Pedestrian and Bicyclist Traffic Control Device Evaluation Methods

PUBLICATION NO. FHWA-HRT-11-035 MAY 2011

U.S. Department of Transportation
Federal Highway Administration

Research, Development, and Technology
Turner-Fairbank Highway Research Center
6300 Georgetown Pike
McLean, VA 22101-2296

FOREWORD

The overall goal of the Federal Highway Administration (FHWA) Pedestrian and Bicycle Safety Research Program is to increase pedestrian and bicycle safety and mobility. From better crosswalks, sidewalks, and pedestrian technologies to growing educational and safety programs, FHWA's Pedestrian and Bicycle Safety Research Program strives to create a more walkable future.

This study was part of a larger FHWA research study to quantify the effectiveness of engineering countermeasures in improving safety and operations for pedestrians and bicyclists. The project focused on existing and new engineering countermeasures that have not yet been comprehensively evaluated. The project involved data collection and analysis to determine whether these countermeasures reduced fatalities and injuries or increased appropriate driving behaviors. The purpose of this report is to describe methods that practitioners can use to conduct reliable evaluations of pedestrian and bicyclist traffic control devices.

This report will be of interest to engineers, planners, and other practitioners who implement pedestrian and bicycle treatments and to city, State, and local authorities responsible for public safety.

Monique R. Evans
Director, Office of Safety
Research and Development

Notice

This document is disseminated under the sponsorship of the U.S. Department of Transportation in the interest of information exchange. The U.S. Government assumes no liability for the use of the information contained in this document. This report does not constitute a standard, specification, or regulation.

The U.S. Government does not endorse products or manufacturers. Trademarks or manufacturers' names appear in this report only because they are considered essential to the objective of the document.

Quality Assurance Statement

The Federal Highway Administration (FHWA) provides high-quality information to serve Government, industry, and the public in a manner that promotes public understanding. Standards and policies are used to ensure and maximize the quality, objectivity, utility, and integrity of its information. FHWA periodically reviews quality issues and adjusts its programs and processes to ensure continuous quality improvement.

TECHNICAL REPORT DOCUMENTATION PAGE

1. Report No. FHWA-HRT-11-035	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Pedestrian and Bicyclist Traffic Control Device Evaluation Methods		5. Report Date May 2011
		6. Performing Organization Code
7. Author(s) Susan T. Chrysler, Kay Fitzpatrick, Marcus A. Brewer, and Mike Cynecki		8. Performing Organization Report No.
9. Performing Organization Name and Address Texas Transportation Institute The Texas A&M University System College Station, TX 77843-3135		10. Work Unit No. (TRAIS)
		11. Contract or Grant No. DTH61-01-C-00049 Task Order #25
12. Sponsoring Agency Name and Address Office of Safety Research and Development Federal Highway Administration 6300 Georgetown Pike McLean, VA 22101-2296		13. Type of Report and Period Covered Technical Report: January 2009–November 2010
		14. Sponsoring Agency Code
15. Supplementary Notes The Contracting Officer's Technical Representative (COTR) was Ann Do, HRDS-30.		

16. Abstract
This report offers traffic engineering practitioners information on how to evaluate roadway traffic control devices used by pedestrians and bicyclists. Though presented in the context of devices meant for pedestrian and bicyclist facilities, the guidance provided can be applied in a more general sense to evaluations of traffic control devices in all settings. The evaluation methods report is designed for practitioners (State transportation departments and county or city engineers and planners) but could also be helpful to traffic safety students and researchers. Personnel without specialized statistical analysis skills should be able to use the report. It presents a detailed plan for practitioners to follow from the initial problem identification stages to documenting the evaluation effort. The first step of any evaluation is to clearly formulate the research question by identifying the motorist, pedestrian, or bicyclist behavior that poses a safety or operational problem. Candidate traffic control devices and other countermeasures can then be identified as potential solutions to that problem. The evaluation methods described in this report include user surveys or interviews, visibility studies, driving performance studies, observational traffic studies, and crash analyses. The selection of the appropriate evaluation method will consider cost, time, research aims, and available research equipment and staff.

17. Key Words Evaluation methods, Guidelines, MUTCD, Pedestrian treatment, Bicyclist treatment, Traffic control device		18. Distribution Statement No restrictions. This document is available to the public through NTIS: National Technical Information Service 5301 Shawnee Road Alexandria, VA 22312 http://www.ntis.gov	
19. Security Classif.(of this report) Unclassified	20. Security Classif.(of this page) Unclassified	21. No. of Pages 92	22. Price

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

SI* (MODERN METRIC) CONVERSION FACTORS

APPROXIMATE CONVERSIONS TO SI UNITS

Symbol	When You Know	Multiply By	To Find	Symbol
		LENGTH		
in	inches	25.4	millimeters	mm
ft	feet	0.305	meters	m
yd	yards	0.914	meters	m
mi	miles	1.61	kilometers	km
		AREA		
in^2	square inches	645.2	square millimeters	mm^2
ft^2	square feet	0.093	square meters	m^2
yd^2	square yard	0.836	square meters	m^2
ac	acres	0.405	hectares	ha
mi^2	square miles	2.59	square kilometers	km^2
		VOLUME		
fl oz	fluid ounces	29.57	milliliters	mL
gal	gallons	3.785	liters	L
ft^3	cubic feet	0.028	cubic meters	m^3
yd^3	cubic yards	0.765	cubic meters	m^3
		NOTE: volumes greater than 1000 L shall be shown in m^3		
		MASS		
oz	ounces	28.35	grams	g
lb	pounds	0.454	kilograms	kg
T	short tons (2000 lb)	0.907	megagrams (or "metric ton")	Mg (or "t")
		TEMPERATURE (exact degrees)		
^oF	Fahrenheit	5 (F-32)/9 or (F-32)/1.8	Celsius	^oC
		ILLUMINATION		
fc	foot-candles	10.76	lux	lx
fl	foot-Lamberts	3.426	candela/m^2	cd/m^2
		FORCE and PRESSURE or STRESS		
lbf	poundforce	4.45	newtons	N
lbf/in^2	poundforce per square inch	6.89	kilopascals	kPa

APPROXIMATE CONVERSIONS FROM SI UNITS

Symbol	When You Know	Multiply By	To Find	Symbol
		LENGTH		
mm	millimeters	0.039	inches	in
m	meters	3.28	feet	ft
m	meters	1.09	yards	yd
km	kilometers	0.621	miles	mi
		AREA		
mm^2	square millimeters	0.0016	square inches	in^2
m^2	square meters	10.764	square feet	ft^2
m^2	square meters	1.195	square yards	yd^2
ha	hectares	2.47	acres	ac
km^2	square kilometers	0.386	square miles	mi^2
		VOLUME		
mL	milliliters	0.034	fluid ounces	fl oz
L	liters	0.264	gallons	gal
m^3	cubic meters	35.314	cubic feet	ft^3
m^3	cubic meters	1.307	cubic yards	yd^3
		MASS		
g	grams	0.035	ounces	oz
kg	kilograms	2.202	pounds	lb
Mg (or "t")	megagrams (or "metric ton")	1.103	short tons (2000 lb)	T
		TEMPERATURE (exact degrees)		
^oC	Celsius	1.8C+32	Fahrenheit	^oF
		ILLUMINATION		
lx	lux	0.0929	foot-candles	fc
cd/m^2	candela/m^2	0.2919	foot-Lamberts	fl
		FORCE and PRESSURE or STRESS		
N	newtons	0.225	poundforce	lbf
kPa	kilopascals	0.145	poundforce per square inch	lbf/in^2

*SI is the symbol for the International System of Units. Appropriate rounding should be made to comply with Section 4 of ASTM E380.
(Revised March 2003)

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AASHTO	American Association of State Highway and Transportation Officials
EB	Empirical Bayes
FHWA	Federal Highway Administration
HAWK	High intensity Activated crossWalK
LIDAR	Light detection and ranging
MOE	Measure of effectiveness
MUTCD	*Manual on Uniform Traffic Control Devices*
ITE	Institute of Transportation Engineers
NCUTCD	National Committee on Uniform Traffic Control Devices
RRFB	Rectangular rapid-flashing beacon
TRDP	Technical report documentation page

EXECUTIVE SUMMARY

Walking and cycling are modes of transportation that can help improve air quality and quality of life. Unfortunately, they are also hazardous because pedestrians and bicyclists are vulnerable when their movements interact with vehicular traffic. Traffic control devices are a low-cost safety solution that inform, warn, and regulate all road users. Effective traffic control devices can promote walking and cycling by providing a feeling of security to users. As new traffic control device technologies and applications are introduced, cost-benefit evaluations are necessary, and these evaluations are often conducted by local and State agencies.

This report is intended to educate practicing engineers, planners, and public works employees at the local, county, and State levels in conducting or overseeing an evaluation of traffic control device effectiveness. The guidance provided, though presented in the context of devices meant for pedestrian and bicyclist facilities, can be applied in a more general sense to evaluations of traffic control devices in any setting.

The Federal Highway Administration (FHWA), through its *Manual on Uniform Traffic Control Devices* (MUTCD), requires evaluations of the effectiveness of new traffic control devices.(1) When determining whether these countermeasures are effective, most engineers and planners rely on anecdotal observations or their professional judgment. In some cases, a limited quantitative safety evaluation is conducted. However, these evaluations are typically limited in scope, experimental design, and statistical rigor because many State and local agencies lack research funds or the specialized knowledge of experimental design and statistics necessary to conduct reliable evaluations of new traffic control devices or other traffic features.

This report is organized as follows:

- Chapter 1 presents a brief overview of the evaluation process and discusses the use of surrogate measures of safety.
- Chapter 2 provides details on the process used by FHWA to make changes to MUTCD. It discusses the distinction between interpretation and experimentation as well as the process to request experimentation.
- Chapter 3 presents six steps to plan an evaluation of a new traffic control device.
- Chapter 4 presents information on how to conduct an evaluation as well as information on sample size and statistical analysis. It focuses primarily on measures of effectiveness (MOEs), such as speed and volume counts, which should be familiar to most traffic engineers.
- Chapter 5 describes how to properly document the evaluation effort in a research report.
- Chapter 6 lists additional resources for practitioners to use to conduct, analyze, and report on evaluations.

- Appendix A provides an example of the process for an actual pedestrian crossing treatment. The steps are as follows:
    - Planning step 1: Problem identification—What is the safety or traffic operations issue?
    - Planning step 2: Evaluation question—What is the research question?
    - Planning step 3: Measures of effectiveness—How will performance be assessed?
    - Planning step 4: Evaluation designs—What is the study approach?
    - Planning step 5: Evaluation methods—How are users, traffic, or crashes measured?
    - Planning step 6: Components of the evaluation plan—How can time, budget, and practicality be balanced to execute the plan?
- Appendix B presents more detailed information on statistical analysis.
- Appendix C presents additional MOEs that focus on human behavior.

CHAPTER 1. INTRODUCTION

OVERVIEW

State, county, and city transportation agencies implement a variety of innovative countermeasures to improve pedestrian and bicyclist safety (see figure 1). These countermeasures include the following:

- Traffic control devices.
- Geometric designs.
- Educational programs.
- Enforcement activities.

© Texas Transportation Institute

Figure 1. Photo. Bar pair crosswalk intended to improve pedestrian and bicyclist safety.

When determining whether these countermeasures are effective, most engineers and planners rely on anecdotal observations or their professional judgment. In some cases, a limited quantitative safety evaluation is conducted. However, these evaluations are often limited in terms of scope, experimental design, and/or statistical rigor because many State and local agencies lack research funds or sufficient knowledge of experimental design and statistics to conduct proper evaluations of new traffic control devices or other traffic features.

If better safety evaluation methods for pedestrian and bicyclist countermeasures were available to State and local agencies, local evaluations could provide sound safety effectiveness data for decisionmaking on national standards like MUTCD or American Association of State Highway and Transportation Officials (AASHTO) guides and programs.

One problem with evaluating bicyclist and pedestrian safety treatments is the difficulty of using direct safety outcomes such as pedestrian and bicyclist crashes. These types of crashes are relatively infrequent and therefore occur in small quantities. Consequently, any pedestrian or bicyclist safety evaluation that uses a crash-based analysis would require tens or hundreds, if not thousands, of study sites to obtain statistical significance. This is often impractical or impossible in typical experiments involving innovative devices. Additionally, local agencies rarely have the resources necessary to install and evaluate innovative traffic control devices at more than a few locations. As a result, surrogate MOEs are often the only feasible approach to evaluate the effectiveness of a device. A significant focus of this report is to provide an approach for establishing when a particular surrogate measure is appropriate.

OBJECTIVE

The purpose of this report is to assist practitioners in conducting sound evaluations on the effectiveness of pedestrian and bicyclist traffic control devices. The report is designed for practitioners (State transportation departments, county or city engineers, and other traffic officials). Additionally, personnel without specialized statistical analysis skills should be able to use the report.

The report includes (but is not limited to) the following information:

- When to conduct an evaluation.
- Why evaluations are important and who will use them.
- What practitioners should know before requesting permission to conduct an evaluation of an innovative traffic control device.
- When and how to write a request for experimentation.
- Examples of evaluation reports.
- MOEs (including crash-based measures, user understanding, compliance and behavior, and other surrogate measures).
- Evaluation design (before-after with/without comparison sites, cross sectional, etc.).
- Data collection (exposure, behavior, compliance, traffic data, crashes, etc.).
- Common statistical analysis techniques to determine the likelihood that the measured results are a true reflection of the device’s effectiveness rather than based on chance.
- Common evaluation errors, such as regression to the mean, and how to avoid them.

This report provides basic information on each of these elements so that a sound evaluation can be conducted. A reliable evaluation, regardless of whether the results are positive or negative, will allow other practitioners to learn from and build on the experience of the implementation of a new traffic control device, placement, or application. Sound evaluations can also lead to the dismissal or elimination of a traffic control device either because it does not provide positive driver or pedestrian results or because it does not offer an improvement over existing traffic control options. One such example is the use of fluorescent yellow-green crosswalk markings. Evaluation of these markings showed that they offered no advantage over standard white crosswalk markings. Therefore, they were not adopted for further use.

FHWA and the National Committee on Uniform Traffic Control Devices (NCUTCD) rely on evaluations to determine if a new device or treatment should be adopted in MUTCD or if it should be considered for further evaluation. Although an evaluation conducted by one agency may not be sufficient to determine if a device should be adopted, the results may be combined with other evaluations completed by other agencies. Inadequate, incomplete, or improper evaluations will lead to poor decisionmaking by agencies and other practitioners.

KEY COMPONENTS

Key components or documents include the following:

- The 2009 MUTCD is the legal standard that must be used by road managers nationwide in order to install and maintain traffic control devices on all public streets, highways, bikeways, and private roads open to public traffic (see figure 2).[1]

© Texas Transportation Institute

Figure 2. Photo. The 2009 MUTCD.

- NCUTCD is an organization whose purpose is to assist with the development of standards, guidelines, and warrants for traffic control devices and practices used to regulate, warn, and guide traffic on streets and highways. NCUTCD makes recommendations to FHWA and other agencies for revisions to MUTCD and other national standards. NCUTCD develops public and professional awareness of the principles of safe traffic control devices and practices and provides a forum for qualified individuals with diverse backgrounds and viewpoints to exchange information.[2]

- FHWA specializes in highway transportation and performs research in many areas including transportation operations, highway materials, construction methods, and safety.

The FHWA MUTCD team administers MUTCD, including the experimentation process, and is available to assist with technical questions or comments about MUTCD. Contact information for the team is available at http://mutcd.fhwa.dot.gov/team.htm.

CHAPTER 2. WHEN TO PERFORM AN EVALUATION

Evaluations are required for any new traffic control device and for applications of devices that do not comply with the provisions of the 2009 MUTCD.[1] However, an agency may perform an evaluation in other instances as well. The 2009 MUTCD contains alternative versions of some devices as well as option and guidance statements that require engineering judgment in order to be applied. An agency may also want to perform an evaluation among several MUTCD-compliant alternatives in order to select an application in its jurisdiction. While this report focuses on evaluating new devices or applications, the same principles of research design and execution would apply to evaluations of existing devices.

MUTCD REQUIREMENTS

MUTCD Section 1A.10 requires any agency seeking to use a nonstandard traffic control device or a nonstandard application or placement of a traffic control device to obtain permission from FHWA to evaluate that device.[1] A measure of engineering judgment is permitted in MUTCD, allowing some flexibility for traffic engineers (see MUTCD Section 1A.09), but this does not extend to the use of nonstandard devices without requesting experimentation or interpretation.[1]

At times, it may be difficult to determine if a variant of a traffic control device is consistent with MUTCD, such as a nonstandard warning sign design for a unique situation not explicitly included in MUTCD (see figure 3). Minor sign design variants, applications, or placement variations that are not contained in MUTCD do not need to be tested as long as they do not conflict with a standard provision in MUTCD. Any variation from a standard normally requires permission from FHWA for testing along with an evaluation plan. If an agency is uncertain whether testing is necessary, it should contact FHWA for an interpretation of MUTCD or to determine if an evaluation is needed. Communication submitted electronically receives more prompt attention than mail-in communication (see the resources section of this report for contact information).

Figure 3. Illustration. Example of signs included in MUTCD.

If an agency has a question about a new application for a standard device, the official meaning of a standard device, or allowed variations of a standard device, a request for an official interpretation can be made to FHWA, which keeps a public database of all requests for (see chapter 6 of this report).

MUTCD Section 1A.10 states that the following information should be contained in a request for interpretation:[1]

- A concise statement of the interpretation being sought.
- A description of the condition that provoked the need for an interpretation.
- Any illustration to help clarify the request.
- Any supporting research data that are pertinent to the item to be interpreted.

If an agency wants to use a nonstandard traffic control device or make use of a nonstandard traffic control device application or placement that goes beyond a simple interpretation, it must submit a request for experimentation for that device and then conduct a formal evaluation under permission granted by FHWA. Devices that have been evaluated by another agency and rejected by FHWA should not be tested unless there is a new variation or application that warrants further evaluation. If an agency can show that a prior evaluation of a rejected traffic control device was flawed or incomplete, FHWA may consider further testing. If a new traffic control device or application is undergoing testing by another agency and the testing is incomplete, a second agency can seek permission to conduct further testing and submit an evaluation plan.

Examples of bicyclist and pedestrian traffic control devices that were not in conformance with the 2003 MUTCD and that underwent such experimentation include the following:[3]

- The shared lane pavement marking for bicyclists (also known as a sharrow), which was adopted in the 2009 MUTCD (see figure 4).[1]
- The pedestrian hybrid beacon (also known as a High intensity Activated crossWalK (HAWK) beacon), which was adopted in the 2009 MUTCD (see figure 5).[1]
- The rectangular rapid-flashing beacon (RRFB) for uncontrolled crosswalks, which was not in conformance with the 2003 MUTCD but has undergone further testing and received interim approval (see figure 6).[3]

Figure 4. Photo. Example of a shared lane marking.

Figure 5. Photo. Example of a HAWK.

Figure 6. Photo. Example of an RRFB.

Only a public agency (or private toll authority responsible for the operation of a road) can request permission to experiment with a new traffic control device or application. The agency can partner with a manufacturer, vendor, consultant, or research agency to test a device and conduct the evaluation. In fact, for a public agency lacking the expertise in statistical analysis techniques, making a partnership with another entity is highly desirable. The use of a consultant, research organization, or university skilled in this area might greatly enhance the ability to conduct an accurate and meaningful evaluation. The process for requesting and conducting experimentation for new traffic control devices or new applications of traffic control devices is shown in figure 7. It is important to note that if a new device or application will undergo only laboratory or closed-track testing and will not be placed on a road open to public travel, there is no requirement for FHWA experimentation approval.

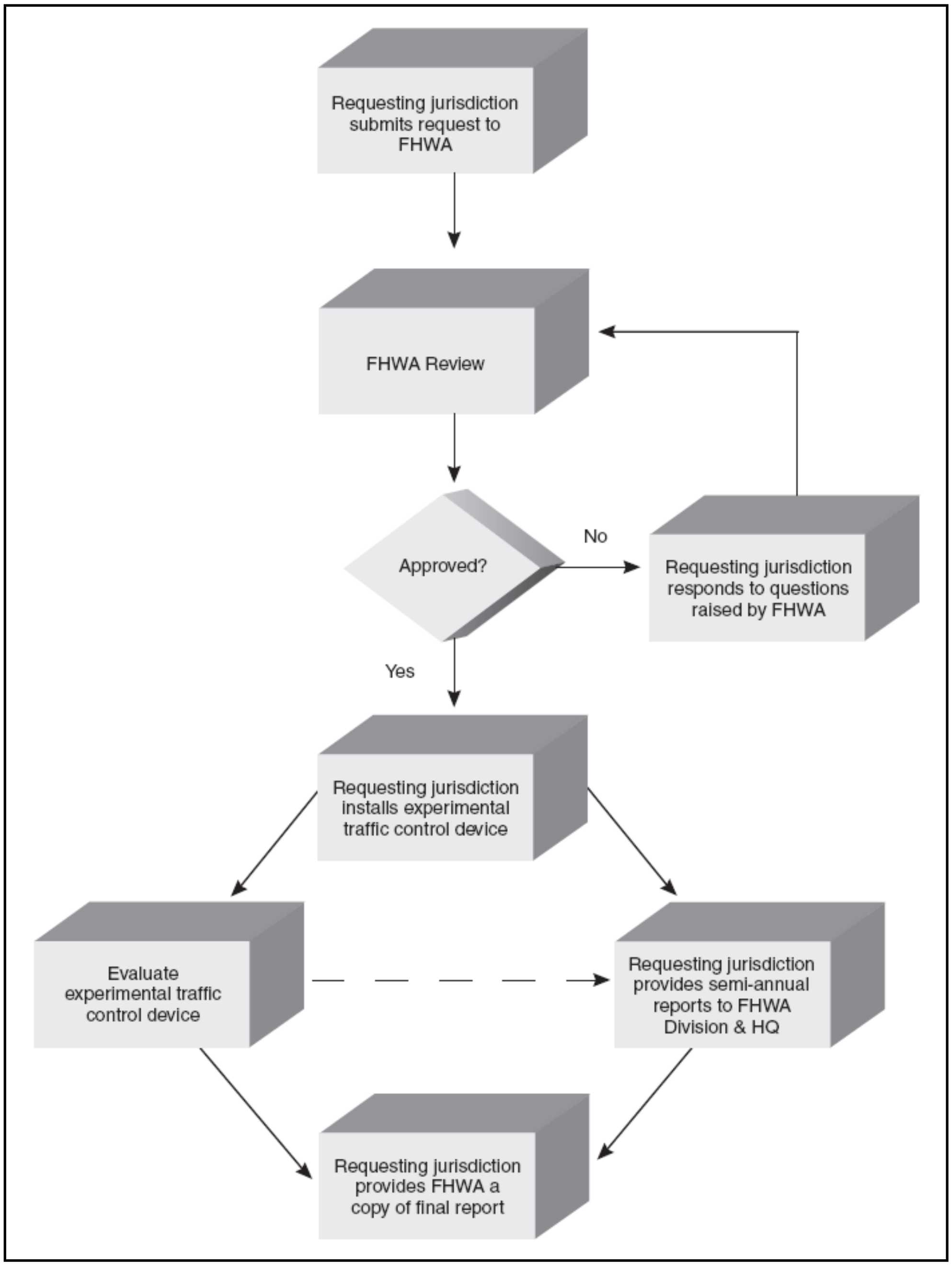


Figure 7. Illustration. Process for requesting and conducting experimentations for new traffic control devices.[(1)]

REQUESTS FOR PERMISSION TO EXPERIMENT

Requests to FHWA for permission to experiment with a new traffic control device or application should "include consideration of field deployment for the purpose of testing or evaluating a new traffic control device, its application or manner of use, or a provision not specifically described" in MUTCD (Section 1A.10, Support statement paragraph 08).[(1)] Chapter 6 includes information on how to access examples of requests to experiment, letters requesting interpretations, and evaluations.

MUTCD Section 1A.10 states that requests should contain the following information:[1]

A. A statement indicating the nature of the problem.

B. A description of the proposed change to the traffic control device or application of the traffic control device, how it was developed, the manner in which it deviates from the standard, and how it is expected to be an improvement over existing standards.

C. Any illustration that would be helpful to understand the traffic control device or use of the traffic control device.

D. Any supporting data explaining how the traffic control device was developed, if it has been tried, in what ways it was found to be adequate or inadequate, and how this choice of device or application was derived.

E. A legally binding statement certifying that the concept of the traffic control device is not protected by a patent or copyright. (An example of a traffic control device concept would be countdown pedestrian signals in general. Ordinarily an entire general concept would not be patented or copyrighted, but if it were it would not be acceptable for experimentation unless the patent or copyright owner signs a waiver of rights acceptable to the FHWA. An example of a patented or copyrighted specific device within the general concept of countdown pedestrian signals would be a manufacturer's design for its specific brand of countdown signal, including the design details of the housing or electronics that are unique to that manufacturer's product. As long as the general concept is not patented or copyrighted, it is acceptable for experimentation to incorporate the use of one or more patented devices of one or several manufacturers.)

F. The time period and location(s) of the experiment.

G. A detailed research or evaluation plan that must provide for close monitoring of the experimentation, especially in the early stages of its field implementation. The evaluation plan should include before and after studies as well as quantitative data describing the performance of the experimental device.

H. An agreement to restore the site of the experiment to a condition that complies with
the provisions of this Manual within 3 months following the end of the time period of
the experiment. This agreement must also provide that the agency sponsoring the experimentation will terminate the experimentation at any time that it determines significant safety concerns are directly or indirectly attributable to the

experimentation. The FHWA's Office of Transportation Operations has the right to terminate approval of the experimentation at any time if there is an indication of safety concerns. If, as a result of the experimentation, a request is made that this Manual be changed to include the device or application being experimented with, the device or application will be permitted to remain in place until an official rulemaking action has occurred.

I. An agreement to provide semi-annual progress reports for the duration of the experimentation, and an agreement to provide a copy of the final results of the experimentation to the FHWA's Office of Transportation Operations within 3 months following completion of the experimentation. The FHWA's Office of Transportation Operations has the right to terminate approval of the experimentation if reports are not provided in accordance with this schedule.

TRACKING THE REQUEST

If FHWA requests that the testing agency terminates the experiment at any time during the experimental period, the experimental device or application must be removed, and the location(s) must be returned to a condition that is in compliance with the 2009 MUTCD.[(1)] This can be costly and can create a public relations problem or potential liability concern. Agencies must be aware that FHWA can terminate an experiment or eventually adopt the new traffic control device in a manner that is inconsistent with the proposed experimental application during the final rulemaking process. This can cause adverse political fallout or embarrassment to an agency that has not considered this possibility and has not informed agency management and political leaders.

An experiment cannot proceed unless the FHWA Director of the Office of Transportation Operations grants the requesting agency approval for the experiment. This means the installation of a new traffic control device or new application or placement that is not consistent with the current MUTCD cannot occur without FHWA approval. FHWA may request additional information or clarification or request a modification in the traffic control device or experimental design before granting or denying a request for experimentation.

INTERIM APPROVALS

FHWA may grant an interim approval for the use of a device. Interim approvals are granted by an official memorandum, which is posted on the MUTCD Web site. They are typically given between revision cycles of MUTCD, and the device is incorporated into the next revision of the manual.

Traffic control devices that have already been given interim approval by FHWA do not require a formal evaluation, but the FHWA Office of Transportation Operations must be contacted for permission to use the device under interim approval. While a formal evaluation is not required, it is always beneficial to other agencies and the traffic engineering community to conduct an evaluation to assess driver, pedestrian, or bicyclist performance, behavior, compliance, and safety with the new device. FHWA may modify the application or design of a new traffic control device under interim approval when the final rulemaking occurs.

CHAPTER 3. PLANNING THE EVALUATION

OVERVIEW

Planning an evaluation involves a series of steps and decisions. A successful evaluation that is thorough and acceptable requires careful consideration. The problem that the proposed traffic control device will resolve must be clearly defined. The evaluation design and method employed must be appropriate for the condition being addressed. Additionally, the evaluation should clearly consider what will be varied and what characteristics will be held constant in order to assess the treatment's influence on a MOE. The steps discussed in this chapter are included in table 1.

Table 1. Steps to plan an evaluation of traffic control devices.

Step	Name of Step	Question Answered
1	Problem identification	What is the safety or traffic operations issue?
2	Research question	What is the research question?
3	Measures of effectiveness	How will performance be assessed?
4	Evaluation designs	What is the study approach?
5	Evaluation methods	How will users, traffic operations, or crashes be measured?
6	Selecting components to the evaluation plan	How can time, budget, and practicality be balanced to execute the plan?

PLANNING STEP 1: PROBLEM IDENTIFICATION

The evaluation process begins when a safety or traffic operations issue is observed. Because this report discusses the evaluation of traffic control devices, the process focuses on issues involving a traffic control device or where a traffic control device can be part of the solution. Certainly, some safety and operations problems cannot be solved by traffic control devices and may require a change in alignment, roadside treatments, or enforcement.

The safety or operations problem may have been identified by citizen complaint, police citations and safety concerns, a routine review of traffic operations, or a formal roadway safety audit. Once the issue is identified, candidate countermeasures can be examined by reading research reports or talking with vendors or other professionals.

The first step in planning an evaluation is to identify the characteristics of the problem that the countermeasure will address. The characteristics of the problem can be identified by the methods shown in table 2.

Table 2. Methods to identify problem characteristics.

Observations of Pedestrian or Bicyclist Movements	Observations of Vehicular Movements	Historical Records
• Law enforcement • Local residents or business owners • Surveillance cameras	• Traffic counters • Surveillance cameras • Traffic survey	• Crash records • Citation records • Citizen complaint logs

The characteristics of the problem should help classify the problem in one of the following areas:

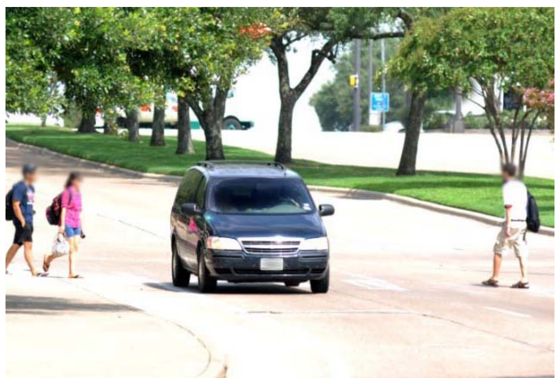
- **Visibility**: The road user did not see the traffic control device (see figure 8).
- **Comprehension**: The road user did not understand the traffic control device (see figure 9).
- **Compliance**: The road user ignored the traffic control device (see figure 10).
- **Opportunity**: The road user needs a change in a traffic pattern to facilitate crossings (see figure 11).

© Texas Transportation Institute

Figure 8. Photo. Driver's view of traffic control devices obstructed by utility poles and trees.

© Texas Transportation Institute

Figure 9. Photo. Driver confused by pavement markings remaining from a work zone.



© Texas Transportation Institute

Figure 10. Photo. Driver not yielding to pedestrians in a crosswalk.

© Texas Transportation Institute

Figure 11. Photo. Pedestrian cannot find acceptable gap due to high traffic volume.

Understanding the characteristics of the problem helps determine how to proceed. When available traffic engineering solutions from MUTCD are not sufficient and a new traffic control device is developed, experimentation may be necessary. Within this process, the next step is to formulate the research question.

PLANNING STEP 2: RESEARCH QUESTION

Once the problem and potential countermeasure(s) has been identified, a research or evaluation question (also called a hypothesis) should be formulated. The hypothesis should be devised so that it can be tested in a study. Generally, it follows the thinking, "If an independent variable changes, will there also be a change in a certain dependent variable?" An independent variable is one that the experimenter manipulates—its value is independent of any response from the people or vehicles being tested. A dependent variable is the component that is being measured—its value is dependent on responses from the people or vehicles being tested. It may be helpful to think of the independent variable as the treatment being evaluated and the dependent variable as the MOE used. The null hypothesis in scientific reasoning is the assumption that the independent variable has no (or null) effect on the dependent variable. The null hypothesis would answer the research question in the negative: "There is no change when this independent variable changes."

Examples of research questions include the following:

- Will the new traffic control device increase the number of vehicles that yield to pedestrians?
- Is the new traffic control device easier for people to understand than an existing traffic control device that is available in MUTCD?

Chapter 6 of the Institute of Transportation Engineers (ITE) *Traffic Engineering Handbook* provides information on statistical analyses applicable to the traffic engineering profession.[4] It includes a discussion on hypothesis testing. Notable scientific reasoning and statistical analysis starts with a presumption that the treatment has no effect. The statistical tests evaluate whether this presumption is accurate given the data. In other words, one should enter an evaluation as a skeptic and let the data and statistical analysis determine whether a treatment is effective.

It is often easiest to think about research questions in terms of the alternative hypothesis, also called the research hypothesis. This is a simple statement of the expected effect of the treatment. Useful research hypotheses avoid qualitative terms such as "better" and instead use language that describes the direction of change of a MOE. Some examples of research hypotheses for bicyclist and pedestrian traffic control device evaluations are as follows:

- The number of vehicles yielding to pedestrians will be greater when there is a flashing beacon present compared to when there is no beacon present.
- The number of bicycle-motor vehicle conflicts within 100 ft of the intersection will decrease after the new pavement markings are present compared to the existing markings.
- The proportion of pedestrians crossing at the marked crosswalk to those crossing at unmarked locations will increase after the new pavement markings are present compared to the existing markings.
- The speed of approaching vehicles will be lower at locations that have the advance warning signs installed compared to locations without advance warning signs.

PLANNING STEP 3: MEASURES OF EFFECTIVENESS

A key step in developing an evaluation plan is to determine appropriate MOEs. A selected MOE should reflect the user behavior that is anticipated to change due to the traffic control device or that is somehow related to the problem being addressed. Thought must be given to the likely cause of the problem and how that cause relates to features of the potential traffic control device countermeasure.

MOEs are observable behaviors or events that can be influenced by a countermeasure or treatment. Characteristics of MOEs include the following:

- **Valid**: A MOE can measure what it is supposed to measure.
- **Sensitive**: A MOE can discriminate between performance changes in the before and after conditions.
- **Collectable**: A MOE can be measured with available resources.
- **Reliable**: A MOE can measure the same phenomenon each time it is used.

In many research reports, the MOE is the dependent variable. The dependent variable typically refers to responses made by people or motor vehicles (i.e., the values of the variable depend on a person's response). This is contrasted to the independent variable, which is independent of a person's response. The independent variable is the element changed by the experimenter during the evaluation—it is the treatment.

Examples of MOEs are listed in table 3. A traffic control device may be installed to improve operations or safety. Safety devices are intended to improve safety through the reduction of crashes. Because crashes occur relatively infrequently, they can be an unstable MOE. Also, some treatments are designed to only alter certain behaviors, so they will not influence all crash types and may result in a nonmeasurable change in total crashes.

Table 3. Examples of MOEs for field evaluations.

Crashes	Pedestrian Behaviors	Traffic Operations
• Number of crashes • Number of crashes by type • Crash rate	• Pedestrian compliance (with signals, crosswalk, etc.) • Inadequate looking • Hesitation or backup • Conflicts with through or turning vehicles • Pedestrian activation of device (percent)	• Volume (by road user group or by movement) • Speed (by road user group) • Deceleration pattern on approach • Turn counts • Delay (by road user group) • Percent stopping or yielding • Conflicts

Because pedestrian and bicyclist crashes are rare, a countermeasure may not have a notable effect on crashes within a reasonable study period. In such cases, a surrogate measure may be used instead of crashes. Surrogate measures include behaviors by bicyclists, pedestrians, or motor vehicles that are related to crash likelihood.

Appendix A of this report gives examples of the process used to select a MOE.

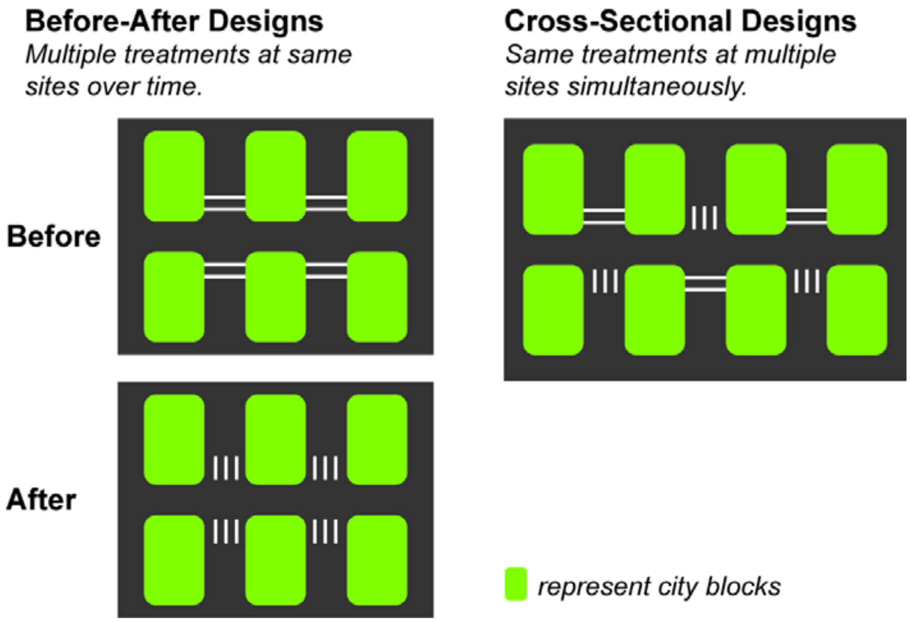
PLANNING STEP 4: EVALUATION DESIGNS

The effectiveness of a traffic control device can be measured in two basic ways: over time or across different locations. A longitudinal study measures the effect of a treatment over time. In traffic safety, these studies are referred to as before-after studies. A cross sectional study takes a portion of time and looks at treatment effects across different locations. Either of these two evaluation designs can be used when measuring traffic responses in observational studies or when measuring human responses in controlled experiments. Observational studies are different than controlled experiments in that the investigator may have limited or no control over the assignment of treatments to sites or the characteristics of the sites (e.g., traffic volume, development, etc.).

Table 4, figure 12, and figure 13 illustrate the analogy between the two basic evaluation approaches when applied to testing. The before-after design uses the same locations to test treatments over time, and the cross sectional design tests multiple treatments at matched sites during a single time period. Appendix B of this report gives examples of experimental design and the basics of statistical analysis.

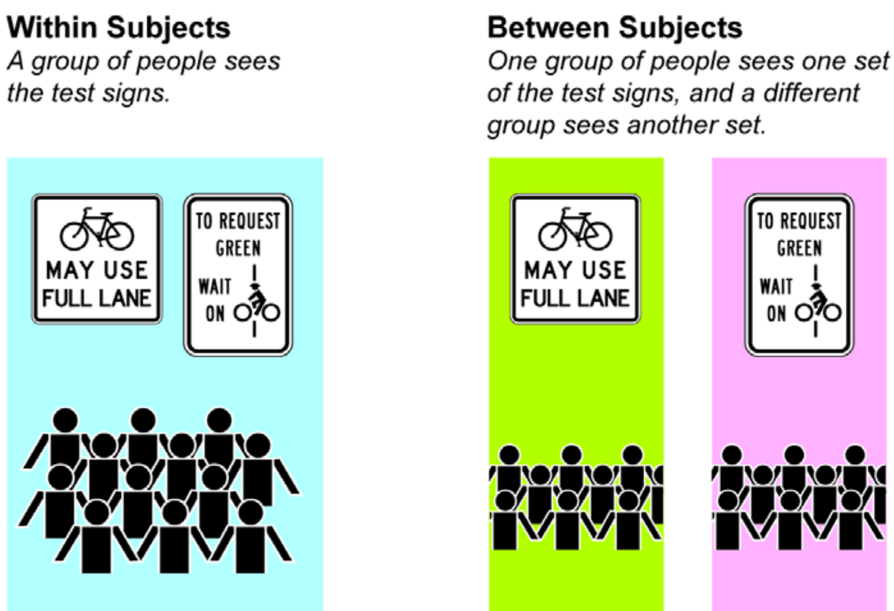
Table 4. Overview of evaluation designs.

Study Type	Longitudinal	Cross Sectional
Observational studies	A traffic control device is installed in the field at one or more sites. Road user reactions to the traffic control devices are compared before and after the device's installation.	A group of sites with a traffic control device is compared to a group of sites without the traffic control device.
Controlled experiments	All candidate treatments are shown to all participants over time.	A group of participants views one device, and another group views a different device.



© Texas Transportation Institute

Figure 12. Illustration. Observational study evaluation designs for crosswalk marking patterns.



© Texas Transportation Institute

Figure 13. Illustration. Controlled experiment designs within subjects and between subjects.

PLANNING STEP 5: EVALUATION METHODS

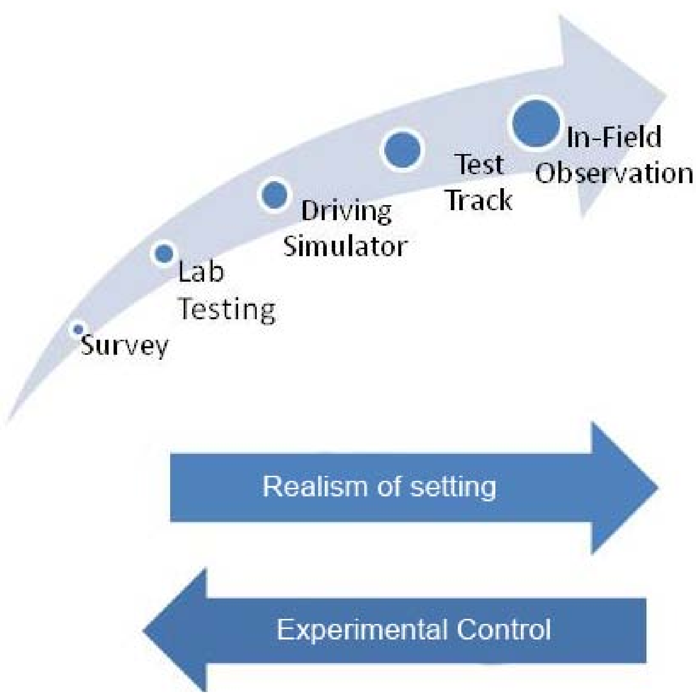
The exact methodology and MOEs employed to evaluate a user's reaction to a traffic control device is dependent on the characteristics of the proposed treatment and the analysis of the problem as described in the previous section.

At a high level, the effectiveness of a countermeasure can be evaluated by examining its effect on one or more of the following:

- Users (motorists, pedestrians, and bicyclists).
- Traffic operations.
- Crashes.

These three areas form the basic structure of this section of the report. Within each of these sections, details are provided that present the benefits and drawbacks to each method.

Methods focused on traffic operations and user behavior have advantages and disadvantages that generally fall along a continuum as shown in figure 14. As the realism of the setting of an evaluation method increases, the amount of experimental control possible typically decreases, and confounding variables begin to be introduced. Laboratory methods, such as comprehension tests, can have a high degree of experimental control but a lower degree of realism. A particular traffic control device may perform well in a laboratory comprehension test, but when placed in the field, it may show a low degree of driver compliance. Evaluation methods are discussed in more detail in appendix C of this report.



© Texas Transportation Institute

Figure 14. Illustration. Tradeoff between realism of the setting and experimental control.

When selecting a research method, the tradeoffs shown in figure 14 must be considered. For example, if an agency wants to evaluate a new symbol sign for advance signs for shared use trail crossings, it could choose any of the methods in figure 14. A survey asking people what they think the new symbol means could be conducted in a short amount of time. At the other end of the continuum, an agency could install a sign with the new symbol and observe yielding behavior at the crossing when bicycles are present. It may take several months to capture enough occurrences of motor vehicle-bicycle conflicts to have a valid statistical sample. Additionally, in the field study, factors such as weather and traffic conditions have to be considered when interpreting the results. All of these tradeoffs, along with cost and the need for specialized equipment or facilities, must be carefully weighed when selecting a study method.

PLANNING STEP 6: SELECTING COMPONENTS TO THE EVALUATION PLAN

The evaluation plan must directly relate to what is being tested. The previous sections presented different study designs and methods. The selection of the evaluation method must weigh cost considerations against which approach will best answer the research question formulated.

Several choices must be made when developing an evaluation plan. The choices include the MOEs, study design (e.g., before-after versus cross sectional), and study method (examining users versus traffic operations versus crashes). Within the study method decision, specific details of the data collection instruments and procedures must be made, most of which require consideration of how one choice affects other features of the overall evaluation plan.

Budget and time affect all evaluation plan developments. Tradeoffs between sample size, MOEs, and budget must be considered. Because these tradeoffs exist, care must be taken when selecting the appropriate methodology to answer the research question at hand. Elements to consider while selecting the appropriate method include the following:

- Crash characteristics common to the problem.
- Driver, pedestrian, or bicyclist behaviors related to crashes, conflicts, or disruptions to traffic operations.
- Availability of necessary measuring equipment and measurement locations.
- Relationship of a MOE to the problem behavior identified.
- Use of a carefully planned succession of evaluations to control costs.

Table 5 relates a sample of study methods to factors contributing to an observed problem. The table illustrates the applicable study method for the given contributing factor. For example, if the contributing factor is the visibility of the device, a survey would not answer whether the proposed traffic control device would improve conditions. After refining the traffic control device using laboratory methods, a field test could address the question of how the traffic control device would perform in the real world. Using laboratory methods allows for greater control over confounding factors and allows researchers to refine the traffic control device prior to field installation.

Table 5. Evaluation methods applied to specific contributing factors.

Evaluation Method	Visibility	Comprehension	Compliance	Opportunity
Examining Users				
Interactive structured interviews		X		
Focus groups		X		
Non-interactive surveys		X		
Controlled experiments		X		
Traditional laboratory		X		
Driving simulator		X		X
Test track	X	X		X
Roads closed to public travel	X	X		
Roads open to public travel	X	X	X	X
Examining Traffic Operations				
Volume				X
Road user behavior (e.g., selected paths)			X	
Erratic maneuvers or conflicts			X	
Compliance, violations, and risk taking			X	
Citations			X	
Speed			X	
Delay or travel time				X
Gap				X
Other potential/unique measures				X
Examining Crashes				
Crash history	X	X	X	X

CHAPTER 4. CONDUCTING THE EVALUATION

DATA COLLECTION PLAN

A detailed data collection plan highlighting data needs, procedures, and schedules is critical to a successful study.

The selected MOEs determine the types of data needed. Chapter 3 of this report provides an overview of different methods for evaluating a treatment. Having established a MOE of interest, the data collection plan should describe the exact type of data needed and the locations where the data should be collected (see figure 15). For example, if the traffic control device is to improve pedestrian compliance, then the action of each pedestrian on each approach leg treated needs to be considered. If the evaluation is to also consider whether the pedestrian is crossing within a crosswalk or prior to the crosswalk, then the data collection plan also needs to indicate whether the pedestrian waited for the appropriate signal and where the pedestrian crossed (e.g., within the crosswalk, 25 ft prior to the crosswalk, 50 ft prior to the crosswalk, etc.).

© Texas Transportation Institute

Figure 15. Photo. Plans specifying locations where data will be collected.

Developing a standardized worksheet for the data collectors will help provide consistency in the data collection and data reduction process. The data collection plan should include the proposed worksheets.

For some evaluations, videotaping the area of interest helps reduce the data after the events (see figure 16). If the data collection is to include videotaping of the site, the data collection plan needs to clearly indicate the field of view needed for each camera.

© Texas Transportation Institute

Figure 16. Photo. Videotaping an area of interest.

To ensure usability, especially if more than one group will be participating in data collection, the draft data collection plan for videotaping should be provided to someone who is not familiar with the study for testing. If agency policy prohibits videotaping, live scoring of surveillance cameras without archival storage of the video can be used instead of replaying an episode to rescore it.[5]

The *Guide to Good Statistical Practice in the Transportation Field* provides the following advice regarding data collection:[6]

- The design of the data collection plan is one of the most critical phases in developing a data system. The accuracy of the data and the estimates derived from the data are heavily dependent on the data collection design.
- Accuracy is dependent on proper sample design, making use of sampling complexity to minimize variance. The data collection process itself will also determine the accuracy and completeness of the raw data.
- Data collection from 100 percent of the target group is usually the most accurate approach but is not always feasible due to cost, time, and other resource restrictions. It is also often more accurate than the data requirement demand and can be a waste of resources.
- A probability sample is an efficient way to automatically select a data source representative of the target group with the accuracy determined by the size of the sample.
- For most statistical situations, it is usually important to be able to estimate the variance along with the mean or total.
- Sample design should be based on established sampling theory, making use of multistaging, stratification, and clustering to enhance efficiency and accuracy.

SAMPLE SIZE

To satisfy statistical requirements, it is necessary that a sufficient sample size be obtained. The suggested minimum for some typical measurements is shown in table 6.[7] However, a qualified statistician should be consulted to determine the appropriate sample size for the study. Many research institutes, universities, and statistical consulting firms offer statistical consulting services. These professionals should be consulted before the study is conducted to ensure that the appropriate experimental control and sample size are present for a valid statistical analysis.

Table 6. Suggested minimum sample sizes.[7]

MOE	Minimum Sample Size
Speeds	100 motor vehicles
Pedestrian-motor vehicle conflicts	30 each type
Pedestrian survey	100 pedestrians
Pedestrian compliance	50 pedestrians

If data are collected in the same manner within different evaluations, it is possible to compare or merge results across regions. The ability to document traffic control devices resulting in improvements in more than one area provides added justification for including the device in future editions of MUTCD.

DATA COLLECTION PERIODS

The time and date of data collection should also be included in the data collection plan. For example, the following may be needed:

- Time of day (e.g., morning or afternoon, or peak versus nonpeak period).
- Day of the week (e.g., some trails and nearby intersections have more activity on Saturdays and Sundays than during the week).
- Time of the year (e.g., when school is in session).

For evaluations using a before-after design, data must be collected before and after the installation of the traffic control device. As a result, data collection must be planned before the countermeasure is installed. The data collection plan should indicate when the before and after data will be collected. For many pedestrian-related traffic control devices, the number of pedestrians at a site may change because of the introduction of a device that favors pedestrians. For those situations, it is critical that before data, including the number of pedestrians, are collected.

The periods of data collection should be the same for the before and after periods. The before and after data should be collected on the same day of the week, time of day, and season to avoid potential confounding from other variables (e.g., volume differences between day and night or between peak and nonpeak traffic). It is also important to allow a sufficient acclimation period after the installation of the countermeasure to allow for any novelty effects. An acclimation period of 1–2 months is usually sufficient for most countermeasures. In some cases, it may be necessary to observe a site immediately for any initial signs of road user confusion. This initial confusion may not be a good indicator of the long-term effects of the treatment.

STATISTICAL ANALYSES

Testing for statistical significance quantitatively determines the likelihood that an observed change was caused by the installation of the countermeasure rather than by chance. Note that testing for statistical significance is only part of the process. It does not demonstrate the practical importance of the difference (see the following section).

In addition, the cost of the device needs to be considered. For example, a new flashing beacon may have a statistically significant effect on traffic speed or on yielding; however, the cost of the beacon versus the benefits of the speed reduction or yielding must still be considered. Appendix B of this report includes additional information concerning statistical analysis.

PRACTICAL DIFFERENCE ANALYSES

In many situations, the difference between two sets of measures may be statistically significant but not of practical difference. The value of the practical difference is dependent on what is measured and how it is measured (technique and equipment). It could also be dependent on the sponsor (e.g., how to value a change with respect to quality of life or feelings of safety). At a minimum, the practical difference needs to be greater than the measuring error of the device used. For example, when speed is measured with tube counters, the accuracy is about ±4 or 5 percent. Because light detection and ranging (LIDAR) equipment is used in enforcement, accuracy is more critical. For LIDAR guns, the guideline is that measurements are accurate within 1 mi/h. Therefore, a difference of at least 1 mi/h (if LIDAR was used) or 4 mi/h (if a tube counter was used) is needed before a speed change should be termed practically different.

EVALUATING THE RESULTS

Cost-Benefit Analysis

A cost-benefit analysis is useful in determining if a particular treatment is cost effective. It is important to be able to justify the cost of a treatment, whether based on a reduction in delay, a reduction in crashes, improvements in services, or another MOE. The dollar amounts used to place a value on reductions in delay or crashes have changed over time. Table 7 lists 2001 crash costs included in the 2010 *Highway Safety Manual* taken from a 2005 FHWA report.[8,9]

Table 7. 2001 crash cost.[8]

Crash Code	Crash Severity	Human Capital Cost	Comprehensive Social Cost
K	Fatal	1,245,600	4,008,900
A	Incapacitating injury	111,400	216,000
B	Nonincapacitating injury	41,900	79,000
C	Possible injury	28,400	44,900
PDO	Property damage only	6,400	7,400

Another measure of benefit is delay savings. One method used to calculate delay savings, along with the methodology for raising the dollar values to reflect the current year's costs, is found in the Texas Transportation Institute's *2009 Annual Urban Mobility Report*.[10] Delay savings calculations require simulation or field study data. Appendix C of this report contains more information about calculating delay savings.

Regardless of the MOE, the perceived benefits of a countermeasure are estimated or forecasted and then compared to the cost of installing and maintaining the treatment. Typically, the expected benefit (savings) is divided by the cost and expressed as a ratio. If the result is better than 1:1, the treatment is deemed to be cost effective.

As with crash evaluations, it is important to consider the MOE when conducting a cost-benefit analysis. For example, the benefit realized by preventing one fatal or serious injury crash could be enough to justify paying for the treatment based on the simple ratio of benefit to cost. Therefore, care must be taken to not overemphasize the effect of one crash at a low-crash location. If a crash surrogate MOE is used, it is not practical to use a cost-benefit analysis.

Life Cycle Costs

Related to the discussion of cost-benefit analysis is the concept of evaluating life cycle costs of a treatment. Using this method, the entire cost of installing, operating, and maintaining a treatment is evaluated over its entire expected service life, which can be compared to the anticipated reduction in crashes, delay, or other cost that would be incurred if no changes were made. When using this method, it is critical that the practitioner has an accurate estimate of the service life of the treatment being considered and a reliable method of obtaining credible estimates of the benefits to be gained and the costs of maintenance over that period of time.

CHAPTER 5. SUGGESTED SECTIONS FOR AN EVALUATION REPORT

INTRODUCTION

A report that records the details of an evaluation needs to describe the treatment, the issues that led to the evaluation (or the problem or question that the evaluation addressed), the evaluation procedure, and the conclusions and recommendations that resulted from the evaluation. This chapter contains an overview of suggested sections for an evaluation document. These sections include the following:

- Front material.
- An introductory section or chapter.
- Background information on the study, problem statement, or issue to be resolved.
- A description of the countermeasure or treatment evaluated.
- A description of the evaluation process.
- A description of the findings and conclusions.
- A list of relevant recommendations, comments, and discussions.
- Any necessary appendices.

The following sections address components of each of these sections. Not all of the components of these sections may be required for every report. The descriptions can be used as guides on how to appropriately describe the tasks and efforts associated with an evaluation. The practitioner can use the components that are relevant, omit those that are unnecessary, and include others that may be needed based on the requirements of the evaluation, the sponsoring agency, and the nature of the project.

FRONT MATERIAL

The elements at the beginning of a document can be collectively called the front material. The necessary items form a function of the funding source for the evaluation. The order of the elements may also vary depending on the requirements of the sponsoring or performing agency. However, each is commonly found in a technical report and helps introduce the reader to the report. Some of these elements are described in further detail in the following sections. Front material includes items such as the following:

- Front cover (outside and inside, as needed).
- Foreword (optional).

- Notices, disclaimers, or other official statements (as required by the sponsoring agency and/or the agency conducting the evaluation).

- Technical report documentation page (TRDP) (required for federally funded projects).

- Abstract or executive summary with conclusions (an abstract is a part of the TRDP).

- Common unit conversion factors (required for projects funded by certain agencies).

- List of commonly used abbreviations and symbols (required for projects funded by certain agencies; may be tailored to include certain abbreviations specific to the project and the report).

- Acknowledgments (optional).

- Table of contents.

- List of tables.

- List of figures.

Cover

How the report is being published, whether as a hardback, softback, or Web-only document, will help determine if outside and inside covers are necessary. The outside cover is generally more elaborate, using cover art or photos along with textual information (see figure 17). The inside cover page has no to few graphics and more text. In general, a cover has several required elements, including the following:

- Report title.

- List of authors.

- Contact information.

- Date of report.

- Report number or reference to experimentation number or project title.

- Sponsoring agency.

© Texas Transportation Institute

Figure 17. Photo. Example of a report cover.

TRDP

The TRDP has a predetermined format in which the authors provide the required information. This page is required in reports on projects that are federally funded. Refer to the front of this document for an example of a completed TRDP.

Abstract or Executive Summary

An abstract or executive summary provides readers with a concise account of the problem, treatment, and evaluation as well as a summary of the conclusions. In many cases, the abstract is 250 words or less, but there is not always a defined limit. Executive summaries tend to be longer, sometimes a full page or more. The TRDP includes an abstract.

Table of Contents

The front material should contain a table of contents to show the document organization and to provide a reference for readers looking for specific information within the report.

List of Tables and List of Figures

A list of tables and a list of figures are supplemental to the table of contents and help readers find key illustrations of findings and conclusions.

INTRODUCTORY SECTION OR CHAPTER

The beginning of the body of the report should contain an introductory section or chapter. This introduction typically includes the following elements:

- A basic description of the observed problem that necessitated the evaluation.
- A list of the objectives of the evaluation (based on MOEs, discussion of alternatives, or other metrics to determine if the evaluation was successful in accomplishing its purpose).
- A description of the organization of the document (commonly a list of the chapters and appendices).

- Detailed background information related to the study (depending on the nature of the project and the length of the report).

BACKGROUND

If the nature of the project, the treatment, and the issue addressed in the evaluation can be summarized briefly, then a background section could be included in the introductory chapter. However, for a detailed literature review, summary of the current state of practice, or other expanded account of related information, it may be necessary to put the background material in its own chapter. Common elements of the background section include the following:

- Purpose of the study/problem statement.
- Description of the sponsor of the project.
- Sponsor's experimentation number.
- Location where the study was conducted.
- Project objectives.
- Literature review (with references) or references to similar evaluations/reports.
- Comparison of the study to previous evaluations.
- Definitions of terms and acronyms (if needed).
- Identification of critical actions and appropriate MOEs (particularly with respect to pedestrian/bicycle treatments).
- Discussion of previous evaluations of similar devices (if available).

Possible questions that could be asked in developing the background include the following:

- How does the treatment relate to MUTCD?
- Why was the tested device chosen?
- Why will existing methods, measures, and devices not work?
- How will the device improve operations, delay, safety, or cost (or another MOE)?

COUNTERMEASURE

After the need for the project has been established and the treatment(s) or countermeasure(s) has been introduced in the background section, a more detailed description is necessary to acquaint readers with its various features and attributes according to the following outline:

- Description of the device.
    - Dimensions.
    - Type of retroreflective sheeting (if applicable).
    - Timing of flashing if the device includes active flashing elements.
    - Version number of any operating software.
- Pictures (color is preferred over black and white), diagrams, and layouts (see figure 18).
- Costs.
    - Capital costs.
    - Installation costs.
    - Maintenance costs.
    - Life-cycle costs.
- Practicality of maintenance.
- MOEs.
- Expected benefit.
- Target user group (e.g., blind, deaf, elderly, bicyclists, etc.).

© Texas Transportation Institute

Figure 18. Photo. Example of test devices in the field.

EVALUATION PROCESS

The procedures and methods used in the evaluation must be clearly described in detail for two reasons: (1) to properly document all of the personnel, equipment, and tasks used to conduct the evaluation and (2) to allow others to reproduce the results when installing the treatments at other locations.

Elements that can be used to describe the evaluation process include the following:

- Problem identification (clearly state why the new traffic control device is needed).
- MOEs used in the study.
    - Measures must make sense with respect to the problem being addressed and the proposed traffic control device countermeasure.
    - Action of the user of interest (during the time period of interest (e.g., night versus day)).
    - Characteristics of subjects being observed.
        - Age.
        - Disability.
        - Clothing for pedestrians (especially at night).
- Evaluation design used (e.g., before-after, cross sectional, within subjects, or between subjects).
- Evaluation method used (e.g., field, lab, survey, etc.).
    - Survey
        - Demographic description/ethnic factors.
        - Sample size.
        - Participant recruitment method.
    - Treatment and comparison site characteristics.
        - Season of the year when data were collected for all study periods.
        - Operations (24-h volume and operations speed).
        - Geometry of sites.
        - Posted speeds versus actual speeds.

        - Road users (frequency and the percent who are tourists or unfamiliar drivers).
        - Environment (rural or urban).
        - Distance to street for shared use trails.
        - Population characteristics.
        - Lighted versus not lighted.
        - Zoning (rural versus urban).
        - Laws applicable to a specific location (e.g., yield to pedestrians or stop for pedestrians).
        - Composition of users (e.g., high number of tourists or elderly users).
    - Crashes.
        - Number of years of data.
        - How crashes were identified.
- Findings.
    - Statistical analysis.
    - Practical differences.
    - Cost-benefit.
- Discussions.
    - Confounding factors.
    - Conflicting measures.
    - Comprehension concerns.
    - Limitations of study.
- Recommendations.

FINDINGS AND RESULTS

After describing the evaluation process and analysis methodology, the findings and results must be clearly stated. Some examples of commonly used methods for presenting findings include the following:

- A table that summarizes the entire study. Table 1 in the executive summary of *Safety Effects of Marked versus Unmarked Crosswalks at Uncontrolled Locations: Executive Summary and Recommended Guidelines* is an example of condensing a study's findings into one table.[11]

- A worksheet that presents a method for using the findings. Appendix A of the Transit Cooperative Research Program study on pedestrian treatments contains a worksheet that can be used to select a pedestrian treatment for a site.[12]

- A table that lists all countermeasures and shows the evaluator's opinion on the effectiveness of the countermeasures (e.g., check boxes under a range that goes from high to low).

- A discussion of unique qualities of the study location(s) and whether the results are transferable to other locations.

- A graph that illustrates pedestrian delay for each countermeasure evaluated.

- A photograph of pedestrian using treatment device (or not using it, depending on the effectiveness of the device).

There are several principles that practitioners should apply when presenting results as follows:

- Use the appropriate level of precision (e.g., report a 1.2 percent increase instead of a 1.207 percent increase).

- Include text to accompany any tables, figures, and photographs to describe the information contained in these graphical representations.

- Emphasize the reporting of results that are relevant to the MOEs defined in the evaluation process. Other results may also be reported, but they should not overshadow those that relate to the MOEs and should help answer questions posed in the project's problem statement.

- Report results objectively; do not allow a bias toward or against a particular treatment affect the interpretation of results or how they are reported to the reader.

- Discuss how results are similar to and different from other implementations (both within the United States and abroad).

CONCLUSIONS AND RECOMMENDATIONS

Based on the findings and results, the evaluator makes conclusions and recommendations for further action, revisions to existing policies or guidance documents, additional installations of the treatment, or discontinuation of the treatment. This may be included in the report as a part of the section on findings and results, or it may be its own section or chapter in the report. Some items to consider when developing conclusions and recommendations are as follows:

- Conclusions.
    - Describe what the evaluators determined that the findings and results mean.
    - Provide realistic and common-sense conclusions.
    - Discuss personnel and costs.
        - What quantities of personnel and costs were used to do the study?
        - How would someone else replicate the study, and what do they need?
    - Use clear language and confirm that conclusions do not contradict one another (do not confuse the conclusions).
- Recommendations.
    - Applicable types of sites: Where would the treatment work?
    - Any changes to the evaluation method that would improve the results of the treatment.
    - Additional research needed to improve or refine the conclusions.

COMMENTS AND DISCUSSION

Following the listing of relevant conclusions and recommendations, it is often appropriate to document other comments or topics of discussion that arose during the study. Some items for comments and discussion include the following:

- Areas for additional research including activities that would advance the knowledge of the profession.
- Aspects of the study that did not go as planned.
- Limitations of the study (e.g., where the results are applicable, transferable, etc).
- Applicability of results to other areas or, conversely, limitations of the device in given areas.

APPENDIX

In some reports, it is appropriate to include one or more appendices to provide additional information that benefits the readers but does not lend itself to inclusion in the main body of the report. Appendices often include large, multipage items that the report authors can refer to for a detailed description, leaving the most pertinent information in the main body. Some examples of items included in appendices are as follows:

- Reproduction of the survey questionnaire and listing of answers.
- Items that would make the body of the report inordinately long. One possible report format is “short report with a big appendix.” This entails moving multiple tables into the appendices and only including key information in the main report.
- Draft language of revisions to guidelines, documents, or policies when appropriate.
- A more detailed tabulation and presentation of data.

CHAPTER 6. SOURCES FOR ADDITIONAL INFORMATION

CONTACTING FHWA

The FHWA MUTCD team is available to assist with technical questions or comments about MUTCD in general or about a particular section or figure in MUTCD. Contact information for members of the team is available at http://mutcd.fhwa.dot.gov/team.htm.

An agency can contact FHWA at MUTCDofficialrequest@dot.gov for an interpretation of MUTCD or to determine if an evaluation is needed. Communications will receive quicker attention if they are submitted electronically.

The FHWA Office of Safety maintains a helpful Pedestrian and Bicycle Safety Web site that includes safety facts, information on roadway safety audits, research reports, and other resources: http://safety.fhwa.dot.gov/ped_bike/.

EXAMPLES OF EXPERIMENT OR INTERPRETATION REQUESTS

Examples of requests to experiment and letters requesting interpretation are available from the the American Traffic Safety Services Association at http://www.atssa.com/cs/root/news_pr/2009_interpretation_letters.

This site offers the experimentation and interpretation letters for 2009 and 2010. Earlier letters dating back to 2004 are also available on the Web site (click the archives button).

MUTCD RESOURCES

The MUTCD Web site includes a Resource page. It provides information on several topics including the following:

- Official rulings.
- Interim approvals.
- Official MUTCD interpretations issued by FHWA.

The main page is available at http://mutcd.fhwa.dot.gov/res-resources.htm. FHWA MUTCD official rulings older than 2 years are at http://mutcd.fhwa.dot.gov/official_rul.htm.

EXAMPLES OF EVALUATION REPORTS

The following lists several examples of evaluation reports:

1. Fitzpatrick, K. and Park, E.S. (2010). *Safety Effectiveness of the Hawk Pedestrian Crossing Treatment*, Report No. FHWA-HRT-10-042, Federal Highway Administration, Washington, DC. Obtained from: http://www.fhwa.dot.gov/publications/research/safety/10042/10042.pdf.

2. Fitzpatrick, K. and Park, E.S. (2010). TechBrief: *Safety Effectiveness of the HAWK Pedestrian Crossing Treatment*, Report No. FHWA-HRT-10-045, Federal Highway Administration, Washington, DC. Obtained from: http://www.fhwa.dot.gov/publications/research/safety/10045/10045.pdf.

3. Shurbutt, J. and Van Houten, R. (2010). *Effects of Yellow Rectangular Rapid-Flashing Beacons on Yielding at Multilane Uncontrolled Crosswalks*, Report No. FHWA-HRT-10-043, Federal Highway Administration, Washington, DC. Obtained from: http://www.fhwa.dot.gov/publications/research/safety/pedbike/10043/10043.pdf.

4. Shurbutt, J. and Van Houten, R. (2010). TechBrief: *Effects of Yellow Rectangular Rapid-Flashing Beacons on Yielding at Multilane Uncontrolled Crosswalks*, Report No. FHWA-HRT-10-046, Federal Highway Administration, Washington, DC. Obtained from: http://www.fhwa.dot.gov/publications/research/safety/pedbike/10046/10046.pdf.

5. Hunter, W.W., Thomas, L., Srinivassan, R., and Martell, C.A. (2010). *Evaluation of Shared Lane Markings*, Report No. FHWA-HRT-10-041, Federal Highway Administration, Washington, DC.

6. Hunter, W.W., Thomas, L., Srinivassan, R., and Martell, C.A. (2010). TechBrief: *Evaluation of Shared Lane Markings*, Report No. FHWA-HRT-10-044, Federal Highway Administration, Washington, DC. Obtained from: http://fhwa.dot.gov/publications/research/safety/pedbike/10044/10044.pdf.

7. Fitzpatrick, K., Chrysler, S.T., Iragavarapu, V., and Park, E.S. (2010). *Crosswalk Marking Field Visibility Study*, Report No. FHWA-HRT-10-068, Federal Highway Administration, Washington, DC.

8. Fitzpatrick, K., Chrysler, S.T., Iragavarapu, V., and Park, E.S. (2010). TechBrief: *Crosswalk Marking Field Visibility Study*, Report No. FHWA-HRT-10-067, Federal Highway Administration, Washington, DC. Obtained from: http://fhwa.dot.gov/publications/research/safety/pedbike/10067/10067.pdf.

9. Zegeer, C.V., Steward, J.R., Huang, H.H., and Lagerwey, P.A. (2002). *Safety Effects of Marked Versus Unmarked Crosswalks at Uncontroleld Locations: Executive Summary and Recommended Guidelines*, Report No. FHWA-RD-01-075, Federal Highway Administration, Washington, DC. Obtained from: http://drusilla.hsrc.unc.edu/cms/downloads/Effects_Un_MarkedCrosswalks_Summary.pdf.

10. Clark, K.L., Hummer, J.E., and Dutt, N. (1996). “Field Evaluation of Fluorescent Strong Yellow-Green Pedestrian Warning Signs,” *Transportation Research Record 1538*, Transportation Research Board, Washington, DC. Obtained from: http://trb.metapress.com/content/46528322163j3731/fulltext.pdf.

11. Carlson, P.J. and Miles, J.D. (2006). *Driver Understanding of Red Retroflective Raised Pavement Markers,* Report No. FHWA-CFL/TD-06-008, Federal Highway Administration, Washington, DC. Obtained from: http://tcd.tamu.edu/documents/Driver_Understanding_of_Red_RPMs.pdf.

12. Singer, J.P. and Lerner, N.D. (2005). *Countdown Pedestrian Signals: A Comparison of Alternative Pedestrian Change Interval Displays*, Final Report, Federal Highway Administration, Washington, DC. Obtained from: http://www.atssa.com/galleries/default-file/Ped_Countdown_Report.pdf

13. Markowitz, F., Sciortino, S., Fleck, J.L., and Lee, B.M..(2006). “Pedestrian Countdown Signals: Experience with an Extensive Pilot Installation,” *ITE Journal*. Obtained from: http://www.ite.org/ .

RESOURCES FOR HUMAN-SUBJECT PROTECTION

Some of the research methods identified in appendix C of this report involve direct contact with roadway users through surveys or interviews. When people are the subject of research, certain regulations regarding their protection may apply. Below lists resources pertaining to human-subject protection. In general, observations of public behavior are not subject to these regulations.

1. U.S. Department of Health and Human Services. *Office for Human Research Protections (OHRP),* Office of Human Subject Protection, Code of Federal Regulations and Institutional Review Board Information, Washington, DC. Obtained from: http://www.hhs.gov/ohrp/.

2. U.S. Department of Health and Human Services. *Human Subjects Research*, Health Resources and Services Administration, Protecting Human Subjects Training, Washington, DC. Obtained from: http://www.hrsa.gov/humansubjects/.

RESOURCES FOR EVALUATION METHODS

Table 8 lists resources pertaining to evaluation methods. These materials may be available through a local university library.

Table 8. Resources for evaluation methods.

Bibliographical Information	Description
ITE. (2009). *Traffic Engineering Handbook, 6th Edition,* Institute of Transportation Engineers, Washington, DC. Obtained from: http://www.ite.org/emodules/scriptcontent/Orders/index.cfm.	Chapters of particular interest include chapter 6, "Probability and Statistics," and chapter 8, "Traffic Engineering Studies."
ITE. (2000). *Manual of Transportation Engineering Studies*, Institute of Transportation Engineers, Washington, DC. Obtained from: http://www.ite.org/emodules/scriptcontent/Orders/index.cfm.	The manual shows how to conduct several transportation engineering studies in the field; discusses experimental design, survey design, statistical analyses, data presentation techniques, and report writing concepts; provides guidelines for both oral and written presentation of study results; and includes useful forms for various transportation studies.
Knoblauch, R.L. and Crigler, K.L. (1987). *Model Pedestrian Safety Program: User's Guide Supplement*, Federal Highway Administration, Washington, DC.	This supplement to the *Model Pedestrian Safety Program: User's Guide* provides information on specific pedestrian safety countermeasures. The countermeasures are grouped into the three major areas of engineering countermeasures, education countermeasures, and enforcement countermeasures.
American Association of State Highway and Transportation Officials. (2010). *Highway Safety Manual*, AASHTO, Washington, DC. Obtained from: http://www.highwaysafetymanual.org/Pages/default.aspx.	This manual provides evidence-based crash modification factors for many countermeasures. It also provides guidance on conducting crash-based evaluations and cost/benefit calculations.
American Association of State Highway and Transportation Officials. *SafetyAnalyst*, AASHTO, Washington, DC. Obtained from: http://www.safetyanalyst.org.	This software tool assists practitioners with identifying safety improvement needs and identifying improvement projects with good cost effectiveness.
Trochim, W.M.K.. *Research Methods Knowledge Base*, Cornell University, Ithaca, NY. Obtained from: http://www.socialresearchmethods.net/kb/index.php.	This Web site provides resources for research design and analysis in the behavioral sciences.
Spiegelman, C.H., Park, E. S., and Rilett, L.R. (2010). *Transportation Statistics and Microsimulation*. Obtained from: http://www.crcpress.com.	This textbook is useful for researchers and scientists.

Hauer, E. (1997). *Observational Before-After Studies in Road Safety*, Pergamon Press, Oxford, England.(13)	This textbook is useful to researchers and scientists. However, the concepts are complex, and some of the recommended experimental designs (e.g., empirical Bayes (EB)) are difficult to use to evaluate pedestrian and bicyclist safety because of sample size limitations.

RESOURCES TO BECOME INVOLVED WITH THE MUTCD

NCUTCD assists in the development of standards, guidelines, and warrants for traffic control devices and practices used to regulate, warn, and guide traffic on streets and highways. NCUTCD recommends proposed revisions and interpretations of MUTCD and other accepted national standards to FHWA and other appropriate agencies. NCUTCD develops public and professional awareness of the principles of safe traffic control devices and practices and provides a forum for qualified individuals with diverse backgrounds and viewpoints to exchange professional information. Membership characteristics of NCUTCD are listed in this section. Practitioners interested in becoming active in NCUTCD can work through one of these organizations with their delegates to NCUTCD. Information about NCUTCD can be found at http://www.ncutcd.org.

Membership Characteristics of NCUTCD

Sponsoring Organizations

NCUTCD is supported by 21 sponsoring organizations, each of which appoints one or more members and associate members, who constitute the voting delegates to NCUTCD. Sponsoring organizations include the following:

- Advocates for Highway and Auto Safety.
- AASHTO.
- American Automobile Association.
- American Public Transportation Association.
- American Public Works Association.
- American Railway Engineering and Maintenance of Way Association.
- American Road and Transportation Builders Association.
- American Society of Civil Engineers.
- American Traffic Safety Services Association.
- Association of American Railroads.

- Association of Bicycle and Pedestrian Professionals.
- American Highway Users Alliance.
- Governors Highway Safety Association.
- Human Factors Resources.
- ITE.
- International Association of Chiefs of Police.
- International Bridge, Tunnel and Turnpike Association.
- International Municipal Signal Association.
- League of American Bicyclists.
- National Association of County Engineers.
- National Safety Council.

Individual Participation

Over 200 individuals are directly involved in NCUTCD activities in the following categories of membership:

- Members are appointed by sponsors and serve as voting members of the council and as voting members of a technical committee. In total, 39 individuals serve as voting members of the council and of the technical committee. Only members are eligible to hold office.
- Associate members are appointed by the sponsors and serve as voting members of a technical committee. They can serve as the member's alternative on the council.
- Technical members are nominated by the technical committee chairperson and approved for membership on each technical committee by the executive board. They are eligible to vote in technical committee proceedings, to serve as secretary of a technical committee, and to be appointed as chair of a working group or task force.[2]

APPENDIX A. PROCESS FOR SELECTING A MEASURE OF EFFECTIVENESS

When planning an evaluation, the first two steps are to identify the problem and pose the research question. In the next step, the appropriate MOE is selected. This appendix provides an example of this process.

Suppose an agency has identified a site where drivers are not yielding to pedestrians at a crosswalk. The following are possible questions to determine why drivers are not yielding (see table 9):

- Are drivers not able to see the traffic control device? Are the markings faded or is sign clutter limiting the visibility of the sign? If so, a solution that addresses the visibility issue is needed, and the MOE needs to determine if the traffic control device is more visible.

- Do drivers not understand the message (e.g., are they supposed to yield to pedestrians in that community)? If so, the MOE should relate to the comprehension of the traffic control device.

- Are drivers ignoring the message because they do not want to stop and they know the likelihood of a traffic citation is slim? If so, increased enforcement may be a better treatment than a novel traffic control device.

- Are drivers not yielding because pedestrians are afraid to start their crossing because vehicular volumes are too high to provide adequate gaps? If so, the problem is one of opportunity. It might be solved by refuge islands or curb extensions to shorten the crosswalk length or by a pedestrian hybrid beacon that requires drivers to stop.

Table 9. Example of process of selecting MOEs: Vehicle not yielding to pedestrian at crosswalk.

Contributing Factor	Potential Countermeasure	MOE Suggestions
Visibility of crosswalk	New advance warning sign	• Detection distance • Yielding to pedestrians • Speed on approach (e.g., deceleration starts at a greater distance upstream)
	Sign in more locations (e.g., overhead or in street)	• Yielding to pedestrians
	Addition of light-emitting diodes to sign	• Detection distance • Yielding to pedestrians • Speed on approach
	Addition of light-emitting diode pavement markers	• Detection distance • Yielding to pedestrians • Speed on approach
Comprehension of traffic control device	New or revised pedestrian crossing sign	• Test people's understanding of meaning of device • Yielding to pedestrians
Visibility of pedestrian	Roadway lighting turns on when pedestrian activates system, system accompanied by signs	• Yielding to pedestrians at night
Lack of compliance	New sign and increased enforcement	• Test people's understanding of meaning of device • Yielding to pedestrians
Opportunity	New sign, beacon, or signal	• Crashes (pedestrian and/or vehicle) • Delay to pedestrians or roadway traffic. • Yielding to pedestrians • Pedestrian volume at site

APPENDIX B. EXPERIMENTAL DESIGN AND STATISTICAL ANALYSIS BASICS

EXPERIMENTAL DESIGN

As presented in Hauer's book on *Observational Before-After Studies in Road Safety*, one of the main sources of factual knowledge about the effect of highway and traffic engineering measures on safety is observational study.[13] The term "observational" is used to emphasize that the evaluation is not an experiment deliberately designed to answer a question that can be carefully controlled in a laboratory. Rather, in the transportation environment, the elements that cannot be controlled must be accounted for due to the dynamics of the location. The two basic types of observational evaluations are before-after and cross sectional.

Controlled experiments observe behavior under more controlled circumstances. In these types of evaluations, the participants know they are being studied, and the experimental designs carefully control extraneous factors. The basic types of controlled experiment designs discussed in this document are within subjects and between subjects.

Before-After Evaluations

In a before-after evaluation design, two measurements are taken, one before and one after the treatment is implemented. *Effectiveness* is defined as the difference in the two measurements over time. The before data must be collected prior to the installation of a countermeasure. Measurements are taken at all sites before installation of the device. After the device is installed at a site, identical measurements are again taken at all sites.

Depending on the type of treatment, a learning period may be necessary to provide road users with time to fully understand the appropriate behavior for the device. An evaluation may also need to collect data after a long period to ensure that the device is creating a long-term behavior change rather than a change due to the novelty effect of the device. While the amount of time needed for learning and to avoid novelty is debatable, previous evaluations have used 2 months for learning and collected data at 6-month intervals to ensure that the benefit of the device is long term.

While the before-after evaluation design is straightforward and easily applied, it has shortcomings. It is vulnerable to changes that occur during the time it takes to complete the evaluation (e.g., traffic volumes or composition). The effects of such variables must be considered in the evaluation.

The goal in a before-and-after evaluation is to have only one component that changes at the site over time—the application of the treatment itself. Therefore, all other conditions at the treated and comparison sites must be monitored or controlled. The following list includes examples of variables that need to be considered:

- Weather conditions.
- Illumination level.

- Traffic volume.
- Traffic mix.
- Calendar time, especially for school crossing treatments.
- Geometric changes (e.g., addition of curbs, etc.).
- Other traffic control devices (e.g., additional signs).
- User familiarity or unfamiliarity.
- Pedestrian age and gender.
- Pedestrian volume.

Table 10 shows evaluation plan considerations for an evaluation of the effectiveness of placing a flashing beacon on a pedestrian crossing advance warning sign.

Table 10. Example of before-after evaluation design considerations.

Design Consideration	Example
Evaluation design	Observational evaluation, before-after
Evaluation question	Will the addition of a flashing beacon improve vehicle yielding?
Research hypothesis	The number of vehicles yielding will be greater when there is a flashing beacon present compared to when there is no beacon present
Independent variable	Beacon presence two levels of this variable: beacon absent/beacon present
MOE	Percentage of vehicles yielding at the crosswalk when a pedestrian is present
Other independent variables controlled	Time of day is same between periods, day of the week is same between periods, no changes to roadway geometry
Other variables to be considered in evaluation	Traffic volume (motor vehicles, bicyclists, and pedestrians)

The evaluation design may also consider a comparison (or control) group. A comparison group consists of sites that are similar to the treated sites in characteristics, such as traffic conditions and roadway geometry, but do not have the treatment installed. The comparison group concept is strongly encouraged in before-after evaluations, especially for crash evaluations. In addition to crash evaluations, a comparison group is highly desirable with a before-after evaluation design when behavioral and operational measures are used. A comparison group is similar to a placebo control group in medical research during which the same measurements are taken before and after a placebo treatment.

For safety evaluations in which crashes are used as a MOE, simply comparing results before and after installation is not sufficient. This method requires advanced statistics; local agencies unfamiliar with the method should consider using an expert in this evaluation method. Use of a comparison group is mandatory with crash MOEs. A comparison group has sites that are similar in location, traffic conditions, and roadway geometry to the treated sites but do not have the treatment installed. The use of a comparison group improves the reliability of the results. The same data are collected at the comparison sites and the treated sites, and results are then compared not only between time periods but also among sites to determine the effectiveness of a particular treatment. Current practice is to use an EB method in safety evaluations. The EB method uses a derived crash prediction for the after period assuming the treatment had not been applied and compares this predicted value to the observed crash frequency for the after period with the treatment installed. This method is used in the *Highway Safety Manual* and the FHWA *SafetyAnalyst* approach (see chapter 6).[6,14]

Cross Sectional Evaluation

An observational cross sectional evaluation is a research design in which a site only has one treatment. In general, an observational cross sectional evaluation estimates the safety or operational effect of an element that is different between two groups of sites. The sites should be similar except for the difference of interest—the treatment. For example, an agency may identify 10 intersections that have similar traffic volume, roadway geometry, land use, and lighting but differ in the pattern of crosswalk markings used (see figure 12). Cross sectional studies are different from controlled experiments in that the investigator cannot control the assignment of treatments to sites.

Cross sectional evaluation design has been used to estimate the safety effect of differences between the treatments. Hauer notes in *Cause and Effect in Observational Cross-Sectional Studies on Road Safety* that

> … the question of whether causal interpretation of cross-sectional studies is at all possible is of central importance for road safety. The reason is that opportunities to do observational before-after studies about, say, the safety effect of change in horizontal curvature, road grade, lane width, median slope, etc. are few and imperfect. This is so, partly, because when a road is rebuilt usually several of its attributes are changed at once and it is difficult to assign the result to any single causal factor. In addition, the rebuilding of a road often changes it to such an extent that it may not be regarded the same unit after reconstruction. In contrast, opportunities for observational cross-sectional studies are plentiful.[15]

Hauer points out the dangers in interpreting the results of cross sectional and before-after designs.[15] Sites that are well matched for cross sectional designs are often hard to find. Likewise, in before-after studies, assuring that the only thing that changed at the site is the traffic control device under evaluation is often difficult.

Comparison of Observational Before-After and Cross Sectional Evaluation Designs

The before-after evaluation identifies the change resulting when a treatment has been applied to a group. The cross sectional evaluation compares the safety or operations of one group having some common features to the safety of a different group not having those features in order to assess the safety effect of that feature.

Hauer provides the following examples for these different types of studies:[13]

- **Observational before-after evaluation**: Circumstances where the entities that are changed by the treatment retain many of their original attributes. For example, replacement of a stop sign with a yield sign would leave the intersection geometry and setting unchanged. Additionally, the introduction of a seat belt law would not modify drivers' travel patterns, vehicle performance, or the road network.

- **Observational before-after with comparison group evaluation**: Circumstances that are similar to the above observational before-after evaluation with the addition of a comparison group used to provide corrections for changes in conditions over time, such as weather, vehicle mix, driver behavior, crash reporting practices, a citywide public relation campaign, etc.

- **Observational cross sectional evaluation**: Circumstances when the treatment substantially alters the entity. For example, a rural two-lane road is to be rebuilt into a four-lane divided road with a substantially modified alignment.

Table 11 provides definitions for different evaluation design types.

Table 11. Overview of evaluation designs.

	Time		
Evaluation Design	**Before Treatment Is Introduced**	**Introduce Treatment**	**After Treatment Is Introduced**
Before-after	Evaluation sites	Treatment 1	Evaluation sites
Before-after with comparison	Evaluation sites	Treatment 1	Evaluation sites
	Comparison sites	None	Comparison sites
Cross sectional (controls)	None	Treatment 1	Evaluation sites
	None	None	Comparison sites
Cross sectional	None	Treatment 1	Evaluation sites subgroup A
	None	Treatment 2	Evaluation sites subgroup B

Hauer provides information on how to conduct and interpret observational before-after studies.[13] He notes that the use of the comparison group method is deceptively similar to that of randomized experiments, which are popular in agriculture, medicine, and other fields of research. However, he notes that there is a crucial distinction. In a randomized experiment, the decision as to which entities get treated and which are left as control is made at random. Therefore, were the experiment repeated a very large number of times, each time with a random assignment of entities to treatment and control, the influence of the causal factors on both groups of entities would tend to be equal. As a result, when the assignment to treatment is made at random, it is legitimate to speak of a statistical experiment that involves a control group. In contrast, when the assignments of entities to a treatment group are not made at random, even if both groups of entities are very large, they may differ systematically with respect to some causal factors. Therefore, even with large groups of entities, there is no assurance that the expected number of crashes in the treatment group (had treatment not been administered) would have changed in the same manner as in the comparison group. For this reason, when entities are not assigned to treatment at random, the terms "experiment" or a "control group" should not be used. To mark the distinction, it is prudent to use the terms "observational studies" (not experiments) and of "comparison groups" (not control groups).

Because most evaluations of a proposed traffic control device involve the introduction of a new device, the before-after evaluation rather than the cross sectional evaluation is the typical approach. If the evaluation uses crashes or occurs over several months, the preferred approach would also use a comparison group.

Within Subjects or Between Subjects

While the before-after (with or without comparison group) and cross sectional evaluation designs generally apply to field installation evaluations, there are other experimental design considerations when performing controlled experiments such as surveys, laboratory, and test track studies in which candidate countermeasures are shown to drivers or pedestrians. The main experimental design feature to consider is whether all of the different candidate treatments will be shown to all the participants (within subjects) or whether some subgroup of participants will see some subgroup of treatments (between subjects). There are statistical power advantages to a within subjects design, which are beyond the scope of this report. By having each person see each treatment, direct comparisons of the treatments can be made within each individual as well as across individuals. A within subjects design for studies of human behavior is comparable to a before-after evaluation where the presence of a treatment is varied within a single site.

There are also practical advantages to a within subjects treatment. Consider an evaluation that is assessing the visibility benefits of installing a flashing beacon on top of an advance warning sign for a pedestrian crossing. Research participants will stand one block away and rate how easy it is to see the sign on a scale of 1 to 5. Treatment A (the standard sign) is installed on the northbound approach, and treatment B (the sign plus the flashing beacon) is installed on the southbound approach to the crosswalk. A total of 40 people volunteer to participate and complete the ratings. They are randomly assigned to group 1 or group 2. A within subjects design would have all 40 people look at both treatments. In contrast, for a between subjects design, group 1 would look at treatment A and group 2 would look at treatment B. If treatment B receives higher ratings, the researcher could not be sure if it was because treatment B is better or because the people in group 2 just happened to be people who tend to give high ratings or have good eyesight.

Problems with administering within subjects designs include treatment order effects and learning. In the example above, if everyone sees treatment A before treatment B, bias could be introduced in the ratings because people naturally compare B to A. One experimental control that can be introduced is counterbalancing in which the order of presentation is balanced across the two groups of people. Half the people would see treatment A first, and half would see treatment B first. This way, any order effects are spread out across the two groups.

It is often not feasible to do a within subjects design, especially when evaluating treatments in the field. Additionally, performing a within subjects evaluation would require having the participants return at a later time after a new treatment has been installed. In these cases, a between subjects evaluation is acceptable, but a higher total number of subjects may be required to ensure adequate statistical power. A between subjects design for human subject research is comparable to a cross sectional design for traffic studies where one site (or group of subjects) gets treatment A, and a different site (or group of subjects) gets treatment B.

Table 12 shows a selection of evaluation plan considerations for a within subjects evaluation on the visibility of a flashing beacon on a pedestrian crossing advance warning sign.

Table 12. Example of within subjects evaluation design considerations.

Design Consideration	Example
Evaluation design	Controlled experiment, within subjects
Evaluation question	Will the addition of a flashing beacon improve the visibility of a pedestrian crossing sign?
Null hypothesis	The visibility ratings will not be different when there is a flashing beacon present compared to when there is a beacon present
Independent variable	Beacon presence; two levels of this variable: beacon absent/beacon present
MOE	Visibility rating on a scale of 1 to 5
Other independent variables controlled by experimenter	Time of day, illumination, with order counterbalanced
Other variables to be considered in evaluation	Equal split in gender and age groups of participants

Factors Affecting the Validity of Results

In the example provided above, the order in which subjects rated the treatments was used to illustrate a factor that can affect the validity of the results, which is an example of confounding. Extraneous factors that vary consistently with a treatment are confounded with the treatment. In the example, the order was confounded with the treatment; treatment A always came first. Confounding makes it difficult to interpret results. Researchers may question whether treatment A received low ratings because it was hard to see or simply because it was the first treatment the people saw so they were inclined to give midrange scores to anything that came first. Confounding is sometimes unavoidable, especially in field studies. With proper planning and consideration, confounding variables can often be controlled or eliminated.

Many factors can influence the results of an evaluation if not adequately considered during the evaluation design. The following is an overview of the factors that need to be accounted or controlled for in the evaluation design:

- Changes over time.
  - Consider any changes in volume between periods for both traffic and pedestrians.
  - Consider changes to the mix of users over time as land use changes (e.g., a new school opens) or use of a facility changes (e.g., lower pedestrian volumes in winter).
  - Identify and consider any other changes that may be occurring (e.g., changes in reporting thresholds for property damage only crashes over subsequent years).

- Presence of recording equipment/observers.
    - Will road users be able to see the equipment or observer? If so, how will it affect their behavior? (If recording equipment or observers can be seen, drivers or pedestrians may be on their best behavior.)
- Instrumentation/measurement procedures.
    - Is the measuring instrument calibrated?
    - Were the rating scales and scoring criteria used consistently?
- Selection of comparison groups.
    - Identify the important factors to match between treatment and comparison groups must relate both to the treatment and the MOEs.

STATISTICAL ANALYSIS

As discussed by Knoblauch and Crigler, appropriate statistical analyses are required to determine if any differences between the before and after data are due to the treatment or to chance.(7) In most cases, one of the following three types of data will be collected:

- **Continuous**: Data that have no distinct intervals between possible values are continuous. Examples include vehicle speed and lateral placement.
- **Dichotomous**: Data that are identified by only two categories (i.e., the occurrence or nonoccurrence of a behavior are dichotomous). Examples include pedestrian compliance and pedestrian-motor vehicle conflicts.
- **Counts of events**: Data are on the number of occurrences. Examples include crash counts and the number of vehicles yielding. Note that for EB before-after crash studies, additional measures are needed beyond simple counts of crashes.

The actual statistical analysis performed will depend on the type of data collected. Table 13, adapted from Knoblauch and Crigler, presents combinations of types of data, recommended statistical tests, and comments regarding the output or use of the tests.(7)

Table 13. Sample applications of statistical techniques.

Data Type	Parameter(s) of Interest	Recommended Tests/Procedures	Comments
Continuous	Two means	*t*-test for difference in means	Assumes data are normally distributed and samples are independent
Continuous	Two means	*z*-test for difference in means	Sample sizes of 30 or more are required
Continuous	Two variances	*F*-test for difference in variances	Assumes data are normally distributed and samples are independent
Continuous	More than two means	Analysis of variance for testing equality of more than two means	Assumes data are normally distributed, variances are equal, and samples are independent
Dichotomous	Two proportions	*z*-test for difference in proportions	Assumes the sample sizes are large enough
Categorical data (more than two categories)	More than two proportions	Chi-square test of the equality of more than two proportions or of the independence of two categorical variables	Used when comparing more than two proportions, e.g., a two-by-two or larger contingency table; particularly used for testing cross-tabulated questionnaire data
Count data	Regression coefficient	Poisson regression or negative binomial regression	Used for assessing crash reduction

APPENDIX C. EVALUATION METHODS

EXAMINING USERS

Roadway users include pedestrians, bicyclists, and motorists of all ages and experience. Traffic control devices and operations should serve all roadway users. The most comprehensive evaluations test the widest variety and greatest number of users.

Who to Test

Decisions about who to test need to be made in the context of the problems identified. For example, if the problem is that school children are not obeying a signal, then the school children should be tested, not adults. Members of the general public are typically recruited to participate in evaluation studies. Efforts should be made to recruit people with different reading abilities, education levels, and visual abilities. If pedestrian users are specifically being recruited, people with mobility or sensory disabilities should be included in the subject group.

Ethics

Professional ethics standards require protection of human research subjects. Academic institutions that conduct human-subject research typically must have their research approved by an Institutional Review Board. Even if an agency is not bound by Federal regulations, the following basic principles should be followed by any professional:

- Voluntary participation (not coerced).
- Safety of research participants.
- Respect for participants.
- Equal treatment of all participants.
- Confidentiality of data.
- Privacy of personal information.

Free training that covers the basic principles of protecting research participants is available via the Internet (see chapter 6 of this report). Practitioners should consult with their risk management or health and safety officers about any applicable local or State regulations and policies concerning recruitment of general public participants for research studies. This is particularly true when minors are involved. Studies that receive Federal funding may be subject to human-subject protection regulations established by the Health and Human Services Department (see chapter 6 of this report).

Recruitment Methods

If surveys or individual testing of road users is needed, research participants can be recruited from the general public in many ways. Regardless of the recruitment method, efforts should be made to recruit a variety of people of different ages, genders, educational backgrounds, and familiarity with the local transportation system. In keeping with the professional ethics standards concerning human subjects, participants should not be forced to participate through personal influence, peer pressures, or supervisory positions.

Participant recruiting method ideas for studies for which testing will be conducted by appointment or during a specific period at a specific location are as follows:

- Posted flyers in study areas (see figure 19). Local businesses, libraries, churches, etc., are often willing to post flyers. This is a particularly effective method when seeking individuals who are familiar with a specific location.

- Ads in local media. Most local-access cable television and radio accept public service announcements free of charge if a public agency is performing the evaluation.

- Internet advertising through Craigslist® or through city/county/State social media outlets on Facebook®, Twitter®, MySpace®, etc.

- Direct email recruiting through friends, colleagues, community service groups (e.g., Lions Club or American Legion), etc. Special care must be taken to avoid coercion and to ensure diversity. Since these types of groups often are "birds of a feather," special efforts may be needed to reach out to a broad range of people in terms of age, race, and education level. By using email or the Internet, the subject sample is immediately biased toward those with access to a computer. This may eliminate lower income or lower education-level participants.

- Direct contact with groups representing a specific targeted demographic feature. Parent-teacher organizations, Girl and Boy Scout troops, YMCA/YWCA, church youth groups, and community recreation centers are ways to reach children. Permission should always be granted by a parent or guardian before a child participates in a study. Groups representing people with disabilities or offering services to those people may also be willing to advertise a study. Senior activity centers, church groups, and retired military groups often have a large contact list of older road users who would likely participate in a study.

© Texas Transportation Institute

Figure 19. Photo. Example of flyers and cards used to recruit study participants.

Research participants can also be recruited and tested immediately in the same location. These types of tests are often called intercept surveys because participants are intercepted in the course of their normal day and asked to participate in the evaluation (see figure 20). These types of intercepts may happen at transit stops, driver's license bureaus, hotel lobbies, and on streets. This type of recruitment method typically works best if the survey or other testing is very short in duration. People are generally unwilling to participate if completion of the survey or other test will be time consuming.

© Texas Transportation Institute

Figure 20. Photo. Researcher and participant in an intercept survey.

Self-Selection Bias

Regardless of the recruitment method, there is always the concern that people who agree to participate in a survey are somehow different than those who refuse. Behavioral researchers use the term self-selection bias to describe this effect. For instance, if researchers are standing near a school asking passersby to participate in a survey about the crosswalk, people who think the crosswalk needs improvement may be more likely to agree to participate than those who think the crosswalk is good. Random selection of participants and random assignment of treatments to participants are the best ways to avoid this bias.

How and Where to Conduct the Test

The questions of how and where testing will occur is separate from the question of what to measure, which is addressed in the next section. This section describes the different ways and places to conduct testing, regardless of the test question content.

Interviews, surveys, and other types of testing that can be performed in an office-type setting are described in the following sections. The location of the testing is often a crucial factor in recruiting participants, particularly if they are not regular drivers. Often, an agency can contact local churches, senior centers, recreation centers, and public libraries to secure a meeting room for free. These types of settings offer a convenient, safe, and familiar location where people are willing to come even if they are a bit apprehensive about volunteering for a study. The location for office testing should be relatively quiet, have access to restrooms, and be compliant with the Americans with Disabilities Act.

Interactive Structured Interviews

Interviews involve give and take between the experimenter and the participant. An interview script should be prepared to assure consistency across successive interviews or across different interviewers. The advantage to interviews is that the interviewer can ask follow-up questions to assure that the participant understands the question or task correctly. Interviews can take place as follows:

- Telephone interviews can be conducted by trained interviewers working from a script to ensure consistency. Telephone interviews are generally quite expensive because of the low response rate. In addition, the sample of people available with land-line, directory-listed numbers may not be representative of the general population as more people are exclusively using mobile phones.

- In-person interviews are most typical. Again, interviewers must be trained to remain impartial and consistent during the interview so as not to skew the results. In-person interviews can be conducted in a number of locations as follows:

    - On the street with pedestrians near the treatment.

    - At other intercept locations, such as a driver’s license office or shopping mall.

    - In a laboratory or office setting.

    - In a motor vehicle near the treatment.

Regardless of the location or setting, interviews can be prearranged by recruiting participants through the methods previously discussed or can be intercept surveys for which experimenters stop people. Care should be taken to avoid bias in selection in intercept surveys, as the experimenters may tend to stop people like themselves in appearance, race, or gender. The concern is that these unknown personality differences would also influence responses to traffic control devices while driving.

Focus Groups

Focus groups are small discussion groups of 8 to 10 people that are convened to discuss a particular topic (see figure 21). They can be held in offices, hotel conference rooms, or community centers. In a focus group, a trained facilitator presents photos or other examples of candidate treatments and asks a series of structured, open-ended questions to the group for discussion. Focus groups can be particularly useful at the start of a study when candidate treatments are being designed or a large group of candidates needs to be reduced to a smaller set that will be deployed in an observational study.

© Texas Transportation Institute

Figure 21. Photo. Focus group discussing and providing feedback.

The results of focus groups are often difficult to summarize because the discussion can be quite unwieldy. An experienced facilitator knows techniques to keep the discussion on point and to avoid one vocal person from overtaking the discussion. One danger of focus groups is "group think" where the opinion of the group changes over time and throughout the discussion due to the reinforcing of each other's ideas. The opinion of a group may shift and, in the end, not be the same as if each individual was asked separately. Focus groups can be a cost-effective method to hear from ordinary people about their concerns and troubles with a particular pedestrian or bicycle area. They are best used in the problem identification phase of an evaluation.

Noninteractive Surveys

As opposed to interviews, surveys (as used in this report) are questionnaires that are filled out by the research participants without interaction with the experimenter. Again, there are multiple methods to deliver and receive surveys, and care needs to be taken to avoid biasing who responds to these surveys. Noninteractive surveys include the following:

- Online or Web-based surveys offer a convenient and fast way to reach many potential survey respondents (see figure 22). The main drawback is that there is no way of knowing who is responding, particularly if the survey is anonymous. Another downside to the Internet survey is that the exposure time of the stimuli cannot be controlled. Computer technology varies so widely and transmittal times are so inconsistent that the researcher cannot guarantee that everyone sees each question, device, or illustration for the same amount of time.

- Email surveys are another convenient way of circulating a survey. Concerns about subject self-selection are high with this method.

- Preaddressed mail-back postcards handed to drivers or pedestrians in the area of a treatment is one way to solicit responses to a treatment in a specific location. Care must be taken to avoid bias in who does or does not receive or accept a postcard. Mailed surveys are also plagued with low response rates, sometimes as low as 5 percent or less. Coupling the mailing with other city services, such as utility bills, can save time and money and possibly increase the response rate. As with online surveys, it is not possible to guarantee that the person completing the survey was the addressee.

- In-person completion of the survey, either on paper or on a computer at any of the locations described above, is a common method. In-person completion can be coupled with Web-based surveys to balance the inequity of the subject sample introduced by requiring a computer. For instance, a Web-based survey may be advertised to the community, and then several additional in-person sessions may be offered at libraries, homeless shelters, or senior centers specifically to reach those without computer access. Staff should be available to assist those who are not comfortable with computers or who may not have sufficient reading skills to complete the survey alone. In-person surveys offer the benefit of being able to confirm demographic information about the person through observation.

© Texas Transportation Institute

Figure 22. Photo. Research participant filling out a survey on paper.

Controlled Experiments

Self-administered paper- and computer-based surveys often permit unlimited viewing time of the test device, which may inflate comprehension scores. Some surveys, using computer or slideshow presentations, may limit the amount of time the stimulus is viewed. One study directly compared these methods to driving simulation and found that a limited viewing time of sign images shown in roadway scenes produced comprehension scores equivalent to those seen in a driving simulator.(16)

The use of a laboratory to provide controlled conditions for evaluating a particular treatment can be beneficial (see figure 23). Laboratories can offer an approximation of real-world conditions without the risk of installing an unproven device at a location of interest. The following include some examples of studies for which laboratory-based evaluations may be appropriate:

- Visibility or comprehension of signs, signals, and markings.
- Comprehension or reception of unique or novel traffic control devices.
- Nighttime conditions that may provide added risk if tested at actual study locations.

Figure 23. Photo. Participant evaluating signs in a laboratory.

There are a variety of laboratories, small and large, indoor and outdoor, but all are capable of controlling conditions to limit the effect of external influences to focus on the variable of interest and the treatment being considered. The following list includes some examples of laboratories, which are discussed further in this appendix:

- Driving simulators.
- Test tracks or other full-scale driving environments.
- Traditional laboratory rooms outfitted for evaluating a particular characteristic of a treatment.

It is important to remember that the location of the study does not necessarily dictate what is being measured. For instance, an evaluation may take place in a driving simulator, but if the evaluation only looks at preference ratings, the study is still qualitative (studying opinion), not quantitative (studying some objective measure of driving performance like speed).

Traditional Laboratory

A materials laboratory can be beneficial in evaluating certain characteristics of a treatment, such as retroreflectivity of a sign or durability of equipment under extreme weather conditions. These laboratories are typically outfitted with specialized equipment to conduct specific scientific tests, often to a set of testing standards developed by FHWA, American Society for Testing and Materials International, or the National Electrical Manufacturers Association. Other laboratories may be equipped with instruments and projection equipment to test human responses to various features of a treatment, such as the audibility of a chirping pedestrian signal or the visibility of a pedestrian signal head (see figure 24).

© Texas Transportation Institute

Figure 24. Photo. Laboratory photometric measurement alley used for human factors studies.

Driving Simulator

A driving simulator provides a safe and controlled environment to further explore comprehension and compliance in response to novel traffic control devices (see figure 25). In the driving simulator environment, it is possible to test multiple variations of the design and placement of a new device that may not be feasible in the field. In addition, a wider variety of roadway geometries and traffic conditions can be tested than are typically possible in a test track study or fiscally practical in a field evaluation. For instance, factors that limit sight distance can easily be introduced, and cross traffic density can be manipulated.

Figure 25. Photo. Driving simulator testing comprehension and compliance for traffic control devices.

Typically, a driving simulator offers a library of different roadway cross sections and interchanges. Using this library, simulator scenarios, or “worlds,” are created to represent a typical roadway design. The worlds can be constructed such that traffic is programmed to interact with the research participant, and events can be initiated by hidden location or time-dependent triggers. Generally, the protocol is limited to 75 min to reduce the likelihood of simulator sickness and to avoid any effects of fatigue or boredom.

The following are examples of how driver performance can be assessed:

- Distance from a stop bar when the final stop was executed.
- Distance from the intersection at which the throttle was released.
- Deceleration rate to the intersection.
- Gap acceptance.
- Verbal responses to questions concerning comprehension of the device.

A problem with driving simulator research is that the simulation facility is typically in a fixed location, often on a university campus. This limits the subject population to those who have the means to travel to the simulation lab, who are not intimidated by visiting a university, and who live in the same city as the simulation lab. In addition, many simulators induce motion sickness, particularly in older women. This may further limit the broadness of the subject sample. In general, due to the cost of using the equipment, the number of participants tested in a simulator study is much smaller than in a survey or field test.

Test Track

Also called a proving grounds or full-scale driving environment, a test track can provide a safe and controlled atmosphere to explore the comprehension and compliance of drivers, pedestrians, or bicyclists in response to novel traffic control devices (see figure 26). These facilities offer the ability to install treatments at full scale and test their effectiveness with real pedestrians, bicycles, and motor vehicles while minimizing the risk of crashes or other calamities that would be possible in a real-world scenario. Signs and beacons can be installed; crosswalks can be marked; and streets, curbs, and paths can be built or laid out to simulate the conditions at an actual site with minimal risk to the general public.

© Texas Transportation Institute

Figure 26. Photo. Test track near Pecos, TX.

Evaluations using a test track typically have a sizeable cost for using the facility and purchasing and installing the treatment to be studied, but the results can be close to approximating the results expected in a real-world scenario.

Roads Closed to Public Travel

Public roads closed to travel can offer a less expensive alternative to test tracks (see figure 27). Proper road closures and notification to neighboring residences and businesses must be in place prior to testing. Full-scale test treatments can be installed, and drivers or pedestrians can evaluate them without concerns of interference of other traffic.

© Texas Transportation Institute

Note: Photo for illustration only, proper MUTCD temporary road closure signs should be used.

Figure 27. Photo. Public road closed to travel.

Roads Open to Public Travel

Driving performance and opinion studies can also be carried out on public roads in traffic. Full-scale test treatments can be installed at multiple locations, and a route can be developed that allows viewing of all of the treatments. The research subjects can be the driver or the passenger in a vehicle driving a prescribed course or could be a pedestrian observing the treatments. This method requires safety precautions, and there are liability concerns in the event that the experimental vehicle or the pedestrian subject is involved in a crash. When conducting visibility studies on the open road, the vehicle must be equipped with some sort of distance-measuring instrument to mark where participants could see the treatments. Also, the headlights of the test vehicle should be properly cleaned and aimed prior to the study, and the same vehicle should be used for all nighttime runs to provide the same illumination to the device.

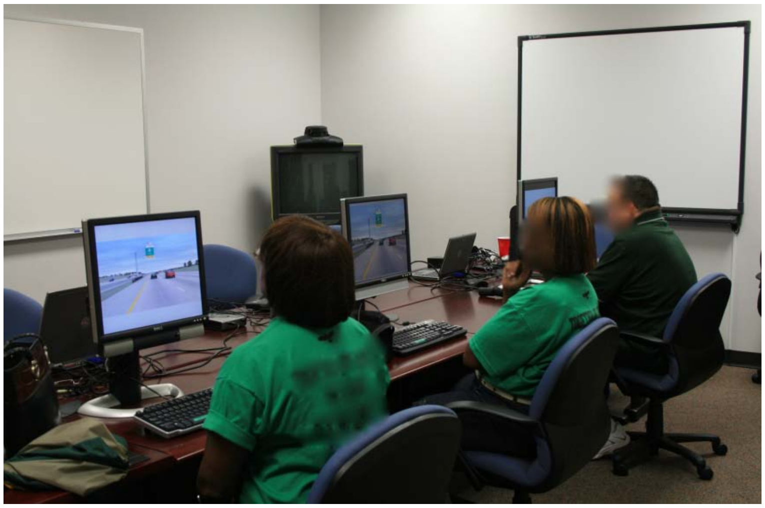
One downside to conducting testing on open roads is that the other traffic (both pedestrian and vehicular) is beyond the control of the experimenters. Other extraneous factors such as weather and ambient light must also be considered.

What to Measure

Measurement techniques can be categorized as qualitative or quantitative. In qualitative tests, the responses are not numerical, nor can they be ranked. They are simply comments or feedback on some quality of the test stimulus. Qualitative research through interviews, focus groups, and surveys can give very detailed information on a subject. However, the generalizability, or the ability to generalize the findings of a study to other similar locations or subject populations, is limited in qualitative research. Generalizability demands inferential statistical tests, random selection of participants, and random assignment of treatments to participants. In many qualitative studies, none of these assumptions are met, and the results have limited generalizability. Quantitative research, on the other hand, uses objective measures and scales that are repeatable and reliable. The most robust quantitative methods use instruments that have an absolute zero, for example 0-s response time or 0-ft detection distance.

Opinion and Preference

The opinion or preference survey is among the most common methods used to evaluate the effectiveness of a pedestrian or bicycle traffic control device. The survey is intended to record the opinions of users or potential users of a countermeasure, which the practitioner can use in conjunction with other data (e.g., safety, operations, etc.) to estimate the countermeasure's effectiveness (see figure 28).



© Texas Transportation Institute

Figure 28. Photo. Potential users of a countermeasure record opinions in a survey.

When creating a survey instrument, the practitioner must consider the following issues:

- What do I want to learn (e.g., is the message on a sign understandable, do pedestrians tend to cross the street at this midblock location, etc.)?

- How am I going to conduct the survey to attain my objective?

Surveys can be very labor intensive to collect and process. If surveys are conducted in person, surveyors must be employed to conduct the survey, and appropriate locations must be identified. Staffing is also a critical component of a phone-based survey. Online, emailed, and mailed surveys are less labor intensive to administer but may not have as high a return rate because they are less personal and easier for the potential respondent to ignore. Regardless of the method used, the answers must be collected, processed, categorized, and analyzed to determine the key findings.

The length of the survey is important. While longer surveys are more thorough in asking questions the practitioner wants to answer, long or complex surveys may not be returned or completed. Shorter surveys provide more limited information but are more likely to be completed and returned. Similarly, questions are more likely to be answered if the responses are short. Multiple-choice questions, yes/no questions, scale questions (e.g., on a scale from 1 to 10), and other questions requiring one-word answers will probably yield more responses than essay or other open-ended questions. Also, while open-ended questions allow respondents to describe whatever comes to their mind regarding the subject, those answers are much more difficult to analyze, requiring the practitioner to make subjective evaluations of responses to categorize them.

The content of the questions must be considered for a number of reasons. Obviously, the questions should ask for the information that the practitioner wants to know, but there are some ways that are better than others for accomplishing that task. To that end, the practitioner should consider two issues: how questions should be asked and how responses should be scored.

As previously discussed, the manner in which a question is asked has a potential effect on its response rate and can also affect the answers received. The following are two principles to consider when asking survey questions:

- Questions must be worded so as not to introduce a bias to the respondent. For example, if the surveyors want to know how well the message on a sign is understood, they could ask, “How can the wording on this sign be improved?” or they could ask, “Of the following four choices, what do you think this sign means?” Both ask for feedback on the sign’s message, but the former could lead respondents into recommending changes they do not really need, while the latter may indicate that comprehension of the sign is reasonable.

- Questions should be worded so as not to be biased against certain groups of respondents. For example, if evaluating the perceived benefit of a separate bicycle trail, rather than asking, “Do you think that there are too many pedestrians on this street to ride your bicycle safely?” the surveyors could ask, “In your opinion, which of the following scenarios is likely if a separate bicycle trail is installed along this corridor: improved safety and/or mobility for pedestrians; improved safety and/or mobility for bicyclists; both of the above; or neither of the above?"

The way in which responses to a question are scored also has an effect on the usefulness of a survey. Some principles are as follows:

- Just as the questions should not be biased, the survey administrator should also not be biased. Although a particular outcome may be desired when evaluating a given countermeasure, it is important to view the results objectively and neutrally. Otherwise, some key findings from the survey may be overlooked or dismissed because they do not fit the surveyor's predetermined desired outcome.

- The practitioner must have specific criteria to define what the results of the survey reveal. For example, is the survey intended to help determine whether a sign should be installed, or is it to help determine whether that sign should have a particular message? The way a survey will be scored should be determined before the survey is conducted so that the questions ask for the information needed to properly evaluate the required criteria.

- In many cases, a survey can be asked before and after a countermeasure is installed to compare actual results with anticipated results. In these cases, the surveyor should ask pedestrians for their opinion in the same way both before and after countermeasure installation. If the survey is changed, the results cannot be directly compared.

- The surveyor should consider the transferability of the results. That is, can the results of the survey at one location be applied to other locations, or are they site specific? Care should be taken before applying survey results to locations, countermeasures, or populations for which they were not originally evaluated. For instance, a study of the effectiveness of fluorescent yellow-green shared use trail crossing signs conducted through surveying the preference of school-age bicyclists may reveal a positive opinion. However, that survey finding does not suggest that engineering measures of slower automobile speeds on the roads that cross the trails will also result from the countermeasure.

- As it relates specifically to this category of treatments, the practitioner should consider the attitude toward pedestrians (or bicyclists) in the area in which the survey is conducted and in which the treatment is proposed. Related to generalizability, a particular treatment may show promise of improvement, but if the location in question is not amenable to pedestrian/bike treatments in general, a survey will likely reflect those attitudes in the responses it receives.

Larger sample sizes generally tend to improve statistical confidence in the results, but it is not the only consideration. Just because a large sample size can quickly be achieved with a survey does not necessarily make it the better study method. Experiences with previous traffic control device evaluations have identified the following concerns:

- The survey needs to target the correct audience.

- It may be important to survey both pedestrians and motorists even if the treatment is for pedestrians.

- A survey instrument should make use of video clips, proposed signs or messages, computer animations, or other illustrations to further describe a situation and allow the respondent to more easily understand the questions that the practitioner is asking.

- Surveys should be performed with people in a similar situation to the one that the practitioner wishes to evaluate. Do not ask questions on paper illustrated only with pictures when the actual location, device, or scenario can be shown either from a recording in live action or in person.

An example of the principle in the last bullet is as follows: A survey is commissioned to determine how far a pedestrian would walk to take advantage of a device that could aid in crossing a busy street. The practitioner could ask a random sample of people in an office building, “Would you walk 100 ft to take advantage of the device?” or the practitioner could stop the same number of people at a midblock location on the street (100 ft away from the location of the proposed device) and ask if they would be willing to walk to the next intersection if that intersection had the device. The latter will produce more reliable and realistic answers because the respondent is making a decision in the same scenario that the practitioner wants to evaluate.

Comprehension

Survey methods can also be used to quantify actual comprehension of the meaning of a device rather than merely asking about an opinion or preference (see figure 29). Comprehension studies can be open-ended questions, (e.g., “What do you think this sign means?”), true/false, or multiple choice. For multiple-choice questions, care must be taken in the wording so that all options are plausible. Questions should be pretested by people with limited education to assess the readability of the questions by all people.

© Texas Transportation Institute

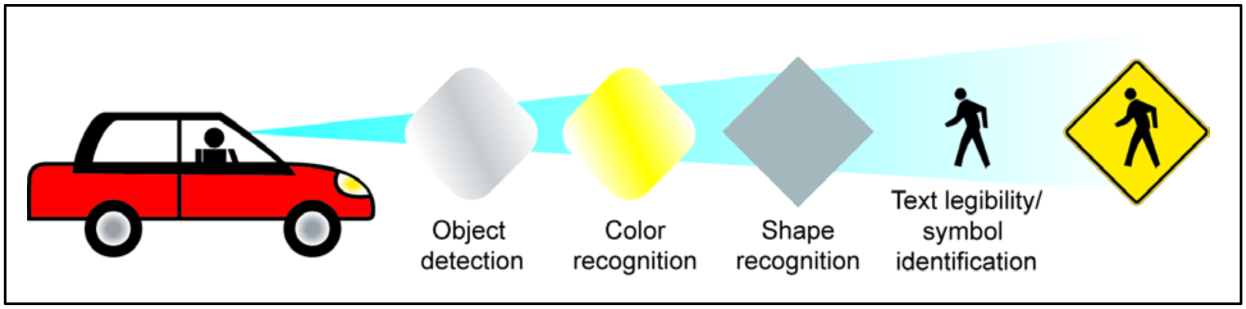
Figure 29. Photo. Participants in a comprehension study provide their understanding of the meaning of a device.

Traffic control devices, in particular signs and marking patterns, are best tested by presenting them in a roadway context through illustrations, digitally edited photographs, or computer animations. Past research has shown that people respond better to sign testing in context as opposed to viewing a line drawing of a sign in isolation.[16] Care should be taken when developing illustrations and photographs to avoid including any extraneous information in the background that subjects may use to help them answer the question. For instance, a researcher may take an actual photograph of a location slated to receive a new pedestrian crossing sign. With digital editing software, the test sign can be inserted into the scene. Participants in comprehension tests will scour the photograph for clues, so if there is a speed limit sign visible in the scene, they may answer based on their experience with roads operating at that speed. Likewise, if there is a local business identifiable in the scene, the participants may recognize the location and answer based on their experience with that location.

Visibility

Visibility studies are often used to assess the effectiveness of traffic control devices for drivers and pedestrians. Visibility studies often measure the distance at which participants can identify some visual characteristic of a treatment. Most often, the order in which the different features of an object appear are as follows (see figure 30):

- **Detection**: People can pick the treatment out of a busy scene or can detect its presence at a great distance. The treatment has attracted their attention. Conspicuity refers to how attention-grabbing a target is. The conspicuity of a target is dependent on its surroundings. For example, a red traffic signal is conspicuous against a blue sky but may be hard to see at night on a hill with red vehicle tail lights in the scene.
- **Color recognition**: People can correctly identify the color of the target.
- **Shape recognition:** People can correctly identify the shape of the target.
- **Text legibility/symbol identification**: People can correctly identify the symbol or read the text. Legibility of a symbol or text does not necessarily mean that the person understands the intended message of the sign.



© Texas Transportation Institute

Figure 30. Illustration. Distance at which drivers can discern characteristics of a traffic control device.

Visibility studies can be performed in a moving motor vehicle where the subject calls out his or her answer and an experimenter marks a distance on a distance measurement instrument. A natural lag occurs in this type of response between the time a subject speaks and the time an experimenter hits the button. Another approach is to have the subject perform the visibility test from a stopped motor vehicle at a number of preselected distances. The distribution of the number of people getting the correct answer at each of the preselected distances can then be derived. In either of these methods, research participants have a natural tendency to be conservative in their responses. In other words, they may have an idea as to what a sign says, but they will wait until they are absolutely certain before they respond. This response bias tends to decrease the absolute value of visibility responses. For this reason, it is best to interpret visibility responses as a comparison between two treatments tested rather than looking at the absolute values of the responses.

Driving Performance

Other measures of driving performance can be measured if the test vehicle is appropriately equipped (see figure 31). These types of instrumented vehicles are typically operated by larger research institutions. Sensors on throttle and brake pedals can measure pedal activation. When paired with global positioning system instrumentation, the location of the vehicle can be recorded. Cameras can be mounted on dashboards to record driver eye glances. Specialized equipment that minutely measures driver glances can be used. This type of eye-tracking equipment can also be either mounted on the dashboard or worn by the driver. This equipment is generally quite expensive and requires special calibration and labor-intensive data coding.

© Texas Transportation Institute

Figure 31. Photo. Instrumented vehicle used for measurements during in-field evaluations.

EXAMINING TRAFFIC OPERATIONS

In contrast to using laboratories to evaluate treatments under controlled conditions, a number of in-field evaluations can be conducted to determine the effectiveness of a treatment using various MOEs. In-field evaluations are often used when a suitable laboratory environment cannot be created or when controlled evaluations have been satisfactorily completed and a real-world test is needed to confirm the results. The practitioner must remember that there are many ways to determine the effectiveness of a treatment, so specific MOEs must be defined prior to establishing the methodology of the in-field evaluation. Some MOEs (and examples of reportable results) are listed below. Sources of information on data collection techniques and procedures are listed in chapter 6 of this report.

Several techniques are available to collect noncrash traffic operational data, including manual and video methods. Each of these has advantages and disadvantages. The manual method is generally less costly because data collection and initial data reduction occur at the same time. Video requires staff resources to collect the data, and additional staff resources are required to watch the video and reduce the data. An advantage of video is that data reduction methods can be refined after initial data are collected since the video can be watched again to collect additional or different data. Also, questionable data can be validated using video, while the manual method may not provide that opportunity.

Volume

An evaluation may investigate changes in traffic or pedestrian volume to determine whether a new device affected the volume at a location (see figure 32). A new pedestrian treatment (e.g., marked crosswalk, countdown signals, etc.) may increase pedestrian traffic at a location. A warning device may also decrease traffic within a particular corridor.

© Texas Transportation Institute

Figure 32. Photo. Pedestrian volume counters at the location of a new device.

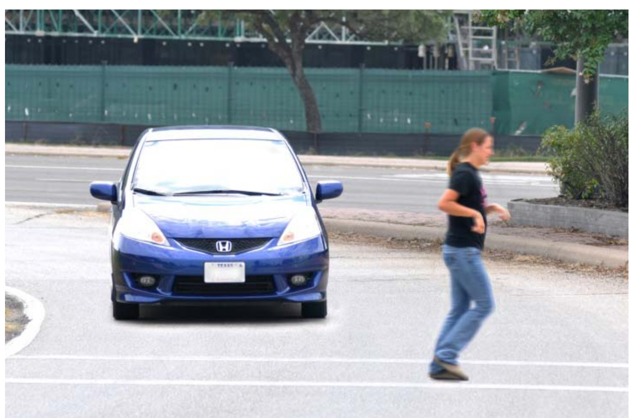
Road User Behavior

Strictly speaking, measuring behavior is a difficult concept. However, specific types of behavior and their results can be evaluated through the use of well-defined MOEs. For example, to measure the acceptance of a new pedestrian pathway, an evaluation could examine changes in pedestrian travel paths before and after the treatment is installed. While the evaluation does not specifically ask pedestrians whether they like the new treatment, if use of the new pathway is high and pedestrian traffic on adjacent routes has decreased, one can reasonably infer that response to the new pathway is positive and the treatment is being accepted.

Evaluations of behavior should account for local conditions. As discussed previously, the evaluator should pay special attention to changes in conditions when using a before-after evaluation. For example, if the after evaluation of a new shared use trail is conducted during rainy weather, the resulting volume count may be too low to accurately reflect the true demand for the facility and would indicate a lower acceptance of the trail than is accurate. Similarly, extra attention should be considered to observed holidays. Counting pedestrian crossing maneuvers in front of a large office building on a holiday may produce counts that are lower than normal and would not provide an accurate representation of demand for the crossing.

It is also important to consider user profiles. That is, the evaluation should consider who uses the treatment and who avoids it. A treatment that works well in one location might not work as well in another location if the potential user group is substantially different.

The evaluation must also clearly define how the road user behavior is to be measured (see figure 33). For example, an operational definition of an aborted crossing might be "pedestrian returns to curb after having both feet in roadway." The definition selected may influence how the data are collected. Technicians may not be able to determine if both feet were in the roadway from the video. Therefore, on-site data collection may be required. An example of an operational definition of running might be "crossing an entire traffic lane in three or less frames of film" (or within a set amount of time). Such a concise statement of the elements used to define a behavior is invaluable for comparing the results between the before and after periods and between different studies. Good operational definitions also allow others to evaluate performance of a traffic control device in different regions of the country. To be able to verify the performance of the device, the MOE needs to be clearly defined.



© Texas Transportation Institute

Figure 33. Photo. Pedestrian running through crosswalk to avoid oncoming vehicle.

Erratic Maneuvers or Conflicts

A number of traffic-related measures can be evaluated to gain an appreciation for the comprehension or acceptance of a device. If a device is commonly violated or if the number of conflicts or erratic maneuvers increases at the treatment site, it is likely that the treatment is not well understood or is being disregarded (see figure 34).

© Texas Transportation Institute

Figure 34. Photo. Pedestrians and driver experiencing a near miss.

For this type of evaluation, the study could record near misses, reversals, and conflicts near or as a result of the treatment. This directly reflects the comprehension of the treatment and provides an indication of users' confusion. If collisions are narrowly avoided or if users commonly and abruptly change their intended course through the treatment area, it is likely that the treatment is not well understood or is not providing useful information.

Compliance, Violations, and Risk Taking

Compliance or violation can be the measure for an evaluation of respect for a particular traffic control device. The user groups studied could be motorists, pedestrians, bicyclists, or any combination of the three, depending on the type of treatment. A compliance evaluation records violations (e.g., non-compliance with the treatment, violations of traffic law, etc.) and typically reports the results as the frequency of violation or the rate of violation as a percentage of the total number of opportunities to comply. For example, motorist compliance with a new pedestrian signal or beacon could be measured in terms of a compliance rate. This rate is expressed as the percentage of yielding vehicles, out of all approaching vehicles that should have yielded (e.g., a 90 percent compliance rate means that of 100 vehicles approaching an activated beacon, 90 of them yielded to the crossing pedestrian). Other examples include bicyclists stopping at a stop sign or pedestrian compliance with the intent of a device (e.g., whether pedestrians are pushing a call button).

One approach can also measure risk-taking behavior. This approach is related to behavioral evaluations previously discussed and attempts to measure users' willingness to travel through a treatment area under less-than-ideal conditions. This approach requires very specific definitions of what constitutes a risk at that location (e.g., crossing a street within 2 s of an oncoming vehicle). Risk-taking can be measured in terms of frequency (i.e., how often risky behavior is observed) or in terms of change (i.e., the difference in frequency before and after the treatment is installed).

Compliance and risk evaluations can be labor intensive and complex. It is therefore important to create detailed evaluation procedures for data collectors to follow. The means of observation (e.g., in-field real-time evaluation versus recorded video surveillance), the definition of the events of interest, and the ability of the data collectors to repeatedly follow the procedures are all critical components of this type of evaluation. If an event is qualitative, there must be specific scoring criteria that data collectors use for each event. Privacy concerns must also be considered, particularly if recorded video is to be used. Recorded images cannot be used to obtain personally identifiable information and cannot be used for purposes outside the evaluation without the permission of the subjects being recorded. Data collectors should be properly trained in procedures and protocols before commencing such a study.

Citations

If the data are available, examining traffic citations relevant to the countermeasure proposed and issued at the treatment site may be useful. Using citations as the sole evaluation method is not desirable because the number of citations is often heavily influenced by the enforcement levels used in a region.

Some of these citations can also be documented from crash reports if that is already part of the evaluation. Again, care must be taken in interpreting the citation record because the actual violation cited may vary depending on the severity of the crash. For citations that were not issued in the aftermath of a crash, obtaining details would probably require a separate request to the law enforcement agency of interest and would likely require a number of assurances to protect the privacy of those involved. Those assurances (e.g., that no one outside of the study would have access to the data, that no one is personally identified in the study results, etc.) are typically similar, if not identical, to those required to use crash data, but citation data may be even more sensitive.

Citation data can enable the practitioner to gain a perspective on operations at a site that would not be available during a typical in-field evaluation, or even through crash data, because citations are not necessarily associated with a condition, time of day, or crash. However, if there is an identifiable pattern of citations issued, that information may contribute to a more informed decision on the most appropriate countermeasure.

Speed

In a discussion on speed, it is important to accurately define speed. ITE's *Manual of Transportation Engineering Studies* describes a number of speed study methods.[17] It gives several definitions of speed, depending on the purpose of the study to be used, as follows:

- *Speed* is the rate of movement of a vehicle in distance per unit of time. Common units are miles per hour (mi/h), feet per second (ft/s), kilometers per hour (km/h), and meters per second (m/s).

- *Spot speed* is the instantaneous measure of vehicle speed at a specific location on a roadway.

- *Time-mean speed* is the arithmetic mean or average of several spot-speed measurements. It is the sum of the measured spot speeds divided by the number of measurements.

- *Space-mean speed* is another type of average speed. It is the length of a segment divided by the mean travel time of several vehicles or trips over the segment. Studies involving space-mean speed are less common than those using time-mean speed.

- *Median spot speed* is the middle value in a series of spot speeds that have been ranked in order of magnitude. It is the midpoint of a range of speeds.

- *Modal spot speed* is the value that occurs most frequently in a sample of spot-speed measurements. It is the most typical value in a range of speeds.

- *85th percentile speed* is the speed at or below which 85 percent of the total observed values fall in a sample of measured spot speeds. Other values besides 85 percent can be used for an *i*th *percentile spot speed*—the value at or below which *i* percent of the spot speeds occur. Typical engineering practice is to use the 85th percentile.

- *Pace* is the specified increment of spot speed, usually 10 mi/h, that includes the greatest number of speed measurements.

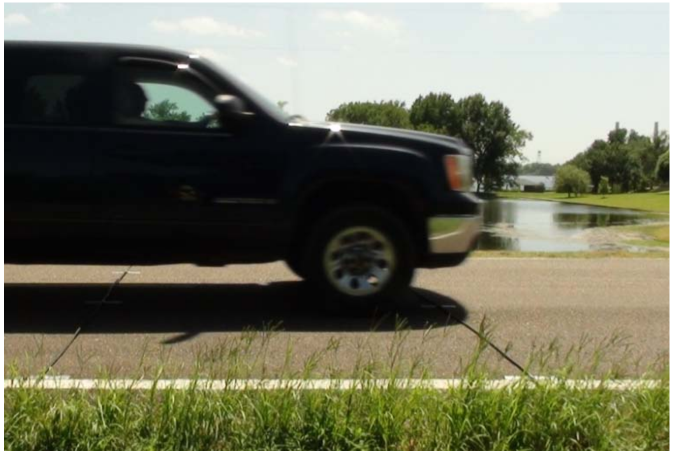
- *Standard deviation* is a commonly used measure of the spread of individual speeds around the mean. It is the square root of the sum of squares of the deviations of the individual spot speeds from the mean speed divided by the number of measurements less one.

Speed data are collected by one of two general approaches: indirect and direct measurements. Indirect measurements provide an estimate of speed because they are actually measurements of time for a vehicle to travel a known distance between two points. The more commonly used method is direct measurement. LIDAR or radar guns measure speeds of selected target vehicles. On-pavement sensors record speeds of every vehicle traveling through the study site.

In a simple spot-speed study, speed values are collected and recorded at key locations (e.g., incremental locations on an approach to a new pedestrian signal). However, when those speed values are analyzed and results are reported, it is important to report more than just an average. The variability of the speed (e.g., 85th percentile, standard deviation, mode, minimum, and maximum) should also be reported. These statistics provide a better overall picture of speed characteristics at the study site. Another important speed value to report in pedestrian- or bike-related studies is the average speed +5 or +10 mi/h, an estimation of whether a collision at that speed is within the range of survivability for a pedestrian or cyclist can be an important measure of a treatment's effectiveness.

Another speed-related study involves identifying the location where braking is initiated by approaching vehicles. This measure produces an indication of how well drivers see (and comprehend) a treatment (such as a pedestrian signal) from a distance. Early, gradual braking means that approaching drivers are aware of and understand the treatment and are responding to it, as opposed to a pattern of later panic braking. This type of study typically requires the use of one or more video cameras and a study site that has a series of well-defined visual landmarks to aid the evaluator in determining the location at which each vehicle's brake lights are activated.

Related to braking location is a study that records the deceleration profile of approaching vehicles (see figure 35). This method helps the evaluator to understand whether the typical driver uses gradual or panic braking, in addition to determining the exact location where braking begins. This study requires the use of a radar or LIDAR gun to collect a series of closely spaced spot speeds and their corresponding locations to generate the target vehicles' speed-distance profiles, which can then be converted to deceleration profiles.



© Texas Transportation Institute

Figure 35. Photo. Road tubes used to determine the deceleration profile of approaching vehicle.

As with all types of field studies, safety is a paramount consideration when conducting spot-speed studies. The measurement of speeds often involves workers being in proximity to the roadway or within the actual travel lanes. Thus, workers must use care and vigilance at all times while working near the roadway. Workers should park their vehicles off the traveled way, wear appropriate protective clothing (e.g., reflective vests, hard hats, etc.), and act in a manner that does not distract motorists or adversely affect traffic speeds. Workers should conduct their in-roadway work under low-volume conditions when possible, and they should use lane-closure procedures, traffic control assistance, and warning devices appropriate for conditions.

Furthermore, workers utilizing LIDAR or radar guns to collect speeds should take extra precautions to be inconspicuous during the data collection in an effort to minimize any effects their presence may have on approaching drivers. If drivers are under the impression that the researchers are law enforcement officers and are conducting speed enforcement, the drivers will likely travel slower, brake earlier, or make other changes in their behavior, adversely affecting the study results.

Delay or Travel Time

Delay may be an appropriate MOE for evaluating treatments such as changes to signal timing or restrictions of movement (e.g., no left turn during selected hours). If, for example, a pedestrian-related treatment is to be installed with minimal impact on the adjacent vehicular traffic stream, an evaluation of delay can provide an indication of the treatment's effects on vehicular traffic. Delay or travel time can also be used to determine if a traffic control device reduces a pedestrian's trip time.

Delay can be determined through multiple methods, but the two most common are in-field evaluations and computer simulation. In-field evaluations require on-site staff to observe and record data, or conditions may be recorded and viewed off site. If a treatment site has high pedestrian volumes, it is likely that a procedure using video-recording will be necessary to document the delay for all road users passing through the site. There are a variety of computer modeling and microsimulation software packages, each with its own capabilities, strengths, and weaknesses. A computer-based delay study requires selecting an appropriate software package that can evaluate the treatment under consideration

Gap

Gap-related studies are important when analyzing traffic movements that conflict with higher-priority movements, including movements controlled by stop signs, right turns on red, permissive left turns, and midblock pedestrian crossings (see figure 36). Gaps at school crossings are another application. The definitions related to gap studies are important and may include the following:

- *Gap* is the time or distance between the back of one vehicle and the front of the trailing vehicle.

- *Headway* is the time or distance between a point (front, center, or back) of one vehicle and the same point of the trailing vehicle.

- *Lag* is the time or distance between the arrival of a vehicle or pedestrian at a minor street and the front of the next conflicting major-street vehicle arriving at that minor street.

© Texas Transportation Institute

Figure 36. Photo. Pedestrians using a gap in traffic to cross the street.

Other Potential/Unique Measures

There may be other MOEs that are appropriate for a given location or treatment depending on the desired outcome of the treatment. These measures may be less common, but rarity should not automatically exclude a potential measure if it is appropriate for the conditions. Unique measures may be related to behavioral measures discussed previously, but they likely will have a distinctive characteristic related to the specific nature of a location or a treatment. Below are some examples of some less common measures that can apply to pedestrian or bicycle treatment evaluations:

- **Right turn on red**: A study of a new marked crosswalk could evaluate the frequency of this maneuver before and after the treatment was installed, or it could determine the frequency of complete stops versus rolling stops.

- **Crossing maneuvers late in the pedestrian clearance interval**: An evaluation of new countdown signals could record the amount of time prior to the onset of steady "DON'T WALK" when pedestrians begin crossing maneuvers.

- **Bicycle lane volume**: Comprehension, effectiveness, and acceptance of a newly installed bicycle lane can be evaluated by comparing the number of bicyclists using the new bicycle lane to the number using the vehicle lane or sidewalk before and after its installation.

EXAMINING CRASHES

As mentioned previously, a study of the crash history at one or more treatment sites can be used to evaluate safety as long as certain steps are taken into account for variables that may change over time. There are several study approaches using this methodology in a before-after with comparison group study, and a few general principles are appropriate, as follows:

- Counting only fatal or fatal/injury crashes frequently results in sample sizes that are too small to make conclusions from a statistical analysis.

- Unless a unique characteristic of the evaluation must consider the actual installation period, exclude a specified time immediately after the installation from the evaluation. It is prudent to begin the after period 30, 60, or 90 days following the conclusion of installation rather than on the day after installation is complete. This allows road users to adapt to the new treatment.

- Where possible, at least 3 years of crash data in each time period studied, before and after, should be included.

- The practitioner should identify the appropriate jurisdiction or authority that maintains the records for crashes at the study sites. In addition, how the records are stored can be of value. For example, if crashes are only located to the nearest 0.1 mi then a search of all crashes occurring at 0.45 mi may return zero crashes when the search should have been from 0.4 to 0.5 mi.

- The practitioner needs to determine the locations that should be included in the study. For an intersection, simply gathering the crashes that occurred at the specific point where the streets cross may be insufficient. Intersection treatments can affect crashes along the approach legs. A pedestrian treatment may also affect pedestrian crossings that occur at midblock. Therefore, the number of crashes occurring both at midblock and at the intersection should be examined.

The following are additional observations about crash studies that practitioners should consider for pedestrian treatments:

- Typically, the severity of pedestrian crashes is so high compared to other motor vehicle crashes that the elimination of a few pedestrian crashes will result in a high safety dividend and a high benefit/cost ratio. Use these results appropriately for the treatment being evaluated.

- Pedestrian crash rates need to account for both pedestrian volumes and motor vehicle volumes.

- Percent reduction in crashes is a readily understandable measure of change in crashes for a given treatment. Accompanying the percent reduction with other data, such as the number of crashes (frequency) and crash rate (typically crashes per million vehicle miles or million entering vehicles with pedestrian volumes considered) will provide an increased understanding of a treatment's effectiveness.

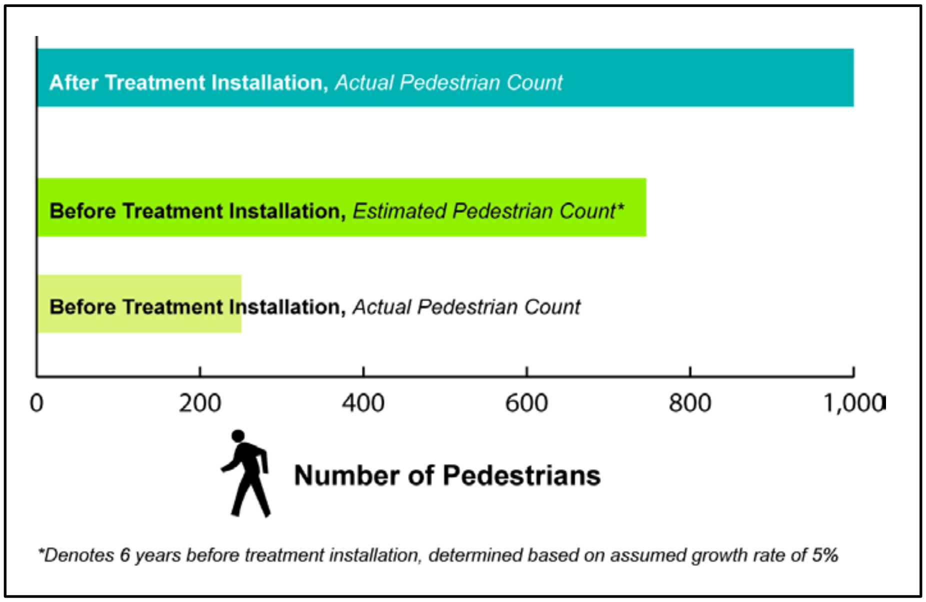
- Research into how best to conduct crash evaluations has increased in recent years. The use of an EB method is an example of a technique developed to specifically address issues associated with examining road crashes. Understanding and appropriately employing these techniques is critical to identifying findings that are both accurate and defensible.

THE IMPORTANCE OF PEDESTRIAN VOLUMES

Unlike vehicle crashes, crash rates for pedestrians are typically not used since pedestrian volumes are usually not known. It is uncommon for agencies to invest resources to collect the systemwide pedestrian counts that are needed to develop pedestrian crash rates. This lack of exposure data emphasizes the need to collect pedestrian volumes both before and after treatment installation so that changes in crash rates can be accurately reported.

While an assumed growth rate has frequently been used to adjust traffic volumes from year to year, this approach has serious limitations with respect to treatments that could increase the pedestrian volume. A constant growth rate used to adjust a current pedestrian volume to a previous-year pedestrian volume could result in an inappropriately high pedestrian volume for the years prior to the installation of the treatment. Computed crash rates would therefore be unreliable.

For example, consider a pedestrian count conducted after a treatment was installed was 1,000 pedestrians per day, and the city had been seeing a 5 percent annual growth rate. The number of pedestrians that existed at that site 6 years previously would be 746 pedestrians ($1{,}000 \times 1.05^{-6}$). However, if the treatment resulted in additional pedestrians using the location, this cannot be accurately determined from the count. Any computation of previous site volumes would be inaccurate. So if there were really only 250 pedestrians in the before condition, then the rate should have been 1 crash in 250 rather than 1 crash in 746 pedestrians. Figure 37 shows a graphic representation of this example.



© Texas Transportation Institute

Figure 37. Graph. Estimated versus actual number of pedestrians at a crossing.

REFERENCES

1. Federal Highway Administration. (2009). *Manual on Uniform Traffic Control Devices*, Washington, DC. Obtained from: http://mutcd.fhwa.dot.gov/kno_2009.htm. Site last accessed February 11, 2010.

2. National Committee on Uniform Traffic Control Devices. (2011). *Purpose*, Washington, DC. Obtained from: http://www.ncutcd.org/purpose.shtml. Site last accessed February 11, 2010.

3. U.S. Department of Transportation. (2003). *Manual on Uniform Traffic Control Devices,* Federal Highway Administration, Washington, DC. Obtained from: http://mutcd.fhwa.dot.gov/kno_2003r1r2.htm. Site last accessed April 25, 2011.

4. Pline, J.L. (2001) *Traffic Control Devices Handbook*, ITE Publication IR-112, Institute of Transportation Engineers, Washington, DC.

5. Emo, A.K. and Do, A.H. (2008). "Evaluating and Validating a New Pedestrian Exposure to Risk Metric," *Human Factors and Ergonomics Society Annual Meeting Proceedings, 52*(23), 1845–1849, Santa Monica, CA.

6. Bureau of Transportation Statistics. (2003). *Guide to Good Statistical Practice in the Transportation Field*, Federal Highway Administration, Washington, DC. Obtained from: http://www.bts.gov/publications/guide_to_good_statistical_practice_in_the_transportation_field/pdf/entire.pdf. Site last accessed March 16, 2010.

7. Knoblauch, R.L. and Crigler, K.L. (1987). *Model Pedestrian Safety Program: User's Guide*, Report No. FHWA-RD-87-039, Federal Highway Administration, McLean, VA.

8. American Association of State Highway and Transportation Officials. (2010). *Highway Safety Manual 1st Edition*, AASHTO, Washington, DC.

9. Council, F., Zaloshnja, E., Miller, T., and Persaud, B. (2005). *Crash Cost Estimates by Maximum Police-Reported Injury Severity Within Selected Crash Geometries*, Report No. FHWA-HRT-05-051, Federal Highway Administration, McLean, VA.

10. Schrank, D. and Lomax, T. (2009). *Annual Urban Mobility Report*, Texas Transportation Institute, College Station, TX. Obtained from: http://tti.tamu.edu/documents/mobility_report_2009_wappx.pdf. Site last accessed July 26, 2010.

11. Zegeer, C.V., Steward, J.R., Huang, H.H., and Lagerwey, P.A. (2002). *Safety Effects of Marked Versus Unmarked Crosswalks at Uncontrolled Locations: Executive Summary and Recommended Guidelines*, Report No. FHWA-RD-01-075, Federal Highway Administration, Washington, DC. Obtained from: http://drusilla.hsrc.unc.edu/cms/downloads/Effects_Un_MarkedCrosswalks_Summary.pdf. Site last accessed March 24, 2010.

12. Fitzpatrick, K., Turner, S., Brewer, M., Carlson, P., Lalani, N., Ullman, B., Trout, N., Park, E.S., Lord, D., and Whitacre, J. (2006). *Improving Pedestrian Safety at Unsignalized Crossings*, TCRP Report No. 112, National Cooperative Highway Research Program Report No. 562, Transportation Research Board, Washington, DC. Obtained from: http://onlinepubs.trb.org/onlinepubs/nchrp/nchrp_rpt_562.pdf. Site last accessed March 24, 2010.

13. Hauer, E. (1997). *Observational Before-After Studies in Road Safety*, Pergamon Press, Oxford, England.

14. American Association of State Highway Officials. *SafetyAnalyst* software tool, AASHTO, Washington, DC. Obtained from: http://www.safetyanalyst.org/. Site last accessed April 25, 2011.

15. Hauer, E. (2005). *Cause and Effect in Observational Cross-Sectional Studies on Road Safety*, Draft HSIS Report, Federal Highway Administration, McLean, VA.

16. Chrysler, S.T., Wright, J., and Williams, A. (2004). *3D Visualization as a Tool to Evaluate Sign Comprehension*, Report No. 167721-1, Southwest Region University Transportation Center, Texas A&M University, College Station, TX. Obtained from: http://swutc.tamu.edu/publications/technicalreports/473700-00021-1. Site last accessed March 24, 2010.

17. ITE. (1994). *Manual of Transportation Engineering Studies*, Institute of Transportation Engineers, Washington, DC.

HRDS-30/05-11(WEB)E