, observers also used the form when they stood at the same street corners at the same time each day to record belt use of the first 50 vehicles to pass.

SAMPLE - SEAT BELT SURVEY - DATA COLLECTION FORM

Site number: _____ State:_____ County:_____

Date: _____ Day of week:_____ Start Time: _____

	Driver				Front Outboard Passenger		
	VEHICLE TYPE C-Pass Car T-Pick Up S-SUV V-Van	RACE W-White B-Black U-Unknown	SEX M-Male F-Female U-Unknown	SEAT BELTED Y-Yes N-No	RACE W-White B-Black U-Unknown	SEX M-Male F-Female U-Unknown	SEAT BELTED Y-Yes N-No
1							
2							

Example seat belt observation form[4]

STEP 5

Implement Your Program

Many people would be surprised to see implementation as a step in the evaluation process. But remember, you should be monitoring how your project is going right from the start, rather than waiting until everything is over. You should be keeping track of project costs and other process data that could indicate if program activity is at expected levels. You might do periodic opinion polls or awareness surveys to see if the public is paying attention to your public information campaigns. You should also gather feedback at any training programs or public presentations. You may discover that there is a serious problem that should be fixed before any other contacts are made. If you include any media events in your program, you definitely want to track the amount of media coverage you receive. This information is much easier to capture in real time rather than to recreate the records weeks or months later.

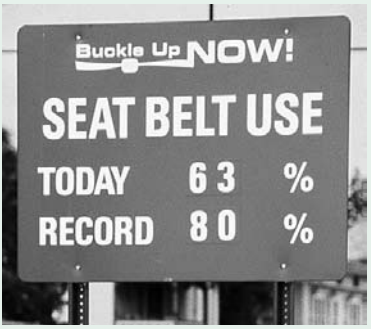
It is critical that you document whatever you learn during program implementation. You may have planned for weekly sobriety checkpoints at five locations in the county, with support from the State Patrol. Due to unexpected budget cuts, the State Patrol can only support one location per week. This may be a problem that you can't fix, but you need to consider it when you analyze the citation data. Based on this development you may want to adjust your program objective or extend the duration of the implementation phase. You will definitely want to document how actual implementation differed from your plan and what impact you believe this change could have.

The Buckle Up NOW sTEP campaign began with a well publicized news conference led by the Chemung County Sheriff. A local communications firm coordinated publicity in newspapers and radio ads for the entire three weeks. They developed posters and flyers to place around town. Signs saying, "We enforce seat belt laws" were on every police car.

The first week, the message announced the upcoming enforcement campaign. Checkpoints began in the second week and the publicity message moved into no-excuses high visibility enforcement.

Daily belt use observations—using modified but standardized techniques and samples—were conducted to provide information to the community. Feedback signs and news releases conveyed the most recent belt use rates (see photos on following page).

There were 32 checkpoints during the three week program. The enforcement goal was to make it virtually impossible to drive without getting a citation when not wearing a seat belt. A final end-of-program media event was held announcing program results and thanking all participants in the community.



Images courtesy of the Insurance Institute for Highway Safety.

It is critical that you document whatever you learn during program implementation.

STEP 6

Gather and Analyze Data

While the work involved in planning an evaluation is critical to success, gathering data is the most labor intensive aspect of the program evaluation, and analyzing it may be the most complex. As a manager, your biggest concern during the data collection phases is that the effort is adequately staffed and that everyone has been trained on the correct procedures to follow. Your evaluation specialist should also keep you informed about any changes that have to be made because of some external event that could influence the outcome. For example you may have collected baseline data on child safety seat use outside of a child care center. One year later, when you are looking to see if your campaign had an effect, you discover that the center has closed. Your evaluation specialist will need to find a suitable alternate site so that you don't miss any data.

Meaningful in this context means that both you and your funding sources will be satisfied that the program really made a difference.

During the analysis phase, your main focus should be on becoming comfortable with

the statistics. Your evaluation specialist will determine what statistical tests, if any, are appropriate and should be able to explain them all to you in terms that you understand.

When you start to get results from your evaluator, there is one thing that you should keep in mind. If something is "statistically" significant, that doesn't mean it is also "programmatically significant" or *meaningful*. Meaningful in this context means that both you and your funding sources will be satisfied that the program really made a difference.

Your evaluator may tell you that there is a statistically significant decrease in the number of repeat DWI offenders following implementation of your mandatory sentencing program. She can report with a high degree of confidence that this change is not due to chance. However, when you look at the actual numbers, you discover that the total number of repeat offenders only dropped by ten. While your evaluator is pleased to prove that your program was a success, you may worry that your funding source may view this result with less enthusiasm. Regardless, consider how you will report your findings to your funder so that the results are clear and truthful.

Causal or Correlated?

A word of caution about statistical analyses and how they are reported: Your evaluator will very carefully choose the right words to describe the outcomes observed and their relationship to the countermeasure implemented. Distinctions will be made between a causal relationship (Implementing A caused outcome B) and a correlation (A was implemented and B happened, and they appear to be connected). The distinction is an important one, and should not be lost in the excitement of success. If your evaluator does not use the term causal relationship, it is because she does not believe that a causal relationship can be proven with the data available. Even though correlation is harder to explain than causation, don't undermine the validity of your effort by slipping into sloppy terminology.

Collected data included:

- Seat belt observational surveys conducted before, during and after the enforcement campaign
- Reports by participating enforcement agencies on the amount and type of enforcement carried out
- Tabulations of the amount and type of media purchased for the campaign as well as the amount and type of earned media generated
- Public awareness surveys assessed who noticed the enforcement and media efforts, and how much support existed for seat belt laws

STEP 7

Report Results

The results are in, and your program was a big success. Before you celebrate, however, you need to pay attention to a very important step in the process. Successful evaluation needs to be clearly communicated to key audiences.

Your purpose in reporting evaluation results are two-fold:

1. You want to convince your funding source that they should continue funding your traffic safety program, and maybe even increase their support.
2. You want to generate support for your program among the media, the general public and among the other organizations you would like to take a more active role in traffic safety.

As program manager you will need to report your results to your funding source and to the media, at a minimum. If other organizations were involved in implementation, you should share the results with them, along with appropriate thanks for their participation.

The presentation of your results will vary depending on your audience. Create a detailed report for your funding source to demon-

strate that you take evaluation seriously. It must include a short, concise Executive Summary, which hits the high points and emphasizes the conclusions. The detailed report should include an accounting of how your program funds were spent and follow a standard research format, with the following sections:

Table of Contents

Executive Summary
No more than three pages in length, ideally shorter.

Background
Why the study was conducted and the questions it attempts to answer. It should include the objectives for the program being evaluated and the criteria for success.

Methods
Complete descriptions of the design, procedures, techniques and other details that describe how data was collected and analyzed. Questionnaires and data collection forms should be included in an appendix.

Findings
The outcomes of the evaluation, presented with easy-to-understand tables, graphs and explanations.

Discussions and Conclusion
Interpretation of the findings, how they relate to the purpose of the evaluation and the objective of the program being implemented.

Recommendations for Action
Discussion of changes that should be made to the program to increase effectiveness. This section could also include proposals for continued or even increased funding, based on the results provided.

The evaluation specialist should be principally responsible for the Methods and Findings sections and should have major input on all other sections.

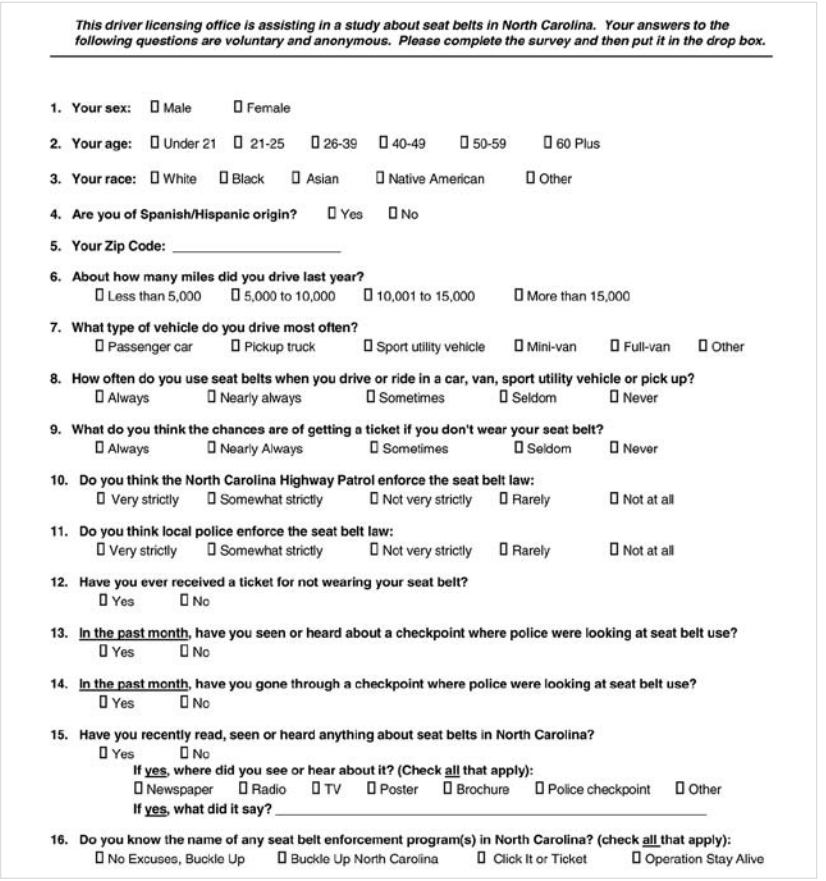
Your report to the media, and through them to the general public, should be very different. It can be issued as a press release which specifies what was done, and why, and what the results were. This information should focus on the impact the program will have on the general public. Will they be seeing more enforcement on the street? Will their children be safer walking to school? A clear table or graph of the most significant findings should be included if possible. Your audience will understand percentages as well as the concept of risk when applied to traffic safety. Try to include a discussion of the average person's risk of being involved in a crash, and how that risk may have changed as a result of your program.

Each experience should provide important lessons learned that can save you time, money, and frustration in the future.

Once you have communicated your results to everyone, you need to turn your attention to what changes should be made before you implement the program the next time. You should review all the documentation on what went right, and what obstacles were encountered, so that you can do some contingency planning the next time. You should also review your performance against your budget and milestone schedule to determine if you need to request more money or allow more time in the future. Did you have enough data collectors? Did the media understand what you were doing? Did you get enough cooperation from the local police or school system? All of the factors should be reviewed and incorporated into your planning for future implementation of this same project or any others. Each experience should provide important lessons learned that can save you time, money, and frustration in the future.

Seat belt use rates during the program rose from 63 percent at baseline, to 75 percent after the first week of enforcement, to 84 percent after the second week, and to 90 percent in just three weeks. In addition, Elmira's opinion surveys found that 90 percent of the respondents were aware of the safety belt program, public perception that the belt law was being enforced "very strictly" increased from 34 percent before "Buckle Up NOW!" to 77 percent after the program, and 79 percent supported enforcement to increase seat belt use. Several articles and reports were written to publicize the results to the general public, funders and other agencies which had interest in transportation safety.

Sample Driver Licensing Office Survey[4]



This driver licensing office is assisting in a study about seat belts in North Carolina. Your answers to the following questions are voluntary and anonymous. Please complete the survey and then put it in the drop box.

1. **Your sex:** ☐ Male ☐ Female
2. **Your age:** ☐ Under 21 ☐ 21-25 ☐ 26-39 ☐ 40-49 ☐ 50-59 ☐ 60 Plus
3. **Your race:** ☐ White ☐ Black ☐ Asian ☐ Native American ☐ Other
4. **Are you of Spanish/Hispanic origin?** ☐ Yes ☐ No
5. **Your Zip Code:** _____
6. **About how many miles did you drive last year?**
   ☐ Less than 5,000 ☐ 5,000 to 10,000 ☐ 10,001 to 15,000 ☐ More than 15,000
7. **What type of vehicle do you drive most often?**
   ☐ Passenger car ☐ Pickup truck ☐ Sport utility vehicle ☐ Mini-van ☐ Full-van ☐ Other
8. **How often do you use seat belts when you drive or ride in a car, van, sport utility vehicle or pick up?**
   ☐ Always ☐ Nearly always ☐ Sometimes ☐ Seldom ☐ Never
9. **What do you think the chances are of getting a ticket if you don't wear your seat belt?**
   ☐ Always ☐ Nearly Always ☐ Sometimes ☐ Seldom ☐ Never
10. **Do you think the North Carolina Highway Patrol enforce the seat belt law:**
    ☐ Very strictly ☐ Somewhat strictly ☐ Not very strictly ☐ Rarely ☐ Not at all
11. **Do you think local police enforce the seat belt law:**
    ☐ Very strictly ☐ Somewhat strictly ☐ Not very strictly ☐ Rarely ☐ Not at all
12. **Have you ever received a ticket for not wearing your seat belt?**
    ☐ Yes ☐ No
13. **In the past month, have you seen or heard about a checkpoint where police were looking at seat belt use?**
    ☐ Yes ☐ No
14. **In the past month, have you gone through a checkpoint where police were looking at seat belt use?**
    ☐ Yes ☐ No
15. **Have you recently read, seen or heard anything about seat belts in North Carolina?**
    ☐ Yes ☐ No
    If yes, where did you see or hear about it? (Check all that apply):
    ☐ Newspaper ☐ Radio ☐ TV ☐ Poster ☐ Brochure ☐ Police checkpoint ☐ Other
    If yes, what did it say? _____
16. **Do you know the name of any seat belt enforcement program(s) in North Carolina? (check all that apply):**
    ☐ No Excuses, Buckle Up ☐ Buckle Up North Carolina ☐ Click It or Ticket ☐ Operation Stay Alive

[1] NHTSA. (2001). *Achieving a High Seat Belt Use Rate: A Guide for Selective Traffic Enforcement Programs* (DOT Report No. DOT HS 809 244). Washington, DC: NHTSA.

[2] Insurance Institute for Highway Safety. (2000). *90% Belt Use - Successful STEP: How to Boost Belt Use Like They Did in Elmira, New York*. Arlington, VA: IIHS.

[3] Adapted from: Devon County Council. (1999). *Speed Management Strategy for Devon*. Accessed at: http://www.devon.gov.uk/index/transport/roads/road_safety/safer_speeds.htm

[4] NHTSA. (2002). *Evaluation of Click It or Ticket Model Programs* (DOT Report No. HS 809 498). Washington, DC: Preusser Research Group.

CHAPTER 5

Getting Help

The previous chapter provided a brief overview of the steps involved in evaluation. There are some complex parts to evaluation that can't all be explained in this brief Guide but where an evaluator will be able to help. The key is to be able to describe the help you need and find the person who best fits the project.

What an Evaluation Specialist Can (and Should) Do for You

An evaluation specialist is able to:

- ☑ Design the evaluation
- ☑ Recruit and train the data collectors
- ☑ Collect the data
- ☑ Provide interim feedback during the conduct of the program
- ☑ Analyze the data and present the findings
- ☑ Provide input to you as you draw conclusions

Since the evaluation should be designed right along with implementation, you want the evaluation specialist working with you at the beginning, when you establish your program objectives. Your evaluation specialist can help you focus on what can be measured and what questions you will be able to answer with the evaluation. The evaluator will also be able to counsel you about the problems you might encounter when gathering or analyzing particular types of data. For example, the evaluator will recognize the potential effect of seasonal differences in driving levels, or the impact political changes might have on enforcement levels. Beyond recognizing the potential problems, the evaluator will also know how to deal with them.

With an experienced evaluator on board, your job is to stay in touch with the evaluation to make sure that the processes and products are meeting your expectations. You will also need to listen if your eval-

uator identifies problems with your plans. An evaluator is trained to be objective and it's worthwhile to take seriously any concerns brought to light.

Before you start looking for an evaluator, you should prepare a clear statement of work with the specific tasks to be accomplished, a description of all deliverables, and a schedule for task completion. This document will serve as the foundation for your relationship with your evaluator. It should be as specific as possible so that there are no misunderstandings down the road.

What to Look for in an Evaluator

If you are like most managers, you get a little nervous delegating responsibility for a critical activity to someone you don't know very well. The anxiety increases significantly if the activity involved is highly technical and is outside your area of expertise, because you know you can't just step in and take over data analysis if there is a problem. The way to overcome this anxiety is to have a very clear understanding of what to look for in an evaluation specialist. The following is a list of criteria you can use for selecting an evaluation specialist.

When you're hiring an evaluation specialist, look for someone who:

Provides clear written and verbal explanations of evaluation processes and results.

Ask the evaluator to briefly explain a recent project to you. Watch for the use of technical jargon. If you can't understand the response during the interview, you won't understand the final report. Ask several questions to measure the evaluator's patience and ability to explain concepts. This individual may need to represent you someday in a meeting. How will he or she do?

Request samples of previous evaluation reports to see if they are written in understandable terms. Also pay attention to how numbers are graphically presented. Do the graphs and charts highlight key issues and make sense to you?

Has previous evaluation experience, particularly in the use of behavioral observations, public opinion questionnaires and analysis of data records.

This is another reason to request examples of previous evaluation reports. Read the descriptions of the types of data collected and look for examples that resemble the type of work you will be doing. Talk to program managers for these projects to assess the evaluator's performance.

Fully understands research design and statistical techniques and when they should and should not be applied.

Good research design is always needed, even on simple evaluations. Statistical analyses may or may not be appropriate, depending on what you are trying to measure.

Can get access to data and data collectors.

You want your evaluator to make your job easier for you, not harder. A good evaluator will already know how to access a variety of relevant data sources. You should not have to locate the most appropriate sources or negotiate access to them. Look for someone with experience working with the following types of data, if appropriate to your evaluation:

- Crash records
- Police reports
- Court records
- Medical records

The evaluator should also have experience dealing with data collectors who may include students, volunteers from the community, and temporary workers.

When a manager is intimidated by the whole concept of evaluation, it's tempting not to ask the nagging questions for fear of looking uninformed. However, while you are delegating the tasks associated with evaluation, you cannot delegate your own responsibility for managing all aspects of your program. Therefore, it is critical to hire an evaluation specialist with whom you are completely comfortable. You should feel free to ask any questions that occur to you, no matter how fundamental, and you should understand every answer that is given. If you don't have that relationship with a potential evaluator, keep looking.

Sources of Evaluation Assistance

With your criteria for selecting an evaluator in hand, you are almost ready to locate the right person for the job. Check your agency's policies regarding hiring and paying for evaluation services. Once you find someone, you want to be able to quickly bring them on-board.

There are a variety of sources for evaluation expertise. The following are good places to look for evaluators or ask for leads:

Your agency

Explore the resources in other departments of your own agency and in other agencies within your jurisdiction. However, don't make the mistake the manager in the opening scenario did. Examine an in-house evaluator with the same rigor that you would an outside consultant. A few college statistics courses do not qualify someone as an evaluator.

Transportation institutes

Many universities have institutes that specialize in transportation, including traffic safety, and may have experienced evaluators on staff.

NHTSA

Search the NHTSA Web site (www.nhtsa.dot.gov) for information, as well as contact details for your NHTSA regional office (see "contact NHTSA" on the Web site). The regional offices may have assistance directly available or can provide a list of potential evaluators.

State Highway Safety Office

Each State has an office that coordinates programs on a variety of highway safety issues. Find State offices by searching the Governors' Highway Safety Association Web site (www.ghsa.org).

Local colleges and universities

Check with the psychology, public health, epidemiology, and education departments at nearby universities. You should be cautious recruiting in a math department because of their focus on theoretical statistics rather than applied statistics.

Expect to pay for the evaluation services you obtain from local universities, but they may be an excellent source of low-cost data collectors and graduate students, as well.

Be sure to find out about the time needed to establish a contract and then clearly specify milestones and due dates so that your project doesn't get lost in the shuffle.

Private consultants

There are numerous consultants in the private sector with evaluation expertise. Unlike universities, private consultants do not have the schedule limitations caused by semester and summer breaks. Keep in mind that the firm you hire should have traffic safety evaluation experience, and should be familiar with the data sources that you will be using.

Depending on your procurement requirements, you may need to conduct a competition to hire a private consultant. This may take several months to complete. Be sure to allow enough time for this process so the evaluator can be involved from the project's start.

Working with an Evaluator

Remember the statement of work you developed before hiring the evaluator? It includes tasks, deliverables and deadlines. This should be referred to often as you work together. You will also need to prepare a budget for your evaluation, and this should be negotiated with your evaluator.

Meet regularly, either in person or over the phone. This demonstrates that you are interested in what's happening and that you want to be involved in any major decisions that need to be made. At the same time, avoid the temptation to micro-manage. You should be focusing your attention on the overall implementation, not the details of the training for data collectors. If there is a problem, your evaluator should tell you about it. Finally, think about how many reports you want the consultant to provide and the type of information you want included. Frequent progress reports are essential if they are the only form of communication you have with your consultant, but if you can meet face to face every week, progress reports may be an extra burden that takes time away from your evaluation effort.

CHAPTER 6

Closing Comments

We want to end with four last thoughts on program evaluation.

It doesn't have to be hard. 1

If you start out with the intention of keeping your evaluation as simple and straightforward as possible, you are much more likely to have useable results. Resist anyone who tries to expand the focus or complicate the design unless the evaluator explains why a more complex design would be beneficial or necessary. Keep the level of evaluation consistent with the size of the program and the objectives you are trying to meet.

It doesn't have to be expensive. 2

First, re-read #1, and keep your design as simple as you can. Second, take advantage of the resources that exist in your community. You might be able to convince a university professor to take your evaluation on as a master's thesis project for a student. Maybe you can hire an evaluator and recruit volunteer data collectors from local citizens organizations. Work with your evaluator to identify activities on which you can economize, and which areas are worth spending a little extra.

Investing in evaluation can save you time and dollars over the long haul. 3

With the information you learn from a worthwhile evaluation, you can focus your resources on the most critical problems and the most effective countermeasures. You will also be able to adjust programs mid-stream to improve effectiveness. And most importantly, you will be much more likely to convince your funding sources that their dollars have been well-spent, which means that you are a good investment for the future.

4 It's never too late to start.

We have spent a lot of time stressing that evaluation should be built into a project right from the start, and not left until the final act of your program performance. However, if you are in the middle of a project right now and are eager to try out your new evaluation mentality, go right ahead. You certainly can check to see if implementation is going as planned and how resources are being spent. An evaluator should be able to help you review what baseline data exists and develop some simple performance measures that you can use to assess how the program did in meeting its objectives. It's even not too late to write some SMART objectives to clarify for everyone what you expect the outcomes to be.

The purpose of this Guide was to convince you that evaluation does not have to be intimidating. You will only truly be convinced when you apply the information you have read here to evaluate a program of your very own. What are you waiting for?

Once upon a time there was a project manager who was faced with a problem. The head of her department informed her that there were two new projects being planned as part of a national effort to reduce night-time collisions. Two county supervisors each had their own favorite solution. However, the funding source informed the department that the money they were providing could only go toward one new initiative. The department head refused to choose one project over another without empirical proof to justify her decision. So, the responsibility of pilot testing each approach and recommending one project over another was placed on the shoulders of the manager.

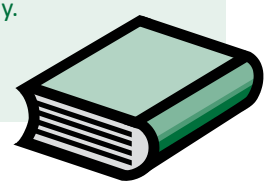
Remembering her training in evaluation management, the manager decided to approach this problem with an evaluation mentality. She was determined to save herself as much wasted time and effort as possible, so she decided to build evaluation procedures into each of the projects right from the start. With the assistance of a carefully selected profes-

sional evaluator, she asked five essential questions to put herself in the right mind-frame: "What do I know about the safety problems involved in night driving? What is the objective of each of these projects? How would I measure results? How can I collect the data I need? What are my criteria for success?"

Feeling like they had a firm grasp on each project, the manager and evaluator settled on reasonable objectives for each pilot test according to the SMART guidelines and created a plan for measuring results. They hired assistants to collect appropriate baseline data according to each project's focus. Next, the pilot programs were implemented according to the carefully outlined schedule. In the following weeks, the collected data was analyzed and the report was carefully drawn up.

The big day arrived. In the conference room gathered the department head, the two supervisors, and the funding representative, all anxious to hear the results. Calmly and confidently, the manager presented her findings. While one approach indicated modest success, she explained, the other program clearly surpassed it, raising safe night driving behaviors by 50 percent. Impressed by the convincing results, the funding representative heartily agreed to fund the successful project for three years. The department head recommended the manager for a long-overdue promotion. The victorious supervisor patted himself on the back for having though of such a brilliant idea. And even the not-so-triumphant supervisor took the news well, reassured that the outcomes had resulted from an impartial and professional study.

And they all lived happily ever after…



Glossary of Terms That Evaluators Use

Baseline – Information that explains conditions before a program begins. This may be related to seat belt use, attitudes towards the likelihood of arrest for impaired driving, motor vehicle injury deaths, or other information related to the traffic safety problem that needs to be solved. Data is collected that may describe attitudes, behaviors or outcomes before a program begins and this same information is collected after the program in order to identify changes. These changes help explain the value of a program, so having baseline data is extremely important.

Before and After Design – An evaluation design that assesses the change in an outcome measure as the difference between pre-program levels and post-program levels. An evaluation of a school-age pedestrian safety program, for example, might observe street crossing behaviors before and after the educational program had been implemented. An increase in the proportion of children observed demonstrating desired behaviors would provide evidence of program effectiveness. This design is sensitive to historical effects, however. If something else happened between the two assessment periods that might affect the observed behavior, then the outcome can not be unequivocally attributed to the program. In this example, the outcome would be confounded if the local news media gave extensive coverage to a child killed or injured by a hit-and-run driver. This design is stronger if a comparison group is also assessed at the same time periods as the treatment group.

Bias – A potential characteristic of non-random samples that affects the program's outcome. For example, an evaluation of a driver improvement program that is provided to volunteers cannot determine how well the program conveys information because volunteers may have different motivations than "average" drivers. Researchers prefer to use random samples whenever possible to avoid bias.

Confounding Factors (or Variables) – Events other than those being investigated that may also have an effect on the outcomes of the program. For example, the results of an evaluation of a speed enforcement program could be confounded by the highway department making engineering changes in the same areas as the enforcement efforts.

Comparison Group and Treatment Group – In order to demonstrate a program's effects, evaluators may compare a group that receives a countermeasure with an equivalent group that does not. The group getting the countermeasure is the "treatment" or "experimental" group and the other is the "comparison" or "control" group.

Correlation – A mathematical technique that assesses the extent to which one variable increases (or decreases) in value as another variable changes in value. Temperature in Fahrenheit and temperature in Celsius is perfectly correlated—as one goes up, so does the other. If one event causes another, they are correlated, but two variables that are highly correlated are not necessarily causally connected—they might both be caused by a third, unmeasured, variable.

Cost-Benefit Analysis – A process comparing the cost of a program with the savings resulting from the outcomes of the program. While it is often difficult to identify and enumerate all the costs and benefits, the process can be meaningfully applied to a single program. For example, a law requiring motorcycle riders to wear protective helmets has limited enforcement costs compared with fairly large benefits in health care expenses and welfare benefits avoided.

Cost-Effectiveness Analysis – A process for determining the relative benefit of alternative programs by comparing the amount each program costs with the extent to which each affects a common measure of effectiveness. In this analysis, the outcomes of the program need not be converted to actual dollars saved. In comparing two approaches to increasing seat belt use rates, for example, one could

calculate the cost of increasing belt use by, for example, 5 percentage points for each program.

Data Records – Information obtained from sources such as police crash records, driver records or other stored records. The findings may be used during problem identification or as a data source to ascertain changes over time. This is also called "archival data."

Earned Media – Media coverage about a program that typically requires proactive outreach, often in the form of a media advisory or news release, to news outlets that reach a program's target audience. Using appropriate media hooks such as personal impact (like the announcement of checkpoints for an upcoming holiday weekend) increase the effectiveness of your proactive efforts, resulting in a higher likelihood that your program will get the desired positive coverage.

Evaluation Design – The plan for conducting an evaluation in a way that permits the evaluator to rule out the possibility that other factors (other than the program) caused the observed outcomes. This plan should include a clear statement of the objectives of the program, how success will be measured, what populations will be exposed to the treatment, how treatment and comparison groups will be constituted, and how the data will be collected, analyzed, and reported.

Field Test – A study of a limited-scale implementation of a new program in a setting similar to where it is likely to be used. Field test sites are generally recruited from candidates showing a high level of interest in participation; a quality that sometimes provides an "ideal" environment rather than a "representative" one. This is not all bad, as it shows the potential benefit of a countermeasure unfettered by implementation problems.

Impact (or Outcome) Evaluation – An evaluation that determines the extent to which a program achieved its stated outcome objectives. For example, an impact evaluation of a program designed to

reduce pedestrian crossings against red lights could compare the observed post-program change in the number of pedestrians crossing on the red and green cycles at selected intersections with an appropriate comparison group.

Outcome Objectives – A specification of the events that would mark the successful achievement of the program's goals. These should be easily and unambiguously measured and closely related to the issues addressed by the program. While all traffic safety programs hope to reduce the number of traffic fatalities, reduction of fatalities is not often closely related to the program's activities. Rather, appropriate objectives should be related to changes such as increasing use of seat belts, reducing the number of drinking drivers, improving street-crossing behavior or increasing helmet use. Objectives may specify the populations of interest (e.g., decrease driving after drinking among Native Americans living in Nevada); and, in an ideal world, objectives should state a quantifiable level of change (e.g., increase belt use by pickup truck drivers on 2-lane rural roads in Iowa by 10 percentage points).

Paid Media – Media visibility that is purchased. High visibility enforcement campaigns sometimes include paying for placement of television or print ads with the campaign slogan and key messages. Well-planned, paid media placements assure a certain amount of air time or exposure for a program's messages.

Problem Identification – A deliberate process to describe and better understand how people are being injured or killed on the road. Usually includes reviewing crash and injury statistics and examining demographics (like age groups or ethnic groups), communities or situations where crashes most often occur. In order to know whether a program is associated with an improvement in a highway safety problem, the times and situations with greatest risk of injury or death need to be uncovered. Problem identification should occur before a program is developed and be clearly understood by the evaluators.

Process Evaluation – An assessment of the extent to which a program was implemented or conducted according to plan. It can indicate whether (and how) a program actually reached its intended target audience with the appropriate messages the desired number of times through the selected media. Process evaluations are most useful in troubleshooting unsuccessful programs delivering proven countermeasures. This is also called "administrative evaluation."

Proxy Measure – A measurement that is used in place of the ideal measurement to help determine program results. Proxy measures are used when it's difficult or impossible to measure the desired change in a direct manner. This may be because the opportunity to detect change does not happen very often or in very large numbers. An example of this would be in a community where injuries from bicycle crashes are rare or difficult to measure. Bicycle helmet use may be measured as a proxy for success in reducing injuries and deaths due to bicycle crashes.

Quasi-Experimental Design – A system of procedures for ruling out alternative explanations for study results when study groups could not be formed by random assignment. While random assignment to groups is the preferred method for ruling out bias in samples, many real-world situations do not permit random assignment. Consequently, evaluators must turn to other techniques (such as additional comparison groups, multiple levels of treatment, comparisons over long time periods) to dismiss threats to the validity of the study.

Random Sample (or Assignment) – A subset of a population chosen in such a way that each member of the population has equal probability of selection. Random samples permit the use of certain statistical procedures that provide measures of the potential error in estimates of means (averages) and differences between means of two groups. A simple system for making random selections is to create an alphabetical listing of population members and selecting every nth name. If the population list contained 1,000 names and the evaluator needed a sample of 100, she would select every 10th name.

Reliability – An assessment of the extent to which a measurement system will give the same results if used to assess the same events on repeated occasions. A measure can be reliable, however, without being valid. For example, a weekly count of citations for driving while intoxicated may be highly repeatable. However, it is not a valid measure for evaluating a program designed to reduce the incidence of impaired driving because it is so dependent on other factors, including police motivation, program funding, and department priorities.

Representative Sample – A group of individuals deliberately chosen from a particular population to try to emulate the characteristics of the target population as a whole. When random sampling is not possible, use of a representative sample, with careful attention to defining the relevant population characteristics may be an acceptable option. Focus groups are usually constituted using representative samples. For example, participants may be selected to match the following characteristics: 60 percent male, 40 percent female; ages 21 through 30; primary vehicle is pickup truck; drives more than 10,000 miles per year; graduated from high school and attended college for 2 or fewer years.

Statistical Significance – An estimate of the probability that the differences observed between treatment and comparison groups occurred by chance alone (i.e., that the treatment had no effect). The probability level below which results are said to be significant is somewhat arbitrary, but is usually .05 (5 chances in 100) or .10 (1 chance in 10). Statistical significance can be obtained with extremely small differences if the size of the groups is sufficiently large. While statistical significance can tell you if the results are not likely due to chance events, it cannot tell you if the size of the difference is programmatically meaningful (that is, worth the effort).

Treatment Group – See comparison group.

Valid/Validity – An assessment of the extent to which a measurement system actually measures what it is supposed to measure. For example, observed seat belt use is a much more valid measure of compliance with belt-use laws than is self-report on a survey. However, there are some circumstances (such as nighttime, fogged windows, high-speed locations) under which observations are not very reliable.

Case Examples

The following two examples of program evaluations are intended to give you a sense of how the steps to evaluation are put into practice. As you read the descriptions, you may also want to take note of the role of professional evaluators, the kinds of data sources used and the information the evaluation revealed about the programs.

Tipsy Taxi Offers Ride Home to Reduce Impaired Driving

Aspen, Colorado

As part of an effort to reduce impaired driving in the Aspen area, the Pitkin County Sheriff's Office, with assistance from the Aspen and Snowmass Police Departments and the local restaurant association, initiated a program to promote and provide a free taxi ride home. Rides are usually requested by a law enforcement officer or bartender who identifies someone who appears intoxicated and in need of transportation. The Tipsy Taxi service, which complements education and enforcement efforts, is available every day and during all hours.

There were two main evaluation interests: (1) to describe how Tipsy Taxi operated and (2) determine whether the program decreased the number of crashes where alcohol was a factor. To measure the first element, the evaluators visited the site and conducted interviews to understand the program's procedures. To examine the program's relationship to alcohol-related crashes, the evaluation team identified two communities with similar characteristics to the Aspen area but that did not have the Tipsy Taxi service that would serve as comparison sites. Finding similar communities would help the evaluators make the case that any differences in crash statistics could be attributable to Tipsy Taxi.

When considering data sources, the evaluators ideally wanted to use alcohol-related fatal crashes. However, because of the small number of these types of crashes in the study area, being able to detect a change would have been very difficult. Using a method tested by other researchers, they used nighttime crashes and injury crashes as proxy measures. They compared several years of data from the Aspen area and the two comparison communities, noting when Tipsy Taxi first began.

The evaluators analyzed the data and found a reduction in nighttime, injury and fatal crashes in the Tipsy Taxi service area. Analysis revealed a statistically significant reduction in injury crashes while there was no reduction in the comparison communities. It was concluded that it is possible to operate a ride service program that helps reduce alcohol-related crashes when conducted as one component of a multi-faceted approach to reduce impaired driving.

Search the NHTSA Web site (www.nhtsa.dot.gov) for the full report: NHTSA. (2000). *Evaluation of a Full-Time Ride Service Program: Aspen, Colorado's Tipsy Taxi Service*. (DOT Report No. HS 809155). Washington, DC: Lacey, J.H., Jones, R.K., and E.W. Anderson.

Heed the Speed Campaign Aims to Slow Down Drivers in Neighborhoods

Peoria and Phoenix, Arizona

Speeding vehicles, particularly in residential areas, pose a risk of severe injury to pedestrians. The Heed the Speed Campaign was implemented to determine whether education and enforcement strategies, when used in conjunction with physical traffic calming measures, afforded a greater reduction in speeds compared to traffic calming alone. Traffic calming measures included speed tables, speed humps and two types of pavement markings. The education aspect of the campaign included lawn and street signs printed with the slogan (Heed the Speed) and logo, as well as materials, press releases, and media coverage that conveyed messages about the relationship between speed and pedestrian injury severity. Enforcement strategies included the use of additional patrol hours, speed trailers, photo speed enforcement and the formation of neighborhood watch groups.

With input from local transportation personnel, evaluators identified a total of ten neighborhood roadways within the cities of Peoria and Phoenix, Arizona, for the program. The roadways each fit into one of three categories: (1) streets with no existing or planned traffic calming measures; (2) streets with existing traffic calming; and (3) streets with traffic calming measures implemented during the study. These categories offered the benefit of being able to study enforcement and education effects with and without roadway treatments.

Main data sources included a neighborhood survey conducted before and after the campaign regarding awareness of the different campaign components and perceptions about changes in traffic speeds and speed data collected three to five times during the program. In addition, law enforcement officers gathered additional information from drivers stopped during the additional patrols, nearby streets were monitored for changes in traffic volumes and interviews were conducted with project and neighborhood representatives to gather feedback.

Survey findings revealed an increase in awareness of all the campaign's components and, in all but one study road, the perception of reduced speeds. Speed data revealed a statistically significant reduction in speeds on all roadways except for one, which had pre-program average speeds below the speed limit. Because of the brevity of the study period, evaluators could not determine whether the speed reductions would last over time.

Search the NHTSA Web site (www.nhtsa.dot.gov) for the full report: NHTSA. (2006). *Pilot Test Of Heed The Speed, A Program To Reduce Speeds In Residential Neighborhoods*. (DOT Report No. DOT HS 810 648). Washington, DC: Bloomberg, Richard D. and Arlene M. Cleven.

Resources

This list is intended to serve as a starting point for finding information that is beyond the scope of this Guide.

Program and Countermeasure Selection Guides

These resources describe countermeasures or programs that have already been evaluated. Some include downloadable materials and guidance on how to conduct a program.

Achieving a High Seat Belt Use Rate: A Guide for Selective Traffic Enforcement Programs

National Highway Traffic Safety Administration (NHTSA) and the Insurance Institute for Highway Safety

Describes how a 90 percent belt use rate was achieved using a three week progression that began with public education of upcoming enforcement and then initiation of citations. This "how-to" guide explains how other communities can replicate the campaign. (2001)

www.nhtsa.gov/people/injury/research/ACHIEVE.pdf

Countermeasures That Work

National Highway Traffic Safety Administration

Review of literature and reports that describe the evidence of effectiveness of widely-used highway safety countermeasures. (2008)

www.nhtsa.gov/staticfiles/DOT/NHTSA/Traffic%20Injury%20Control/Articles/Associated%20Files/HS810891.pdf

Note: This publication is updated annually. Search the NHTSA Web site (www.nhtsa.dot.gov) for the most recent edition.

Guide to Community Preventive Services: Motor Vehicle Occupant Injury

Centers for Disease Control and Prevention

Describes countermeasures that have been evaluated and the results of

those evaluations. Identifies "recommended" strategies that have strong evidence for their positive effect.
www.thecommunityguide.org/mvoi/

NCHRP Report 500: Guidance for Implementation of the AASHTO Strategic Highway Safety Plan

American Association of State Highway and Transportation Officials (AASHTO)

Organized around the emphasis areas outlined in the AASHTO Strategic Highway Safety Plan, this series of guides provides background information and recommended strategies to address a specific type of crash or contributing factor. (As of 2008, new volumes were still being added.)
http://safety.transportation.org/guides.aspx

Stop Impaired Driving Campaign

National Highway Traffic Safety Administration (NHTSA)

Program toolkits, planners and information for a wide range of audiences concerned with the prevention of impaired driving.
www.stopimpairedriving.org

Traffic Safety Marketing

National Highway Traffic Safety Administration (NHTSA)

Planning information, materials and tips on conducting a high visibility enforcement campaign.
www.trafficsafetymarketing.gov

Tween Traffic Safety

Automotive Coalition for Traffic Safety

Focused on increasing belt use among children ages eight to twelve years, this guide describes developmental implications, current trends and strategies. Provides specific steps for how and what data to collect before, during and after the program and suggests development of a logic model. (2006)
www.tweensafety.org/cms/researchUploads/Tween%20Booklet.pdf

Detailed Evaluation Guides

These resources provide step-by-step background and methods for how to evaluate a program when a professional evaluator will not be used.

Demonstrating Your Program's Worth

Centers for Disease Control and Prevention

Describes stages of evaluation and provides a worksheet to determine which stage of evaluation is appropriate. Provides explanation of data collection methods and an introduction to research design. (2000)

www.cdc.gov/ncipc/pub-res/demonstr.htm

Evaluation Guide for Community Safe Routes to School Programs

National Highway Traffic Safety Administration

Provides step-by-step instructions to plan and conduct a program evaluation. Includes a worksheet to assist with the evaluation process. While intended for Safe Routes to School program coordinators, the information is easily transferable to highway safety programs. (2008)

www.saferoutesinfo.org/guide/evaluation/

User Friendly Handbook for Project Evaluation

National Science Foundation

Comprehensive explanation of each stage of evaluation, methods, analysis and reporting. It also highlights advantages and disadvantages of different data collection methods. Designed for National Science Foundation education managers, but focuses on the theory that is applicable to evaluation of any program. (2002)

www.nsf.gov/pubs/2002/nsf02057/start.htm

DOT HS 811 061
December 2008