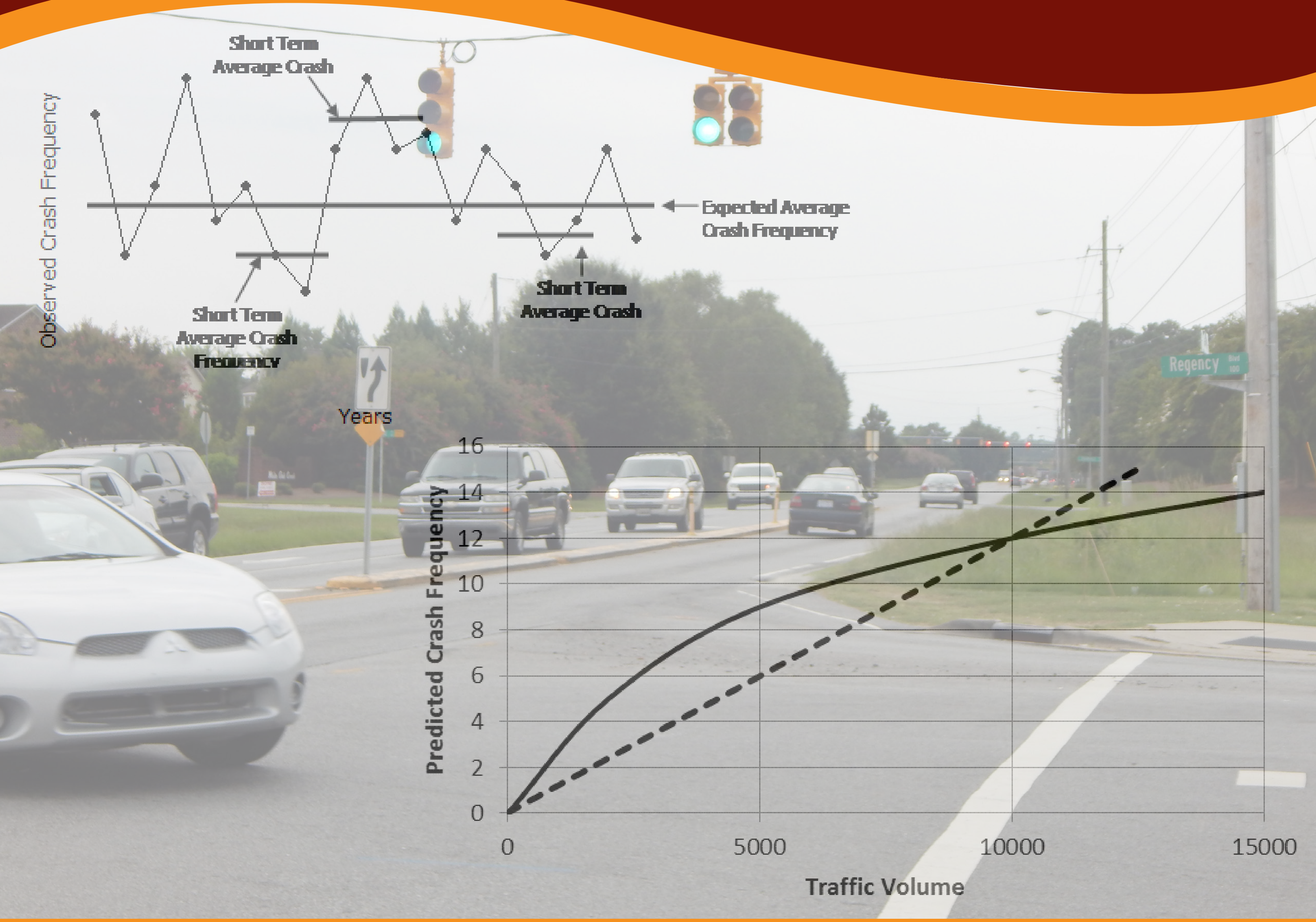
Reliability of Safety Management Methods

Safety Effectiveness Evaluation



FHWA-SA-16-040

March 2016

U.S. Department of Transportation
Federal Highway Administration

Safe Roads for a Safer Future
Investment in roadway safety saves lives
http://safety.fhwa.dot.gov

Notice

Quality Assurance Statement

Acknowledgements

The project team would like to thank the following individuals who served as members on the Technical Working Group for this effort.

Technical Working Group

Stuart Thompson, FHWA Office of Safety

Ray Krammes, FHWA Office of Safety

Yanira Rivera, FHWA Office of Safety

Karen Scurry, FHWA Office of Safety

Craig Thor, FHWA Office of Safety Research and Development

Matt Warren, Oklahoma Department of Transportation

TECHNICAL DOCUMENTATION PAGE

1. Report No. FHWA-SA-16-040	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Reliability of Safety Management Methods: Safety Effectiveness Evaluation		5. Report Date March 2016 6. Performing Organization Code
7. Author(s) Raghavan Srinivasan, Frank Gross, and Geni Bahar		8. Performing Organization Report No.
9. Performing Organization Name and Address VHB 8300 Boone Boulevard Suite 700 Vienna, VA 22182		10. Work Unit No. 11. Contract or Grant No. DTFH61-10-D-00022-T-13005
12. Sponsoring Agency Name and Address Federal Highway Administration Office of Safety 1200 New Jersey Ave., SE Washington, DC 20590		13. Type of Report and Period Final Report, March 2015 – February 2016 14. Sponsoring Agency Code FHWA
15. Supplementary Notes The Contracting Officer's Task Manager (COTM) was Stuart Thompson.		

16. Abstract

High quality data and reliable analytical methods are the foundation of data-driven decision-making. The Reliability of Safety Management Methods series includes five information guides that identify opportunities to employ more reliable methods to support decisions throughout the Roadway Safety Management process. Four of the guides focus on specific components of the Roadway Safety Management process: network screening, diagnosis, countermeasure selection, and safety effectiveness evaluation. The fifth guide focuses on the systemic approach to safety management, which describes a complimentary approach to the traditional methods related to network screening, diagnosis, and countermeasure selection. The purpose of the Reliability of Safety Management Methods series is to demonstrate the value of more reliable methods in these activities, and demonstrate limitations of traditional (less reliable) methods.

The Reliability of Safety Management Methods in Safety Effectiveness Evaluation guide describes various methods and the latest tools to support Safety Effectiveness Evaluation. The target audience includes data analysts and project managers involved in projects that impact highway safety. The objectives of this guide are to 1) raise awareness of more reliable methods, and 2) demonstrate through examples the value of more reliable methods in Safety Effectiveness Evaluation. This guide compares more reliable evaluation methods to traditional methods which are more susceptible to bias and may result in less reliable estimates and less effective decisions. Readers will understand the value of and be prepared to select more reliable methods in Safety Effectiveness Evaluation.

The guide includes five sections and an appendix. The first section introduces the Roadway Safety Management process and Safety Effectiveness Evaluation. The second section provides an overview of various methods for conducting Safety Effectiveness Evaluation, including a discussion of the associated strengths and limitations. The strengths and limitations focus on the ability (and inability) of the methods to account for issues in Safety Effectiveness Evaluation that can lead to less reliable results. The third section demonstrates the value of applying more reliable methods in Safety Effectiveness Evaluation. Simulated and empirical examples highlight the shortcomings of less reliable methods, which lead to less reliable results and conclusions. The next section summarizes the data requirements to employ the various methods. The final section describes available tools and resources to support Safety Effectiveness Evaluation. The Appendix presents further details on the simulated and empirical examples used to demonstrate the value of applying more reliable evaluation methods.

17. Key Words: Roadway, Safety, Data, Analysis, Effectiveness, Evaluation	18. Distribution Statement No restrictions.		
19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 35	22. Price

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

Table of Contents

List of Tables

List of Figures

List of Acronyms

AASHTO	American Association of State Highway and Transportation Officials
CMF	Crash Modification Factor
CPM	Crash prediction model
DOT	Department of Transportation
EB	Empirical Bayes
FHWA	Federal Highway Administration
RTM	Regression-to-the-mean
SE	Standard error
SPF	Safety Performance Function

Preface

High quality data and reliable analytical methods are the foundation of data-driven decision-making. The Reliability of Safety Management Methods series includes five information guides that identify opportunities to employ more reliable methods to support decisions throughout the Roadway Safety Management process. Four of the guides focus on specific components of the Roadway Safety Management process: network screening, diagnosis, countermeasure selection, and safety effectiveness evaluation. The fifth guide focuses on the systemic approach to safety management, which describes a complimentary approach to the traditional methods related to network screening, diagnosis, and countermeasure selection. The purpose of the Reliability of Safety Management Methods series is to demonstrate the value of more reliable methods in these activities, and demonstrate limitations of traditional (less reliable) methods.

The Reliability of Safety Management Methods in Safety Effectiveness Evaluation guide describes various methods and the latest tools to support Safety Effectiveness Evaluation. The target audience includes data analysts and project managers involved in projects that impact highway safety. The objectives of this guide are to 1) raise awareness of more reliable methods, and 2) demonstrate through examples the value of more reliable methods in Safety Effectiveness Evaluation. This guide compares more reliable evaluation methods to traditional methods which are more susceptible to bias and may result in less reliable estimates and less effective decisions. Readers will understand the value of and be prepared to select more reliable methods in Safety Effectiveness Evaluation.

The remainder of this guide includes five sections and an appendix. The first section introduces the Roadway Safety Management process and Safety Effectiveness Evaluation. The second section provides an overview of various methods for conducting Safety Effectiveness Evaluation, including a discussion of the associated strengths and limitations. The strengths and limitations focus on the ability (and inability) of the methods to account for issues in Safety Effectiveness Evaluation that can lead to less reliable results. The third section demonstrates the value of applying more reliable methods in Safety Effectiveness Evaluation. Simulated and empirical examples highlight the shortcomings of less reliable methods, which lead to less reliable results and conclusions. The next section summarizes the data requirements to employ the various methods. The final section describes available tools and resources to support Safety Effectiveness Evaluation. The Appendix presents further details on the simulated and empirical examples used to demonstrate the value of applying more reliable evaluation methods.

1. INTRODUCTION TO SAFETY EFFECTIVENESS EVALUATION

The Roadway Safety Management process is a six-step process as shown in Figure 1 and outlined in the Highway Safety Manual. (1) Safety Effectiveness Evaluation is the final step in the Roadway Safety Management process.

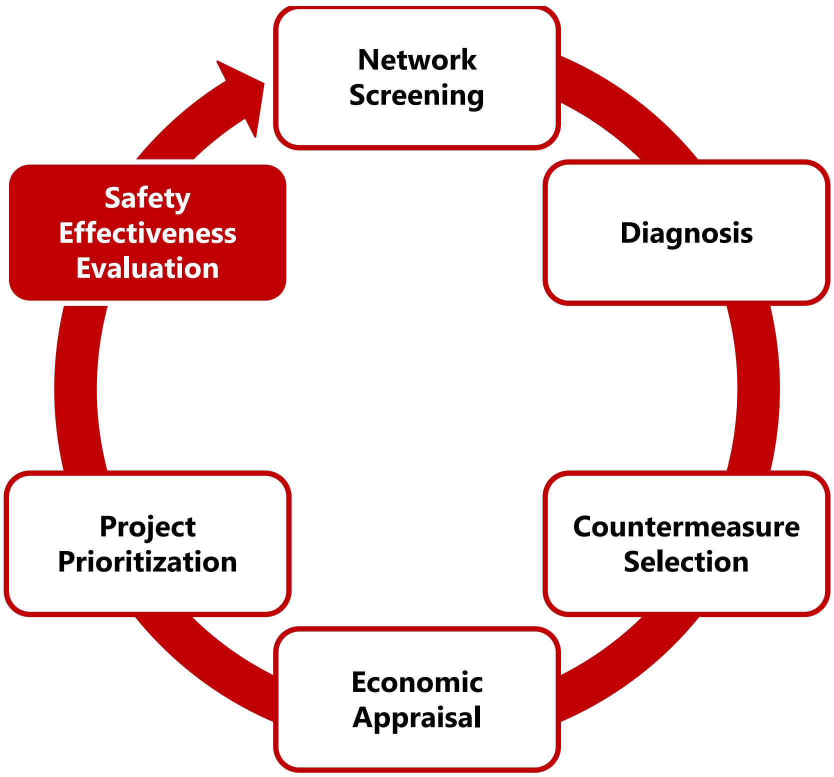


Figure 1. Diagram. Schematic of Roadway Safety Management process.

The objective of Safety Effectiveness Evaluation is to determine how a particular treatment (or group of treatments) has affected the safety performance (crash frequency and severity) at the treated locations. Agencies often use the results of such evaluations in future decisions about allocation of funds and changes to policies. In some cases, this evaluation leads to a crash modification factor (CMF[1]).

As discussed in the Highway Safety Manual, agencies can evaluate the safety effectiveness of a specific project implemented at one site or similar projects implemented at a group of sites. If the goal is to develop a CMF, then the analyst will need data from a group of sites.

[1] A CMF is a multiplicative factor that indicates the expected change in crashes associated with a treatment. A CMF of 1.0 indicates the treatment is not expected to change the crash frequency at a particular location. A CMF less than 1.0 indicates the treatment is expected to reduce the number of crashes. A CMF greater than 1.0 indicates the treatment is expected to increase the number of crashes.

2. OVERVIEW OF SAFETY EFFECTIVENESS EVALUATION METHODS

The Highway Safety Manual lists three basic study designs for Safety Effectiveness Evaluation:

- Experimental before-after studies.
- Observational cross-sectional studies.
- Observational before-after studies.

In experimental studies, researchers randomly select sites for treatment and control, and then administer the treatment to the treatment group. In observational studies, researchers do not randomly select sites for treatment. Instead, researchers are limited to sites already selected for treatment based on other reasons including safety concerns. Observational studies are more common than experimental studies in highway safety because agencies do not use random selection to identify sites for treatment. Hence, further discussion will focus on observational studies.

There are two broad classifications of observational studies: cross-sectional studies and before-after studies. Cross-sectional studies include those where "one is comparing the safety of one group of entities having some common feature (say, STOP controlled intersections) to the safety of a different group of entities not having that feature (say, YIELD controlled intersections), in order to assess the safety effect of that feature (STOP versus YIELD signs)." [2] Before-after studies include "all techniques by which one may study the safety effect of some change that has been implemented on a group of entities (road sections, intersections, drivers, vehicles, neighborhoods, etc.)." [2] Following is an overview of both observational cross-sectional and before-after studies.

OBSERVATIONAL CROSS-SECTIONAL STUDIES

In observational cross-sectional studies, analysts develop CMFs by comparing the safety of a group of sites *with* a feature to the safety of a group of sites *without* that feature. The CMF is the ratio of the average crash frequency of sites *with* the feature to the average crash frequency of sites *without* the feature. For this method to work, the two groups of sites should be similar in their characteristics except for the feature. In practice, this is difficult to accomplish and multiple variable regression models are typically used. These cross-sectional regression models are also called safety performance functions (SPFs) or crash prediction models (CPMs). SPFs or CPMs are mathematical equations that relate crash frequency and site characteristics. Analysts use the coefficients of the variables from these equations to estimate the CMF associated with a treatment. For example, suppose the intent is to estimate the CMF for shoulder width based on the following SPF that predicts the number of crashes per mile per year on rural two-lane roads in mountainous areas with paved shoulders in North Carolina. [3]

$$Y = \exp\left(0.8727 + 0.4414 * ln\left(\frac{AADT}{10{,}000}\right) + 0.4293 * \frac{AADT}{10{,}000} - 0.0164 * SW\right)$$

Figure 2. Equation. CMF calculation.

where:

Y = predicted number of crashes.
AADT = annual average daily traffic.
SW = width of the paved shoulder, in feet.

The estimated CMF for changing the shoulder width from three feet to six feet is the ratio of the predicted number of crashes when the shoulder width is six feet to the predicted number of crashes when the shoulder width is three feet. The following equation shows this calculation.

$$CMF = \frac{\exp\left[0.8727 + 0.4414 * \ln\left(\frac{AADT}{10{,}000}\right) + 0.4293 * \left(\frac{AADT}{10{,}000}\right) - 0.0164 * 6\right]}{\exp\left[0.8727 + 0.4414 * \ln\left(\frac{AADT}{10{,}000}\right) + 0.4293 * \left(\frac{AADT}{10{,}000}\right) - 0.0164 * 3\right]}$$

Figure 3. Equation. CMF for increasing shoulder width from three to six feet.

This ratio simplifies to the following equation. [Note the multiplier in the equation (-0.0164) is the coefficient for the shoulder width variable from the cross-sectional regression model.]:

$$CMF = \exp[-0.0164 * (6 - 3)] = 0.952$$

Figure 4. Equation. Simplified CMF for increasing shoulder width from three to six feet.

The primary criticism of CMFs estimated from cross-sectional studies is the potential for confounding. Confounding occurs when the method does not account for extraneous variables correlated with both the dependent variable (crash frequency) and the independent variables in the model. For example, if the intent is to estimate the safety effects of chevrons on horizontal curves and the curves with chevrons also have the worst roadside hazards, but the information on roadside hazards is not included in the model, then a cross-sectional study may incorrectly conclude that chevrons are associated with an increase in run-off-road crashes. Refer to Gross and Carter for further discussion about the statistical issues associated with estimating CMFs from observational cross-sectional studies. [4,5]

OBSERVATIONAL BEFORE-AFTER STUDIES

In observational before-after studies, analysts develop CMFs by comparing the frequency and severity of crashes before and after implementation of the treatment. The key to before-after studies is accounting for site selection bias and other changes over time such as changes in traffic volume and other temporal trends. The following is a brief explanation of site selection bias and how this leads to issues related to regression-to-the-mean (RTM).

Transportation agencies often select sites with high crash counts for treatment. This results in site selection bias because the agency selects sites based on crash history, not at random. When this is the case, there is potential for RTM. RTM describes the situation when periods with relatively high crash frequencies are followed by periods with relatively low crash frequencies simply due to the variability of crashes. [Note RTM also implies that periods with relatively low crash frequencies are likely to be followed by periods with relatively high crash frequencies.]

Figure 5 illustrates RTM, comparing the difference between short-term average and long-term average crash history. If an agency selects sites based on high short-term average crash history, then crashes at those sites may be lower in the following years due to RTM, even if the agency does not treat those sites. If RTM is not properly accounted for, then the results of a Safety Effectiveness Evaluation may incorrectly overestimate or underestimate the treatment effect.

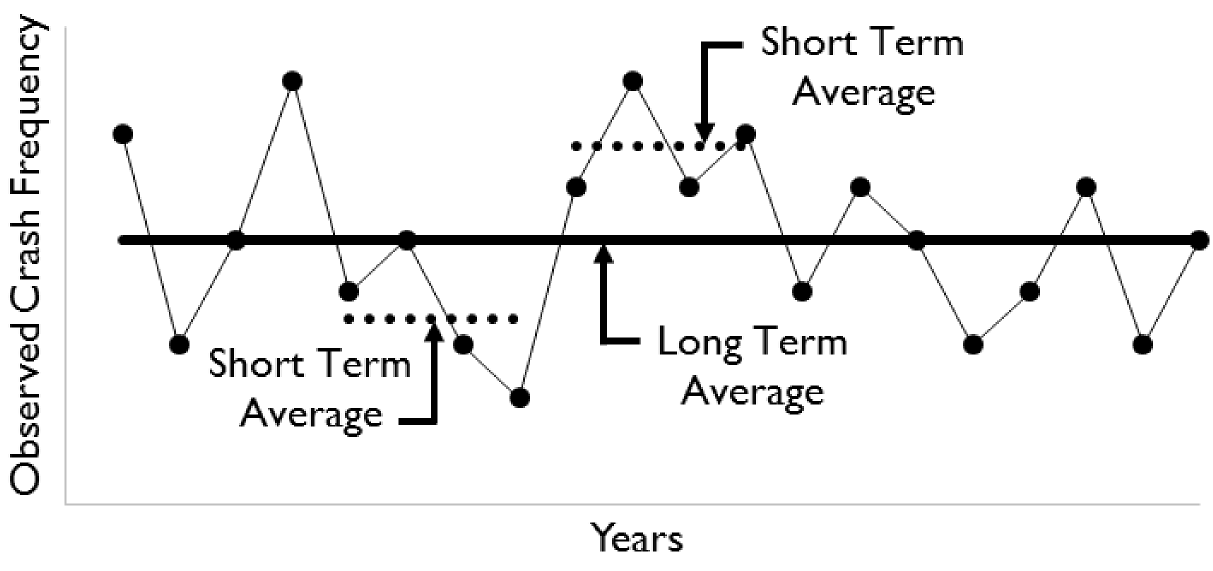


Figure 5. Chart. Illustration of RTM comparing short- and long-term averages.

In addition to RTM, before-after evaluations need to account for changes in traffic volume and other temporal trends that can impact crashes. Following is a discussion of four common methods for conducting before-after studies and the ability (or inability) to use each method to address RTM, changes in traffic volume, and temporal trends:

- Simple before-after study.
- Before-after study with traffic volume correction.
- Before-after study with comparison group.
- Empirical-Bayes before-after study.

Simple Before-After Study

A simple before-after study is a basic comparison of crashes before and after treatment. The safety effect of a treatment is assessed by directly comparing the crash frequency in the after period with the crash frequency in the before period. The simple before-after study design does not account for possible bias due to RTM and does not account for temporal effects or trends such as changes in traffic volume, changes in driver behavior, and changes in crash reporting.

Before-After Study with Traffic Volume Correction

A before-after study with traffic volume correction is a variation of the simple before-after study that accounts for changes in traffic volume over time. For example, comparing the crash rates (i.e., crashes per some measure of exposure such as vehicle miles traveled) before and after treatment rather than the crash counts helps to account for changes in traffic volume.

The traffic volume correction may be a linear or nonlinear trend. The use of crash rates implicitly assumes the relationship between crash frequency and traffic volume is linear; however, many studies have shown the relationship between crash frequency and traffic volume is nonlinear. Further, the use of crash rates may not account for the annual variation in traffic volume within the before and after periods. SPFs are a more reliable method to account for changes in traffic volume because they reflect the nonlinear relationship between crash frequency and traffic volume.

Figure 6 illustrates the difference between a linear and nonlinear trend to define the relationship between crash frequency and traffic volume. Hypothetically, if the traffic volume increases from 5,000 to 10,000 vehicles per day, the nonlinear trend from Figure 6 predicts a 25 percent increase in crashes (i.e., 9 crashes at 5,000 vehicles per day versus 12 crashes at 10,000 vehicles per day). Using the linear trend, this same increase in traffic volume is associated with a 50 percent increase in predicted crashes. Nonlinear traffic volume correction methods such as SPFs are more appropriate than linear traffic volume correction methods such as crash rates.

Note the traffic volume correction does not account for possible bias due to RTM, and does not account for temporal effects or trends such as changes in driver behavior and changes in crash reporting.

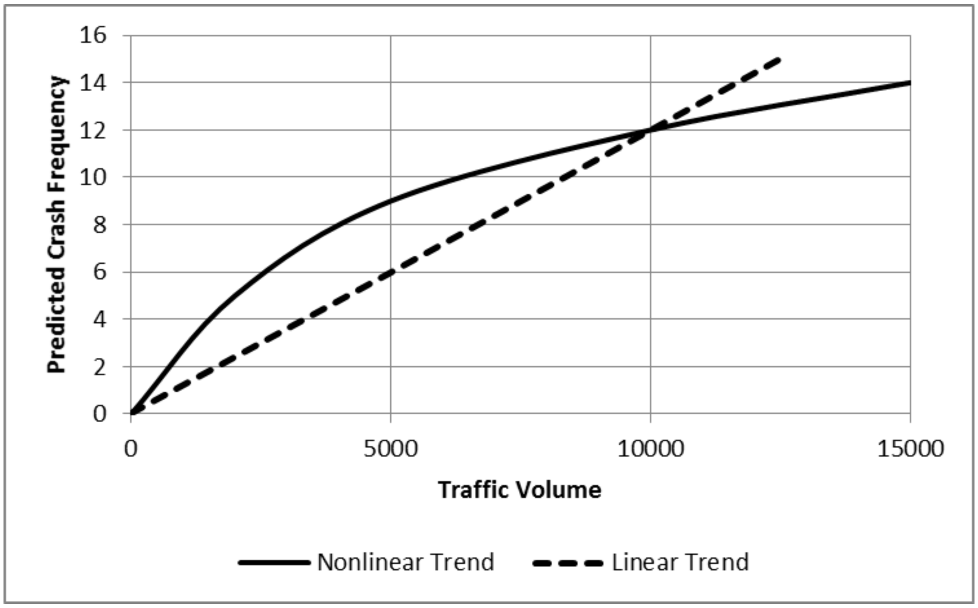


Figure 6. Graph. Relationships between crash frequency and traffic volume.

Before-After Study with Comparison Group

The before-after study with comparison group incorporates information from an untreated group of sites to account for temporal effects and changes in traffic volume. One way to apply this method is to use the comparison group to calculate a *comparison ratio,* which is the ratio of observed crash frequency in the after period to that in the before period. (4) The observed crash frequency in the before period at the treatment sites is multiplied by this comparison ratio to estimate the number of crashes at the treatment sites in the after period had the treatment not occurred. The estimated crashes at the treatment sites in the after period (had the treatment not occurred) is then compared with the observed crashes at the treatment sites in the after period to determine the safety impact of the treatment.

For this approach to work, the assumption is the trends in the crash counts in the treatment and comparison groups are similar. Hauer proposes a test that makes use of a sequence of sample odds ratios to determine if the trends in the two groups are indeed similar. (2) Analysts typically select comparison sites from the same jurisdiction as the treatment sites to increase the likelihood that comparison sites will have similar trends as the treatment sites.

Another possible approach for applying this method is to develop SPFs using data from the comparison group. Using SPFs, the ratio of the predicted crash frequency in the after period to the predicted crash frequency in the before period is used to estimate the comparison ratio. By estimating an SPF, this approach accounts for changes in traffic volume from the before to the after period, and the nonlinear relationship between crash frequency and traffic volume. The comparison group method does not account for RTM. Hence, it may be a viable approach if sites are not selected for treatment based on crash history (e.g., blanket installation of a treatment), which reduces the concern for possible bias due to RTM.

Empirical-Bayes Before-After Study

The Empirical-Bayes (EB) method addresses bias due to RTM, accounts for changes in traffic volume, and accounts for temporal effects. [2] The intent of the EB method is to estimate the expected number of crashes that would have occurred had there been no treatment, and compare that with the number of observed crashes after treatment. The following steps describe how to estimate the expected number of crashes that would have occurred had there been no treatment:

1. **Identify Reference Group**: Identify a group of sites without the treatment, but similar to the treatment sites in terms of the major factors affecting crash risk, including traffic volume and other site characteristics.
2. **Develop SPFs**: Using data from the reference sites, estimate an SPF relating crashes to independent variables such as traffic volume and other site characteristics. As discussed in the following steps, the EB method incorporates information from SPFs to predict crashes based on traffic volume and site characteristics. By selecting a reference group that is similar to the treatment group in terms of the major risk factors, analysts can reduce the possible bias due to confounding factors.

   It is important to note there are at least two variations of the EB method. For the purpose of this document, *EB using SPFs* is the title of the first variation and *EB using Method of Moments* is the title of the second variation. The method of moments is appropriate when SPFs are not available. Hauer discusses both methods. [2]
3. **Develop Annual Calibration Factors**: Agencies must calibrate SPFs annually to account for the temporal effects on safety (e.g., variation in weather, demography, and crash reporting). The annual calibration factor is the ratio of the observed crashes to the predicted crashes from the SPF based on data from the reference group.
4. **Estimate Predicted Crashes**: Use the SPFs, annual calibration factors, and traffic volume data to estimate the predicted number of crashes for each year in the before and after periods at each treatment site.
5. **Estimate Ratio of Predicted Crashes**: Using the results of step 4, compute the ratio of total predicted crashes after treatment to total predicted crashes before treatment.
6. **Estimate Expected Crashes Before Treatment**: Using the EB method, compute the expected crashes in the before period at each treatment site as the weighted sum of observed crashes before treatment and predicted crashes before treatment from step 4.
7. **Estimate Expected Crashes After Treatment**: For each treatment site, estimate the expected crashes after treatment as the product of the expected crashes before treatment (step 6) and the ratio of predicted crashes (step 5). This is the expected number of crashes that would have occurred had there been no treatment. The variance of this expected number of crashes is also estimated in this step.

CMFs are developed by comparing the expected number of crashes that would have occurred had there been no treatment to the observed crashes with treatment. For details, refer to Hauer or Gross. [2,4]

3. DEMONSTRATING THE VALUE OF MORE RELIABLE METHODS

At this time, there is general agreement within the safety research community that properly designed and conducted before-after studies provide more reliable results than cross-sectional studies for Safety Effectiveness Evaluation. Further, the EB before-after method provides more reliable results than the comparison group method, and the comparison group method provides more reliable results than the simple before-after method.

This section demonstrates the value of applying more reliable methods in Safety Effectiveness Evaluation to account for RTM bias, changes in traffic volume, the nonlinear relationship between crash frequency and traffic volume, and temporal effects. Simulated and empirical examples highlight the shortcomings of less reliable methods, which may lead to less reliable results and conclusions. For interested readers, the Appendix presents further details on the simulated and empirical examples. Note the examples illustrate the general comparative results of the methods. Different data and relationships within the data will produce different results. In general, the examples demonstrate the value of applying more reliable evaluation methods.

EXAMPLE 1: ACCOUNTING FOR REGRESSION-TO-THE-MEAN BIAS

The first example demonstrates the importance of using the EB method to account for RTM bias in before-after studies. RTM is a result of the random fluctuation in crashes over time. Sites with randomly high crashes in one period are likely to have fewer crashes in the subsequent period, even in the absence of other changes. Similarly, sites with randomly low crashes in one period are likely to have more crashes in the subsequent period. If RTM is present and not properly accounted for, then the analyst will incorrectly overestimate or underestimate the treatment effect. Refer to the appendix for details related to example 1.

Using 10 years of simulated data, this example is able to establish the ground truth for the treatment effect. Specifically, there is no treatment, so the ground truth is a CMF of 1.0 (i.e., no effect). The before period is defined as the first five years of simulated data, and the after period is defined as the last five years of simulated data. Four scenarios define the hypothetical selection of sites for further diagnosis and treatment. Scenarios 1 and 2 are two groups of high-crash sites. Scenarios 3 and 4 are two groups of low-crash sites.

Table 1 presents the CMFs estimated from the simple and EB before-after methods. A reliable method is one that yields a CMF not statistically different from the ground truth (i.e., the 95% confidence interval includes 1.0). In all four scenarios, the CMF estimated from the EB method is not statistically different from 1.0. Conversely, in three of the four scenarios, the CMF estimated from the simple before-after method is statistically different from 1.0.

It is clear the simple before-after method may overestimate or underestimate the safety effect of a treatment because it does not account for potential bias due to RTM. The EB method is able to account for potential bias due to RTM and produce a CMF not statistically different from the ground truth established in this simulation study. These results demonstrate the reliability of the EB method compared to the simple before-after method.

Table 1. Results of before-after evaluation using simulated data.

Scenario	Crashes in Before Period	Crashes in After Period	Evaluation Method	CMF	S.E. of CMF	95% Confidence Interval of CMF
1: Top 50 High-Crash Sites	338	237	Simple	**0.699**	0.059	(0.583, 0.815)
			EB	0.988	0.090	(0.812, 1.165)
2: Top 100 High-Crash Sites	618	499	Simple	**0.806**	0.048	(0.711, 0.901)
			EB	1.056	0.068	(0.923, 1.188)
3: Top 50 Low-Crash Sites	180	219	Simple	1.210	0.121	(0.973, 1.447)
			EB	0.906	0.084	(0.741, 1.071)
4: Top 100 Low-Crash Sites	355	459	Simple	**1.289**	0.091	(1.111, 1.467)
			EB	0.951	0.062	(0.830, 1.072)

Note: Bold indicates the CMF is statistically different from the ground truth (i.e., 1.0) at the 5% significance level.

EXAMPLE 2: ACCOUNTING FOR CHANGES IN TRAFFIC VOLUME

The second example further supports the use of the EB method to account for RTM bias, and demonstrates the importance of properly accounting for changes in traffic volume in before-after safety evaluations. Research studies have established that traffic volume is the most important predictor for most crash types. If the traffic volume changes, then the expected crash frequency changes accordingly. Further, research studies have shown the relationship between crashes and traffic volume is nonlinear. As such, methods based on nonlinear traffic volume correction such as SPFs are more reliable than methods based on linear traffic volume correction such as crash rates to account for changes in traffic volume.

The following example includes a safety evaluation of improved delineation at 89 horizontal curves on rural, two-lane roads in Connecticut. Specifically, it compares the results from simple before-after, before-after with linear traffic volume correction, before-after with nonlinear traffic volume correction, and EB before-after methods. The last two methods employ SPFs and properly account for the nonlinear relationship between crash frequency and traffic volume. Further, the EB method accounts for potential RTM bias and general temporal effects. While the ground truth is unknown for this example, it assumes the EB before-after method provides the most reliable estimate as demonstrated in example 1. Refer to the appendix for details related to example 2.

Table 2 presents the CMFs and associated standard errors estimated from the various methods. All CMFs are statistically significant at the five percent significance level. The simple before-after and before-after with traffic volume correction methods consistently estimate larger treatment effects (i.e., smaller CMFs) than the EB method. This suggests RTM bias is present and unaccounted for by the simple before-after and before-after with traffic volume correction methods. These differences may be due, in part, to changes in traffic volume and general temporal effects. Note the simple before-after method does not account for changes in traffic volume or temporal effects. While the before-after with traffic volume correction methods account for changes in traffic volume, they do not account for temporal effects.

The before-after with traffic volume correction methods (both linear and nonlinear) consistently produce larger treatment effects than the simple before-after method. This suggests there are changes in traffic volume, which are unaccounted for by the simple before-after method. While both the linear and nonlinear traffic volume correction methods produce similar results in this case, the use of SPFs (nonlinear correction) is more appropriate than a linear correction to account for changes in traffic volume.

The differences among the methods are likely due to potential bias due to RTM, changes in traffic volume, and trends during the study period. These results demonstrate the potential differences in estimates obtained from various methods, and reinforce the need to apply more reliable methods such as the EB method when conducting before-after evaluations. Otherwise, the results of the evaluation may be less accurate and less reliable.

Table 2. Results of curve delineation evaluation from Connecticut.

Crash Type	Simple	Traffic Volume Correction using Linear Trend	Traffic Volume Correction using SPFs (Nonlinear)	EB
	CMF (S.E.)	CMF (S.E.)	CMF (S.E.)	CMF (S.E.)
Total	**0.760** (0.073)	**0.700** (0.067)	**0.689** (0.066)	**0.822** (0.077)
Lane departure	**0.774** (0.081)	**0.712** (0.074)	**0.718** (0.075)	**0.823** (0.084)
Injury and fatal	**0.579** (0.101)	**0.537** (0.094)	**0.515** (0.091)	**0.747** (0.127)
Total crashes in dark conditions	**0.532** (0.088)	**0.489** (0.081)	**0.483** (0.080)	**0.647** (0.105)
Lane departure in dark conditions	**0.528** (0.094)	**0.486** (0.087)	**0.483** (0.086)	**0.658** (0.115)

Note: Bold indicates the CMF is statistically different from 1.0 at the 5% significance level.

EXAMPLE 3: ACCOUNTING FOR TEMPORAL TRENDS

The third example further supports the use of the EB method to account for RTM bias, and demonstrates the importance of properly accounting for changes in traffic volume and other temporal trends in before-after studies. Temporal trends are changes over time, such as changes in driver behavior and changes in crash reporting. It is important to account for temporal trends affecting safety performance; otherwise, the results of the safety evaluation also reflect these other changes and not the actual effect of the treatment.

The following example includes a safety evaluation of installing red light cameras at 24 signalized intersections in Arizona. Specifically, it compares the results from simple before-after, before-after with nonlinear traffic volume correction, before-after with comparison group, and EB before-after methods. The before-after with nonlinear traffic volume correction and EB before-after methods properly account for the nonlinear relationship between crash frequency and traffic volume. The before-after with comparison group and EB before-after methods account for temporal effects. Further, the EB before-after method accounts for RTM bias. While the ground truth is unknown for this example, it assumes the EB before-after method provides the most reliable estimate as demonstrated in example 1. Refer to the appendix for details related to example 3.

Table 3 presents the CMFs and associated standard errors estimated from the various methods. The four methods produced different results. The simple before-after method results in larger treatment effects than the EB method because it does not account for RTM, changes in traffic volume, or trends in the study period. While the results from the before-after with nonlinear traffic volume correction method accounts for changes in traffic volume, the results are inconsistent with the EB method because it does not account for possible bias due to RTM or temporal effects. The comparison group method accounts for temporal effects, but the results are inconsistent with the EB method because it does not account for possible bias due to RTM or changes in traffic volume at the treatment sites.

The differences among the methods are likely due to trends during the study period, changes in traffic volume, and potential bias due to RTM. These results demonstrate the potential differences in estimates obtained from various methods, and reinforce the need to apply more reliable methods such as the EB method when conducting before-after evaluations. Otherwise, the results of the evaluation may be less accurate and less reliable.

Table 3. Results of red light camera evaluation from Arizona.

Approach	Jurisdiction	Evaluation Method	CMF (S.E.)		
			Angle	Left-turn	Rear-end
Target Approaches	Phoenix	Simple	**0.61** (0.16)	0.96 (0.11)	**1.58** (0.24)
		Comparison Group	**0.58** (0.13)	0.90 (0.08)	**1.51** (0.17)
	Scottsdale	Simple	0.80 (0.14)	**0.52** (0.06)	**1.57** (0.18)
		Traffic Volume Correction	0.83 (0.15)	**0.52** (0.06)	**1.67** (0.22)
		EB	0.80 (0.14)	**0.55** (0.06)	**1.41** (0.11)
All Approaches	Phoenix	Simple	0.90 (0.15)	1.06 (0.09)	**1.26** (0.13)
		Comparison Group	0.86 (0.12)	0.99 (0.07)	**1.20** (0.10)
	Scottsdale	Simple	**0.69** (0.09)	**0.59** (0.06)	**1.48** (0.10)
		Traffic Volume Correction	**0.70** (0.09)	**0.59** (0.06)	**1.62** (0.13)
		EB	**0.83** (0.08)	**0.60** (0.05)	**1.45** (0.06)

Note: Bold indicates the CMF is statistically different from 1.0 at the 5% significance level.

EXAMPLE 4: A DECADE OF EMPIRICAL EVIDENCE IN SAFETY EFFECTIVENESS EVALUATIONS

The fourth example is a compilation of safety effectiveness estimates from seven studies spanning more than a decade. Specifically, it compares the results from the before-after with linear traffic volume correction and EB before-after methods. The estimates from the before-after with linear traffic volume correction do not control for potential bias due to RTM or time trends in crash occurrence unrelated to the treatment. While the before-after with linear traffic volume correction method accounts for changes in traffic volume, it assumes a linear relationship between crash frequency and traffic volume, which is typically invalid. While the ground truth is unknown for this example, it assumes the EB before-after method provides the most reliable estimate as demonstrated in example 1. Accordingly, this example further supports the conclusions from the previous sections, namely the importance of using the EB method to properly account for RTM bias, changes in traffic volume, and general temporal effects in before-after safety evaluations.

Table 4 presents 62 CMFs for various crash types and severities, including 31 CMFs based on the before-after with linear traffic volume correction method and 31 CMFs based on the EB before-after method. The last column in Table 4 presents the ratio of percent reduction estimated from the two methods, computed as the percent reduction in crashes from the before-after with linear traffic volume correction method divided by the percent reduction from the EB method. [Note the percent reduction is 100*(1-CMF).] A ratio greater than 1.0 indicates the before-after with linear traffic volume correction method estimates a larger effect than the EB method. A ratio less than 1.0 indicates the before-after with linear traffic volume correction method estimates a smaller effect than the EB method. Refer to the appendix for details related to example 4.

In 23 of the 31 comparisons (74 percent), the percent reduction estimated from the before-after with linear traffic volume correction is greater than the percent reduction estimated from the EB method. Of these, 10 of the 23 estimates from the before-after with linear traffic volume correction are 1.0 to 1.25 times as large as the estimates from the EB method. The remaining 13 of the 23 estimates from the before-after with linear traffic volume correction are more than 1.25 times as large as the estimates from the EB method.

Table 4. Results of a decade of EB before-after evaluations.

Crash Type	Crashes/mile year (segments) or crashes/year (intersections) in after period		CMF		Ratio of percent change[1]
	Expected crashes from linear traffic volume correction	Observed crashes	Before-after with linear traffic volume correction	EB	
Two-way left-turn lanes on rural two-lane roads (6)					
Total	8.66	5.89	0.68	0.64	0.89
Injury	1.50	0.74	0.49	0.65	1.45
Rear-end	3.62	1.67	0.46	0.53	1.15
Offset left turn lanes (Wisconsin) (7)					
Total	7.31	4.7	0.64	0.66	1.06
Injury	3.32	1.9	0.57	0.64	1.20
Left-turn opposing	3.13	1.9	0.58	0.62	1.12
Rear-end	2.09	1.5	0.72	0.68	0.89
Improve curve delineation (8)					
Injury	2.90	1.89	0.64	0.82	2.00
Dark	2.70	1.51	0.56	0.73	1.60
Lane departure dark	2.38	1.34	0.56	0.75	1.72
Centerline plus shoulder rumble strips on two-lane rural roads (9)					
Total	0.567	0.463	0.82	0.80	0.92
Injury	0.254	0.183	0.72	0.77	1.21
Run-off-road	0.189	0.123	0.65	0.74	1.35
Head-on	0.024	0.014	0.58	0.63	1.15
Sideswipe	0.031	0.015	0.49	0.77	2.20

Crash Type	Crashes/mile year (segments) or crashes/year (intersections) in after period		CMF		Ratio of percent change[1]
	Expected crashes from linear traffic volume correction	Observed crashes	Before-after with linear traffic volume correction	EB	
Wet-reflective pavement markings (10)					
Total multilane	5.78	3.66	0.63	0.83	2.10
Injury multilane	1.93	1.10	0.57	0.60	1.06
Injury freeway	1.70	1.07	0.63	0.88	3.14
Run-off-road multilane	0.99	0.62	0.69	0.54	0.68
Wet road multilane	1.25	0.62	0.50	0.75	2.02
Wet road freeway	1.12	0.76	0.67	0.86	2.34
Nighttime multilane	1.50	1.02	0.68	0.70	1.06
Red light indicator lights (11)					
Total	9.47	8.37	0.88	0.94	1.91
Injury	4.85	4.03	0.83	0.86	1.18
Right angle	1.86	1.55	0.83	0.91	1.79
Left-turn	0.92	0.51	0.55	0.60	1.12
Nighttime	3.02	2.50	0.83	0.89	1.61
Disobey signal	0.77	0.59	0.76	0.71	0.83
Edgeline rumble stripes on rural horizontal curves (12)					
Total	2.96	2.36	0.80	0.74	0.77
Injury	1.21	0.95	0.78	0.72	0.79
Run-off-road	1.92	1.75	0.91	0.83	0.53

[1]*Ratio of percent change = (1 – CMF from linear traffic volume correction)/(1 – CMF from EB method).*

The differences among the methods are likely due to trends during the study period, changes in traffic volume, and potential bias due to RTM not properly accounted for by the before-after with linear traffic volume correction. This is compelling evidence of the potential differences in safety effectiveness estimates from different methods.

Other evaluations provide similar support for the use of the EB method in preference to the simple before-after methods. For example, Harwood concluded the following based on a comparison of methods to estimate the safety effectiveness of installing intersection left and right turn lanes: (13)

> "The EB approach to observational before-after evaluations of safety improvements appears to perform effectively. Comparisons of the EB approach to the [yoked comparison] and [comparison group] approaches found that the EB approach was more likely to provide statistically significant effectiveness measures. Furthermore, the effectiveness measures obtained from the EB approach were generally smaller than those from the other approaches; this may have resulted from reduced effect of the regression-to-the-mean phenomenon; compensation for regression-to-the-mean is highly desirable in providing accurate evaluation results."

These results demonstrate the potential differences in estimates obtained from various methods, and reinforce the need to apply more reliable methods such as the EB method when conducting before-after evaluations. Otherwise, the results of the evaluation may be less accurate and less reliable.

SUMMARY OF METHODS FOR CONDUCTING SAFETY EFFECTIVENESS EVALUATIONS

The examples presented in this information guide demonstrate the value of applying more reliable methods in conducting before-after safety evaluations. More reliable methods are those that account for potential bias due to RTM, changes in traffic volume, the nonlinear relationship between crash frequency and traffic volume, and general temporal effects. Table 5 presents the five general methods compared in this guide, and indicates the ability of each to account for the potential sources of bias.

The simple before-after method may overestimate or underestimate the safety effect of a treatment because it does not account for potential bias due to RTM, changes in traffic volume, and general temporal effects. The before-after with traffic volume correction methods account for changes in traffic volume, but do not account for possible bias due to RTM or temporal effects. The comparison group method accounts for temporal effects, but does not account for possible bias due to RTM and does not completely account for changes in traffic volume at the treatment sites. The EB method accounts for all sources of potential bias listed in Table 5.

The results from methods that do not properly account for potential sources of bias are less reliable and may result in less effective decisions. The examples presented throughout this guide demonstrate the potential magnitude of differences in results obtained from various methods, and reinforce the need to apply more reliable methods such as the EB method when conducting before-after evaluations. Otherwise, the results of the evaluation may be less accurate and less reliable.

Table 5. Summary of sources of bias accounted for by before-after methods.

Method	RTM	Changes in Traffic Volume	Nonlinear Relationship	Temporal Trends
Simple Before-After				
Before-After with Linear Traffic Volume Correction		•		
Before-After with Non-Linear Traffic Volume Correction		•	•	
Before-After with Comparison Group		•		•
Empirical Bayes Before-After	•	•	•	•

4. DATA REQUIREMENTS FOR SAFETY EVALUATIONS

Table 6 summarizes the data requirements for observational before-after methods in Safety Effectiveness Evaluation. The following is a description of each data element.

- **Treatment Details**: Identify the treatment(s) for evaluation, including the specific type of treatment, treatment locations, and implementation date.
- **Crash Data**: Summarize the crashes before and after treatment for each site included in the analysis. It is often useful to evaluate treatments with respect to total crashes as well as specific crash types (e.g., run-off-road) and crash severities (e.g., fatal and injury).
- **Traffic Volume Data**: Summarize the traffic volume before and after treatment for each site included in the analysis. It is desirable to obtain at least one traffic volume estimate in the before period and one in the after period for each site. For years where traffic volumes are not available, consider estimating the value based on linear interpolation.
- **Reference Group**: Identify a group of sites without the treatment, but similar to the treatment sites in terms of the major factors affecting crash risk including traffic volume and other site characteristics.
- **Comparison Group**: Identify a group of sites without the treatment, but nearby the treatment sites to account for temporal factors affecting crash risk such as changes in crash reporting, weather, and driver populations.
- **SPFs**: Calibrate an existing SPF or develop a new SPF using data from the reference sites.

Table 6. Data requirements for observational before-after methods.

Method	Treatment Details	Crash Data	Traffic Volume Data	Reference or Comparison Group	SPF
Simple Before-After	●	●			
Before-After with Linear Traffic Volume Correction	●	●	●		
Before-After with Non-Linear Traffic Volume Correction	●	●	●		●
Before-After with Comparison Group	●	●		●	
Empirical Bayes Before-After	●	●	●	●	●

5. TOOLS AND RESOURCES FOR SAFETY EVALUATIONS

Tools and resources are available to support Safety Effectiveness Evaluation, including guides and software. Some guides provide a discussion of the evaluation process or specific methods, while other guides relate to specific components of the process. For example, implementation of the EB method requires the estimation or calibration of SPFs, and guides are available to explain how to develop and calibrate SPFs.

The FHWA Roadway Safety Data and Analysis Toolbox is a web-based repository of safety data and analysis tools. Use the Toolbox to identify an appropriate tool for your Safety Effectiveness Evaluation needs. A Primer is available to understand the overall scope and functionality of the Toolbox as well as the roles, responsibilities, and tasks supported by tools in the Toolbox.

USING THE ROADWAY SAFETY DATA AND ANALYSIS TOOLBOX

There are two primary options for searching the Toolbox. The first is a predefined query using the four large icons in the upper right of Figure 7 (Manage, Analyze, Collect, and Research). The second is an advanced search option where users can search keywords and apply filters to customize their search as shown in the lower left of Figure 7.

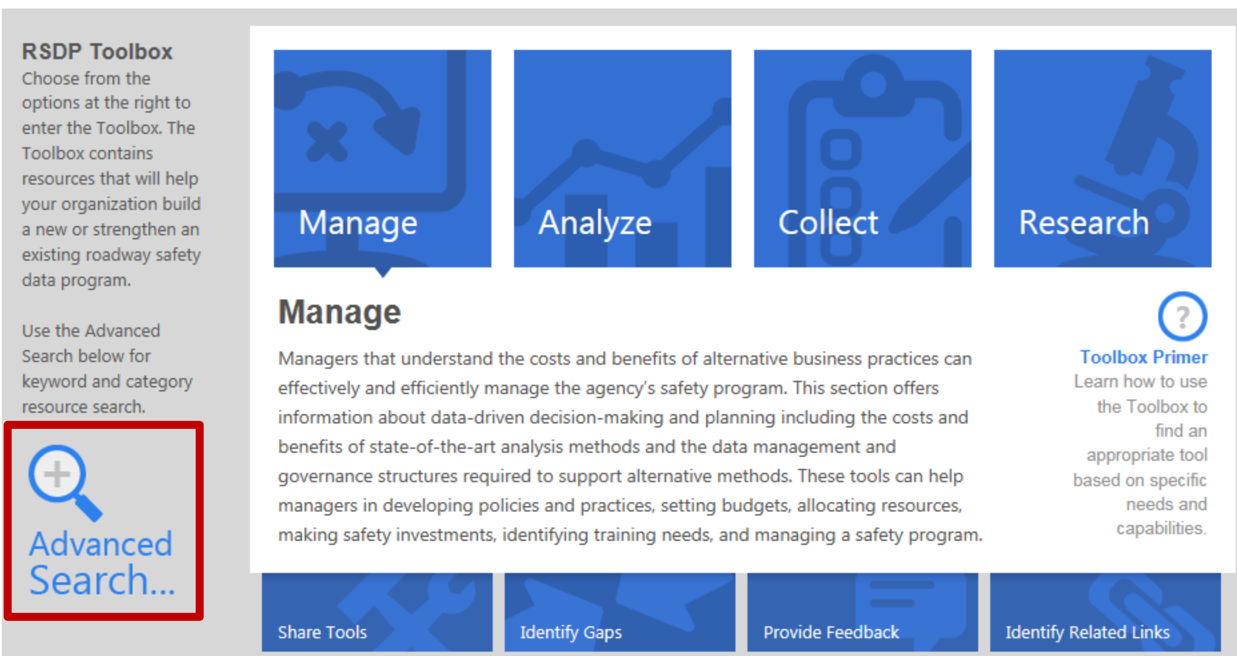


Figure 7. Graphic. Screenshot of Roadway Safety Data and Analysis Toolbox.

The following is a brief demonstration of the stepwise process to identify an appropriate tool to support Safety Effectiveness Evaluation.

1. Click the 'Advanced Search' icon, highlighted in the lower left of Figure 7.
2. From the advanced search page (Figure 8), leave the keyword blank and click the search button. This returns a list of all tools in the Toolbox.

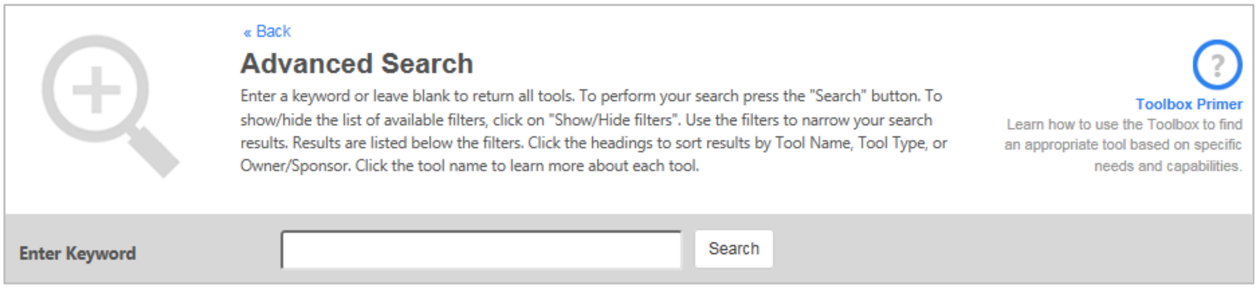


Figure 8. Graphic. Screenshot of advanced search feature.

3. Click the 'Show/Hide Filters' button, highlighted in the upper left of Figure 9. This reveals a list of filters to refine the general search.
4. Use the 'Safety Management Process' filter to select 'Safety Effectiveness Evaluation' as the primary area of interest as shown in Figure 9. Apply additional filters as needed to refine the results. For example, apply the 'Tool Type' filter to narrow the list of tools to application guides, information guides, software, information sources, or databases.

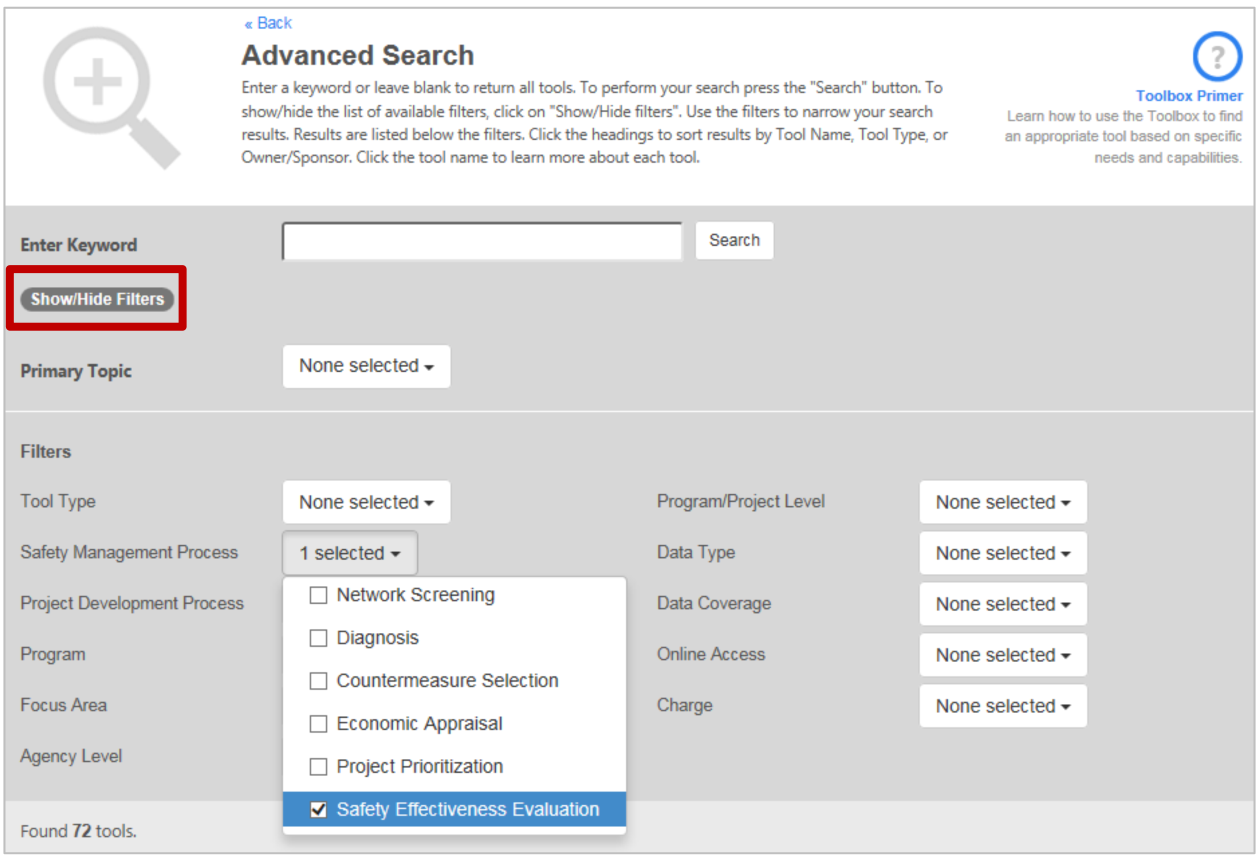


Figure 9. Graphic. Screenshot of filter options from advanced search page.

Using the stepwise process described in this section, the Toolbox returns guides such as *A Guide to Developing Quality Crash Modification Factors*, *Recommended Protocols for Developing Crash Modification Factors*, and *The Art of Appropriate Evaluation: A Guide for Highway Safety Program Managers*. Related software tools from the Toolbox include AASHTOWare Safety Analyst™.

REFERENCES

1) American Association of State Highway and Transportation Officials (AASHTO), Highway Safety Manual, First Edition, Washington, DC, 2010.
2) Hauer, E. (1997), Observational Before-After Studies in Road Safety, Pergamon, Elsevier Science Ltd.
3) Srinivasan, R. and D. Carter, Development of Safety Performance Functions for North Carolina, Report FHWA/NC/2010-09, North Carolina Department of Transportation, December 2011.
4) Gross, F., B. Persaud, and C. Lyon, A Guide to Developing Quality Crash Modification Factors, Report FHWA-SA-10-032, Federal Highway Administration, Washington, DC, 2010.
5) Carter, D., R. Srinivasan, F. Gross, and F. Council, Recommended Protocols for Developing Crash Modification Factors, Final Report from NCHRP Project 20-07 (Task 314), 2012. Available at: http://www.cmfclearinghouse.org/collateral/CMF_Protocols.pdf.
6) Persaud, B. and C. Lyon (2007), Empirical Bayes Before-After Safety Studies: Lessons Learned From Two Decades of Experience and Future Directions, Accident Analysis and Prevention, Volume 39, No. 3, pp, 546-555, Elsevier Science Ltd.
7) Persaud, B., C. Lyon, K. Eccles, N. Lefler, and F. Gross, Safety Evaluation of Offset Improvements for Left-Turn Lanes, Report FHWA-HRT-09-035, Federal Highway Administration, Washington, DC, 2009.
8) Srinivasan, R., J. Baek, D. Carter, B. Persaud, C. Lyon, K. Eccles, F. Gross, and N. Lefler, Safety Evaluation of Improved Curve Delineation. Report FHWA-HRT-09-045, Federal Highway Administration, Washington, DC, 2009.
9) Lyon, C., B. Persaud, and K. Eccles, Safety Evaluation of Centerline Plus Shoulder Rumble Strips, Report FHWA-HRT-15-048, Federal Highway Administration, Washington, DC, 2015.
10) Lyon, C., B. Persaud, and K. Eccles, Safety Evaluation of Wet Reflective Pavement Markers, Report FHWA-HRT-15-065, Federal Highway Administration, Washington, DC, 2015.
11) Himes, S., F. Gross, B. Persaud, and K. Eccles, Safety Evaluation of Red Light Indicator Lights, Draft Report, Federal Highway Administration, Washington, DC, 2015.
12) Gross, F., S. Himes, B. Persaud, and K. Eccles, Safety Evaluation of Edgeline Rumble StripEs (ELRS) on Rural Horizontal Curves, Draft Report, Federal Highway Administration, Washington, DC, 2015.
13) Harwood, D., K. Bauer, I. Potts, D. Torbic, K. Richard, E.R. Kohlman Rabbani, E. Hauer, L. Elefteriadou, and M. Griffith, Safety Effectiveness of Intersection Left- and Right-Turn Lanes, Journal of the Transportation Research Board, Transportation Research Record 1840, pp. 131-139, 2003.

APPENDIX: EXAMPLE DETAILS

EXAMPLE 1: COMPARISON OF SIMPLE AND EB BEFORE-AFTER METHODS USING SIMULATED DATA

Objective

The objective of this study was to use a simulated dataset to illustrate the importance of using the EB method to account for RTM bias in before-after studies. The example compares results from the EB before-after and simple before-after methods.

Description of Methods

The earlier section titled, *Overview of Safety Effectiveness Evaluation Methods*, describes the simple and EB before-after methods. The researchers considered using an actual crash and roadway dataset for this demonstration, but determined it would be difficult to identify and control for other changes over time. Instead, they used a simulated dataset to establish the 'ground truth' (the actual CMF) and ensure there are no other changes over time. The results compare CMFs from the simple and EB methods to the ground truth. A reliable method is one that yields a CMF not statistically different from the ground truth. For relative comparison, the method that yields a CMF closest to the ground truth is the more reliable method for conducting before-after evaluations.

Approach

The researcher used the following approach to develop the simulated dataset and conduct the before-after evaluation:

1. Using Microsoft Excel and assuming a Poisson distribution, the researchers randomly generated crash counts for 500 sites for 10 years. In generating the crash counts, they assumed the mean crash rate per year for these sites was 1.0. This mean crash rate is consistent with a similar before-after evaluation of four-legged minor road stop-controlled intersections in North Carolina (Srinivasan et al., 2014).
2. The first five years (i.e., year 1, year 2, year 3, year 4, and year 5) were assumed to be the before period and the last five years (i.e., year 6, year 7, year 8, year 9, and year 10) were assumed to be the after period. Table 7 provides a sample of the dataset to illustrate the data structure for the before and after periods.

Table 7. Sample dataset to illustrate data structure for example 1.

Site	Before Period					After Period				
	Yr. 1	Yr. 2	Yr. 3	Yr. 4	Yr. 5	Yr. 6	Yr. 7	Yr. 8	Yr. 9	Yr. 10
1	0	1	0	1	4	0	2	0	0	2
2	1	0	0	0	2	1	1	1	1	1
3	0	1	0	2	2	0	2	2	1	0

3. The analysis included four scenarios. In scenarios 1 and 2, the selection included high-crash sites for further diagnosis and treatment. In scenarios 3 and 4, the selection included low-crash sites for further diagnosis and treatment. Scenarios 1 and 2 are obviously more common; however, scenarios 3 and 4 can occur under specific circumstances when an agency suspects the treatment could increase crashes (e.g., an agency may decide to increase the speed limit for low-crash sites only). For all scenarios, the assumption is that sites receive a treatment for which the CMF is 1.0 (this is the ground truth because there is no treatment applied to the sites in the simulated dataset). Table 8 provides a description of the four scenarios examined.

Table 8. Description of scenarios from example 1.

Scenario	Description
1	The dataset was sorted in decreasing order based on the last year of the before period (i.e., year 5). The top 50 sites (i.e., the 50 sites with the most crashes in year 5) represent treatment sites. In other words, these are the high crash sites selected for treatment. The remaining 450 sites represent reference sites that receive no treatment.
2	The dataset was sorted in decreasing order based on the last year of the before period (i.e., year 5). The top 100 sites (i.e., the 100 sites with the most crashes in year 5) represent treatment sites. In other words, these are the high crash sites selected for treatment. The remaining 400 sites represent reference sites that receive no treatment.
3	The dataset was sorted in increasing order based on the last year of the before period (i.e., year 5). The top 50 sites (i.e., the 50 sites with the fewest crashes in year 5) represent treatment sites. In other words, these are the low crash sites selected for treatment. The remaining 450 sites represent reference sites that receive no treatment.
4	The dataset was sorted in increasing order based on the last year of the before period (i.e., year 5). The top 100 sites (i.e., the 100 sites with the fewest crashes in year 5) represent treatment sites. In other words, these are the low crash sites selected for treatment. The remaining 400 sites represent reference sites that receive no treatment.

4. Using the equations and procedure discussed in Hauer (1997), the researchers conducted a simple before-after evaluation and estimated a CMF. Similarly, using data from the reference sites and treatment sites, the researchers applied the EB method using the method of moments. The next section presents and compares the results.

Discussion of Results

Table 9 presents the results of both the simple before-after and EB before-after methods. For each scenario, the table shows the number of crashes in the before and after periods at the treated sites, the CMF and the standard error (S.E.) of the CMF, and the 95 percent confidence interval for the CMFs from each method. If the 95 percent confidence interval includes 1.0 (the ground truth CMF), then the CMF estimate from the given method is not statistically different from 1.0 at the five percent significance level (these results are shown in bold).

Table 9. Results of before-after evaluation using simulated data.

Scenario	Crashes in Before Period	Crashes in After Period	Evaluation Method	CMF	S.E. of CMF	95% Confidence Interval of CMF
1	338	237	Simple	**0.699**	0.059	(0.583, 0.815)
			EB	0.988	0.090	(0.812, 1.165)
2	618	499	Simple	**0.806**	0.048	(0.711, 0.901)
			EB	1.056	0.068	(0.923, 1.188)
3	180	219	Simple	1.210	0.121	(0.973, 1.447)
			EB	0.906	0.084	(0.741, 1.071)
4	355	459	Simple	**1.289**	0.091	(1.111, 1.467)
			EB	0.951	0.062	(0.830, 1.072)

Note: Bold indicates the CMF is statistically different from the ground truth (i.e., 1.0) at the 5% significance level.

In all four scenarios, the CMF estimated from the EB method is not statistically different from 1.0. Conversely, in three of the four scenarios, the CMF estimated from the simple before-after method is statistically different from 1.0. Specifically, the simple before-after method produces a CMF that indicates a statistically significant reduction in crashes for the first two scenarios. In the final scenario, the simple before-after method produces a CMF that indicates a statistically significant increase in crashes.

It is clear the simple before-after method may overestimate or underestimate the safety effect of the treatment because it does not account for potential bias due to RTM. The EB method is able to account for potential bias due to RTM and produce a CMF not statistically different from the ground truth established in this simulation study.

Example 1 References

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

Srinivasan, R., Lan, B., and Carter, D. (2014), *Safety Evaluation of Signal Installation with and Without Left Turn Lanes on Two Lane Roads in Rural and Suburban Areas*, Report FHWA/NC/2013-11, North Carolina Department of Transportation.

EXAMPLE 2: COMPARISON OF SIMPLE BEFORE-AFTER, BEFORE-AFTER WITH TRAFFIC VOLUME CORRECTION, AND EB BEFORE-AFTER METHODS FROM AN EVALUATION OF IMPROVED CURVE DELINEATION

Objective

The objective of this study was to evaluate the safety effectiveness of improved delineation at horizontal curves on rural, two-lane roads. The researchers employed an EB before-after method, including data from 89 treated curves in Connecticut (Srinivasan et al., 2009). Treatments varied by site and included new chevrons, horizontal arrows, and advance warning signs as well as the improvement of existing signs using fluorescent yellow sheeting.

The original study only included the results from the EB before-after evaluation. For the purpose of this example, the researchers applied the simple before-after method and two variations of the before-after with traffic volume correction method to the same data to estimate CMFs for comparison with the results from the EB before-after method.

Description of Methods

The earlier section titled, *Overview of Safety Effectiveness Evaluation Methods*, describes the before-after methods. For the before-after with traffic volume correction, two variations illustrate potential differences. First, a linear correction is used to account for changes in traffic volume (i.e., ratio of traffic volume in after period to the traffic volume in the before period). Second, a nonlinear correction is applied, using SPFs to account for changes in traffic volume. Note the EB method also uses the SPFs to account for changes in traffic volume. The SPFs were estimated using data from a reference group of 334 horizontal curves on rural, two-lane roads in Connecticut. As part of the EB method, the researchers applied annual calibration factors to account for time trends in the study period.

In summary, the simple before-after method does not account for potential bias due to RTM, changes in traffic volume, or temporal effects. The before-after with traffic volume correction methods do not account for possible bias due to RTM or temporal effects. The before-after with linear traffic volume correction does not account for the nonlinear relationship between crash frequency and traffic volume. The EB method is able to properly account for all of these issues, including potential bias due to RTM, changes in traffic volume, the nonlinear relationship between crash frequency and traffic volume, and general temporal effects. While the ground truth is unknown for this example, it assumes the EB before-after method provides the most reliable estimate as demonstrated in example 1.

Discussion of Results

Table 10 presents the results from each method, including the CMF and standard error of the CMF for each target crash type. Target crashes included total crashes, lane departure crashes, injury and fatal crashes, total crashes during dark conditions, and lane departure crashes during dark conditions. This study excluded intersection crashes.

Table 10. Results of curve delineation evaluation from Connecticut.

Crash Type	Simple	Traffic Volume Correction using Linear Trend	Traffic Volume Correction using SPFs (Nonlinear)	EB
	CMF (S.E.)	CMF (S.E.)	CMF (S.E.)	CMF (S.E.)
Total	**0.760** (0.073)	**0.700** (0.067)	**0.689** (0.066)	**0.822** (0.077)
Lane departure	**0.774** (0.081)	**0.712** (0.074)	**0.718** (0.075)	**0.823** (0.084)
Injury and fatal	**0.579** (0.101)	**0.537** (0.094)	**0.515** (0.091)	**0.747** (0.127)
Total crashes in dark conditions	**0.532** (0.088)	**0.489** (0.081)	**0.483** (0.080)	**0.647** (0.105)
Lane departure in dark conditions	**0.528** (0.094)	**0.486** (0.087)	**0.483** (0.086)	**0.658** (0.115)

Note: Bold indicates the CMF is statistically different from 1.0 at the 5% significance level.

Reviewing the results for each target crash type, it is apparent the different evaluation methods produce different CMFs. In some cases, the differences are small. In other cases, the differences are substantial. The remainder of this section describes the results and the differences among the various methods.

Overall, the evaluation of curve delineation in Connecticut indicated a reduction in all target crash types. These findings are consistent with previous research on this topic. There are, however, differences in the magnitude of CMFs based on the different evaluation methods.

The simple before-after and before-after with traffic volume correction methods consistently produce larger treatment effects than the EB method. This suggests RTM bias is present and unaccounted for by the simple before-after and before-after with traffic volume correction methods. These differences may also be due, in part, to changes in traffic volume and general temporal effects. Note the simple before-after method does not account for changes in traffic volume or temporal effects. While the before-after with traffic volume correction methods account for changes in traffic volume, they do not account for temporal effects.

The before-after with traffic volume correction methods (both linear and nonlinear) consistently produce larger treatment effects than the simple before-after method. This suggests there are changes in traffic volume, which are unaccounted for by the simple before-after method. While both the linear and nonlinear traffic volume correction methods produce similar results in this case, the use of SPFs (nonlinear correction) is more appropriate than a linear correction to account for changes in traffic volume.

The differences among the methods are likely due to potential bias due to RTM, changes in traffic volume, and trends during the study period. These results demonstrate the potential differences in estimates obtained from various methods, and reinforce the need to apply more reliable methods such as the EB method when conducting before-after evaluations. Otherwise, the results of the evaluation may be less accurate and less reliable.

Example 2 References

Srinivasan, R., J. Baek, D. Carter, B. Persaud, C. Lyon, K. Eccles, F. Gross, and N. Lefler, *Safety Evaluation of Improved Curve Delineation*, Report FHWA-HRT-09-045, Federal Highway Administration, 2009.

EXAMPLE 3: COMPARISON OF SIMPLE BEFORE-AFTER, BEFORE-AFTER WITH NONLINEAR TRAFFIC VOLUME CORRECTION, BEFORE-AFTER WITH COMPARISON GROUP, AND EB BEFORE-AFTER METHODS FROM AN EVALUATION OF RED LIGHT CAMERAS

Objective

The objective of this study was to estimate the safety impacts of red light cameras on traffic crashes at signalized intersections in the cities of Phoenix and Scottsdale, Arizona. The evaluation included 10 intersections equipped with red light cameras in Phoenix and 14 intersections in Scottsdale (Shin and Washington, 2007). The example includes the following four evaluation methods.

1. Simple before-after study.
2. Before-after study with nonlinear traffic volume correction.
3. Before-after study with a comparison group.
4. EB before-after method.

Description of Methods

The earlier section titled, *Overview of Safety Effectiveness Evaluation Methods*, describes the before-after methods. The before-after with traffic volume correction method includes a nonlinear trend, using SPFs to account for changes in traffic volume. Note the EB method also uses the SPFs to account for changes in traffic volume.

The simple before-after method includes data from both Phoenix and Scottsdale. The before-after with comparison group method only includes data from Phoenix since the researchers did not find comparison sites in Scottsdale. The before-after study with nonlinear traffic volume correction and the EB before-after method only include data from Scottsdale.

In summary, the simple before-after method does not account for potential bias due to RTM, changes in traffic volume, or temporal effects. The before-after with nonlinear traffic volume correction method does not account for possible bias due to RTM or temporal effects. The before-after with comparison group method does not account for possible bias due to RTM. The EB before-after method is able to properly account for all of these issues, including potential bias due to RTM, changes in traffic volume, the nonlinear relationship between crash frequency and traffic volume, and general temporal effects. While the ground truth is unknown for this example, it assumes the EB before-after method provides the most reliable estimate as demonstrated in example 1.

Discussion of Results

Table 11 presents the results of the evaluation, including the CMF and standard error of the CMF for each target crash type. [Note this is the same as Table 4 from Shin and Washington (2007).] Target crashes included angle, left-turn, and rear-end crashes. The study included angle and left-turn crashes occurring within the intersection. For rear-end crashes, the study included those occurring within 100 feet from the center of the intersection. Crashes involving heavy drinking, influence of drugs, illness, and sleep deprivation/fatigue were not included in the evaluation.

Table 11. Results of red light camera evaluation from Arizona.

Approach	Jurisdiction	Evaluation Method	CMF (S.E.)		
			Angle	Left-turn	Rear-end
Target Approaches	Phoenix	Simple	**0.61** (0.16)	0.96 (0.11)	**1.58** (0.24)
		Comparison Group	**0.58** (0.13)	0.90 (0.08)	**1.51** (0.17)
	Scottsdale	Simple	0.80 (0.14)	**0.52** (0.06)	**1.57** (0.18)
		Traffic Volume Correction	0.83 (0.15)	**0.52** (0.06)	**1.67** (0.22)
		EB	0.80 (0.14)	**0.55** (0.06)	**1.41** (0.11)
All Approaches	Phoenix	Simple	0.90 (0.15)	1.06 (0.09)	**1.26** (0.13)
		Comparison Group	0.86 (0.12)	0.99 (0.07)	**1.20** (0.10)
	Scottsdale	Simple	**0.69** (0.09)	**0.59** (0.06)	**1.48** (0.10)
		Traffic Volume Correction	**0.70** (0.09)	**0.59** (0.06)	**1.62** (0.13)
		EB	**0.83** (0.08)	**0.60** (0.05)	**1.45** (0.06)

Note: Bold indicates the CMF is statistically different from 1.0 at the 5% significance level.

Reviewing the results for each target crash type, it is apparent the different evaluation methods produce different CMFs. In some cases, the differences are small. In other cases, the differences are substantial. The remainder of this section describes the results and the differences among the various methods.

Overall, the evaluation of red light cameras in Phoenix and Scottsdale indicated a reduction in angle and left-turn crashes, and an increase in rear-end crashes. These findings are consistent with previous research on this topic. There are, however, differences in the magnitude of CMFs based on the different evaluation methods.

Compared to the EB method, the simple before-after method results in larger treatment effects in nearly all cases. For example, the CMF for rear-end crashes (for target approaches) in Scottsdale based on the simple before-after method is 1.57, while the CMF based on the more reliable EB method is 1.41. Similarly, the CMF for angle crashes in Scottsdale (for all approaches) is 0.69 based on the simple before-after method, while the CMF based on the more reliable EB method is 0.83. The simple before-after method does not account for possible bias due to RTM, changes in traffic volume, or trends during the study period, which are all addressed by properly applying the EB method.

Comparing the simple before-after method and before-after with nonlinear traffic volume correction method, the results are inconsistent. The results from the simple before-after method indicate a similar treatment effect for angle and left-turn crashes, but a smaller increase in rear-end crashes.

Compared to the EB method, the before-after with nonlinear traffic volume correction method typically produces a greater treatment effect. While the traffic volume correction helps to account for changes in traffic volume, it does not account for potential RTM bias and trends during the study period.

Comparing the comparison group method and the simple before-after method, the results are inconsistent. The results from the simple before-after method indicate a smaller treatment effect for angle and left-turn crashes, but a greater effect for rear-end crashes. While the comparison group method helps to account for trends during the study period, it does not account for changes in traffic volume or possible bias due to RTM.

The differences among the methods are likely due to trends during the study period, changes in traffic volume, and potential bias due to RTM. These results demonstrate the potential differences in estimates obtained from various methods, and reinforce the need to apply more reliable methods such as the EB method when conducting before-after evaluations. Otherwise, the results of the evaluation may be less accurate and less reliable.

Example 3 References

Shin, K. and S. Washington, The Impact of Red Light Cameras on Safety in Arizona, *Accident Analysis and Prevention* 39, pp. 1212-1221, 2007.

EXAMPLE 4: COMPARISON OF BEFORE-AFTER WITH LINEAR TRAFFIC VOLUME CORRECTION AND EB BEFORE-AFTER METHOD FROM VARIOUS FHWA FUNDED STUDIES

Objective

The objective of this study is to provide a comparison of safety effectiveness estimates from seven studies of various treatments. In total, there are 62 CMFs presented for various crash types and severities, including 31 CMFs based on the before-after with linear traffic volume correction method and 31 CMFs based on the EB before-after method. These results update and supplement the results in Table 2 of Persaud and Lyon (2007), which compared 21 safety effectiveness estimates obtained from simple and EB before-after methods.

The original studies included the results from the EB before-after evaluation. For the purpose of this example, the researchers applied the before-after with linear traffic volume correction method to the same data to estimate CMFs for comparison with the results from the EB before-after method.

Description of Methods

The earlier section titled, *Overview of Safety Effectiveness Evaluation Methods*, describes the before-after methods. For the before-after with linear traffic volume correction method, the observed crashes before treatment is multiplied by the ratio of the traffic volume in the after period to traffic volume in the before period. This assumes a linear relationship between crash frequency and traffic volume, which is generally invalid. However, the intent was to replicate a typical before-after evaluation with traffic volume correction in the absence of SPFs. Note the EB method uses SPFs to account for changes in traffic volume. The estimates from the before-after with linear traffic volume correction do not control for potential bias due to RTM or time trends in crash occurrence unrelated to the treatment. While the ground truth is unknown for this example, it assumes the EB before-after method provides the most reliable estimate as demonstrated in example 1.

Discussion of Results

Table 12 presents the results of the seven evaluations, including the CMFs for various target crash types and severities. [Note these results update and supplement the results in Table 2 of Persaud and Lyon (2007).] There are a total of 62 CMF estimates; 31 CMFs based on the before-after with linear traffic volume correction method and 31 CMFs based on the EB before-after method. The last column in Table 12 presents the ratio of percent reduction estimated from the two methods, computed as the percent reduction in crashes from the before-after with linear traffic volume correction method divided by the percent reduction from the EB method. [Note the percent reduction is 100*(1-CMF).] A ratio greater than 1.0 indicates the before-after with linear traffic volume correction method estimates a larger effect than the EB method. A ratio less than 1.0 indicates the before-after with linear traffic volume correction method estimates a smaller effect than the EB method.

Table 12. Results of a decade of EB before-after evaluations.

Crash Type	Crashes/mile year (segments) or crashes/year (intersections) in after period		CMF		Ratio of percent change (Simple/EB)
	Expected from linear traffic volume correction	Observed	Before-after with linear traffic volume correction	EB from study	
Two-way left-turn lanes on rural two-lane roads (Persaud and Lyon, 2007)					
Total	8.66	5.89	0.68	0.64	0.89
Injury	1.50	0.74	0.49	0.65	1.45
Rear-end	3.62	1.67	0.46	0.53	1.15
Offset left turn lanes (Wisconsin) (Persaud et al., 2009)					
Total	7.31	4.7	0.64	0.66	1.06
Injury	3.32	1.9	0.57	0.64	1.20
Left-turn opposing	3.13	1.9	0.58	0.62	1.12
Rear-end	2.09	1.5	0.72	0.68	0.89
Improve curve delineation (Srinivasan et al., 2009)					
Injury	2.90	1.89	0.64	0.82	2.00
Dark	2.70	1.51	0.56	0.73	1.60
Lane departure dark	2.38	1.34	0.56	0.75	1.72
Centerline plus shoulder rumble strips on two-lane rural roads (Lyon et al., 2015)					
Total	0.567	0.463	0.82	0.80	0.92
Injury	0.254	0.183	0.72	0.77	1.21
Run-off-road	0.189	0.123	0.65	0.74	1.35
Head-on	0.024	0.014	0.58	0.63	1.15
Sideswipe	0.031	0.015	0.49	0.77	2.20

Crash Type	Crashes/mile year (segments) or crashes/year (intersections) in after period		CMF		Ratio of percent change (Simple/EB)
	Expected from linear traffic volume correction	Observed	Before-after with linear traffic volume correction	EB from study	
Wet-reflective pavement markings (Lyon et al., 2015)					
Total multilane	5.78	3.66	0.63	0.83	2.10
Injury multilane	1.93	1.10	0.57	0.60	1.06
Injury freeway	1.70	1.07	0.63	0.88	3.14
Run-off-road multilane	0.99	0.62	0.69	0.54	0.68
Wet road multilane	1.25	0.62	0.50	0.75	2.02
Wet road freeway	1.12	0.76	0.67	0.86	2.34
Nighttime multilane	1.50	1.02	0.68	0.70	1.06
Red light indicator lights (Himes et al., 2015)					
Total	9.47	8.37	0.88	0.94	1.91
Injury	4.85	4.03	0.83	0.86	1.18
Right angle	1.86	1.55	0.83	0.91	1.79
Left-turn	0.92	0.51	0.55	0.60	1.12
Nighttime	3.02	2.50	0.83	0.89	1.61
Disobey signal	0.77	0.59	0.76	0.71	0.83
Edgeline rumble stripes on rural horizontal curves (Gross et al., 2015)					
Total	2.96	2.36	0.80	0.74	0.77
Injury	1.21	0.95	0.78	0.72	0.79
Run-off-road	1.92	1.75	0.91	0.83	0.53

In 23 of the 31 comparisons (74 percent), the percent reduction estimated from the before-after with linear traffic volume correction is greater than the percent reduction estimated from the EB method. Of these, 10 of the 23 estimates from the before-after with linear traffic volume correction are 1.0 to 1.25 times as large as the estimates from the EB method. The remaining 13 of the 23 estimates from the before-after with linear traffic volume correction are more than 1.25 times as large as the estimates from the EB method.

The differences among the methods are likely due to trends during the study period, changes in traffic volume, and potential bias due to RTM not properly accounted for by the before-after with linear traffic volume correction. This is compelling evidence of the potential differences in safety effectiveness estimates from different methods.

The researchers considered other evaluations for inclusion in Table 12, but in almost all cases there was insufficient information to facilitate this analysis. Nevertheless, there is support from other evaluations for the use of the EB method in preference to less reliable before-after methods. For example, Harwood et al. (2003) concluded the following based on a comparison of methods to estimate the safety effectiveness of installing intersection left and right turn lanes:

> "The EB approach to observational before-after evaluations of safety improvements appears to perform effectively. Comparisons of the EB approach to the [yoked comparison] and [comparison group] approaches found that the EB approach was more likely to provide statistically significant effectiveness measures. Furthermore, the effectiveness measures obtained from the EB approach were generally smaller than those from the other approaches; this may have resulted from reduced effect of the regression-to-the-mean phenomenon; compensation for regression-to-the-mean is highly desirable in providing accurate evaluation results."

These results demonstrate the potential differences in estimates obtained from various methods, and reinforce the need to apply more reliable methods such as the EB method when conducting before-after evaluations. Otherwise, the results of the evaluation may be less accurate and less reliable.

Example 4 References

Gross, F., S. Himes, B. Persaud, and K. Eccles. *Safety Evaluation of Edgeline Rumble StripEs (ELRS) on Rural Horizontal Curves*. Federal Highway Administration, Draft Report, 2015.

Harwood, D., K. Bauer, I. Potts, D. Torbic, K. Richard, E.R. Kohlman Rabbani, E. Hauer, L. Elefteriadou, and M. Griffith. *Safety Effectiveness of Intersection Left- and Right-Turn Lanes*. Transportation Research Record 1840, pp. 131-139, 2003.

Himes, S., F. Gross, B. Persaud, and K. Eccles. *Safety Evaluation of Red Light Indicator Lights*. Federal Highway Administration, Draft Report, 2015.

Lyon, C., B. Persaud, and K. Eccles. *Safety Evaluation of Centerline Plus Shoulder Rumble Strips*. Report FHWA-HRT-15-048, Federal Highway Administration, 2015. (In press)

Lyon, C., B. Persaud, and K. Eccles. *Safety Evaluation of Wet Reflective Pavement Markers*, Report FHWA-HRT-15-065, Federal Highway Administration, 2015 (In press).

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*, Volume 39, Issue 3, pp, 546-555, 2007.

Persaud, B., C. Lyon, K. Eccles, N. Lefler, and F. Gross. *Safety Evaluation of Offset Improvements for Left-Turn Lanes*. Report FHWA-HRT-09-035, Federal Highway Administration, 2009.

Srinivasan, R., J. Baek, D. Carter, B. Persaud, C. Lyon, K. Eccles, F. Gross, and N. Lefler, *Safety Evaluation of Improved Curve Delineation.* Report FHWA-HRT-09-045, Federal Highway Administration, 2009.

For More Information:

Stuart Thompson
Stuart.Thompson@dot.gov
202-366-8090

Federal Highway Administration
Office of Safety
1200 New Jersey Avenue SE
Washington, DC 20590
http://safety.fhwa.dot.gov/

Publication # FHWA-SA-16-040