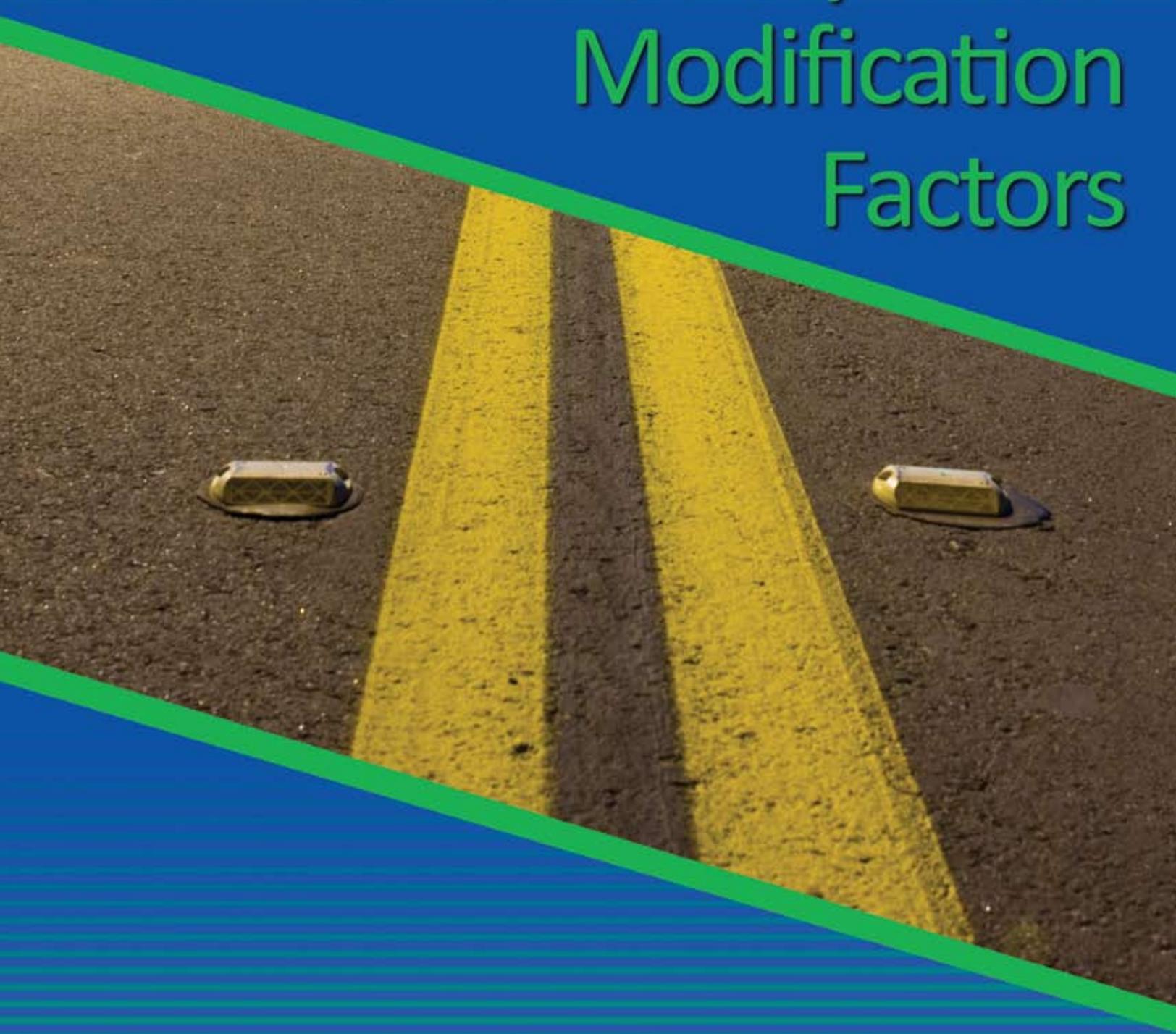


A Guide to Developing Quality Crash Modification Factors



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.

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 DOCUMENTATION PAGE

1. Report No. FHWA-SA-10-032	2. Government Accession No.	3. Recipient's Catalog No.	
4. Title and Subtitle <i>A Guide to Developing Quality Crash Modification Factors</i>		5. Report Date December 2010	
		6. Performing Organization Code	
7. Author(s) Frank Gross, Bhagwant Persaud, and Craig Lyon		8. Performing Organization Report No.	
9. Performing Organization Name and Address. Vanasse Hangen Brustlin, Inc (VHB) 8300 Boone Blvd., Ste. 700 Vienna, VA 22182-2626		10. Work Unit No.	
		11. Contract or Grant No. DTFH61-05-D-00024	
12. Sponsoring Agency Name and Address U.S. Department of Transportation Federal Highway Administration (FHWA) Office of Safety 400 Seventh Street, SW, HSSD Washington, DC 20590		13. Type of Report and Period Covered January 2010- December 2010	
		14. Sponsoring Agency Code FHWA/HSSP	
15. Supplementary Notes. The FHWA Office of Safety Task Order Manager was Karen Yunk. Technical support for the development of the guide was provided by Persaud and Lyon, Inc. and production was performed by Annette Gross.			
16. Abstract. The purpose of this guide is to provide direction to agencies interested in developing crash modification factors (CMFs). Specifically, this guide discusses the process for selecting an appropriate evaluation methodology and the many issues and data considerations related to various methodologies. The guide opens with a background of CMFs, including the definition of CMFs and related terms, purpose and application, and general issues related to CMFs. The guide then introduces various methods for developing CMFs. Discussion of these methods is not intended to provide step-by-step instruction for application. Rather, this guide discusses study designs and methods for developing CMFs, including an overview of each method, sample size considerations, and strengths and weaknesses. A resources section is provided to help users identify an appropriate method for developing CMFs based on the available data and characteristics of the treatment in question. The resources section also includes a discussion of considerations for improving the completeness and consistency in CMF reporting. The guide is written for transportation safety practitioners, consultants, and researchers. These primary users are expected to have experience and/or education in the theory and practice of road safety engineering, including basic analytical procedures and statistical concepts.			
17. Key Words. Crash Modification Factors, Crash Modification Functions, Crash Reduction Factors, Accident Modification Factors, Safety Analysis		18. Distribution Statement No restrictions.	
19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 72	22. Price

TABLE OF CONTENTS

1. OVERVIEW	1
2. BACKGROUND ON CRASH MODIFICATION FACTORS	3
2.1 Definitions	3
2.2 Purpose and Application	5
2.3 General Issues Related to CMFs	10
2.4 Summary	12
3. STUDY DESIGNS TO DEVELOP CMFS	13
3.1 Before-After with Comparison Group Studies	14
3.2 Empirical Bayes Before-After Studies	20
3.3 Full Bayes Studies	26
3.4 Cross-Sectional Studies	28
3.5 Case-Control Studies	30
3.6 Cohort Studies	35
3.7 Alternative Approaches for Developing CMFs	39
3.8 Summary	45
4. RESOURCES	49
4.1 Flow Chart	49
4.2 Sample Problems	51
4.3 Improving the Completeness and Consistency in CMF Reporting ..	56
4.4 Summary	59
5. CONCLUSION	61
REFERENCES	62

1. OVERVIEW

The purpose of this guide is to provide direction to agencies interested in developing crash modification factors (CMFs). Specifically, this guide discusses the process for selecting an appropriate evaluation methodology and the many issues and data considerations related to various methodologies.

The next chapter provides a background of CMFs, including the definition of CMFs and related terms, purpose and application, and general issues related to CMFs. Chapter 3 outlines various methods for developing CMFs. Discussion of these methods is not intended to provide step-by-step instruction for application. Rather, this guide discusses study designs and methods for developing CMFs, including an overview of each method, sample size considerations, and strengths and weaknesses. A resources section is provided in Chapter 4 to help users identify an appropriate method for developing CMFs based on the available data and characteristics of the treatment in question. The resources section also includes a discussion of considerations for improving the completeness and consistency in CMF reporting.

The guide is written for transportation safety practitioners, consultants, and researchers. These primary users are expected to have experience and/or education in the theory and practice of road safety engineering, including basic analytical procedures and statistical concepts.

2. BACKGROUND ON CRASH MODIFICATION FACTORS

This chapter provides background information related to CMFs, including the definition of a CMF and related terms, the purpose and application of CMFs, and a discussion of the general issues associated with CMFs. This information will better inform users how to judge the quality of a CMF and identify important issues to consider in CMF development.

2.1 DEFINITIONS

This section introduces crash modification factors (CMFs), crash modification functions (CMFunctions), and several related terms that are necessary to understand the quality of the CMF. The terms are defined and an example of each is provided. Subsequently, the term CMF is used with the intention of also meaning CMFunctions, unless specifically noted otherwise.

Crash Modification Factor

A CMF is a multiplicative factor used to compute the expected number of crashes after implementing a given countermeasure at a specific site. The CMF is multiplied by the expected crash frequency without treatment. A CMF greater than 1.0 indicates an expected increase in crashes, while a value less than 1.0 indicates an expected reduction in crashes after implementation of a given countermeasure. For example, a CMF of 0.8 indicates an expected safety benefit; specifically, a 20 percent expected reduction in crashes. A CMF of 1.2 indicates an expected degradation in safety; specifically, a 20 percent expected increase in crashes.

EXAMPLE

The CMF for installing a traffic signal at a rural stop-controlled intersection is 0.23 for angle crashes (Harkey et al., 2008). If a specific stop-controlled intersection is to be converted to a signalized intersection and the expected number of crashes at this intersection is 6.24 angle crashes per year, the expected crash frequency after signalization would be equal to $6.24 \times 0.23 = 1.44$ angle crashes per year.

Stated in terms of the expected change in crashes, the CMF indicates a 77 percent (i.e., $100 \times (1 - 0.23)$) expected reduction in angle crashes after the installation of a traffic signal.

Crash Modification Function

A CMFunction is a formula used to compute the CMF for a specific site based on its characteristics. It is not always reasonable to assume a uniform safety effect for all sites with different characteristics (e.g., safety benefits may be greater for sites with high traffic volumes). A countermeasure may also have several levels or potential values (e.g., improving intersection skew angle, or widening a shoulder). A crash modification function allows the CMF to change over the range of a variable or combination of variables.

Where possible, it is preferable to develop CMFunctions as opposed to a single CMF value since safety effectiveness most likely varies based on site characteristics. In practice, however, this is often difficult since more data are required to detect such differences.

EXAMPLE

The CMFunction for improving intersection skew angle at a rural, 4-legged, stop-controlled intersection is a function of the absolute value of intersection angle minus 90 degrees, as shown in Equation 1, where the intersection angle is in degrees (Bonneson et al., 2005).

$$CMF(\text{Intersection Skew}) = \exp(0.0054 * |\text{intersection angle} - 90|) \quad (1)$$

The CMFunction allows the user to calculate the CMF for a specific intersection skew angle compared to a baseline of 90 degrees. For example, if the intersection angle is 120 degrees, as shown below in Figure 1a, the CMF is $\exp(0.0054 * |120 - 90|) = 1.18$. Note that the CMF is the same if the other angle of the intersection is used; $\exp(0.0054 * |60 - 90|) = 1.18$.

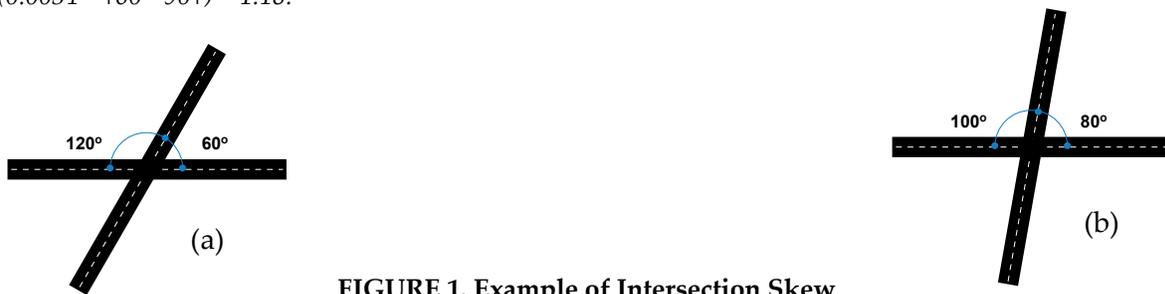


FIGURE 1. Example of Intersection Skew

As the intersection angle approaches 90 degrees, the CMF approaches 1.0. For instance, if the intersection angle is 100 degrees, as shown above in Figure 1b, $\exp(0.0054 * |100 - 90|) = 1.06$.

Standard Error

The standard error is the standard deviation of a sample mean. The standard error provides a measure of certainty (or uncertainty) in the CMF. A relatively small standard error, with respect to the magnitude of the CMF estimate, indicates greater certainty in the estimate of the CMF, while a relatively large standard error indicates less certainty in the estimate of the CMF. The standard error is used in the calculation of the confidence interval. In some cases, the variance of the CMF may be reported instead of the standard error. The standard error is simply the square root of the variance as shown in Equation 2.

$$\text{Standard Error} = \sqrt{\text{Variance}} \quad (2)$$

Confidence Interval

A confidence interval is another measure of the certainty of a CMF. A CMF is simply an estimate of the actual safety effect of a countermeasure based on observations from a sample of sites. The confidence interval provides a range of potential values of the CMF based on the standard error. As the width of the confidence interval increases, there is less certainty in the estimate of the CMF. If the confidence interval does not include 1.0, it can be stated that the CMF is significant at the given confidence level. If, however, the value of 1.0 falls within the confidence interval (i.e., the CMF could be greater than or less than 1.0), it can be stated that the CMF is insignificant at that confidence level. It is important to note insignificant CMFs because the treatment could potentially result in 1) a reduction in crashes, 2) no change, or 3) an increase in crashes. These CMFs should be used with caution (AASHTO, 2010).

A confidence interval is calculated by multiplying the standard error by a factor (i.e., the cumulative probability) and adding and subtracting the resulting value from the CMF estimate. Equation 3 is used for calculating the confidence interval.

$$\text{Confidence Interval} = \text{CMF} \pm (\text{Cumulative Probability} * \text{Standard Error}) \quad (3)$$

The cumulative probability factors for common confidence intervals are shown in Table 1.

TABLE 1. Cumulative Probability Factors

Confidence Interval	Cumulative Probability
99%	2.576
95%	1.960
90%	1.645

EXAMPLE

The CMF for a given countermeasure is 0.761 with a standard error of 0.168. An engineer would like to calculate the 95 percent confidence interval for this CMF.

The first step is to determine the appropriate cumulative probability factor from Table 1, given the desired confidence interval. The factor for a 95 percent confidence interval is 1.960. The 95 percent confidence interval is then calculated by adding and subtracting 1.96 times the standard error of 0.168 from the CMF estimate of 0.761.

$$95\% \text{ Confidence Interval} = 0.761 \pm 1.960(0.168)$$

This gives a confidence interval of 0.432 to 1.090. Note the value of 1.0 lies within the confidence interval. As such, it cannot be stated with 95 percent confidence that the true value of the CMF is not 1.0 (i.e., it cannot be stated with 95 percent confidence that the treatment had any effect).

2.2 PURPOSE AND APPLICATION

Purpose

CMFs can be used by several groups of transportation professionals for various reasons. The primary user groups include highway safety engineers, traffic engineers, highway designers, transportation planners, transportation researchers, and managers and administrators. CMFs can be used to:

- Estimate the safety effects of various countermeasures.
- Compare safety benefits among various alternatives and locations.
- Identify cost-effective strategies and locations in terms of crash effects.
- Check reasonableness of evaluations (i.e., compare new analyses with existing CMFs).
- Check validity of assumptions in cost-benefit analyses.

EXAMPLE

A traffic engineer is considering the following countermeasures for enhancing signal visibility, including the safety effect of each potential measure: increasing lens size, installing signal backplates, or installing dual red indicators in each signal head. The traffic engineer could use CMFs to evaluate the relative cost-effectiveness of each countermeasure and select the most cost-effective improvement or set of improvements.

A highway designer is trying to decide whether to provide paved or gravel shoulders on a two-lane rural road. The highway engineer could use CMFs to compare safety benefits between paved and unpaved shoulders to support the decision.

A transportation planner is considering two alternatives for the long-term design of a corridor. The planner could use CMFs to compare the long-term safety impacts of a series of roundabouts as opposed to a series of signalized and unsignalized intersections for the corridor.

Regression-to-the-mean (RTM) is the natural tendency of observed crashes to regress (return) to the mean in the year following an unusually high or low crash count. RTM effects arise when sites with randomly high short-term crash counts are selected for treatment and experience a subsequent reduction in crashes when these counts regress toward their true long-term mean. In fact, one would expect a reduction in crashes subsequently even if there was no treatment. If a treatment had been installed at these locations, one would tend to over-estimate the effect if the regression-to-the-mean bias is not properly addressed in the analysis.

CMFs provide a general idea of the safety effects of a countermeasure. To apply CMFs, it is necessary to know how many crashes are expected without the countermeasure. Specifically, the annual expected number of crashes without treatment is multiplied by the CMF to estimate the expected number of crashes with treatment. Estimating the expected crashes without treatment is not a trivial task; it is not simply the number of observed crashes before treatment, since this value could be higher or lower than expected due to **regression-to-the-mean**. Also, changes in traffic volume will cause changes in expected crashes. Users are referred to Section 3.2 and the Highway Safety Manual for a discussion of estimating the expected number of crashes without treatment (AASHTO, 2010).

EXAMPLE

*If the expected crashes without treatment equals 10.5 crashes per year and the CMF for installing larger STOP signs is 0.81 (Gan et al., 2005), the expected crashes after installing larger STOP signs = $0.81 * 10.5$ crashes per year = 8.5 crashes per year (a 19 percent reduction).*

It is important to note that a CMF represents the long-term expected change in crash frequency. Also, a CMF may be based on the crash experience at a limited number of study sites. As such, the actual change in crashes observed after treatment will vary by location and by year.

Application

A CMF may be applicable for all crash types and locations (e.g., all crashes for all area types) or only for a specific scenario (e.g., angle crashes at rural signalized intersections). The applicability of a CMF depends upon the underlying study from which the CMF was estimated. In general, the applicability of a CMF may vary by crash severity, crash type, and/or site condition. Each of these is further discussed with appropriate examples in this section.

When evaluating expected changes in crashes it is useful to determine the change in crashes by type and severity, but this should only be done when applicable CMFs are available. For some countermeasures, there may only be one CMF available, which is applied to all crash types and severities. In other cases, there may be multiple CMFs available. The selection of a suitable CMF will require some judgment, but in general the CMF should be selected that most closely matches the scenario at hand (i.e., specific crash type, severity, and site condition).

Crash Severity

A CMF should be selected based on the applicable severity. Crash severity is defined by the most severe outcome of those involved in the crash. While it may be desirable to estimate the change in crashes for a specific injury type (e.g., fatal or injury crashes), CMFs should be applied only to the severity types for which they were developed.

EXAMPLE

The CMF for installing cable median barrier is 1.34 for total crashes (a 34 percent increase) and 0.74 for injury crashes (a 26 percent reduction) (Elvik and Vaa, 2004). Since total crashes are expected to increase while injury crashes decrease this indicates that property damage only (PDO) crashes are expected to increase. It would not be appropriate to apply the CMF for injury crashes to PDO crashes because PDO crashes have been shown to increase after the installation of cable median barrier. Applying the CMF for total crashes to PDO crashes would also underestimate the expected increase in PDOs. When possible, it is beneficial to apply CMFs to specific severities because it produces more precise estimates for use in a benefit-cost analysis.

Crash Type

A CMF should be selected based on the applicable crash type. CMFs are often listed for total crashes and for specific crash types when available. Crash types differ for intersection- and segment-related crashes and may include total, left-turn, right-turn, right-angle, run-off-road, rear-end, sideswipe, head-on, fixed-object, animal, pedestrian, bicycle, and other. CMFs may indicate opposing effects for different crash types for the same treatment. For example, installation of a traffic signal can reduce angle crashes, but often increases rear-end crashes.

EXAMPLE

The CMFs in Table 2 were obtained from the CMF Clearinghouse and NCHRP Report 617. The CMFs illustrate various crash types, severities, and area types that may be assessed when considering the installation of a traffic signal. Note that the values indicate an increase in rear-end crashes and a decrease in angle crashes. It would not be appropriate to apply a CMF to crash types that are different from that associated with the CMF. For example, assume the CMF for total crashes is 0.78 and there is no CMF available for rear-end crashes; it is not appropriate to assume that rear-end crashes will be reduced by 22 percent just because total crashes are reduced. However, it is beneficial to apply CMFs to specific crash types when possible because it produces more precise estimates for use in a benefit-cost analysis.

TABLE 2. Example CMFs for Installing a Traffic Signal

CMF	Crash Type	Crash Severity	Area Type
0.78 ¹	All	All	All
0.85 ²	All	All	Rural
0.83 ²	All	All	Urban
0.62 ²	All	Fatal	All
1.15 ²	All	PDO	All
1.48 ²	Rear-end	All	All
0.71 ²	Angle	All	All
1.58 ³	Rear-end	All	Rural
0.23 ³	Angle	All	Rural
1.38 ⁴	Rear-end	Fatal/Injury	Urban
0.33 ⁴	Angle	Fatal/Injury	Urban

1. Gan et al., 2005
2. Pernia et al., 2002
3. Harkey et al., 2008
4. McGee et al., 2003

Site Condition

A CMF should be selected based on the applicable site condition. Site condition may be described by one or several variables, including area type, geometry, traffic control, traffic volume, functional classification, and/or jurisdiction. CMFs should not be applied in scenarios where the characteristics of the location of interest are different from those associated with the CMF.

EXAMPLE

The CMFs in Table 3 illustrate various types of traffic control, circulating lanes, severities, and area types that may be assessed when considering the conversion of an intersection to a roundabout (Rodegerts et al., 2007).

Note that a CMF is available for estimating the change in crashes when converting a two-way stop-controlled intersection to a roundabout in urban areas. However, there is not a specific CMF given for converting an all-way stop-controlled intersection to a roundabout in an urban area. Therefore, it would not be appropriate to apply the former CMF to the latter situation because the prior traffic control differs.

Similarly, there is a CMF for converting suburban signalized intersections to two-lane roundabouts, but no CMF for single lane roundabouts in this category. Again, it would not be appropriate to apply the former CMF to the latter situation because the number of circulating lanes differs.

TABLE 3. Example CMFs for Converting an Intersection to a Roundabout

Traffic Control Before Roundabout	Area Type	Circulating Lanes	CMF	
			All	Fatal + Injury
All Sites	All	All	0.65	0.24
Signalized	All	All	0.52	0.22
	Suburban	2	0.33	Sample too small
	Urban	All	Effects insignificant	0.40
All way stop	All	All	Effects insignificant	Effects Insignificant
Two way stop	All	All	0.56	0.18
	Rural	1	0.29	0.13
	Urban	All	0.71	0.19
		1	0.60	0.20
		2	Sample too small	Sample too small
	Suburban	All	0.68	0.29
		1	0.22	0.22
		2	0.81	0.32
	Urban/ Suburban	All	0.69	0.26
		1	0.44	0.22
2		0.82	0.28	

2.3 GENERAL ISSUES RELATED TO CMFs

The intent of this section is to identify basic issues related to the development and application of CMFs. Specific discussions include: 1) a word of caution regarding the application of multiple CMFs to a single location, 2) issues related to the use of CMFs derived from high crash locations, 3) considerations related to the use of before-after and cross-sectional data, and 4) an introduction to factors that can significantly affect the quality of CMFs.

Applying Multiple CMFs

Common practice is to multiply the CMFs to estimate the combined effect when multiple countermeasures are implemented at one location. Currently, there is limited research to support the combination of CMFs for this purpose. Although implementing several countermeasures is likely more effective than implementing a single countermeasure, it is unlikely that the full effect of each countermeasure would be realized when implemented concurrently. This is particularly true if the countermeasures target the same crash type (e.g., installing lighting and enhancing pavement markings to address nighttime crashes). Therefore, unless the countermeasures act completely independently and target unique crash types, multiplying several CMFs is likely to overestimate the combined effect. The likelihood of overestimation increases with the number of CMFs that are multiplied. Therefore, caution and engineering judgment should be exercised when estimating the combined effect of multiple countermeasures at a given location. Ideally, a CMF for a combination treatment should be derived directly from a rigorous before-after evaluation of sites where the combination treatment was applied.

CMFs Derived From High Crash Locations

Caution should be used when applying CMFs to a site with an average crash history, if the CMF was derived from applications of the countermeasure at sites with high frequencies of crashes that were correctable by the countermeasure. In such cases, the CMF may over-estimate the effectiveness of the countermeasure at sites with an average crash history. A user can determine the background of a CMF by reviewing the study from which it was developed.

EXAMPLE

The CMF for total crashes is 0.764 for the application of skid treatments. This CMF was derived from an evaluation of skid treatments targeted at road segments with a high frequency of wet weather crashes and low skid numbers. It should not be expected that the same CMF will apply for resurfacing any road segment. In fact, there is evidence to suggest that resurfacing can increase crashes at some locations (Lyon and Persaud, 2008).

Considerations Related to Before-After and Cross-Sectional Designs

Several specific study designs are discussed in Chapter 3 along with their associated strengths and weaknesses. The data used in these study designs can typically be classified as either before-after or cross-sectional.

Before-after designs include a treatment at some period in time and a comparison of the safety performance *before* and *after* treatment for a site or group of sites. **Cross-sectional** designs compare the safety performance of a site or group of sites *with* the treatment of interest to similar sites *without* the treatment at a single point in time. Both before-after and cross-sectional study designs have issues that need to be considered in the development and application of CMFs as outlined below and discussed in more detail in Chapter 3.

Before-After Designs

CMFs derived from before-after data are based on the change in safety performance due to the implementation of some treatment. There are two fundamental issues with deriving quality CMFs from before-after designs.

1. **Sample Size:** The required sample size depends on the magnitude of the treatment effect and the uncertainty of the estimate (i.e., the standard error). Generally, the standard error decreases as the sample size increases. As such, one can reduce the uncertainty of an estimate by increasing the sample size.
2. **Potential Bias:** The observed change in crash experience at treated sites between the periods before and after treatment may be due not only to the countermeasure, but to other factors as well. Other factors include:
 - a. Traffic volume changes.
 - b. Changes in reported crash experience.
 - c. Regression-to-the-mean.

Simple before-after comparisons, also known as naïve before-after studies, do not account for these changes. As a result, CMFs derived from such studies are usually considered unreliable and rated as being of poor quality.

Cross-Sectional Designs

CMFs derived from cross-sectional data are based on a single time period under the assumption that the ratio of average crash frequencies for sites with and without a feature is an estimate of the CMF for implementing that feature. Where there are sufficient applications of a specific countermeasure, the before-after design is preferred. Cross-sectional designs are particularly useful for estimating CMFs where there are insufficient instances where a countermeasure is actually applied. For example, there may be few or no projects where the shoulder is widened from four feet to six feet, yet there are many road segments with a shoulder width of four feet and many with a shoulder width of six feet. In this case, crash data could be collected for the two groups of segments for use in a cross-sectional design, but a before-after design would be less feasible because there are too few actual projects that widen the shoulder from four feet to six feet.

Given the substantive issues associated with cross-sectional and before-after designs, it is not surprising that CMFs derived from different evaluations tend to be highly inconsistent, making the selection of a CMF for application quite challenging. Considering these issues and the difficulty of addressing them in practice, CMFs from cross-sectional designs tend to indicate smaller expected crash reductions than those derived from before-after studies. It is important to recognize the strengths and issues associated with the various methods for developing CMFs. Awareness of these differences will help researchers identify a suitable method for developing a CMF, given the constraints of their specific evaluation.

Factors Affecting the Quality of CMFs

The quality of a CMF is related to the study design from which the CMF was derived and other factors such as sample size, robustness of data, standard error, and the accounting for potential sources of bias. The following discussion summarizes several key points that directly affect the quality of a study and the resulting CMF estimate. As a study addresses more key points, the quality will inherently improve. Determining whether a study is unreliable is challenging because it requires a certain amount of statistical knowledge and experience related to the study methodologies. Studies may be flawed for various reasons, including:

- Not properly accounting for regression-to-the-mean in a before-after study.
- Failure to separate the safety effects of the treatment from the effects of other changes (e.g., traffic volumes, other countermeasures, and crash reporting practices).
- Use of comparison groups that are unsuitable.
- Inappropriate functional form or improper model specification for regression models used in before-after or cross-sectional studies.
- Incorrect interpretation of the accuracy of estimates or presentation of results without statements of accuracy.
- Incorrect interpretation of the results of cross-sectional designs where differences between two groups may be due to factors other than the measure of interest.
- Selective citing of results (the tendency to ignore negative aspects of results such as declining effects over time or increases in non-target crashes).

Key questions to be asked in assessing study quality have been documented by Elvik (2002). The following are relevant questions that one might ask to assess the quality of a study:

- How were units sampled for the study?
- Do the data collected in the study refer directly to the outcome of interest or to aggregated data?
 - Was crash or injury severity specified?
 - Were study results tested for statistical significance or their statistical uncertainty otherwise estimated?
 - Did the study use appropriate techniques for statistical analysis?
 - Can the causal direction between treatment and effect be determined?
 - How well did the study control for **confounding factors**?
 - Did the study have a clearly defined target group, and were effects found in the target group only?
 - Are study results explicable in terms of well-established theory?

A confounding factor is a variable that completely or partially accounts for the apparent association between an outcome and a treatment.

Two recent efforts amalgamating knowledge of CMFs and evaluating their quality are the development of the Highway Safety Manual and the CMF Clearinghouse. These efforts and their approaches to evaluating CMF quality are discussed in Section 4.3.

2.4 SUMMARY

This chapter provided an overview of CMFs and CMFunctions, including issues related to their application and development and a discussion of several factors that affect the quality of CMFs. Chapter 3 builds upon this knowledge by discussing in further detail the various study methodologies available for developing CMFs. This discussion includes an overview of the methods, sample size considerations, and strengths and weaknesses.

3. STUDY DESIGNS TO DEVELOP CMFS

The intended audience for this guide is transportation safety practitioners, consultants, and researchers with experience and/or education in the theory and practice of road safety engineering, including basic analytical procedures and statistical concepts. The objective of this chapter is to identify and provide an overview of several methods for developing CMFs. Specifically, this chapter will help users understand the strengths and issues associated with the various methods, but it is not meant to be a step-by-step guide to applying the methods.

Study designs can fall into one of two general study types. These two types essentially differ in how the data are collected. **Experimental studies** are *planned*, meaning sites that are identified for some treatment are then randomly assigned to either a treatment or to a control group that is left untreated. The groups are identified before implementation of the treatment. **Observational studies** are *not planned*, meaning that data are collected retrospectively by observing the performance of an existing road system, where the treatment has already been implemented at some sites, usually not on the basis of a planned experiment, but on engineering considerations, including safety.

In experimental studies, differences in crash experience between the treatment and control groups, in a period after the treatment, are then attributed to the treatment. This is referred to as an experimental cross-section study. For the equivalent observational study, as noted previously in *Considerations Related to Before-After and Cross-Sectional Designs*, sites with and without the treatment of interest are identified retrospectively rather than in a randomized experimental setting.

For both experimental and observational studies, a **before-after** design is usually preferred to a **cross-sectional design**. For the before-after design, the CMF is estimated from the change in crash frequency between the periods before and after the implementation of a treatment. There is, however, a need to account for changes in safety due to factors other than the treatment of interest. In an experimental study, the planned control group is used for this purpose. However, observational studies are more common in road safety research in view of the ethical concerns with experimentation in road safety. Thus observational study designs are the focus of the rest of this chapter.

In an observational study, an untreated group can be identified retrospectively, and used to account for changes in safety due to factors other than the treatment of interest. There are several types of observational before-after studies, which vary in the use of an untreated group to account for the confounding factors. The naïve before-after study involving a simple before-after comparison of crash counts, without accounting for changes unrelated to a treatment, is not considered a reliable method, as noted in *Considerations Related to Before-After and Cross-Sectional Designs*. Two methods, one simple and one somewhat more complex, are preferred to derive a CMF from a before-after study. The comparison group method is the simpler of the two, while the empirical Bayes method is more complex, but is also more robust. Basic information on the two methods, including variations, is presented below. Of late, a more complex approach, the full Bayes method, has also been proposed by some researchers. Basic information on this method follows the discussion of the two more common methods.

While rigorous before-after methods are usually preferred to cross-sectional methods, there are situations that call for an alternative approach because before-after methods are not practical (i.e., when there are insufficient before-after situations to allow for credible results). Several alternative methods are presented for developing CMFs, including cross-sectional, case-control, cohort, meta-analysis, expert panel, and surrogate measure studies. Each of these methods is presented below with a discussion to help agencies identify when each method might be appropriate. For each method, the guide provides an overview of the method, a more detailed description of the method with supporting examples, a discussion of sample size considerations, and a summary of issues to consider when applying the method. While much of the information presented in this chapter is statistical in nature, the overview outlines each method in a more simplistic nature.

3.1 BEFORE-AFTER WITH COMPARISON GROUP STUDIES

Overview

A before-after with comparison group study uses an untreated **comparison group** of sites similar to the treated ones to account for changes in crashes unrelated to the treatment such as time and traffic volume trends. The comparison group is used to calculate the ratio of observed crash frequency in the after period to that in the before period. The observed crash frequency in the before period at a treatment site group is multiplied by this comparison ratio to provide an estimate of expected crashes at the treatment group had no treatment been applied. This is then compared to the observed crashes in the after period at the treatment site group to estimate the safety effects of the treatment.

A comparison group is a group of sites, which is similar to the treated ones and used to account for changes in crashes unrelated to the treatment such as time and traffic volume trends.

Spill-over occurs when a treatment is implemented at a specific location and the effects of that treatment are observed at nearby untreated locations.

Ideally, the comparison group should be drawn from the same jurisdiction as the treatment group and be similar to the treatment group in terms of geometric and operational characteristics. The difficulty is that the pool available for the comparison group could be too small if most or all sites are treated or at least affected by the treatment. The former may be the case when local policy results in a blanket treatment. The latter may be the case for treatments such as red light cameras which are believed to have significant **spill-over** effects to untreated sites.

This method will not account for regression-to-the-mean unless treatment and comparison sites are also matched on the basis of the observed crash frequency in the before period. Specifically, a control site would need to be matched to each treated site based on the annual crashes in the before period. There are immense practical difficulties in achieving an ideal comparison group to account for regression-to-the-mean (i.e., matching on the basis of crash occurrence) as illustrated in Pendleton (1996). In addition, the necessary assumption that the comparison group is unaffected by the treatment is difficult to test and can be an unreasonable assumption in some situations.

Where there is no regression-to-the-mean and where a suitable comparison group is available, the comparison group methodology can be a simple alternative to the more complex empirical Bayes approach. This may be true in cases where 1) crash frequency is not considered in selecting a site for safety treatment, 2) the safety evaluation is strictly related to a change implemented for operational reasons, or 3) a blanket treatment is applied to all sites of a given type. In practice, except for blanket treatments, it is difficult to ascertain that there is no regression-to-the-mean and only a truly random selection of sites for treatment will ensure that there is no selection bias.

Method

A suitable comparison group is one where the ratios of expected crash counts in the after period to the expected crash counts in the before period are equal for the comparison group and the treatment group, had no treatment been applied. For example, if the expected crash count at treated sites were to increase by 10 percent in the after period without treatment then a perfect comparison group should also show this expected increase of 10 percent. Naturally, it is difficult to achieve a perfect comparison group, since the change in crashes at the treatment sites *without* treatment cannot be known (since there is treatment).

The suitability of a comparison group can be determined by performing a test of comparability for the treatment group and potential comparison groups that is outlined in Hauer (1997). The test of comparability compares a time series of target crashes for a treatment group and a candidate comparison group during a period before the treatment is implemented. If the annual trend in crash frequencies is similar to that of the treatment group

(in the absence of treatment), then a candidate comparison group is a good one. Figure 2 illustrates the idea of similarity and suitability of a comparison group. In this example, crashes in the treatment group are subject to some treatment after the year 2000 and crashes in the comparison group are compared for suitability. It seems evident from visual inspection of Figure 2 that crashes in the two groups track each other well in the period before treatment. Of course this is only a qualitative evaluation.

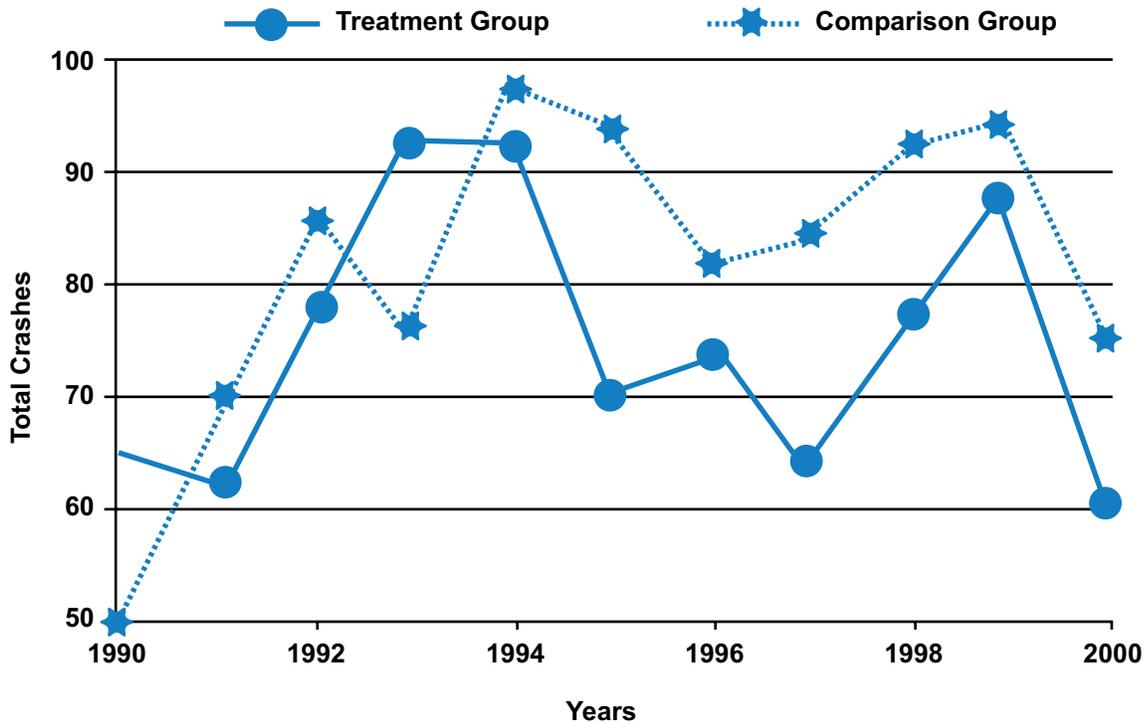


FIGURE 2. Example Time Series Plot of Crashes in Treatment and Comparison Group.

Hauer (1997) proposes the use of a sequence of sample odds ratios to quantitatively assess the suitability of a candidate comparison group. Equation 4 is used to calculate the sample odds ratio. The sample odds ratios are computed for each before-after pair in the time series before the treatment is implemented. From this sequence of sample odds ratios, the sample mean and standard error are determined. If this sample mean is sufficiently close to 1.0 (i.e., subjectively close to 1.0 and the confidence interval includes the value of 1.0) then the candidate reference group is deemed suitable.

$$\text{sample odds ratio} = \frac{(Treatment_{before} Comparison_{after}) / (Treatment_{after} Comparison_{before})}{1 + \frac{1}{Treatment_{after}} + \frac{1}{Comparison_{before}}} \tag{4}$$

Where,

Treatment_{before} = total crashes for the treatment group in year i.

Treatment_{after} = total crashes for the treatment group in year j.

Comparison_{before} = total crashes for the comparison group in year i.

Comparison_{after} = total crashes for the comparison group in year j.

EXAMPLE

Consider the data in Table 4, which represent the time series of crashes in the treatment and comparison groups, in the period before treatment.

TABLE 4. Example Time Series of Crash Data for Treatment and Comparison Group

Group	Year 1	Year 2	Year 3	Year 4
Treatment	100	90	105	110
Comparison	95	98	110	105

The first sample odds ratio is calculated using years 1 and 2 as follows:

$$\text{sample odds ratio} = \frac{(100 * 98) / (90 * 95)}{1 + \frac{1}{90} + \frac{1}{95}} = 1.12$$

The sample odds ratios are similarly computed for each subsequent before-after pair (year 2-3 and year 3-4). In this illustration the odds ratios are 1.12, 0.94 and 0.89, for years 1-2, 2-3, and 3-4, respectively. The mean and standard error of the sample odds ratios are 0.99 and 0.12, respectively.

$$95\% \text{ Confidence Interval} = 0.99 \pm 1.96(0.12) = 0.75 \text{ to } 1.23$$

It can be concluded that the sample mean odds ratio is sufficiently close to 1.0, partly because 0.99 is subjectively very close to 1.0 and also because the standard error is such that even at low levels of confidence the confidence intervals would include the value 1.0. Thus, it can be concluded that the comparison group is a good one.

Additional requirements of a suitable comparison group, as outlined by Hauer (1997), include:

1. The before and after periods for the treatment and comparison group should be the same.
2. There should be reason to believe that the change in factors other than the treatment under study (e.g., traffic volume changes), which influence safety are the same in the treatment and comparison groups.
3. The crash counts must be sufficiently large. (This point is discussed in more detail later.)

The CMF for a given crash type at a treated site is estimated by first summing the observed crashes for both the treatment and comparison groups for the two time periods (assumed equal). The notation for these summations is summarized in Table 5.

TABLE 5. Summary of Notation for Comparison Group Method

Time Period	Treatment Group	Comparison Group
Before	$N_{\text{observed,T,B}}$	$N_{\text{observed,C,B}}$
After	$N_{\text{observed,T,A}}$	$N_{\text{observed,C,A}}$

Where,

$N_{\text{observed},T,B}$ = the observed number of crashes in the before period for the treatment group.

$N_{\text{observed},T,A}$ = the observed number of crashes in the after period for the treatment group.

$N_{\text{observed},C,B}$ = the observed number of crashes in the before period in the comparison group.

$N_{\text{observed},C,A}$ = the observed number of crashes in the after period in the comparison group.

The comparison ratio ($N_{\text{observed},C,A} / N_{\text{observed},C,B}$) indicates how crash counts are expected to change in the absence of treatment (i.e., due to factors other than the treatment of interest). This is estimated from the comparison group as the number of crashes in the after period divided by the number of crashes in the before period. The expected number of crashes for the treatment group that would have occurred in the after period without treatment ($N_{\text{expected},T,A}$) is estimated from Equation 5.

$$N_{\text{expected},T,A} = N_{\text{observed},T,B} (N_{\text{observed},C,A} / N_{\text{observed},C,B}) \quad (5)$$

If the comparison group is suitable, that is, if the crash trends in that group and the treatment group are similar as determined by the test for comparability, and as is evident from time series plots such as that illustrated in Figure 2, the variance of $N_{\text{expected},T,A}$ is estimated approximately from Equation 6.

$$\text{Var}(N_{\text{expected},T,A}) = N_{\text{expected},T,A}^2 (1/N_{\text{observed},T,B} + 1/N_{\text{observed},C,B} + 1/N_{\text{observed},C,A}) \quad (6)$$

This estimate is only an approximation since it applies to an *ideal* comparison group with yearly trends *identical* to the treatment group, a situation that is practically impossible. A more precise estimate can be obtained by applying a modification, which is typically minor, as derived in Hauer (1997). Estimating this modification is not trivial, so it is recommended to estimate the variance assuming an *ideal* comparison group and recognize that this estimate is a conservatively low approximation. In the *ideal* case, the CMF and its variance are estimated from Equation 7 and 8.

$$\text{CMF} = (N_{\text{observed},T,A} / N_{\text{expected},T,A}) / (1 + (\text{Var}(N_{\text{expected},T,A}) / N_{\text{expected},T,A}^2)) \quad (7)$$

$$\text{Variance}(\text{CMF}) = \frac{\text{CMF}^2 [(1/N_{\text{observed},T,A}) + (\text{Var}(N_{\text{expected},T,A})/N_{\text{expected},T,A}^2)]}{[1 + \text{Var}(N_{\text{expected},T,A})/N_{\text{expected},T,A}^2]^2} \quad (8)$$

EXAMPLE

Table 6 below provides crash counts for an example CMF calculation using the comparison group method, with 25 treatment sites and an equal number of 25 comparison sites. For this illustration the comparison group is assumed to be ideal. The assumption of an ideal comparison group is made to simplify the computation of the variance of the CMF, recognizing that this will result in a conservative approximation.

TABLE 6. Example Data for Before-After with Comparison Group Study

Time Period	Treatment Group (25 sites)	Comparison Group (25 sites)
Before	100	84
After	75	80

The comparison ratio in the example is estimated as:

$$\text{Comparison Ratio} = 80/84 = 0.9524$$

The expected number of crashes in the after period in the treatment group that would have occurred without treatment (denoted by $N_{\text{expected},T,A}$) is estimated as the number of crashes in the before period times the comparison ratio:

$$N_{\text{expected},T,A} = 100 \times 0.9524 = 95.24$$

The variance of $N_{\text{expected},T,A}$, which is a measure of its precision and used to derive the CMF and estimate its variance, is estimated as:

$$\text{Var}(N_{\text{expected},T,A}) = 95.24^2(1/100+1/84+1/80) = 312.06$$

The CMF is approximately equal to the after period crash count for the treatment group divided by the expected number without treatment ($N_{\text{expected},T,A}$). It is only approximate because there is a small adjustment based on the value of $N_{\text{expected},T,A}$ and its variance.

$$\text{CMF} = (75/95.24)/(1+(312.06/95.24^2)) = 0.761$$

$$\text{Var}(\text{CMF}) = (0.761^2((1/75)+(312.06/95.24^2)))/(1+312.06/95.24^2)^2 = 0.0258.$$

Taking the square root of the variance, the standard error of the CMF is 0.168.

The 95% confidence interval is $0.761 \pm 1.96 \times 0.168 = 0.432$ to 1.090 .

Exploration of the numbers and results in this example allows several key points to be made.

1. Regression-to-the-mean is likely since:
 - a. Treatment sites tend to be selected because they experienced an unusually high count in the before period.
 - b. One could not reject the hypothesis that this count was randomly high; this is because it is higher than the count in the same period in a comparison group that was equal in size and data.
2. If the comparison group had the same number of crashes in the before period and was similar to the treatment group in all other respects, then either there is no regression-to-the-mean or the method will automatically account for this effect. The result, in either case, would be an unbiased estimate of the safety effect, with respect to regression-to-the-mean.
3. The CMF estimate of 0.761 is not significant at the 95 percent confidence level that many engineers might require as the standard. In this case, the 95 percent confidence interval is 0.432 to 1.090. It cannot be stated with 95 percent confidence that the true value of the CMF is not 1.0 (i.e., it cannot be stated with 95 percent confidence that the treatment had any effect). A larger sample size in the treatment and/or comparison group would have yielded the required confidence in the result, providing the CMF estimate did not become substantially closer to 1.0 after adding the additional data. How sample size affects the degree of certainty in CMF estimates is discussed next.

Sample Size for Comparison Group Studies

When planning a comparison group before-after safety evaluation, it is vital to ensure that enough crashes are included such that the expected change in safety can be statistically detected. Recall that a statistically significant CMF means that one can say with a given level of significance that the confidence interval for the CMF does not include 1.0.

The four variables that impact whether or not a sample is sufficiently large are:

1. The size of the treatment group, in terms of the number of crashes in the before period.
2. The relative duration of the before and after periods.
3. The likely (postulated) CMF value.
4. The size of the comparison group in terms of the number of crashes in the before and after periods.

It is challenging to assess the adequacy of a sample before collecting data because it is necessary to estimate the number of crashes in the sample that is yet to be collected and develop an intelligent guess about the magnitude of the CMF. These variables impact the precision (standard error) with which the CMF is estimated. For a detailed explanation of sample size considerations, as well as estimation methods, see Chapter 9 of Hauer (1997). In that source, a spreadsheet layout is provided for exploring the interaction of sample size related variables. To gain an appreciation for the impacts of these variables, consider the example calculation above and the discussion below.

Impact of treatment group size

If, in the previous example, the treatment sample were tripled to 300 and 225 crashes in the before and after periods, the new calculations show that the expected number of crashes in the after period in the treatment group that would have occurred without treatment, $N_{\text{expected},T,A}$, is 285.72. Note that the treatment sample could be increased by including more sites or more years of data. The comparison ratio remains the same ($80/84 = 0.9524$).

$$N_{\text{expected},T,A} = 300 \times 0.9524 = 285.72$$

$$\text{Var}(N_{\text{expected},T,A}) = 285.72^2(1/300+1/84+1/80) = 2264.42$$

$$\text{CMF} = (225/285.72)/(1+(2264.42/285.72^2)) = 0.766$$

$$\text{Variance} = (0.766^2((1/225)+(2264.42/285.72^2)))/(1+2264.42/285.72^2)^2 = 0.018.$$

Taking the square root of the variance, the standard error of the CMF is 0.134. Still this is not significant at the 95 percent confidence level since the 95 percent confidence interval for the CMF includes 1.0 [95 percent confidence interval = $0.766 \pm 1.96(0.134) = 0.503$ to 1.028]. However, this marginally insignificant result may still be acceptable since it is significant at the 90 percent confidence level. This is because, for a 90 percent confidence interval, the multiplier for the standard error is 1.64 instead of 1.96. Thus, the 90 percent confidence interval is $0.766 \pm 1.64(0.134) = 0.546$ to 0.986 , which is less than 1.0. That is, one is at least 90 percent certain that there is a decrease in crashes resulting from the countermeasure. A larger sample would be required to detect the same level of effect with 95 percent certainty.

A yoked comparison group study is a special case of the comparison group study where a single comparison site is matched to each treatment site based on similar geometric and traffic volume conditions. The strengths and weaknesses of a yoked comparison group study are similar to those of a general comparison group study with a couple of exceptions. The primary benefit of the yoked comparison group, in relation to the general comparison group, is that it does not require as much data (i.e., fewer comparison sites). This is also, however, a weakness of the yoked comparison group. While the investigator may be able to better match treatment and comparison sites in a yoked comparison, this one-to-one matching is unnecessarily restrictive.

Impact of comparison group size

If the comparison group was doubled, in addition to tripling the treatment group sample, the estimated CMF is 0.775 with a standard error of 0.108.

$$\text{Comparison Ratio} = 160/168 = 0.9524$$

$$N_{\text{expected},T,A} = 300 \times 0.9524 = 285.72$$

$$\text{Var}(N_{\text{expected},T,A}) = 285.72^2(1/300+1/168+1/160) = 1268.27$$

$$\text{CMF} = (225/285.72)/(1+(1268.27/285.72^2)) = 0.775$$

$$\text{Variance} = (0.775^2((1/225)+(1268.27/285.72^2)))/(1+1268.27/285.72^2)^2 = 0.011.$$

$$\text{Standard error} = 0.011^{0.5} = 0.108.$$

It can be seen that this result is now significant at the 95 percent confidence level. The message is that if the treatment sample is limited and increasing its size is not an option, then obtaining more comparison sites, which is often a feasible option, will increase confidence in the results. This suggests that one-to-one matching of treatment and comparison sites, as is done in **yoked comparison group** studies, is unnecessarily restrictive.

Impact of size of treatment effect

Suppose in the original example, 69 crashes were recorded in the after period in the treatment group instead of 75, indicating that the CMF is smaller (the crash reduction is larger). The CMF is then estimated to be 0.700 with an upper 95 percent confidence limit of 0.994, which indicates a significant reduction in crashes. Thus, the smaller sample in both the treatment and comparison groups would have been sufficient if the estimated effect was larger.

Issues with Comparison Group Studies

Refer to Section 3.2 for a discussion of issues related to both the comparison group and empirical Bayes before-after study design. It is more convenient to explain the drawbacks of the comparison group method in relation to the empirical Bayes method. As such, it is necessary to present the fundamental principles of the empirical Bayes method before discussing issues related to each.

3.2 EMPIRICAL BAYES BEFORE-AFTER STUDIES

Overview

The objective of the empirical Bayes methodology is to more precisely estimate the number of crashes (denoted as $N_{\text{expected},T,A}$ in the comparison group method) that would have occurred at an individual treated site in the after period had a treatment not been implemented. Similar to the comparison group method, the effect of the safety treatment is estimated by comparing the sum of the estimates of $N_{\text{expected},T,A}$ for all treated sites with the number of crashes actually recorded after treatment.

The advantage of the empirical Bayes approach is that it correctly accounts for observed changes in crash frequencies before and after a treatment that may be due to regression-to-the-mean. In doing so, it also facilitates a better approach than the comparison group method for accounting for changes in safety due to traffic volumes and time trends.

Method

In accounting for regression-to-the-mean, the number of crashes expected in the before period without the treatment ($N_{\text{expected},T,B}$) is a weighted average of information from two sources:

1. The number of crashes observed in the before period at the treated sites ($N_{\text{observed},T,B}$).
2. The number of crashes predicted at the treated sites based on reference sites with similar traffic and physical characteristics ($N_{\text{predicted},T,B}$).

A safety performance function (SPF) is a mathematical equation used to predict the crash experience for a given site based on its traffic and physical characteristics.

To estimate the weights and the number of crashes expected on sites with similar traffic and physical characteristics, a reference group of sites similar to the treated sites is used. This is similar in principle to the use of a comparison group in the comparison group method. However, the point of departure is that data from the untreated “reference” group are used to first estimate a **safety performance function (SPF)** that relates crash experience of the sites to their traffic and physical characteristics. An SPF is a mathematical model that predicts the mean crash frequency for similar locations with the same characteristics. These characteristics typically include traffic volume and may include other variables such as traffic control and geometric characteristics. This SPF is then used to derive the second source of information for the empirical Bayes estimation – the number of crashes predicted at treated sites based on sites with similar operational and geometric characteristics ($N_{\text{predicted},T,B}$).

EXAMPLE

Equations 9 and 10 are examples of SPFs for road segments and intersections.

For road segments:

$$\text{Crashes per year} = \alpha(\text{segment length})(\text{AADT})^\beta \quad (9)$$

For intersections:

$$\text{Crashes per year} = \alpha(\text{Major road entering AADT})^{\beta_1}(\text{Minor road entering AADT})^{\beta_2} \quad (10)$$

Where the AADTs are traffic volumes and α , β , β_1 , and β_2 are numbers estimated during the SPF development.

The empirical Bayes estimate of the expected number of crashes without treatment, $N_{\text{expected},T,B}$, is computed from Equation 11.

$$N_{\text{expected},T,B} = \text{SPF weight}(N_{\text{predicted},T,B}) + (1-\text{SPF weight})(N_{\text{observed},T,B}) \quad (11)$$

The SPF weight is derived using what is called the over-dispersion parameter from the SPF calibration process, but also depends on the number of years of crash data in the period before treatment. There is an inverse relationship between the SPF weight and the over-dispersion parameter. The SPF weight is such that an SPF which explains more of the between-site variability in crash counts, and having a lower over-dispersion parameter, will have a higher weight. Specifically, if the SPF has little over-dispersion, more

weight is placed on the crashes predicted from the SPF ($N_{\text{predicted},T,B}$) and less weight on the observed crash frequency ($N_{\text{observed},T,B}$). However, the weight is reduced if many years of crash data are used.

Figure 3 illustrates how the SPF estimate is weighted with the observed crash count to estimate $N_{\text{expected},T,B}$. As shown in Figure 3, the empirical Bayes estimate falls somewhere between the values from the two information sources ($N_{\text{observed},T,B}$ and $N_{\text{predicted},T,B}$). The regression-to-the-mean effect is the difference between $N_{\text{observed},T,B}$ and $N_{\text{expected},T,B}$.

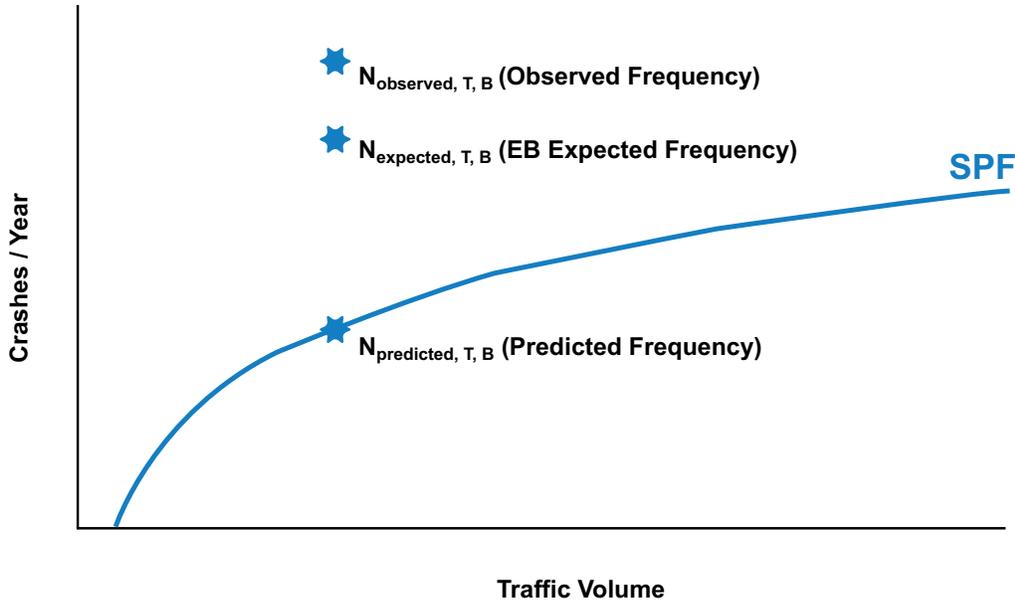


FIGURE 3. Illustration of Regression-to-the-Mean and Empirical Bayes Estimate.

The SPF is not only used to account for regression-to-the-mean, but also to better account for time trends and traffic volume changes compared to the comparison group method. As shown in Figure 3, the SPF allows an estimation of the change in safety that would occur as a result of a change in traffic volume. SPFs can also be calibrated to each year and these calibration factors (multipliers) reflect time trends in the relationship between crash frequency and traffic volume. The same reference group used to develop the SPF is applied to derive these time trend multipliers. For a given year, the multiplier is calculated as the sum of observed crashes divided by the sum of predicted crashes in that year.

The adjusted value of the empirical Bayes estimate, $N_{\text{expected},T,A}$, is the expected number of crashes in the after period without treatment and is calculated from Equation 12.

$$N_{\text{expected},T,A} = N_{\text{expected},T,B} (N_{\text{predicted},T,A} / N_{\text{predicted},T,B}) \quad (12)$$

Where,

$N_{\text{expected},T,B}$ = the unadjusted empirical Bayes estimate

$N_{\text{predicted},T,B}$ = the *predicted* number of crashes estimated by the SPF in the before period

$N_{\text{predicted},T,A}$ = the *predicted* number of crashes estimated by the SPF in the after period

The variance of $N_{\text{expected},T,A}$ is estimated from $N_{\text{expected},T,A}$, the before and after SPF estimates and the SPF weight, from Equation 13.

$$\text{Var}(N_{\text{expected},T,A}) = N_{\text{expected},T,A} (N_{\text{predicted},T,A} / N_{\text{predicted},T,B})(1 - \text{SPF weight}) \quad (13)$$

Recall the comparison group method introduced earlier. To estimate the CMF, the *observed* crashes were summarized for the treatment and comparison groups for both the before and after periods. In the empirical Bayes method, the *predicted* crashes are used in the computation of the CMF. Specifically, the column corresponding to the comparison group, now called a *reference* group, contains the sum of the SPF predictions for the before and after periods. Using the notation from the comparison group method, the corresponding parameters for the empirical Bayes method are shown in Table 7.

TABLE 7. Summary of Notation for Empirical Bayes Method

Time Period	Treatment Group	SPF Prediction (SPFs Developed from Reference Group)
Before	$N_{\text{observed},T,B}$	$N_{\text{predicted},T,B}$
After	$N_{\text{observed},T,A}$	$N_{\text{predicted},T,A}$

Where,

$N_{\text{observed},T,B}$ = the observed number of crashes in the before period for the treatment group.

$N_{\text{observed},T,A}$ = the observed number of crashes in the after period for the treatment group.

$N_{\text{predicted},T,B}$ = the *predicted* number of crashes (i.e., sum of the SPF estimates) in the before period.

$N_{\text{predicted},T,A}$ = the *predicted* number of crashes (i.e., sum of the SPF estimates) in the after period.

As demonstrated in the following example, these parameters are used in an identical manner to the equivalent numbers for the comparison group method to estimate the CMF.

EXAMPLE

Table 8 presents information to support calculations using the empirical Bayes method. For this simplified example, a weight of 0.25 is assumed for the SPF prediction for all sites and there are no traffic volume changes at the treated sites. With these assumptions the calculations may be applied in one step for all sites together. Normally, the calculations of $N_{\text{expected},T,A}$ and $\text{Var}(N_{\text{expected},T,A})$ would be computed for each site individually and then summed to use in the estimation of the CMF and its standard error.

TABLE 8. Example Data for Empirical Bayes Before-After Study

Time Period	Treatment Group Observed Crashes (25 sites)	SPF Estimates for Treatment Group (SPFs Developed from Reference Group)
Before	100	Sum for 25 sites = 81.08
After	75	Sum for 25 sites = 77.36

The empirical Bayes estimate, $N_{\text{expected},T,B}$, is calculated as:

$$N_{\text{expected},T,B} = 0.25 \times 81.08 + (1 - 0.25) \times 100 = 95.27$$

The ratio of after period SPF estimates to before period SPF estimates is now:

$$N_{\text{predicted},T,A} / N_{\text{predicted},T,B} = 77.36 / 81.08 = 0.954$$

The expected number of crashes in the after period in the treatment group that would have occurred without treatment ($N_{\text{expected},T,A}$) is:

$$N_{\text{expected},T,A} = 95.27 * 0.954 = 90.90$$

The variance of $N_{\text{expected},T,A}$ is estimated as:

$$\text{Var}(N_{\text{expected},T,A}) = 90.90 * 0.954 * (1 - 0.25) = 65.05$$

The CMF is approximately equal to the after period crash count divided by the expected number without treatment ($N_{\text{expected},T,A}$). It is only approximate because there is a small adjustment based on the value of $N_{\text{expected},T,A}$ and its variance.

$$\text{CMF} = (75/90.90) / (1 + (65.05/90.90^2)) = 0.819$$

$$\text{Variance} = (0.819^2 * ((1/75) + (65.05/90.90^2))) / (1 + 65.05/90.90^2)^2 = 0.0140.$$

Taking the square root of the variance, the standard error of the CMF is 0.118.

The 95% confidence interval is $0.819 \pm 1.96 * 0.118 = 0.588$ to 1.050 .

As shown in the example above, the estimate of the CMF using the empirical Bayes method is 0.819 with a standard error of 0.118. The CMF estimate is now larger (i.e., the estimated crash reduction is smaller) than for the example for the comparison group study which had a value of 0.761, mainly because the regression-to-the-mean effect has been taken into account. Note also that even though the CMF is larger, its standard error is smaller than the value of 0.168 for the comparison group method. A key feature of the empirical Bayes method is that it reduces uncertainty in CMF estimates because it uses more information and a more rigorous methodology. However, as before, the CMF is not significant at the 95 percent confidence level because the 95 percent confidence interval is 0.588 to 1.050 ($0.819 \pm 1.96(0.118)$).

In the example above, it was assumed that the traffic volume remained constant from the before to the after period. Had there been a traffic volume change, it would be necessary to incorporate this information in the analysis. However, this does not change the general process for estimating the CMF and standard error. The change in traffic volume is accounted for by the SPF (i.e., the predicted crashes in the before and after period, $N_{\text{predicted},T,B}$ and $N_{\text{predicted},T,A}$, respectively). This would only affect the ratio of predicted crashes after to predicted crashes before ($N_{\text{predicted},T,A} / N_{\text{predicted},T,B}$). The predicted crashes ($N_{\text{predicted},T,B}$ and $N_{\text{predicted},T,A}$) and the calculations of $N_{\text{expected},T,A}$ and $\text{Var}(N_{\text{expected},T,A})$ would be computed for each site individually and then summed to use in the calculation of the CMF and its standard error.

Sample Size for Empirical Bayes Studies

When planning an empirical Bayes before-after safety evaluation, it is vital to ensure that enough data are included such that the expected change in safety can be statistically detected. Currently, there is no formal method for determining required sample sizes for the empirical Bayes before-after approach. The method presented in Hauer (1997) pertains to the comparison group method and can be used to approximate the sample size required for an empirical Bayes study. The sample size estimates could be considered conservative in that the empirical Bayes approach reduces uncertainty in the estimate of expected crashes.

Nevertheless, it is informative to explore how a larger sample of treated sites in the above example would have affected the results. Specifically, how large a sample would be needed to yield a statistically significant result? If, in the previous example, the treatment sample were doubled to 200 and 150 crashes in the before

and after periods, the new calculations show a CMF of 0.822, instead of 0.819, with a standard error reduced to 0.084 from 0.118. Note that for these calculations, the SPF estimate for the treatment sites would also double to 162.16 and 154.72 crashes in the before and after periods, respectively.

$$N_{\text{expected},T,B} = 0.25*162.16 + (1-0.25)*200 = 190.54$$

$$N_{\text{predicted},T,A} / N_{\text{predicted},T,B} = 154.72/162.16 = 0.954$$

$$N_{\text{expected},T,A} = 190.54*0.954 = 181.78$$

$$\text{Var}(N_{\text{expected},T,A}) = 181.78*0.954*(1-0.25) = 130.06$$

$$\text{CMF} = (150/181.78)/(1+(130.06/181.78^2)) = 0.822$$

$$\text{Variance} = (0.822^2((1/150)+(130.06/181.78^2)))/(1+130.06/181.78^2) = 0.0071.$$

Taking the square root of the variance, the standard error of the CMF is 0.084. This is now significant at the 95 percent confidence level since the 95 percent confidence interval for the CMF does not include 1.0 [95 percent confidence interval = $0.822 \pm 1.96*0.084 = 0.657$ to 0.987].

Issues with Comparison Group and Empirical Bayes Studies

The observed change in crash experience at treated sites between the periods before and after treatment may be due not only to the countermeasure, but to other factors as well. If these factors are not properly accounted for, there is the potential to bias the results. These other factors include:

1. Traffic volume changes due to general trends or to the countermeasure itself.
2. Changes in reported crash experience due to changes in crash reporting practice, weather, driver behavior, effects of safety programs, etc.
3. Regression-to-the-mean is problematic because safety is expected to change even in the absence of a treatment. A comparison group study will not account for regression-to-the-mean unless treatment and comparison sites are matched on the basis of crash occurrence. As discussed in Section 3.1, there are practical difficulties in matching treatment and comparison sites on the basis of crash occurrence.

In both the comparison group and empirical Bayes before-after methods, untreated sites are used to account for time trends and changes in other factors such as traffic volumes and crash reporting. As such, it is desirable to conduct a test of comparability to evaluate the suitability of the untreated group. This is described in Section 3.1 and detailed in Hauer (1997).

Another issue is that in some cases the treatment may affect the logical reference group. Red light camera programs are a classical example, but there is evidence of this effect for other measures, such as traffic calming, all-way stop installation, and raised pavement markers. In the case of red light cameras (RLC), the actual hope is that there would be a general deterrent or spill-over effect at all signalized intersections, not just those with cameras, especially if the public does not know where the cameras are installed. Ignoring spill-over effects to intersections without RLCs will lead to an underestimation of RLC benefits. This is even more the case if sites with spill-over effects are used as a comparison group because a reduction in crashes due to spill-over would be attributed to time trends and the expected crashes without treatment for the treated sites would be incorrectly adjusted downwards. To resolve this issue in an empirical Bayes evaluation of RLCs (Persaud et al., 2005), the effects of regression-to-the-mean and changes in traffic volume were explicitly accounted for using SPFs relating crashes of different types and severities to traffic flow and other relevant factors for each jurisdiction based on signalized intersections *without* RLCs. Annual SPF multipliers were calibrated to account for the temporal safety effects of other factors (e.g., weather, demography, and crash reporting). This is common practice in applying the empirical Bayes methodology. However, due to the

possibility of spill-over effects at neighboring signalized intersections, it was decided to use a comparison group of unsignalized intersections in the jurisdiction of interest to estimate annual multipliers for the period *after* the first RLC installation.

3.3 FULL BAYES STUDIES

Overview

Full Bayes is not a type of evaluation study on its own. Rather, it is a modeling approach that can be used in the same way as the more common generalized linear modeling approach, typically employed in the empirical Bayes method for before-after studies (see Section 3.2) or in the development of cross-sectional models (see Section 3.4).

In the empirical Bayes approach to before-after studies, a reference group is used to estimate the expected crash frequency and its variance from a calibrated SPF. These estimates of crash frequency are then combined with the observed crash frequency at the treatment site to obtain an improved estimate of a site's long-term expected crash frequency in the absence of treatment.

In the full Bayes approach to before-after studies, a reference population is also used. However, instead of using a point estimate of the expected crash frequency and its variance, a distribution of likely values is generated. This distribution of likely values is then combined with the observed crash frequency to estimate the long-term expected crash frequency. Through the use of a distribution, rather than a point estimate, the expected crash frequency, the variance of the long term crash frequency, and the variance of the estimated CMF can be calculated more accurately.

In the cross-sectional study approach to developing CMFs, data for locations with and without a feature are obtained (see Section 3.4). Generalized linear regression is commonly used to develop a model, relating geometric and operational characteristics to the expected crash experience. Full Bayes modeling is applied in the same principle; however, it is a much more flexible modeling tool as will be discussed.

Method

Both empirical Bayes and full Bayes approaches require the same considerations to control for confounding effects in evaluating the safety effectiveness of treatments. However, there are a number of attractive characteristics of the full Bayes approach. One benefit is that the modeling framework allows for complex model forms to be specified, such as those that include both multiplicative and additive terms. Additive terms are useful for representing point hazards such as driveways. Such model forms are not easily handled in conventional generalized linear modeling approaches.

Another benefit is that the properties of full Bayes models should allow for the estimation of valid models with smaller sample sizes. This may be particularly valuable for relatively rare crash types such as those involving pedestrians or for reference groups with limited sites (such as five-legged intersections).

Perhaps most advantageous is the ability to consider spatial correlation between sites in the full Bayes model formulation. Spatial correlation considers the effect of one location's proximity to other locations on the expected crash frequency. For example, a recent study of county-level injury and fatal crashes in Pennsylvania (Agüero-Valverde and Jovanis, 2006) found spatial correlation to be significant. While the county-level full Bayes models reveal the existence of spatial correlation in crash data, they also provide a mechanism to quantify and reduce the effect of this correlation. For before-after studies, spatial correlation will likely be an issue where both treated and comparison sites are nearby. Considering spatial correlation accommodates the inclusion of sites geographically close to each other. If exposure over time is not known, then the comparison or reference group selected should in fact be as close in proximity as possible to the treated sites since the exposure is more likely to be similar than if the sites are farther away.

Another attractive feature of full Bayes modeling is that it provides the opportunity to incorporate prior knowledge. The prior knowledge could be for the parameter estimates of the model or for the estimated CMF. Where previous research has found reliable estimates of crash prediction models or for CMFs, these estimates can be introduced in the full Bayes approach. This prior knowledge has an impact on the final estimates for the new study. Alternatively, where prior knowledge is not available or otherwise not desired to be used, it is not introduced in the full Bayes approach.

In summary, the strengths of the full Bayes method, relative to the empirical Bayes can be identified as follows:

- The ability to specify complex model forms.
- The potential for estimation of valid crash models with small sample sizes.
- The ability to consider spatial correlation between sites in the model formulation.
- The ability to include prior knowledge on the values of the coefficients in the modeling along with the data collected.

Sample Size Calculations

Sample size considerations for full Bayes modeling are similar to those for cross-sectional studies or before-after studies. For cross-sectional studies, the number of locations required will depend on a number of factors including:

- Average crash frequencies.
- The number of variables desired in the model.
- The level of statistical significance desired in the model.
- The amount of variation in each variable of interest between locations.

It is difficult to estimate sample sizes for cross-sectional models, including the full Bayes modeling process, prior to model development. Determining if the sample size is adequate can only be done once the model output is available. If the variables of interest are not statistically significant, then more data are required. For this reason, the determination of required sample size is an iterative process, although through experience and familiarity with specific databases an educated guess may be possible.

For before-after studies, as indicated earlier, it is vital to ensure that enough data are included such that the expected change in safety can be statistically detected. Although there is no formal method for determining required sample sizes for the full Bayes before-after approach, methods do exist for the before-after with comparison group method. These methods may be applied and could be considered conservative in that the full Bayes approach reduces uncertainty in the estimates of expected crashes. For further information on the sample size estimation approach see *Sample Size Calculations* in Section 3.1, *Before-After with Comparison Group*.

Issues with Full Bayes Studies

The principle issue with the full Bayes method is the complexity of its application as it may require a very high level of statistical training. Moreover, while it has been possible to develop software for application of the empirical Bayes method (e.g., *SafetyAnalyst*¹), this seems to be very difficult for the full Bayes method.

Whether the benefits of the full Bayes method outweigh the increased complexity remains an open question. Limited research to date suggests that the empirical Bayes approach will produce equally reliable results as the full Bayes method where sufficient sites are available to estimate robust safety performance functions for the empirical Bayes approach. Section 4.2 presents a case study and several sample problems to further explain the considerations involved in the selection of a study design. The case study in Section 4.2 specifically addresses the considerations of a before-after study and weighs the strengths and weaknesses of the comparison group, empirical Bayes, and full Bayes methods.

¹SafetyAnalyst. American Association of State Highway and Transportation Officials (AASHTO). Available at: www.safetyanalyst.org

3.4 CROSS-SECTIONAL STUDIES

Overview

Cross-sectional studies look at the crash experience of locations with and without some feature and then attribute the difference in safety to that feature. In its most basic application, the CMF is estimated as the ratio of the average crash frequency for sites with and without the feature. For this approach to be reliable it is important that all locations are similar to each other in all other factors affecting crash risk. In practice this requirement is difficult to meet.

EXAMPLE

To illustrate the basic application, the safety effects of signalization are of interest and crash data for 100 two-way stop-controlled and 100 signalized intersections have been collected. All intersections are in rural environments, have four approach legs and similar traffic volumes. The average crash frequency for two-way stop-controlled intersections is 3.4 crashes per year and the average for signalized intersections is 2.9 crashes per year. The CMF for installing a signal at a two-way stop-controlled intersection is then calculated as:

$$CMF = \frac{2.9}{3.4} = 0.85$$

Cross-sectional studies are particularly useful for estimating CMFs where there are insufficient instances where the treatment was applied to conduct a before-after study. For example, there may be few or no projects where the shoulder is widened from, say, four feet to six feet. However, there would be many road segments with four foot shoulders and many with six foot shoulders. The reason that before-after studies are impractical in such cases is that there are often not enough before-after situations to allow for credible results.

Method

In practice, it is difficult to collect data for enough locations that are alike in all factors affecting crash risk. Hence, cross-sectional analyses are often accomplished through multiple variable regression models. In these models an attempt is made to account for all variables that affect safety. If such attempts are successful, the models can be used to estimate the change in crashes that results from a unit change in a specific variable. The CMF is derived from the model parameters.

EXAMPLE

To illustrate the use of multivariate regression models to derive CMFs, consider the model for crashes on two-lane rural roads developed by Vogt and Bared (1998). This model was developed using data collected from two States, including roadway geometry, traffic volumes and crash data. Data for portions of the road in the vicinity of intersections were not used in developing the model. The model was developed in order to assess the impacts of changes in various road characteristics on expected crashes. The model equation is shown in Equation 14.

$$Y = T * \exp(0.649 + 0.139S - 0.845L - 0.059SW + 0.067R + 0.008DD) \sum WH_i \exp(0.045D_i) \sum WG_k \exp(0.105G_k) \quad (14)$$

Where,

Y = predicted number of non-intersection crashes per year.

T = traffic exposure in millions of vehicle-miles.

L = lane width in feet.

SW = average of left and right shoulder widths in feet.

R = average roadside hazard rating along segment.

DD = driveway density in driveways per mile.

$S = 0$ for Minnesota, 1 for Washington.

D_i = degree of curve in degrees per hundred feet of the i th horizontal curve that overlaps the segment.

WH_i = fraction of the total segment length occupied by the i th horizontal curve.

G_k = absolute grade in percent of the k th uniform grade section that overlaps the segment.

WG_k = fraction of the total segment length occupied by the k th uniform grade section.

From the estimated parameters of the model, CMFs can be inferred. These CMFs represent the changes in mean predicted crash count when the value of a variable is increased by one unit. For example, the CMF for increasing lane width (L) by one foot is equal to:

$$CMF = \exp(-0.845L) = \exp(-0.845*(1)) = 0.92$$

The percentage change would be equal to $(1-0.92) \times 100 = 8$ percent decrease for each one foot increase in lane width.

The regression approach for estimating a CMF is consistent with the belief that the CMF is a function of the traits of the treated unit. A cross-sectional approach can be used to develop a CMF function, and is preferable if the cause-effect relationship with crashes can be determined with confidence.

Sample Size Calculations

The determination of required sample sizes for cross-sectional studies is difficult. For multivariate regression models, the number of locations required will depend on a number of factors including:

- Average crash frequencies.
- The number of variables desired in the model.
- The level of statistical significance desired in the model.
- The amount of variation in each variable of interest between locations.

Determining if the sample size is adequate can only be done once the model output is available. If the variables of interest are not statistically significant, then more data are required. For this reason the determination of required sample size is an iterative process, although through experience and familiarity with specific databases an educated guess will be possible.

Issues with Cross-Sectional Studies

The basic issue with the cross-sectional design is that the comparison is between two distinct groups of sites. As such, the observed difference in crash experience can be due to known or unknown factors, other than the feature of interest. Known factors, such as traffic volume or geometric characteristics, can be controlled for in principle by estimating a multiple variable regression model and inferring the CMF for a feature from its coefficient. However, the issue is not completely resolved since it is difficult to properly account for unknown, or known but unmeasured, factors. For these reasons, caution needs to be exercised in making inferences about CMFs derived from cross-sectional designs. Where there are sufficient applications of a specific countermeasure, the before-after design is clearly preferred.

At present, the science of assembling CMFs from multivariate models is not fully developed. As such, the validation of CMFs determined from such studies is especially important. Such CMFs could be inaccurate for a number of reasons, including inappropriate functional form, omitted variable bias, or correlation of variables. It is common practice to use generalized linear modeling techniques, assuming a negative binomial error structure, to estimate multivariate crash prediction models. However, it is difficult to account for all factors that affect safety using such modeling techniques. For example, intersections with left-turn lanes also

tend to have illumination. If a crash prediction model is used to estimate a CMF for left-turn lanes, and the presence of illumination is not accounted for in the model, the difference in model predictions with and without left-turn lanes could be partly due to illumination differences. Ironically, it is precisely because a variable is found to be correlated with another variable that it may be omitted during the model fitting exercise. Including correlated variables could in fact lead to effects that are counterintuitive (e.g., illumination increases nighttime crashes).

Another reason why the effect of an element that may affect safety cannot be captured in a model is because the sample used to develop the model is too small, or there is little or no variation in the element. For example, the effect of illumination cannot be captured if all locations in a sample are illuminated.

In evaluating CMFs derived from a cross-sectional study the following questions should be considered:

- Is the direction of effect (i.e., expected decrease or increase) in crashes in accord with expectations?
- Does the magnitude of the effect seem reasonable?
- Are the parameters of the model estimated with statistical significance?
- Do different cross-section studies come to similar conclusions?
- Do before-after studies come to similar conclusions?

3.5 CASE-CONTROL STUDIES

Overview

Case-control methods have been used in certain areas of highway safety, but few have focused on the effects of geometric design elements. For example, case-control studies have been applied to investigate the effectiveness of motorcycle-helmet use and the crash risk of hours of service for truck drivers. More recently, the case-control method was employed to estimate CMFs for geometric design elements, including lane and shoulder width (Gross, 2006; Gross and Jovanis, 2007).

Matching is one method used to account for potential confounding variables and involves the random selection of control sites with characteristics similar to the corresponding case site.

Case-control studies are based on cross-sectional data. However, they should not be confused with cross-sectional studies in general. For cross-sectional studies, samples are generally selected based on the presence and absence of a specific characteristic (e.g., lighting) or based on a specific roadway or intersection type, ignoring whether there was a crash there or not. Case-control studies select sites based on outcome status (e.g., crash or no crash) and then determine the prior treatment (or risk factor) status within each outcome group. Additional criteria may be applied in a case-control design when a **matching** scheme is used. Matching cases with controls that are identical in factors which may contribute to crash risk is one method to control for potential confounders.

Case-control studies assess whether exposure to a potential treatment is disproportionately distributed between the cases and controls, thereby indicating the likelihood of an actual benefit from the treatment.

EXAMPLE

A case control study was employed to investigate the safety effects of degree of horizontal curvature. Cases were defined as those curves with a crash and control sites were identified as those curves without a crash. Once the cases and controls were defined, the degree of curve was identified for each site in the two groups.

Method

The likelihood of an actual treatment is expressed as the odds ratio between two levels of a variable. For example, it may be found that the odds of a crash occurring on horizontal curves with a degree of curvature greater than 15 degrees is 1.5 times the odds of a crash occurring on curves less than 15 degrees. The odds ratio is a direct estimate of the CMF. Treatments may take the form of binary variables (e.g. median barrier, roadway lighting, or guardrail) or multi-level variables such as lane width (e.g. 9, 10, 11 and 12 foot lanes). The sample is summarized by treatment and case-control status to calculate the odds ratio. To illustrate the concept of the odds ratio, consider the data in Table 9.

TABLE 9. Tabulation for Simple Case-Control Analysis

Treatment	Number of Cases	Number of Controls
With	A	B
Without	C	D

The odds ratio (CMF) is expressed as the expected increase or decrease in the outcome in question due to the presence of the treatment. An odds ratio greater than 1.0 suggests that the presence of the treatment increases risk, while a value less than 1.0 would suggest a decrease in risk. Using the notation in Table 9 the odds ratio is calculated from Equation 15.

$$\text{Odds Ratio (OR)} = \text{CMF} = \frac{A/B}{C/D} = \frac{AD}{BC} \tag{15}$$

Case-control studies cannot be used to measure the probability of an event (e.g., crash, severe injury, etc.) in terms of expected frequency. They are more often used to show the relative effects of treatments. Statistical analyses, such as multiple logistic regression techniques, are commonly used to clarify these relationships because they are able to examine the risk/benefit associated with one factor while controlling for other factors.

EXAMPLE

Tsai et al., (1995) investigated the effectiveness of helmet use and type for the prevention of head injuries among motorcycle riders in Taipei, Taiwan. A case-control method was used to investigate crash-involved motorcycle riders comparing those with head injuries (cases) to those not suffering head injuries (controls). The case-control method was used to control for confounding variables such as age, gender, and helmet type that may influence the risk of head injury.

Cases and controls were selected from a group of 1351 victims of motorcycle crashes located in one of 15 hospitals in Taipei, Taiwan. This study is unique because a second group of “on-street” controls was also selected. For every daytime (8am-6pm) motorcycle injury, pictures of four motorcycles were taken at the same time of day at the same location. Multiple logistic regression models were used to estimate the odds of head injury associated with the use of different types of helmets as well as other predictors. This study illustrates the application of multiple logistic regressions to estimate odds ratios and the use of covariates to make adjustments for confounders.

The ratio of controls to cases may vary and often depends on the availability of time, budget, and potential sites. Increasing the number of controls will increase the power of the study, especially when there are relatively few cases. Power is defined as the probability that the test will reject a false null hypothesis. In a matched design, controls are sampled randomly and matched to each case based on similar values of the potential confounding variable. Matching provides a balanced design and automatically adjusts the estimates for the effects of variables included in the matching scheme.

The case-control method may be very useful for studying rare events because the number of cases and controls is predetermined. Another advantage of the case-control design is that multiple treatments may be investigated in relation to a single outcome using the same sample. A single sample may be used to investigate any variables that are not included in the selection or matching criteria for cases and controls. Although case-control studies may be used to explore multiple treatments, they can only investigate one outcome per sample. The sampling is conducted separately within the case and control populations based on outcome status and different outcomes will produce different samples.

EXAMPLE

A case-control design was used to investigate the effects of edgeline rumble strips on run-off-road crashes. Cases were defined as segments that experienced a run-off-road crash within the six month study period and controls were defined as segments that did not experience a run-off-road crash during the same study period.

This same sample could be used to investigate the effects of other variables such as lane and shoulder width on run-off-road crashes. However, this sample could not be used to investigate the effects related to other outcomes such as nighttime crashes because the case definition was based on run-off-road crashes. To investigate the effect of edgeline rumble strips on nighttime crashes, a new sample would need to be drawn based on the new case definition.

Sample Size Calculations

In general, the required sample size for a case-control study design may be calculated using Equation 16. See below for the calculation of sample size for a matched study design).

$$n = \frac{(r + 1)(1 + (\lambda - 1)P)^2}{rP^2(P - 1)^2(\lambda - 1)^2} \left[z_{\alpha} \sqrt{(r + 1)p_c(1 - p_c)} + z_{\beta} \sqrt{\frac{\lambda P(1 - P)}{[1 + (\lambda - 1)P]^2} + rP(1 - P)} \right]^2 \tag{16}$$

The common proportion over two groups (p_c) is obtained from Equation 17.

$$p_c = \frac{P}{r + 1} \left(\frac{r\lambda}{1 + (\lambda - 1)P} + 1 \right) \tag{17}$$

Where,

n = total sample size.

r = case:control ratio (number of cases divided by the number of controls).

λ = desired detectable level of effect (i.e., magnitude of the safety effect to be detected).

P = prevalence of the treatment (proportion of the population with the treatment).

z_{α} = z-statistic for significance level (α for a one-sided test or $\alpha/2$ for a two-sided test).

α = statistical significance level.

z_{β} = z-statistic for statistical power $1 - \beta$.

β = probability of a Type II error (false negative rate).

p_c = common proportion over two groups.

EXAMPLE

A case-control design is desired to investigate the effects of edgeline rumble strips on run-off-road crashes on two-lane rural roads. Several States were included in the study population. The rural two-lane roads from each State were divided into 1/2 mile study segments and a six-month crash history was determined for each segment. Cases were defined as

segments that experienced a run-off-road crash within the six month study period and controls were defined as segments that did not experience a run-off-road crash during the same study period. An equal number of cases and controls (i.e., $r=1$) were randomly sampled from the study population of rural two-lane roads in each State. The researcher currently has 8,000 cases and 8,000 controls and would like to know if a larger sample size is required before proceeding with the analysis.

In this example, is the sample sufficient to detect a 10 percent reduction in run-off-road crashes ($\lambda = 0.9$) with 90 percent power using a two-sided 5 percent significance test? The researcher knows that edgeline rumble strips have been installed on approximately 30 percent of the total miles of rural, two-lane roads in the State (i.e., the prevalence of the treatment, P , is 0.3).

$$p_c = \frac{0.3}{1+1} \left(\frac{1(0.9)}{1+(0.9-1)0.3} + 1 \right) = 0.289$$

$$n = \frac{(1+1)(1+(0.9-1)0.3)^2}{1(0.3)^2(0.3-1)^2(0.9-1)^2} \left[1.96\sqrt{(1+1)(0.289)(1-0.289)} + 1.2816 \sqrt{\frac{0.9(0.3)(1-0.3)}{[1+(0.9-1)0.3]^2} + 1(0.3)(1-0.3)} \right]^2$$

$$n = 18,408$$

In this case, the researcher would need approximately 9,204 cases and a similar number of controls to detect a 10 percent reduction in run-off-road crashes with 90 percent power using a two-sided 5 percent significance test. However, the researcher currently has only 8,000 of each. One option is to increase the sample size. Another option is to revisit the assumptions and possibly increase the minimum detectable safety effect and/or relax the level of significance, both of which would reduce the required sample size.

For a matched case-control design, the required sample size is proportional to the expected number of discordant pairs (i.e., case-control pairs with a different treatment status). The required number of discordant pairs (d_p) is based on the desired level of statistical significance, statistical power, and detectable level of effect as shown in Equation 18.

$$d_p = \frac{[z_\alpha(\lambda + 1) + 2z_\beta\sqrt{\lambda}]^2}{(\lambda - 1)^2} \tag{18}$$

Where,

d_p = number of discordant pairs.

z_α = z-statistic for significance level (α for a one-sided test or $\alpha/2$ for a two-sided test).

α = statistical significance level.

z_β = z-statistic for statistical power $1-\beta$.

β = probability of a Type II error (false negative rate).

λ = desired detectable level of effect (i.e., magnitude of the safety effect to be detected).

The required sample size is then equal to the number of discordant pairs divided by the proportion of expected discordant pairs in the sample. The probability of a discordant pair (π_d) for a specific treatment can be determined by examining a sample of case-control data as shown in Equation 19.

$$n = \frac{2d_p}{\pi_d} \tag{19}$$

Where,

n = total sample size.

π_d = probability of a discordant pair.

EXAMPLE

Consider the previous example where a case-control design is desired to investigate the effects of edgeline rumble strips on run-off-road crashes on two-lane rural roads. Previously, an equal number of cases and controls (i.e., $r=1$) were randomly sampled from the study population of rural two-lane roads in each State. Consider now that controls are randomly selected and matched to each case on the basis of potential confounding factors (i.e., traffic volume, speed limit, and presence of horizontal curvature).

Given the same assumptions, how does the required sample size change for a matched design? It is assumed that the probability of a discordant pair is 0.8.

$$d_p = \frac{[1.96(0.9 + 1) + 2(1.2816)\sqrt{0.9}]^2}{(0.9 - 1)^2} = 3,789.22$$

$$n = \frac{2(3,789.22)}{0.8} = 9,473$$

In this case, only 4,737 cases and a similar number of controls would be required. Note that this is substantially less than the sample size required for an unmatched design. A matched design can improve the efficiency of a study design, resulting in fewer required sites. However, the matching process is not a trivial task and can quickly result in a limited sample for analysis if the matching criteria are too restrictive.

Issues with Case-Control Studies

The most important step in a case-control study is defining the cases and controls. Ambiguous or broad definitions for cases and controls may lead to misclassification and will likely produce unclear results. Care must be taken to ensure that cases and controls are representative of the sites of interest. In other words, the chance of being included in the study must not be associated with the treatment of interest.

EXAMPLE

The following is an example of a broad case definition.

Cases = roadway segments that experience at least one crash during the study period.

Controls = segments that do not experience a crash during the specified study period.

These definitions are fairly general and may need to be more specific to include only rural roadway segments or

segments with specific geometric and traffic characteristics. A more specific case definition helps to isolate the effect of the treatment in question. The case and control definition could include exposure levels as well (e.g., traffic volume), but exposure is more commonly accounted for in the analysis.

Case-control studies effectively fix the number of controls based on the number of cases, which may or may not represent the appropriate proportions in the entire population. As such, the case-control method cannot be used to determine relative risk. The odds ratio, however, may be a good approximation of the relative risk on the condition that the outcome is relatively rare. In the case that the outcome is rare, the number at risk may be approximated by the number of controls.

Finally, the case-control method cannot demonstrate causality because there is no time sequence of events in the analysis. Instead, the odds ratio indicates the increased/decreased likelihood of a crash occurring when a treatment (e.g., roadway characteristic) is present. It does not, however, recognize differences between locations with many crashes or a single crash. This is a loss of potentially important information and thus, the true increase in risk could be underestimated.

3.6 COHORT STUDIES

Overview

Cohort study methods have been used in some areas of highway safety, typically related to issues such as seatbelt effectiveness and driver training. To date, these methods have received little attention in the area of highway design, but show promise as alternatives for estimating safety effectiveness (Gross, 2006; Gross and Jovanis, 2008).

Cohort methods are used to estimate relative risk, which indicates the expected percent change in the probability of an outcome given a unit change in the treatment (or risk factor). The relative risk is a direct estimate of the CMF. Sites are assigned into a particular cohort based upon current treatment status and followed over time to observe exposure and event frequency. Cohort studies then assess whether the time at risk is disproportionate between cohorts, which indicates the relative effect of the treatment.

EXAMPLE

Cohorts could be defined as sites with a particular geometric or operational characteristic (e.g., two-lane rural roads with and without centerline rumble strips). The outcome may be defined as a crash. The cohorts would be followed over time (or assessed retrospectively during a specific time period) to identify the time at risk until a crash occurred.

Method

In the simplest approach, the tabulation approach, the risk of an outcome due to a particular level of treatment is calculated as the number of cases divided by the total number at risk, or more completely as the time at risk for cases divided by the total time at risk. Relative risk may be computed for any two levels of a treatment as the ratio of the two risks. There are other more sophisticated analysis approaches which are capable of accounting for confounders, including the time at risk. These include cohort life tables, subject-years approach, and other statistical models. In any case, it is important to adjust for confounding factors because ignoring true confounding variables may lead to incorrect estimates of the relative risk.

To illustrate the tabulation approach, consider Table 10 below which identifies data for two cohorts (i.e., sites with and without some feature). The number of outcomes in each cohort is the number of sites at which a crash was observed. The number of non-outcomes is the number of sites at which no crashes were observed. Total at-risk is the sum of sites for the cohort. The issue with this simple tabulation is that confounders, such as time at risk and other site characteristics that differ between the two cohorts, are not accounted for properly.

TABLE 10. Tabulation for Simple Cohort Analysis

Cohort	Outcomes	Non-Outcomes	Total At-Risk
With	A	B	A+B
Without	C	D	C+D

The relative risk of 'with' compared to 'without' would be calculated using Equation 20.

$$Relative\ Risk = CMF = \frac{A/(A+B)}{C/(C+D)} \quad (20)$$

Similar to case-control studies, matching is an option in cohort designs to account for potential confounding variables. It is more common, however, to adjust for confounding variables during the analysis of a cohort design and matching is reserved for special situations. Specifically, a known powerful confounder or one that is difficult to measure precisely may call for a matching scheme.

EXAMPLE

An agency does not have complete records of the degree of horizontal curvature, but they know which sections of roadway include a curve and which sections do not. It was also determined that horizontal curvature is a potential confounder for the treatment in question. In this case, the presence of horizontal curvature could be used as a matching variable to account for the potential confounding effects, even though the precise radius or degree of curve is unknown.

Pair matching is often used to account for confounding variables when a matching scheme is required. Pair matching involves matching each study site closely with a control site on the specific confounding factor. Frequency matching is another type of matching scheme where each study site or group is matched with controls based on a category of a factor (e.g. age or gender). Frequency matching helps to prevent large imbalances between study groups that may reduce the power of the study.

The cohort method is well suited for studying rare treatments because the sample is selected based on treatment status. Additionally, several outcomes can be studied for a particular treatment. Cohort study designs are methodologically stronger than case-control studies because it is easier to ensure that the groups are defined and selected independently of the outcome of interest. In highway safety, it is often unfeasible (or unethical) to conduct controlled experimental studies. However, it is still necessary to test new strategies and concepts. When newly developed treatments are being considered for implementation, it is prudent to implement the strategy on a relatively small scale and test its effectiveness before deploying on a large scale basis. The cohort method may be useful for evaluating the effectiveness of such strategies.

Sample Size Calculations

For a cohort design, the required sample size may be calculated using Equation 21.

$$n = \frac{r + 1}{r(\lambda - 1)^2 \pi^2} \left[z_{\alpha} \sqrt{(r + 1)p_c(1 - p_c)} + z_{\beta} \sqrt{\lambda \pi(1 - \lambda \pi) + r \pi(1 - \pi)} \right]^2 \quad (21)$$

The common proportion over two groups (p_c) is obtained using Equation 22.

$$p_c = \frac{\pi(r\lambda + 1)}{r + 1} \quad (22)$$

Where,

n = total sample size.

r = ratio of treatment group to reference group.

π = proportion in the reference group where an outcome was observed.

λ = desired detectable relative risk (i.e., magnitude of the safety effect to be detected).

z_{α} = z-statistic for significance level (α for a one-sided test or $\alpha/2$ for a two-sided test).

α = statistical significance level.

z_{β} = z-statistic for statistical power $1 - \beta$.

β = probability of a Type II error (false negative rate).

p_c = common proportion over two groups.

EXAMPLE

An agency is setting up a cohort study and has determined that they would like to estimate the sample size required to detect a 20 percent reduction in crashes ($\lambda = 0.8$) with 90 percent power using a two-sided 10 percent significance test. A statewide database was examined to determine the proportion of crash segments in the reference group. The reference group proportion, π , is calculated as the number of crash segments in the reference group divided by the total number of segments in the reference group. For the existing dataset, the proportion was 0.50. The following calculations illustrate how the required sample size changes as the ratio of treatment group to reference group, r , changes from 1.0 to 0.25.

When $r=1$:

$$p_c = \frac{0.5(1(0.8) + 1)}{1 + 1} = 0.45$$

$$n = \frac{1 + 1}{1(0.8 - 1)^2 0.5^2} \left[1.6449 \sqrt{(1 + 1)(0.45)(1 - 0.45)} + 1.2816 \sqrt{0.8(0.5)(1 - 0.8(0.5)) + 1(0.5)(1 - 0.5)} \right]^2$$

$$n = 200[1.157 + 0.897]^2 = 844$$

When $r=0.25$:

$$p_c = \frac{0.5(0.25(0.8) + 1)}{0.25 + 1} = 0.48$$

$$n = \frac{0.25 + 1}{0.25(0.8 - 1)^2 0.5^2} \left[1.6449\sqrt{(0.25 + 1)(0.48)(1 - 0.48)} + 1.2816\sqrt{0.8(0.5)(1 - 0.8(0.5)) + 0.25(0.5)(1 - 0.5)} \right]^2$$

$$n = 500[0.919 + 0.705]^2 = 1319$$

Assuming the size of the treatment group and reference group are equal ($r=1$), the researcher would require 422 treatment sites and an equal number of reference sites. If the treatment was relatively rare and the size of the reference group was four times the size of the treatment group ($r=0.25$), the researcher would require a total of 1,319 sites (264 treatment sites and 1,055 reference sites).

Issues with Cohort Studies

Cost and time restrictions are often cited as drawbacks to the cohort method. Large samples are often required making studies relatively expensive, particularly if the outcome is rare or coupled with a long follow-up period in the case of prospective studies.

Care must also be taken to ensure that treatments and confounding variables are accounted for properly. Treatments are subject to change during the study period. If the treatment for a particular site changes during the study period then the site is effectively moving from one cohort to another. The time at risk should be allocated proportionally between the respective cohorts.

EXAMPLE

Data for roadway segments are collected starting on the first day of the year and followed for a one year period. After five months, a section of roadway is widened from eleven to twelve feet. Assuming that the construction period lasted one month and no crashes occurred on the improved section, the section would contribute five months of exposure to the eleven foot cohort and six months to the twelve foot cohort. The analysis then reflects the periods of exposure to different treatments and excludes the period under construction so that crashes that occur during the work zone condition are not included in the analysis.

Finally, the cohort method does not recognize differences between locations with many crashes or a single crash. Only the time to the first crash is recorded and analyzed. Thus, the true increase in risk could be underestimated.

3.7 ALTERNATIVE APPROACHES FOR DEVELOPING CMFS

The intent of this section is to identify alternative approaches for developing CMFs for situations where conducting a new crash-based research study is either not feasible or not desired. Specifically, it introduces meta-analysis, expert panels, and the potential use of surrogate measures in safety. These alternative methods for developing CMFs are not equal with respect to the level of rigor and confidence in the results. The three methods identified in this section are listed in order of preference. An overview of each method is presented and the issues related to each method are discussed.

Alternative approaches, as used in this context, refers to the development of CMFs using information other than crashes. The procedure could involve the combination of multiple CMFs that were derived from crash-based studies, but the resulting CMF is not the direct result of a crash-based evaluation.

Meta-Analysis Studies

Overview

Where multiple CMFs exist for the same countermeasure, it is believed that the desired practice is not to merely select the highest rated CMF, but to combine the knowledge from all relevant studies for the same countermeasure. Meta-analysis is a systematic way of combining knowledge on CMFs from multiple previous studies while considering the study quality of each in arriving at a final CMF estimate. Elvik (2005) provides an overview of the meta-analysis process. The process is described in five steps, which are outlined below.

1. Defining the topic of the meta-analysis as precisely as possible.
2. Conducting a systematic search for relevant studies.
3. Defining study inclusion criteria.
4. Determining which data to extract from each study.
5. Converting estimates of effect to a common scale.

For a meta-analysis to be effective, all the studies included should be similar in terms of data used, outcome measure, and study methodology. The study should also provide an estimate of the CMF standard error, or the information made available in order to derive it. Where the expected reduction in crashes may be small, but there are many studies, a meta-analysis may be able to increase the statistical power by combining the individual results into an overall result.

Method

The systematic review of literature includes the following (Elvik, 2005):

1. A systematic and extensive search for relevant studies is performed with the objective of including all studies that have been conducted, even unpublished studies. Ideally speaking, the search for relevant studies should be global, without any restrictions with respect to language, region, or study age.
2. Data are extracted from each study according to a standardized procedure, using a data extraction form. In order to ensure the accuracy of data extraction, two researchers independently extract data from the same studies.
3. Clear study inclusion criteria are formulated. An attempt is made to assess study quality and present the findings of the best studies.
4. Procedures for study retrieval, data extraction, and meta-analysis are reported in detail in order to ensure reproducibility of the review.

The key step in combining results from various studies is converting the estimates of effect into a common measure, for example:

- Proportions of target crash types.
- Crash rates.
- Odds ratios.
- Expected crash reductions.

All meta-analysis techniques are based on the same principle that the estimated CMF is an average of all individual CMF estimates from reviewed studies. The simplest form of an average is the median, but many use a weighted average to combine information from studies that produce a quantitative estimate of effect size. The weight may reflect the precision of the estimates from each study and the suitability of the study methodology. In general, higher weights are given to CMFs from studies with large sample sizes, small variance, and from studies that use the most appropriate methodology to account for confounding factors. An example of a weighting scheme is shown in Equation 23. The Highway Safety Manual includes a variation of this weighting scheme as detailed in Bahar (2010).

$$CMF = \frac{\sum W_i CMF_i}{\sum W_i} \quad (23)$$

Where,

CMF_i = the estimated CMF of study i .

W_i = a statistical weight assigned to study i that depends on the quality of the study.

The weights applied to each individual result are a measure of the certainty of the estimate such that results with greater uncertainty are given less weight. The weights assigned to individual studies may be determined in different ways. One such example is a simple function of the standard error of the estimate, as shown in Equation 24. The Highway Safety Manual uses a slightly different method to determine the weight, as detailed in Bahar (2010).

$$W_i = \frac{1}{SE_i^2} \quad (24)$$

EXAMPLE

A study on the safety effects of median barriers, guardrails, and crash cushions applied the meta-analysis technique to thirty-two individual evaluations (Elvik, 1995). This study used the log odds meta-analysis method in which the estimated mean effect on crashes using all studies from Equation 25.

$$E = \frac{\exp(\sum_{i=1}^n \ln(E_i)(W_i))}{\sum_{i=1}^n W_i} \quad (25)$$

Where,

E_i = the estimated effect of study i .

W_i = statistical weight assigned to study i depending on the number of crashes involved in the study.

This meta-analysis study also tested for publication bias in the individual studies used. Publication bias occurs when research results are not published, often due to the results being counterintuitive (e.g., an increase or no effect on crashes when a decrease was expected). Publication bias was investigated in this case using a graphical procedure called the funnel plot method.

In this method, each study result is plotted on a graph in which the horizontal axis shows the CMF of each result and the vertical axis shows the statistical weight assigned. A study which uses a larger number of crashes receives a higher statistical weight. If there is no publication bias, a scatter-plot of results should resemble an upside down funnel. As sample size increases, the dispersion of estimates should converge since larger sample sizes should give more accurate results. If the tails of the funnel are not symmetrical, then publication bias may exist.

An example of a funnel plot taken from Elvik (1995) is shown in Figure 4. It shows 69 estimates of the safety effects of daytime running lights for cars. Note the use of a logarithmic scale for the horizontal axis. There does not seem to be evidence of publication bias because the plotted points resemble a funnel.

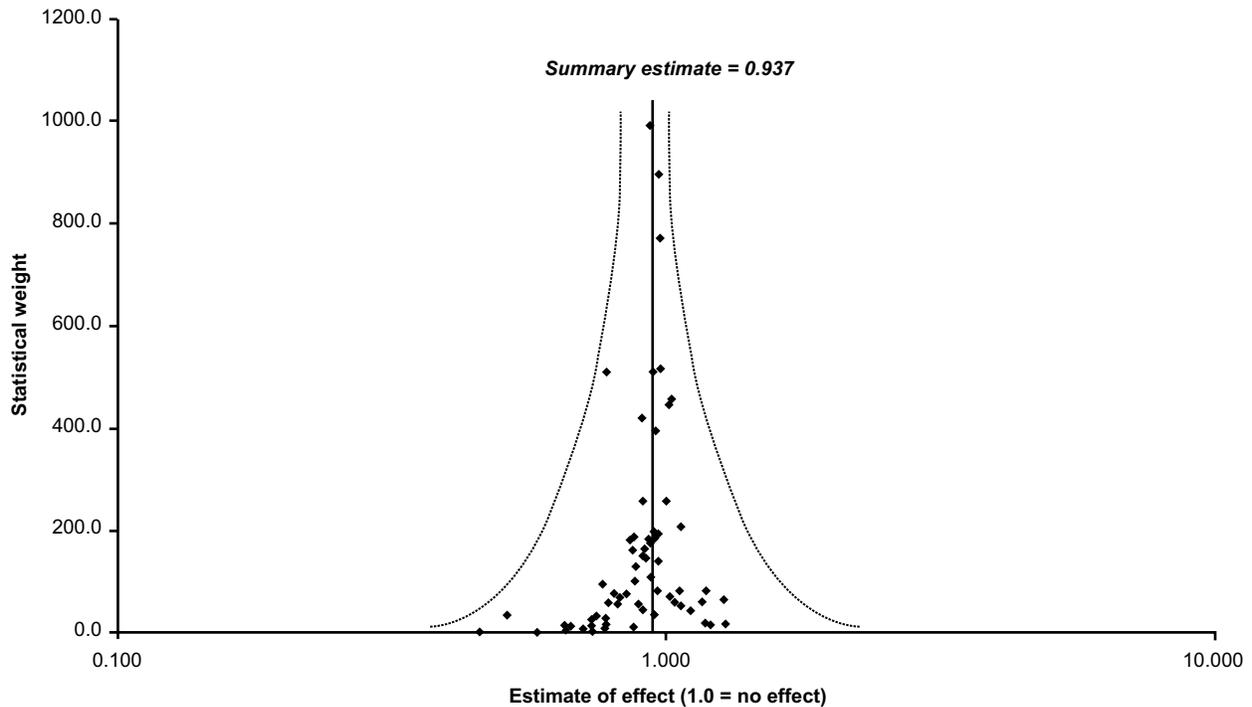


FIGURE 4. Example Funnel Plot.

The reporting results of a meta-analysis should contain several key pieces of information (Egger et al., 2001), including:

1. A list of all studies included in the meta-analysis and a brief presentation of their main findings.
2. A list of studies that were judged to be relevant, but were not included in the meta-analysis, stating explicitly for each study why it was not included.
3. A concise description of how the literature search was performed.
4. A list of all variables coded for each study, as well as frequency distributions for these variables.
5. If study quality has been assessed, a detailed explanation of how this was done should be provided.
6. A funnel plot of estimates of effect and an analysis of the funnel plot with respect to skew, the presence of outliers, and the presence of publication bias.
7. A presentation of the analysis of publication bias made and the possibility of adjusting for publication bias if it was detected.
8. A presentation of the findings of meta-analysis, for all versions of it that were performed.
9. A presentation of the main findings of the sensitivity analysis performed.

Issues with Meta-Analysis Studies

In selecting studies to include in the meta-analysis, it is important to ensure that all studies used are of sufficient quality. It is possible that a study using robust methods, but poorly executing them, could still produce results that would receive a high weighting, especially if the sample size is large.

Sensitivity and publication bias are two issues with meta-analysis studies. As such, it is recommended to perform a sensitivity analysis to determine the impacts of any assumptions and decisions made in arriving at the result. Elvik (2005) provides a systematic approach for performing a sensitivity analysis. Factors to consider in a sensitivity analysis may include:

- The estimate of effect.
- Including or excluding certain studies, particularly if their results appear to be outliers.
- Adjustment of estimates of effect for publication bias.
- Approaches to assessing study quality.
- Choice of statistical weights.

Expert Panel Studies

Overview

Expert panels are assembled to critically evaluate the findings of published and unpublished research. Each panel selects reliable studies and derives CMFs through consensus. In this way an expert panel is similar to the meta-analysis approach but is less formal.

Method

Washington et al. (2008), in offering a critique of the expert panel process, provide a good overview of the expert panel method. The process is described in four steps, which are summarized below.

Step 1: Identify Expert Panelists

Expert panels typically consist of 8-12 members drawn from the practitioner and academic communities. While there are no rules for determining the appropriate panel size, a panel that is too small may not reflect the broad spectrum of opinions in the profession, while a panel that is too large may have difficulty coming to a consensus.

Selection of panelists must consider the relevant knowledge of potential panelists in the subject matter and methodologies applied. Consideration of geographical representation is also appropriate in order to represent diverse needs.

Step 2: Set Panel Meeting Date and Prepare Supporting Panel Materials

Panel meeting duration will depend on the amount of material to review. Two to four full day meetings is typical. Preparation for the meeting requires the assembly of materials for the panel members to review in detail prior to the meeting. This involves compiling copies or summaries of all the relevant research or other related documents for the treatments under consideration.

Typically, an expert panel review will consider between 15 and 30 treatments at a rate of one treatment per hour. The treatment list should be circulated to the panel experts prior to the compilation of the materials, to ensure that important treatments have not been omitted.

Panel members are assigned topics to read in detail and are expected to lead the discussion on this material during the expert panel meeting. As much as is possible, each panel member should be assigned responsibility for the material most closely related to their expertise. If there are a large number of topics an expert may be assigned a set of treatments to review, whereas a small number of treatments may result in overlap among experts.

Step 3: Conduct Expert Panel Meeting

Expert panel meetings are typically held in an open discussion format with a designated panel member leading the discussion for the material they were assigned in Step 2. During the meeting, careful documentation of meeting minutes is essential.

Details of each treatment are discussed along with relevant research results. The aim is to develop a weighted average CMF through an open discussion of all the research and by informally assigning a weight to each estimate of the CMF. The weights are not formally defined or necessarily voted on, but discussion continues until a consensus is reached. The consensus may be that the CMF should be 1.0 (i.e. no effect). In other cases, the consensus may be that there is a lack of a suitable CMF.

Washington et al. (2008) also outline a number of important factors that should be discussed in a systematic way:

1. Relevance of the research to the application being discussed.

For example, was the research conducted in an urban environment when a rural treatment is being sought? Was the research conducted in mountainous terrain when flat terrain is the setting of interest? Typically these questions of relevance surround issues of traffic exposure, driving population, location (e.g. country in which research was conducted), range of conditions examined, and similarity of 'non-treatment' traffic controls.

2. Timeliness of the research.

The age of the research and its relevance in regard to road users, analysis methods, vehicle safety, and injury reporting thresholds is often discussed. The age of the research may be used for discounting the relevance and weighting of the results.

3. Non-ideal conditions of the research design.

The research conditions that may lead to incorrect or weak conclusions such as omitted important variables, included irrelevant variables, **endogeneity**, inappropriate analysis methods, or sampling procedure are discussed. Research studies conducted under non-ideal conditions are typically discounted or given lesser weight in panel deliberations.

4. Sample size and sample representativeness.

Studies with large samples typically are given greater weight than studies using small samples, all else being equal. In addition, studies with greater sampling representativeness (heterogeneity) of the population are given greater weight than studies conducted on more limited or biased samples.

5. Findings and conclusions of the research.

The conclusions of the research are often viewed to make sure the expert panel arrives at the same conclusions as the study authors. While some of the previously listed issues may attract greater attention, studies where the authors over-state or misstate the conclusions are scrutinized.

6. Consensus on research.

Research that confirms prior research, or that represents a substantial body of research that has reached consensus on a topic is more convincing than the lone study. Of course, research quality is important, but assuming equal quality, consensus on the effect of a treatment tends to lend relatively greater credibility.

All of the details necessary to derive a CMF are recorded, including, 1) the value of the factor, 2) the limits of a CMFunction if applicable, 3) the shape of a CMFunction, and 4) any non-linearity, spikes, or discontinuities.

Step 4: Disseminate Results

The results of the panel meeting are distributed to panel members for review and comment. After this opportunity for feedback, the CMFs are described and detailed in a document intended for broader dissemination. The implicit weights and factors that underlie the development of the CMFs are typically not recorded or documented for broad distribution.

Endogeneity occurs when one or more of the variables in a model (or analysis) are dependent on another variable or variables in the same model.

The Delphi method is a systematic, interactive forecasting method which relies on a panel of experts. The Delphi method is based on the principle that forecasts from a structured group of experts are more accurate than those from unstructured groups or individuals.

Issues with Expert Panel Studies

Washington et al. (2008) identify some important questions that need to be addressed with regard to the derivation of CMFs from expert panels. Specifically:

1. Are the results derived from expert panels accurate and precise?
2. Can expert panels be used to derive estimates of uncertainty?
3. Do results across expert panels differ, and if so, how?
4. Can expert panels be made to ensure repeatable and accurate results?
5. Should expert panels follow informal procedures (as they traditionally have) or more formal and structured procedures such as the **Delphi method**?

Washington et al. (2008) argue that traditional face-to-face expert panels do not systematically derive precision estimates of a CMF. For this purpose, it may be more appropriate to employ methods that poll or query experts independently, such as the Delphi method. Among the other shortcomings of expert panels are possible complications arising from interactions and group dynamics, and possible forecasting bias as a result.

Surrogate Measure Studies

Overview

The use of surrogate measures may be required to derive a CMF indirectly, in lieu of using crash data, where treatments have little after period data or are rarely implemented. Typical performance measures in a surrogate evaluation include vehicle speeds, lane departure encroachments, traffic control obedience, stopping behavior, and traffic conflicts. In some cases, a CMF can be estimated by using a model that relates the observed change in the surrogate before and after treatment with an expected change in crash frequency.

Method

The change in the surrogate measure can be evaluated using the same methodologies for evaluating crash changes. As is the case for crash-based evaluations, studies can be experimental or observational. The key to the application of this approach is the availability of a reliable model to relate crash frequency to the surrogate measure. Perhaps the most reliable of the few such models available pertains to the effects of speed on crash experience (Harkey et al., 2008).

EXAMPLE

Table 11 provides factors for estimating the expected change in injury crashes from a change in mean speed. This table is based on results recently published in NCHRP Report 617 (Harkey et al., 2008). The results may be used to estimate a CMF based on the mean speed before treatment and the expected speed reduction. The table provides CMFs for non-fatal injury crashes.

In NCHRP Report 617, the expected crash reductions for fatal crashes are even larger than for non-fatal injury crashes. However, to be conservative, it may be prudent to apply the CMF for injury crashes to fatal crashes as well. Interpolation would be valid for deriving CMFs for speeds not listed in the table. Alternatively, the necessary equations are documented in NCHRP Report 617 and can be used.

TABLE 11. Non-Fatal Injury CMFs for Speed Reduction Treatments

Mean Pre-treatment Speed (mph)	Speed Reduction (mph)							
	8	7	6	5	4	3	2	1
45	0.51	0.57	0.64	0.70	0.76	0.82	0.88	0.94
50	0.56	0.62	0.68	0.73	0.79	0.84	0.89	0.95
55	0.60	0.66	0.71	0.76	0.81	0.86	0.90	0.95
60	0.64	0.68	0.73	0.78	0.82	0.87	0.91	0.96
65	0.66	0.71	0.75	0.79	0.83	0.88	0.92	0.96
70	0.69	0.73	0.77	0.81	0.84	0.88	0.92	0.96

Issues with Surrogate Measure Studies

Where surrogate measures are evaluated using the same methodologies as for conducting crash-based studies the same general issues apply. The critical step in developing CMFs from surrogate measures is establishing the relationship between changes in surrogates with changes in crashes. At present, the approach for this step is relatively undeveloped, a notable exception being speed reduction treatments.

3.8 SUMMARY

In this chapter various study designs were discussed. An overview of each method was presented along with sample size considerations and associated issues. Table 12 highlights the general applicability, strengths, and weaknesses of each study design discussed previously. Chapter 4 provides resources for selecting the most appropriate method based on the data available for developing a CMF.

TABLE 12. Summary of Study Designs for Developing CMFs

Study Design	General Applicability	Strengths	Weaknesses
Before-After with Comparison Group	<p>Treatment is sufficiently similar among treatment sites.</p> <p>Before and after data are available for both treated and untreated sites.</p> <p>Untreated sites are used to account for non-treatment related crash trends.</p>	<p>Simple.</p> <p>Accounts for non-treatment related time trends and changes in traffic volume.</p> <p>Employs SPFs to account for:</p> <p>Regression-to-the-mean.</p> <p>Traffic volume changes over time.</p> <p>Non-treatment related time trends.</p>	<p>Difficult to account for regression-to-the-mean.</p>
Before-After with Empirical Bayes	<p>Treatment is sufficiently similar amongst treatment sites.</p> <p>Before and after data are available for both treated sites and an untreated reference group.</p> <p>A separate comparison group may be required where the treatment has an effect on the reference group.</p>	<p>Reliable results with small sample sizes.</p> <p>Can include prior knowledge, spatial correlation, and complex model forms in the evaluation process.</p>	<p>Relatively complex.</p> <p>Cannot include prior knowledge of treatment.</p> <p>Cannot consider spatial correlation.</p> <p>Cannot specify complex model forms.</p>
Full Bayes	<p>Useful for before-after or cross-section studies when:</p> <p>Complex model forms are required.</p> <p>There is a need to consider spatial correlation among sites.</p> <p>Previous model estimates or CMF estimates are to be introduced in the modeling.</p>	<p>Possible to develop CMFunctions.</p> <p>Allows estimation of CMFs when conversions are rare.</p> <p>Useful for predicting crashes.</p>	<p>CMFs may be inaccurate for a number of reasons including:</p> <p>Inappropriate functional form.</p> <p>Omitted variable bias.</p> <p>Correlation among variables.</p>
Cross-Sectional	<p>Useful when limited before-after data are available.</p> <p>Requires sufficient sites that are similar except for the treatment of interest.</p>		

TABLE 12. Summary of Study Designs for Developing CMFs *continued*

Study Design	General Applicability	Strengths	Weaknesses
Case-Control	<p>Assess whether exposure to a potential treatment is disproportionately distributed between sites with and without the target crash.</p> <p>Indicates the likelihood of an actual treatment through the odds ratio.</p>	<p>Useful for studying rare events because the number of cases and controls is predetermined.</p> <p>Can investigate multiple treatments per sample.</p>	<p>Can only investigate one outcome per sample.</p> <p>Does not differentiate between locations with one crash or multiple crashes.</p> <p>Cannot demonstrate causality.</p>
Cohort	<p>Used to estimate relative risk, which indicates the expected percent change in the probability of an outcome given a unit change in the treatment.</p>	<p>Useful for studying rare treatments because the sample is selected based on treatment status.</p> <p>Can demonstrate causality.</p>	<p>Only analyzes the time to the first crash.</p> <p>Large samples are often required.</p>
Meta-Analysis	<p>Combines knowledge on CMFs from multiple previous studies while considering the study quality in a systematic and quantitative way.</p>	<p>Can be used to develop CMFs when data are not available for recent installations and it is not feasible to install the strategy and collect data.</p> <p>Can combine knowledge from several jurisdictions and studies.</p>	<p>Requires the identification of previous studies for a particular strategy.</p> <p>Requires a formal statistical process.</p> <p>All studies included should be similar in terms of data used, outcome measure, and study methodology.</p>

TABLE 12. Summary of Study Designs for Developing CMFs *continued*

Study Design	General Applicability	Strengths	Weaknesses
Expert Panel	Expert panels are assembled to critically evaluate the findings of published and unpublished research. A CMF recommendation is made based on agreement among panel members.	<p>Can be used to develop CMFs when data are not available for recent installations and it is not feasible to install the strategy and collect data.</p> <p>Can combine knowledge from several jurisdictions and studies.</p> <p>Does not require a formal statistical process.</p>	<p>Traditional expert panels do not systematically derive precision estimates of a CMF.</p> <p>Possible complications may arise from interactions and group dynamics.</p> <p>Possible forecasting bias.</p>
Surrogate Measures	Surrogate measures may be used to derive a CMF where crash data are not available or insufficient (e.g., there is limited after period data or the treatment is rarely implemented).	Can be used to develop CMFs in the absence of crash-based data.	<p>Not a crash-based evaluation.</p> <p>The approach to establish relationships between surrogates and crashes is relatively undeveloped.</p>

4. RESOURCES

This chapter provides several resources for selecting a study design and improving the completeness and consistency of reporting CMFs. Specifically, a flow chart is provided to help users select an appropriate study method to develop a CMF. A case study and several example scenarios are provided as an opportunity to practice using the flow chart. A sample annotated report outline is provided to help improve reporting consistency of CMFs, which will help others to assess the quality of the results.

4.1 FLOW CHART

The following flow chart (Figure 5) guides the selection of the preferred study design based on data availability and project goals. The first step is to determine whether or not a crash-based evaluation will be possible for the treatment of interest (i.e., do you have existing data for the treatment, or can you install and collect data for the treatment). The answer to this question will determine whether a traditional evaluation is appropriate (e.g., before-after, cross-sectional, etc.) or if it will be necessary to develop a CMF using meta-analysis or an expert panel. Several additional questions will guide the user through the thought process to identify an appropriate study design, alternative approach, or to conclude that it is not possible to develop a CMF at present. The use of the flow chart is demonstrated through several examples in Section 4.2.

Note that surrogate measures may be considered for developing CMFs when it is not possible or desirable to conduct a crash-based evaluation. Surrogate measures can be evaluated using the same methodologies for evaluating crash changes. The flow chart can be used to identify an appropriate study design, but the key to the application of surrogate measures is the availability of a reliable model to relate crash frequency to the surrogate measure.

Flow Chart Legend

EB = Empirical Bayes

FB = Full Bayes

CG = Comparison Group

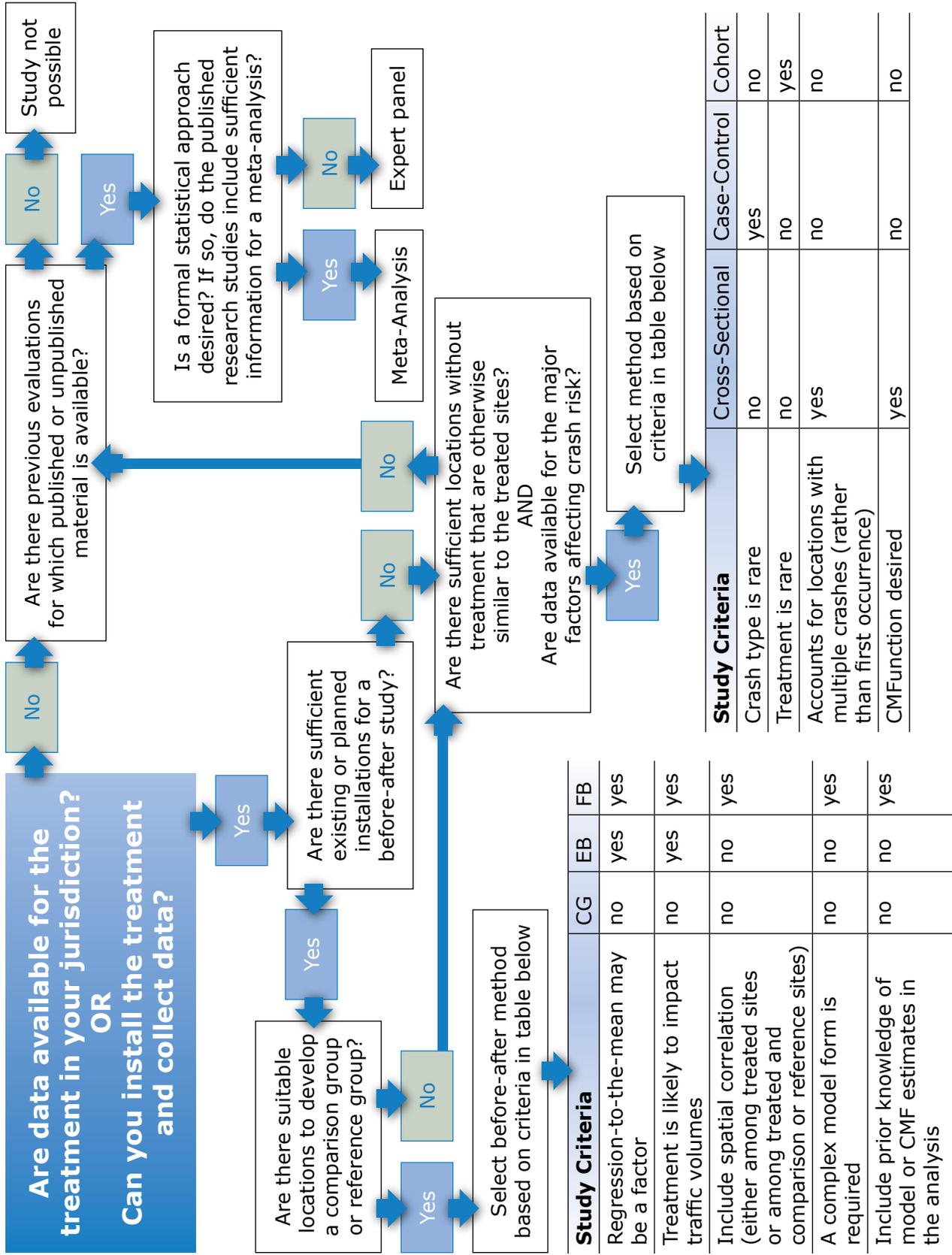


FIGURE 5. Flow Chart for Study Design Selection.

4.2 SAMPLE PROBLEMS

The following sample problems are designed to help the reader think through the process of selecting an appropriate study design. A case study is first provided as an example of the thought process for selecting a study design. Several scenarios are then presented, each indicating the study objective, data availability, and potential limitations. The reader is encouraged to read the scenarios and practice using the flow chart presented in Section 4.1 to identify an appropriate study design for each scenario. Each scenario is followed by a discussion of why a particular study design was chosen.

Case Study: Evaluation of the Safety Effectiveness of Red-Light Cameras

A city has hired a safety consultant to evaluate the safety effects of their red-light camera program. The program has been in place for three years and there are 35 cameras in total. Although previous evaluations of red-light cameras are available, the City wishes to conduct a new study using only their data because they believe that their program is unique when compared to other jurisdictions using red-light running cameras. Data are available for the past six years for both the camera-installed signalized intersections, the other 300 signalized intersections, and for unsignalized intersections in the City. The 35 intersections where cameras were installed were selected because they had a large number of right-angle crashes and they were located on highly traveled routes.

Following the flow chart, data do exist for the treatment sites so either a before-after or cross-sectional/case-control/cohort study may be possible. To determine if there is a sufficient sample size of treated sites for a before-after study, which is the preferred method of evaluation, the sample size estimate procedure discussed in Section 3.1, *Before-After with Comparison Group*, and detailed in Hauer (1997) is applied and it is determined that the 35 intersections with 3 years of before and 3 years of after data should provide a sufficient sample.

A suitable comparison and/or reference group is required to proceed with a before-after study. Not all signalized intersections were treated in the City. As such, there is a potential reference group of up to 300 signalized intersections, which is adequate for developing SPFs required for the empirical Bayes or full Bayes study designs. However, to control for time trends between the before and after periods, the untreated signalized intersections are not considered a good comparison group because of the potential spill-over effects from the red-light cameras. If the cameras do indeed reduce crashes at signalized intersections city-wide then using the untreated signalized sites as a comparison group would underestimate the benefits. The City does have data for unsignalized intersections that should not be subject to spill-over effects from the program and these sites could serve as a comparison group. For empirical Bayes or full Bayes studies, the reference group would be used to control for regression-to-the-mean in the before period and for traffic volume changes. The unsignalized intersection comparison group would be used to control for time trends between the before and after periods.

Suitable comparison and reference groups exist. Hence, we can proceed to the selection of the preferred before-after study design. Several factors influence this decision.

- There is likely to be regression-to-the-mean in the after period because sites were selected in part on a high number of angle crashes. Either the empirical Bayes or full Bayes designs would be preferred because the comparison group design cannot easily control for regression-to-the-mean.
- The added complexity of the full Bayes study design would not be warranted in this case because the City does not want to include any information on previous evaluations of red-light cameras in the analysis.
- The safety consultant does not believe that there are issues of spatial correlation or a necessarily complex model form that would warrant the more complex full Bayes study design.

To summarize, it is necessary to account for regression-to-the-mean because the treated sites were selected in part due to a high number of right-angle crashes. Since data exist for an empirical Bayes or full Bayes study, the comparison group method is ruled out as it cannot easily account for regression-to-the-mean. The empirical Bayes study design is selected because it is not deemed necessary to apply the full Bayes approach which involves a significantly more complex method.

Practice Scenarios

The following scenarios are intended to provide an opportunity to practice using the flow chart from Section 4.1 to select an appropriate study design. Each scenario identifies the need to estimate a CMF and provides background information on the data availability and potential data restrictions. To complete the following exercises, read the entire scenario and use the flow chart from Section 4.1 to select an appropriate study design. Following each scenario is a suggested study design with a detailed explanation of the thought process used in selecting the study design.

Scenario 1

A jurisdiction recently implemented a 1.5 second all-red phase at all traffic signals in their downtown area as a matter of practice, not as a result of a safety issue. They would like to develop a CMF for implementing the all-red phase. This was a system-wide treatment at 16 signalized intersections and there are no other signalized intersections remaining to develop a safety performance function. The signalized intersections are located along two main routes through the downtown area. All signalized intersections are four-legged. There are several two-way stop-controlled intersections that are located along the same two routes, in between the signalized intersections. The stop-controlled intersections are also all four-legged. It is reasonable to believe that the treatment does not have an impact on the safety of stop-controlled intersections in the area. Crash data were obtained for a six-year period, three years before and three years after treatment. The data track relatively well with respect to crash trends when comparing the signalized and stop-controlled intersections.

Discussion of Scenario 1

Selected Study Design: Comparison Group Before-After

Using the flow chart from Section 4.1, the first question to ask is whether or not data are available for the treatment of interest. It was indicated that the 1.5 second all-red phase was implemented at all traffic signals in the jurisdiction, so data are available for a crash-based analysis.

Following the flow chart, it is now necessary to assess whether or not there is a sufficient sample for a before-after study. In this case, the treatment was installed at 16 signalized intersections in a jurisdiction. From this information alone, it is difficult to determine if the sample size is adequate for a before-after study. Section 3.1, *Before-After with Comparison Group*, referred readers to Chapter 9 of Hauer (1997) for sample size estimation procedures. For this case, it is assumed that the sample size is adequate for a before-after study; however, the sample size may not be adequate to detect a change in safety with a high level of confidence.

The next consideration is the availability of a suitable reference group. In this scenario, it was noted that all signalized intersections were treated, leaving no sites for a reference group; however, comparison sites are available from the group of two-way stop-controlled intersections that are located along the treatment corridors. The stop-controlled intersections are similar to the treated signalized intersections (i.e., number of approaches and traffic volume on the major road) and the treatment is not expected to impact safety at the stop-controlled intersections because it is a signal modification. As such, the stop-controlled intersections can be used to account for changes in factors that may affect safety other than the treatment of interest.

The flow chart has indicated that a before-after study design may be appropriate for this evaluation. There are three before-after study designs to choose from, but each is employed for different reasons and involves various levels of complexity.

- In this case, it is not necessary to account for spatial correlations or include prior information about the treatment. It is also not expected that a complex model form will be necessary. As such, the full Bayes method is crossed-off the list. While it could be employed for this evaluation, it would add an unnecessary level of complexity.
- While regression-to-the-mean may be present, it is not a particular concern because the strategy was installed as a blanket treatment (i.e., all signals were treated) and the sites were selected as a matter of practice, not based on crash history. The treatment is an operational measure, but it is not expected to influence the entering traffic volumes at the intersections because it is applied to all sites and involves a minor change to the signal timing. As such, the empirical Bayes method is not necessary.
- The comparison group before-after study design is the logical choice for this scenario.

Scenario 2

A jurisdiction desires to develop a CMF for converting two-way stop-controlled intersections to roundabouts. Due to a large elderly population and the limited number of existing roundabouts in this area of the country, there is a concern that drivers may have difficulty with this new type of intersection. It is believed that the safety benefits in this jurisdiction may possibly be less than those found elsewhere. The jurisdiction is interested in developing a single CMF value that can be applied.

No new roundabout conversions will be undertaken until after the study, so only retrospective data will be available. Data will only be used from this jurisdiction and limited before and after data exist (five years before and two years after) for the 10 converted sites in the jurisdiction. All of the converted sites are similar in terms of area type, number of approaches, number of lanes, and traffic volumes. However, the number of locations is relatively small. Thus, there is concern that the limited data may make an evaluation of crash effects difficult.

The converted sites were selected to improve traffic operations, but were also selected due to a high number of angle crashes, a crash type that is eliminated through roundabouts. Few locations have been converted from a large pool of two-way stop-controlled intersections so a reference group is readily available.

The conversion to roundabouts is likely to change traffic volumes, particularly if the anticipated traffic operation improvements materialize.

Discussion of Scenario 2

Selected Study Design: Empirical Bayes Before-After

Using the flow chart from Section 4.1, the first question to ask is whether or not data are available for the treatment of interest. These data do exist because there are ten existing conversions, so data are available for a crash-based analysis.

Following the flow chart, it is now necessary to assess whether or not there is a sufficient sample for a before-after study. To determine if there is a sufficient sample size of treated sites for a before-after study, which is the preferred method of evaluation, the sample size estimate procedure discussed in Section 3.1, *Before-After with Comparison Group*, and detailed in Hauer (1997) is applied. Although the number of sites and years of after period are small, previous evaluations of roundabout conversions have estimated large crash reductions so the sample size may be more viable than at first glance. Assuming crash reductions of 30 percent to 70 percent, depending on crash severity, it is determined that the 10 intersections with five years of before and two years of after data should provide a sufficient sample.

The next consideration is the availability of a suitable reference group. In this scenario, few locations have been converted and a large reference group exists. There are unlikely to be spill-over effects due to roundabout conversion so an additional comparison group is not required.

The flow chart has indicated that a before-after study design may be appropriate for this evaluation. There are three before-after study designs to choose from, but each is employed for different reasons and each is associated with various levels of complexity.

- In this case, it is not necessary to account for spatial correlations or to include prior information about the treatment. It is also not expected that a complex model form will be necessary. As such, the full Bayes method is crossed-off the list. While it could be employed for this evaluation, it would add an unnecessary level of complexity.
- Traffic volumes at the treated sites are likely to change since the roundabouts are expected to improve traffic operations. The change in traffic volume can be accounted for using either the comparison group or empirical Bayes method.
- Regression-to-the-mean is likely present since the sites were selected in part based on crash history. The comparison group method cannot easily account for this potential bias.
- The empirical Bayes before-after study design is the logical choice for this scenario.

Scenario 3

In the example for selecting an empirical Bayes before-after study, a jurisdiction was interested in developing a CMF for converting two-way stop-control intersections to roundabouts, using only data from their jurisdiction. The empirical Bayes approach is well suited to satisfy all the needs and requirements of that study. However, there was a concern of limited sample size.

Suppose now, that instead of 10 locations there were only five converted sites and these sites had only three years of data before and one year of data after. Also, it is believed that although the safety benefits in this particular jurisdiction may be different from other areas, it is still reasonable to consider the knowledge base of CMFs for similar conversions in other jurisdictions.

Discussion of Scenario 3

Selected Study Design: Full Bayes Before-After

Similar to Scenario 2, the flow chart from Section 4.1 leads to the selection table for a suitable before-after method. Due to the small sample size, it is not anticipated that the results will be reliable, or in technical terms, “statistically significant” on their own. The full Bayes before-after study is selected over the empirical Bayes method because there are two important benefits, even though the analytical complexities are large.

1. Full Bayes can provide statistically significant results with smaller sample sizes of data.
2. Full Bayes modeling can include prior information on CMFs from other jurisdictions.

Scenario 4

There is a desire to estimate a CMF for flattening the curvature of horizontal curves with sharp radii on two-lane rural roads. The agency’s crash data system has recently been updated and crash data are available in the latest format for a five year period. There are records for some curves which have undergone reconstruction but these are few in number and many of these were completed more than 5 years ago. The available dataset consists of approximately 1,000 miles of roadway and 350 curves on rural two-lane roads. A preliminary investigation showed that the average crash rate at horizontal curves is three crashes per curve per year. Data on curve radii are available as are other geometric and traffic volume data.

Discussion of Scenario 4

Selected Study Design: Cross-sectional

Using the flow chart from Section 4.1, the first question to ask is whether or not data are available for the treatment of interest. These data do exist for 350 curves on two-lane rural roads. Data for other geometric and traffic factors affecting crash risk are also available.

Following the flow chart, it is now necessary to assess whether or not there is a sufficient sample for a before-after study. There are very few records for reconstructed curves. Hence, a before-after study is not possible, particularly so because even the ones available will have a wide range of curve radii before and after treatment.

Since a before-after study is not possible, the next consideration is whether or not data exist for similar sites with and without treatment. In this case the answer is yes because the 350 existing curves are expected to have a wide range of horizontal curvature. Since they are all on two-lane rural roads it is also expected that traffic volumes and other geometric variables affecting crash risk will be similar throughout the database of curves.

The flow chart indicates that a cross-sectional, case-control, or cohort study design may be appropriate for this evaluation. While there are three potential study designs to choose from, each is employed for different reasons.

- In this case, with 350 curves and five years of data, neither crashes nor treatment variations are rare.
- It would be expected that as the radii become very small the rate of increase in crash risk would increase. This may warrant the consideration of number of expected crashes with respect to curve radii.
- Data are available for traffic volume and other geometric data affecting crash risk.
- The cross-section study design is the logical choice for this scenario.

Scenario 5

A State is considering options for spending their High Risk Rural Roads funding. A major safety concern is run-off-road crashes on two-lane rural roads. A common issue identified on the most hazardous of these roads is narrow paved shoulders. There are several miles of two-lane, rural roads with narrow shoulders (0 – 2 feet) and several more miles with more substantial shoulders (3 – 4 feet). The State would like to determine a CMF for increasing shoulder width from 0 – 2 feet to 3 – 4 feet. There are few sites in the State where shoulders have been improved in this manner and they do not intend to implement this type of treatment until they can show a positive safety effect on run-off-road crashes.

Another consideration is sample size. While the total number of run-off-road crashes on two-lane, rural roads is relatively high, these crashes are spread-out over the network. Hence, there are several segments that do not experience any crashes over a three year period and several that experience only one or two crashes in three years.

Geometric and traffic volume data are available for these segments, which can be used to control for factors other than the treatment.

Discussion of Scenario 5

Selected Study Design: Case-Control

Using the flow chart from Section 4.1, the first question to ask is whether or not data are available for the treatment of interest. These data do exist for several miles of roadway for both the 0 – 2 feet and 3 – 4 feet shoulder groups.

Following the flow chart, it is now necessary to assess whether or not there is a sufficient sample for a before-after study. Since there are few locations where this improvement has been made, a before-after study is not possible.

A before-after study is not possible, so the next consideration is whether or not data exist for similar sites with and without treatment. In this case, the answer is yes because there are two groups of sites being compared, the 0 – 2 feet and 3 – 4 feet shoulder groups.

The flow chart indicates that a cross-sectional, case-control, or cohort study design may be appropriate for this evaluation. While there are three study designs to choose from, each is employed for different reasons.

- In this case, the crashes are considered to be relatively rare and spread out across the network. Development of a cross-sectional model could be difficult in this instance.
- The cohort method is a potential for this scenario, but it could be problematic if there are several segments that do not experience any crashes during the study period.
- Data are available for traffic volume and other geometric data affecting crash risk.
- The case-control study design is the logical choice for this scenario because sites can be selected to ensure an adequate sample of sites with and without crashes (i.e., cases and controls).

Scenario 6

Consider now the previous scenario, but instead of looking at all two-lane rural roads, the State wishes to develop a separate CMF to be used in mountainous regions. The safety concern is still run-off-road crashes, but crashes are more prevalent on two-lane rural roads in mountainous regions; most segments experience at least one crash per year. There are fewer miles of two-lane rural roads for analysis and there have been no recent projects to upgrade narrow shoulders (0 – 2 feet) to more substantial shoulders (3 – 4 feet). The State would like to determine a CMF for increasing shoulder width from the 0 – 2 feet range to the 3 – 4 feet range. They do not intend to implement this type of treatment until they can show a positive safety effect on run-off-road crashes.

Geometric and traffic volume data are available for these segments, which can be used to control for factors other than the treatment.

Discussion of Scenario 6

Selected Study Design: Cohort

Similar to Scenario 5, the flow chart from Section 4.1 indicates that a cross-sectional, case-control, or cohort study design may be appropriate for this evaluation. In this case however, the treatment is considered rare because the analysis is being restricted to mountainous regions. The one advantage is that crashes are slightly less rare and in fact most segments experience at least one crash.

- In this case, the treatment is rare. Development of a cross-sectional model or case-control analysis could be difficult in this instance.
- Since most segments experience at least one crash, it will be difficult to identify controls for use in a case-control design.
- Data are available for traffic volume and other geometric data affecting crash risk.
- The cohort study design is the logical choice for this scenario.

4.3 IMPROVING THE COMPLETENESS AND CONSISTENCY IN CMF REPORTING

It is the responsibility of the user to determine the quality of a CMF before applying it to a specific situation. However, it is often the case that insufficient information is provided to make an appropriate assessment. Factors and issues affecting the quality of CMFs were presented in Section 2.3. The CMF Clearinghouse (FHWA, 2010) and Highway Safety Manual (AASHTO, 2010) are sources of CMFs from previous studies. Both sources have made a valiant attempt at considering these issues in providing an indication of quality for many CMFs.

The following provides an overview of the evaluation criteria used in the CMF Clearinghouse and Highway Safety Manual to illustrate the level of detail needed to adequately assess the quality of a CMF. Following the discussion of the CMF Clearinghouse and Highway Safety Manual, a sample annotated report outline is provided to help improve the level of detail and consistency in the reporting of CMFs. By following this outline, researchers can ensure that they are reporting the necessary information for users to judge the quality of CMFs derived from their efforts.

CMF Clearinghouse

A five point rating serves as the primary method for indicating the quality of a CMF. Elements that contribute to the overall quality rating include study design, sample size, standard error, potential bias, and data source. Each element is identified from the underlying study and classified as excellent, fair, or poor. Points are assigned to each of the five elements based on the level of rigor (i.e., excellent, fair, or poor) and a final rating is computed based on a weighted point score. When information is not available for a specific element, the element does not contribute any points to the overall score, reducing the overall quality rating.

The overall quality rating reflects the accuracy and precision of the CMF as well as the general applicability of the results. Accuracy indicates how close the CMF is to the true value and depends on the type of study and potential sources of bias. Precision indicates the relative size of the confidence interval based on the sample size and standard error. The applicability of the results depends on the number of jurisdictions included in the evaluation.

The Clearinghouse does not provide a specific indication of the statistical significance of the results because this depends on the desired confidence level. Instead, the user is provided with all of the information necessary to determine statistical significance (i.e., point estimate, standard error, and instructions for computing confidence intervals for various levels of significance).

In order to facilitate the consideration of a new CMF for inclusion in the CMF Clearinghouse, the documentation of the new CMF should include sufficient detail for the five elements used for evaluating its quality. This information includes:

- *Study design*: The study design used to develop the CMF (i.e., comparison-group before-after, cross-sectional using regression models, etc.).
- *Sample size*: The number of sites and crashes in the treatment group and comparison or reference group in all time periods analyzed.
- *Standard error*: The variability of the outcome measure (i.e., the standard error, variance, or confidence interval for the CMF).
- *Potential bias*: Discussion of any potential biases to the data and how they were or were not accounted for. This may include potential spill-over or crash migration issues, traffic volume changes, regression-to-the-mean, and differences in crash reporting over time or between jurisdictions.
- *Data source*: Discussion of the sources of all data and any steps and assumptions made in transforming the raw data for analysis.

Highway Safety Manual

The development of the Highway Safety Manual considered the inclusion of many CMFs for various treatments. The literature review developed a procedure for re-estimating reported CMFs and their standard errors to reflect the quality of the study. This procedure, fully documented in Bahar (2010), involved the following steps:

1. Determine estimate of safety effect of treatment as documented in respective evaluation study publication.
2. Adjust estimate of safety effect to account for potential bias from regression-to-the-mean and changes in traffic volume.
3. Determine ideal standard error of safety effect.
4. Apply method correction factor (MCF) to ideal standard error, based on evaluation study characteristics.
5. Adjust corrected standard error to account for bias from regression-to-the-mean and changes in traffic volume.
6. Combine CMFs when specific criteria are met.

Steps 1 to 5 use information in the original documentation of the CMF to make the adjustments to reported CMF value and the standard error. This information includes:

- Study design used to estimate the CMF.
- Reported CMF and its standard error.
- Selection of treatment sites (i.e., if selected based on high crash counts).
- A summary of years of data used and number of observed crashes in all time periods.
- Changes in traffic volume and how they were or were not accounted for.

For step 6, where multiple CMFs exist for the same countermeasure, it is believed that the desired practice is not to merely select the highest rated CMF, but to combine the knowledge from all relevant studies for the same countermeasure. With this principle in mind, some of the CMFs in the Highway Safety Manual have been derived by combining CMF estimates from multiple studies, as described in Bahar (2010). The process also estimates the level of uncertainty in the combined CMF.

Measures of uncertainty are used to decide whether the CMF is sufficiently robust for inclusion in the Highway Safety Manual. The basis of the inclusion process is an accuracy test, which measures how likely the CMF value would be to substantially change if it were to be updated with knowledge from some future study. CMFs that do not pass the accuracy test are not recommended to be included in the Highway Safety Manual. It is recommended that even if a CMF passes the accuracy test, if the CMF is in conflict with generally accepted knowledge (e.g. the treatment is shown to increase crashes when all other studies of acceptable quality have shown a decrease), then the CMF should be reviewed by an expert panel prior to inclusion.

Sample Annotated Report Outline

The following outline identifies the basic information that should be included in a research report that documents CMFs. Use of this outline will help to improve consistency in the type of information that is reported, allowing a more complete evaluation of the quality of CMFs. Note that an abstract/executive summary, introduction, and conclusion sections are typically included in a report. These are not included in the annotated outline because they provide a summary of the relevant information documented in the body of the report.

Objective – this section should identify the treatment of interest, discuss the reason for conducting the study, and identify the target crash types and severities investigated (e.g., total crashes, injury crashes, angle crashes, etc.).

Background – this section should describe the treatment of interest, including details on its application. For example, a treatment may be applied and investigated on two-lane, undivided, rural roads. Items such as geometric characteristics are important to note so users of the CMF can determine the general applicability of the results.

Literature Review – this section should contain a summary of recent and salient literature related to the treatment of interest. This type of information is useful for comparing the consistency of results from the current study with the results of previous studies. A review of relevant literature is also useful for identifying potential variables to consider in the analysis. There are several sources for identifying CMFs from previous studies, including the CMF Clearinghouse (FHWA, 2010).

Methodology – this section should provide a discussion of the method used to develop the CMF. It is important to identify potential sources of bias in the analysis and how these biases are addressed (and those that cannot be addressed) using the selected method.

Data – this section should provide an overview of the data, including the data source(s), years of data, number of sites (and or miles of sites if applicable), average crashes per year, annual traffic volume, average traffic volume, minimum traffic volume, and maximum traffic volume. Similar to the background section, this information is useful for identifying the applicability of the CMFs developed from these data. It is also useful to provide this information for both the before and after periods when conducting a before-after study.

Results – this section should present the CMFs derived from the underlying study. It is important to include both the estimate of the CMF and the standard error. The standard error is used to calculate the confidence interval and, in general, used to judge the quality and significance of the results.

4.4 SUMMARY

This chapter provided several resources for developing and reporting CMFs. Specifically, a flow chart was provided to help guide readers through the study design selection process. Several sample scenarios were presented to provide the reader with an opportunity to practice using the flow chart. Finally, a sample annotated outline was provided to encourage consistency in the reporting of CMFs and underlying study details. If researchers provide more complete information related to their study and present the information in a consistent format, it will be easier for users to identify and assess the quality of CMFs.

5. CONCLUSION

While there are several available resources related to the identification and application of CMFs, there is relatively little guidance on the development of CMFs. Existing literature related to the development of CMFs mainly focuses on individual methods. This guide fills this void by providing a thorough overview of the CMF development process, including appropriate methods for developing reliable CMFs and issues to consider when applying the various methods. It illustrates that there are a number of methods available to estimate CMFs, and the most appropriate method depends on a number of factors that focus on the type and availability of data. A flowchart is provided to assist agencies in identifying the method that most closely meets their needs. The case study and scenarios demonstrate various situations for which each method might apply. The body of CMFs is ever increasing and the information presented herein will help practitioners, consultants and researchers develop more reliable CMFs.

REFERENCES

- Agüero-Valverde, J. and P.P. Jovanis. Spatial Analysis of Fatal and Injury Crashes in Pennsylvania. *Accident Analysis and Prevention*, Vol. 38, Issue 3, 618-615, 2006.
- American Association of State Highway Transportation Officials (AASHTO). *Highway Safety Manual*, 1st Edition, Washington, DC, 2010.
- Bahar, G., M. Masliah, C. Mollett, and B. Persaud. Integrated Safety Management Process. *NCHRP Report 501*, Transportation Research Board, National Cooperative Highway Research Program, Washington, DC, 2003. Available online at: http://onlinepubs.trb.org/Onlinepubs/nchrp/nchrp_rpt_501.pdf <Accessed May 17, 2010>
- Bahar, G. Methodology for the Development and Inclusion of Crash Modification Factors in the First Edition of the Highway Safety Manual. Transportation Research Board, *Transportation Research Circular*, Number E-C142, April 2010.
- Bonneson, J., K. Zimmerman, and K. Fitzpatrick. *Roadway Safety Design Synthesis*. Texas Transportation Institute for Texas DOT, 2005.
- Carriquiry, A. and M. Pawlovich, *From Empirical Bayes to Full Bayes: Methods for Analyzing Traffic Safety Data*, 2004. Available online at: http://www.iowadot.gov/crashanalysis/pdfs/eb_fb_comparison_whitepaper_october2004.pdf <Accessed May 17, 2010>
- Crash Modification Factors (CMF) Clearinghouse. Federal Highway Administration. Available online at: www.cmfclearinghouse.org <Accessed June 18, 2010>
- Egger, M., G. Davey Smith, and D.G. Altman, eds. *Systematic Reviews in Health Care. Meta-Analysis in Context*. BMJ publishing group, London, UK, 2001.
- Elvik, R. Introductory Guide to Systematic Reviews and Meta-Analysis. *Transportation Research Record 1908*, Washington, DC, 2005.
- Elvik, R. The Safety Value of Guardrails and Crash Cushions: A Meta-Analysis of Evidence from Evaluation Studies. *Accident Analysis and Prevention*, Vol. 27, Issue 4, 523-549, 1995.
- Elvik, R. *Measuring the Quality of Road Safety Evaluation Studies: Mission Impossible?* Paper Presented at Transportation Research Board 81st Annual Meeting, Special Session 539, Washington, DC, 2002. Available upon request, e-mail the author at re@tio.no.
- Elvik, R. and T. Vaa. *Handbook of Road Safety Measures*. Oxford, United Kingdom, Elsevier, 2004.
- Gan, A., J. Shen, and A. Rodriguez. *Update of Florida Crash Reduction Factors and Countermeasures to Improve the Development of District Safety Improvement Projects*. Florida Department of Transportation, 2005.
- Gross, F. *A Dissertation in Civil Engineering: Alternative Methods for Estimating Safety Effectiveness on Rural, Two-Lane Highways: Case-Control and Cohort Methods*. The Pennsylvania State University, December 2006.

Gross, F. and P.P. Jovanis. Estimation of the Safety Effectiveness of Lane and Shoulder Width: The Case-Control Approach. American Society of Civil Engineers, *Journal of Transportation Engineering*, Vol. 133, No. 6, 2007.

Gross, F. and P.P. Jovanis. Estimation of Safety Effectiveness of Changes in Shoulder Width using Case-Control and Cohort Methods. *Transportation Research Record 2019*, Washington, DC, 2008.

Harkey, D., R. Srinivasan, J. Baek, F. Council, K. Eccles, N. Lefler, F. Gross, B. Persaud, C. Lyon, E. Hauer, E. and J. Bonneson. Accident Modification Factors for Traffic Engineering and ITS Improvements. *NCHRP Report 617*, Appendix F, Transportation Research Board, National Cooperative Highway Research Program, Washington, DC, 2008. Available online at: http://www.trb.org/Publications/Blurbs/Accident_Modification_Factors_for_Traffic_Engineer_156844.aspx <Accessed May 17, 2010>

Hauer, E., D. Terry, and M. Griffith. Effect of Resurfacing on Safety of Two-Lane Rural Roads in New York State. *Transportation Research Record 1467*, Washington, DC, 1996.

Hauer, E. *Observational Before-After Studies in Road Safety*. Pergamon Press, Oxford, UK, 1997.

Hauer, E. Cause, Effect, and Regression in Road Safety: A Case Study. *Accident Analysis and Prevention*, Vol. 42, Issue 4, 1128-1135, 2010.

Lan, B., B. Persaud, and C. Lyon. Validation of a Full Bayes Methodology for Observational Before-After Road Safety Studies and Application to Evaluation of Rural Signal Conversions. *Accident Analysis and Prevention*, Vol. 41, Issue 3, Pages 574-580, 2009.

Lyon, C and B. Persaud. Safety Effects of a Targeted Skid Resistance Improvement Program. *Transportation Research Record 2068*. Washington, DC, 2008.

McGee, H., S. Taori, and B.N. Persaud. *NCHRP Report 491: Crash Experience Warrant for Traffic Signals*. Transportation Research Board, National Research Council, Washington, DC, 2003.

Pendleton, O. *Evaluation of Accident Analysis Methodology*. Report No. FHWA-RD-96-039, Federal Highway Administration, Washington, DC, 1996.

Pernia, J.C., J.J. Lu, M.X. Weng, X. Xie, and Z. Yu. *Development of Models to Quantify the Impacts of Signalization on Intersection Crashes*. Florida Department of Transportation, 2002.

Persaud, B. Statistical Methods in Highway Safety Analysis, A Synthesis of Highway Practice. *NCHRP Synthesis 295*, Transportation Research Board, National Cooperative Highway Research Program, Washington, DC, 2001. Available online at: http://onlinepubs.trb.org/onlinepubs/nchrp/nchrp_syn_295.pdf <Accessed May 17, 2010>.

Persaud B., F. Council, C. Lyon and M. Griffith. Multi-Jurisdictional Safety Evaluation of Red Light Cameras. *Transportation Research Record 1922*, Washington, DC, 2005.

Persaud, B. and C. Lyon. Empirical Bayes Before-After Safety Studies: Lessons Learned from Two Decades of Experience and Future Directions. *Accident Analysis and Prevention*, Vol. 39, Issue 3, 546-555, 2007.

Persaud, B., B. Lan, C. Lyon and R. Bhim. Comparison of Empirical Bayes and Full Bayes Approaches for Before-and-After Road Safety Evaluations. Accepted for publication in *Accident Analysis and Prevention* (June 2009).

Rodegerts, L., B. Persaud, and C. Lyon. Roundabouts in the United States. *National Cooperative Highway Research Program (NCHRP) Report 572*, Transportation Research Board, 2007. Available at: http://onlinepubs.trb.org/onlinepubs/nchrp/nchrp_rpt_572.pdf

Tsai Y.J., J.D. Wang and W.F. Huang. Case-control Study of the Effectiveness of Different Types of Helmets for the Prevention of Head Injuries among Motorcycle Riders in Taipei, Taiwan. *American Journal of Epidemiology*, Vol. 142, Issue 9, 974–81, 1995.

Vogt, A., and J.G. Bared. Accident Models for Two-Lane Rural Segments and Intersections. *Transportation Research Record 1635*. Washington, DC, 1998.

Washington, S., D. Lord, and B. Persaud. *The Use of Expert Panels in Highway Safety: A Critique*. Submitted for Publication November 2008. Available online at: https://ceprofs.civil.tamu.edu/dlord/Papers/Washington_et_al._Expert_Panel_Review_Critique.pdf <Accessed May 17, 2010>

Federal Highway Administration
Office of Safety
1200 New Jersey Avenue SE
Washington, DC 20590
<http://safety.fhwa.dot.gov/hsip>
Publication # FHWA-SA-10-032



*Highway Safety Improvement Program
Data Driven Decisions*