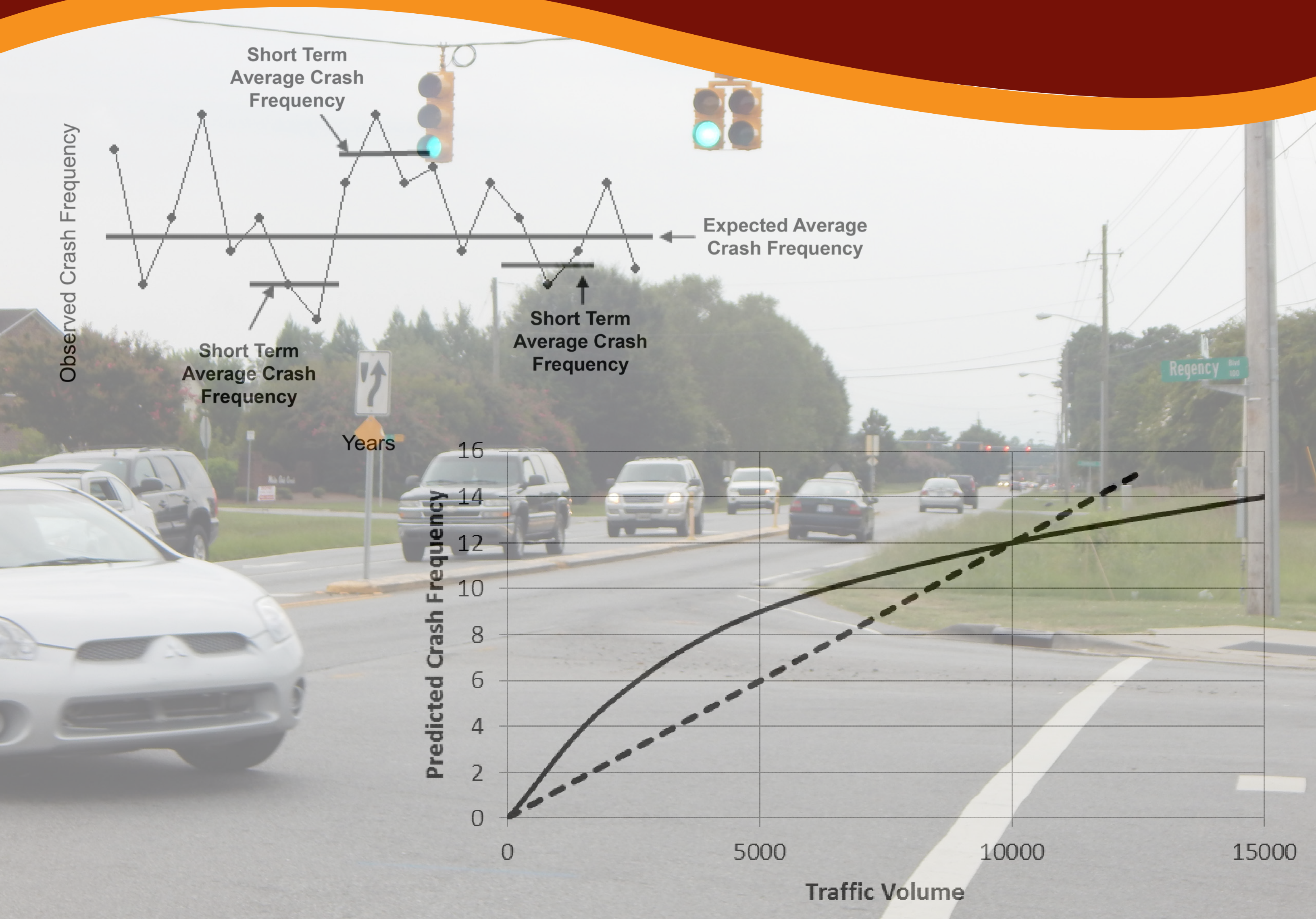
Reliability of Safety Management Methods

Network Screening



FHWA-SA-16-037

October 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

TECHNICAL DOCUMENTATION PAGE

<table>
<tr><td colspan="2">1. Report No.
FHWA-SA-16-037</td><td>2. Government Accession No.</td><td colspan="2">3. Recipient's Catalog No.</td></tr>
<tr><td colspan="3" rowspan="2">4. Title and Subtitle
Reliability of Safety Management Methods: Network Screening</td><td colspan="2">5. Report Date
October 2016</td></tr>
<tr><td colspan="2">6. Performing Organization Code</td></tr>
<tr><td colspan="3">7. Author(s)
Raghavan Srinivasan, Frank Gross, Bo Lan, and Geni Bahar</td><td colspan="2">8. Performing Organization Report No.</td></tr>
<tr><td colspan="3" rowspan="2">9. Performing Organization Name and Address
VHB
8300 Boone Boulevard
Suite 700
Vienna, VA 22182</td><td colspan="2">10. Work Unit No.</td></tr>
<tr><td colspan="2">11. Contract or Grant No.
DTFH61-10-D-00022-T-13005</td></tr>
<tr><td colspan="3" rowspan="2">12. Sponsoring Agency Name and Address
Federal Highway Administration Office of Safety
1200 New Jersey Ave., SE
Washington, DC 20590</td><td colspan="2">13. Type of Report and Period
Final Report, March 2015 – October 2016</td></tr>
<tr><td colspan="2">14. Sponsoring Agency Code
FHWA</td></tr>
<tr><td colspan="5">15. Supplementary Notes
The Contracting Officer’s Task Manager (COTM) was Stuart Thompson.</td></tr>
<tr><td colspan="5">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 methods described in the network screening, diagnosis, and countermeasure selection guides. 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: Network Screening guide describes various methods and the latest tools to support network screening. The target audience includes data analysts, program managers, 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 network screening. This guide compares more reliable crash-based performance measures to traditional measures which are more susceptible to bias and may result in less reliable results and less effective decisions. Readers will understand the value of and be prepared to select more reliable performance measures in network screening.

This guide includes five sections and an appendix. The first section introduces the roadway safety management process and network screening. The second section provides an overview of various performance measures for conducting network screening, including a discussion of the associated strengths and limitations. The strengths and limitations focus on the ability (and inability) of the measures to account for issues in network screening that can lead to less reliable results. The third section demonstrates the value of applying more reliable performance measures in network screening. Empirical examples highlight the shortcomings of less reliable measures, which lead to less reliable results and conclusions. The next section summarizes the data requirements to employ the various measures. The final section describes available tools and resources to support network screening. The appendix presents further details on the empirical examples used to demonstrate the value of applying more reliable performance measures in network screening.</td></tr>
<tr><td colspan="2">17. Key Words:
Roadway, Safety, Data, Analysis, Network, Screening</td><td colspan="3">18. Distribution Statement
No restrictions.</td></tr>
<tr><td colspan="2">19. Security Classif. (of this report)
Unclassified</td><td>20. Security Classif. (of this page)
Unclassified</td><td>21. No. of Pages
68</td><td>22. Price</td></tr>
</table>

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

ACKNOWLEDGMENTS

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

Ning Le, Virginia Department of Transportation

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ACRONYMS

AADT	Annual average daily traffic
AASHTO	American Association of State Highway and Transportation Officials
CMF	Crash Modification Factor
CPM	Crash prediction model
DOT	Department of Transportation
EB	Empirical Bayes
EPDO	Equivalent property damage only
FHWA	Federal Highway Administration
GIS	Geographic information system
HSIP	Highway Safety Improvement Program
LOSS	Level of service of safety
MAE	Mean absolute error
PDO	Property damage only
PSI	Potential for safety improvement
RTM	Regression-to-the-mean
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 methods described in the network screening, diagnosis, and countermeasure selection guides. 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: Network Screening guide describes various methods and the latest tools to support network screening. The target audience includes data analysts, project managers, and program 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 network screening. This guide compares more reliable crash-based performance measures to traditional measures which are more susceptible to bias and may result in less reliable results and less effective decisions. Readers will understand the value of and be prepared to select more reliable performance measures in network screening.

This guide includes five sections and an appendix. The first section introduces the roadway safety management process and network screening. The second section provides an overview of various performance measures for conducting network screening, including a discussion of the associated strengths and limitations. The strengths and limitations focus on the ability (and inability) of the measures to account for issues in network screening that can lead to less reliable results. The third section demonstrates the value of applying more reliable performance measures in network screening. Empirical examples highlight the shortcomings of less reliable measures, which lead to less reliable results and conclusions. The next section summarizes the data requirements to employ the various measures. The final section describes available tools and resources to support network screening. The appendix presents further details on the empirical examples used to demonstrate the value of applying more reliable performance measures in network screening.

1. INTRODUCTION TO NETWORK SCREENING

The roadway safety management process is a six-step process as shown in Figure 1 and outlined in the Highway Safety Manual (AASHTO, 2010). [(1)] Network screening is the first step in the roadway safety management process.

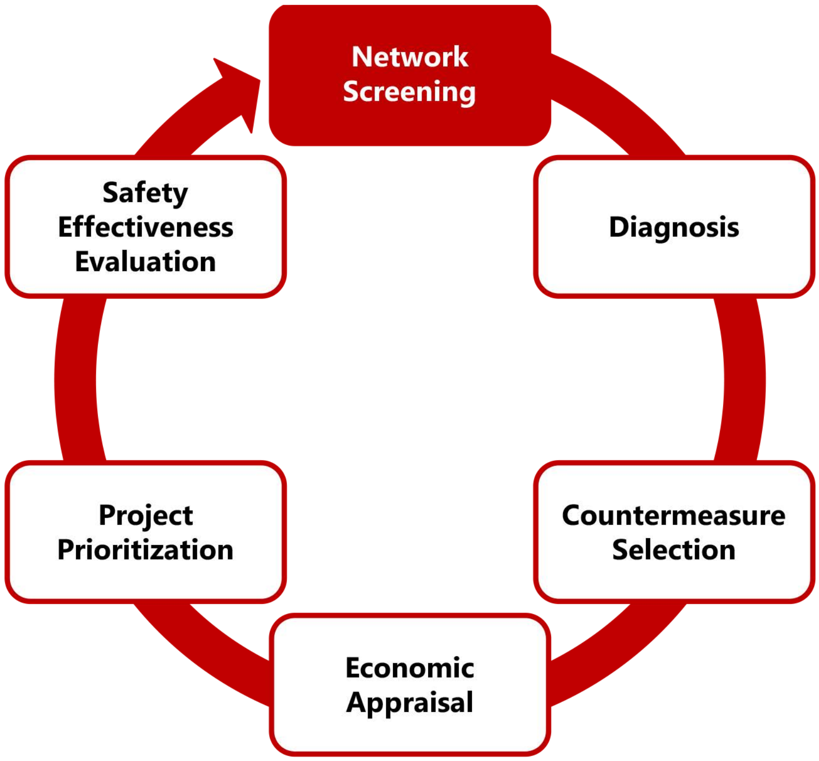


Figure 1. Chart. Schematic of Roadway Safety Management process.

Network screening is the process of identifying sites for further investigation and potential treatment. The intent is to identify sites expected to benefit the most from targeted, cost-effective treatments. This aligns with the purpose of the federal Highway Safety Improvement Program (HSIP), which is to achieve a significant reduction in fatalities and serious injuries on all public roads. [(2)] To achieve this goal, the network screening process should maximize the opportunity to improve safety; otherwise, agencies may allocate resources inefficiently to sites with less potential for improvement while locations with a higher potential for cost-effective safety improvement remain untreated.

There are two general approaches to safety management: 1) selecting and treating sites based on site-specific crash history (referred to as the crash-based approach for this report), and 2) selecting and treating sites based on site-specific geometric and operational attributes (referred to as the systemic approach for this report). These two approaches are complementary and support a comprehensive, network-wide approach to safety management. This report focuses on the site-specific crash-based approach. Refer to the *Reliability of Safety Management Methods: Systemic Safety Analysis* for further discussion of the systemic approach. [(3)]

In the crash-based approach, analysts identify sites based on site-specific, crash-based performance measures. The more rigorous crash-based measures use the Empirical Bayes (EB) method, incorporating crash predictions from safety performance functions (SPFs) and site-specific crash history. SPFs are mathematical equations that relate crash frequency to geometric and operational attributes. Once an agency identifies a list of sites, a multidisciplinary analysis team reviews the site-specific crash history and site characteristics (e.g., geometry, traffic operations, road users, and adjacent land use) in detail to identify target crash types and crash contributing factors. This provides the foundation for the identification and selection of appropriate countermeasures to mitigate the specific safety issues (e.g., crash patterns and contributing factors) at each site.

In the systemic approach, analysts identify sites based on site-specific geometric and operational attributes rather than observed crashes. The first step is to select the focus crash type(s) and identify site-specific geometric and operational attributes (i.e., risk factors) associated with an increased risk of the focus crashes. Risk factors are site-specific attributes common across locations with the focus crash type(s). For example, sharp horizontal curves are a common feature (i.e., risk factor) associated with roadway departure crashes. Analysts identify risk factors by analyzing crash data from their jurisdiction or by reviewing previous research studies. Using a list of risk factors as a guide, agencies identify sites with those specific characteristics, and then develop targeted treatments to address or mitigate the specific risk factors at the specific locations. Agencies can apply crash history, if available, and other thresholds to reduce the list of sites based on available resources and program objectives.

The systemic approach has several attractive features. First, an agency can employ the systemic approach without crash data. This is useful when an agency does not have reliable crash data at the site level or when crashes are underreported. Second, the systemic approach is useful for treating safety issues where crashes are highly dispersed such as on rural and local roads with low traffic volumes. Specifically, agencies can use the systemic approach to address existing and potential safety issues across a large portion of the network (e.g., shoulder rumble strips on all rural, two-lane roads with a certain shoulder width and traffic volume level).

While the systemic approach does not require site-specific crash history, it builds on the fundamental concept of predicting crashes. Specifically, analysts predict crashes based on crash history, traffic volume, and other geometric and operational attributes. For a given site, the left-hand side of the predictive equation is the predicted crashes and the right-hand side is the site-specific attributes. The systemic approach focuses on the right-hand side of the predictive equation to identify risk factors (i.e., attributes that increase the risk of a crash). Agencies then identify locations with those attributes for potential improvement. The following website provides further details, examples, and resources related to the systemic approach to safety (http://safety.fhwa.dot.gov/systemic/).

The crash-based and systemic approaches are complementary and typically result in different levels of effects. For sites identified and addressed by the crash-based approach, there is potential for substantial safety benefits at the treatment locations; however, the benefits are typically localized to the specific locations that receive treatment. The systemic approach identifies and addresses safety issues based on risk factors, often resulting in safety treatments at more locations; however, the results at any one location may not be as pronounced in comparison with sites treated using the crash-based approach. As such, the use of the crash-based and systemic approaches provide a complementary approach to address network-wide safety issues at locations with a high potential for improvement.

The remainder of this guide focuses primarily on the crash-based approach to safety due to the availability of examples and current state-of-the-practice. While there are general discussions of the benefits of the systemic approach, a companion guide, *Reliability of Safety Management Methods: Systemic Safety Programs*, provides further information about this approach. [3]

2. OVERVIEW OF NETWORK SCREENING

The first edition of the Highway Safety Manual (AASHTO, 2010) identifies five major steps in network screening: [1]

1. **Establish focus**: This step establishes the reason for network screening. For example, the focus of the screening process may be to identify the sites where treatments or modifications could reduce the number of crashes or to identify the sites that may benefit the most from a particular countermeasure (e.g., a shoulder rumble strip program).
2. **Identify network and establish reference population**: This step identifies the portion of the network for screening (e.g., stop-controlled intersections in rural areas, signalized intersections in urban areas, freeway segments, etc.).
3. **Select performance measures**: Analysts can use a variety of performance measures to evaluate the potential to reduce the frequency and severity of crashes at a site. The Highway Safety Manual lists 13 possible performance measures, which differ with respect to the data requirements, analytic requirements, and statistical rigor. This guide focuses on the selection of an appropriate network screening performance measure.
4. **Select screening method**: Depending on the facility and site type identified for screening (e.g., intersections, segments, or corridors), and the availability of data, different screening methods are possible. For intersections, the simple ranking method is appropriate. For segments, options include simple ranking, sliding window, and peak searching algorithms. In general, the sliding window and peak searching methods are preferred over the simple ranking method for segments.
5. **Screen and evaluate results**: The final step is to conduct the network screening and evaluate the results.

The key to effective network screening is selecting an appropriate performance measure. More reliable performance measures account for regression-to-the-mean (RTM), differences in traffic volume, and crash severity. Table 1 lists the 13 performance measures discussed in the Highway Safety Manual with an indication of the ability to account for potential bias due to RTM and differences in traffic volume. While some measures account for crash severity directly (e.g., relative severity index), analysts can adapt any of the measures to account for crash severity. Following Table 1 is a brief explanation of the three primary issues in network screening:

- Regression-to-the-mean.
- Differences in traffic volume.
- Differences in crash severity.

Table 1. Performance measures for network screening. (1)

Performance Measure	Accounts for RTM Bias	Accounts for Traffic Volume
1. Average crash frequency	No	No
2. Crash rate	No	Yes
3. Equivalent property damage only (EPDO) average crash frequency	No	No
4. Relative severity index	No	No
5. Critical rate (rate quality control)	No	Yes
6. Excess predicted average crash frequency using method of moments	No	No
7. Level of service of safety (LOSS)	No[1]	Yes
8. Excess predicted average crash frequency using safety performance functions (SPFs)	No	Yes
9. Probability of specific crash types exceeding threshold proportion	Not affected by RTM bias[2]	No
10. Excess proportion of specific crash types	Not affected by RTM bias[2]	No
11. Expected average crash frequency with Empirical Bayes (EB) adjustments	Yes	Yes
12. EPDO average crash frequency with EB adjustment	Yes	Yes
13. Excess expected average crash frequency with EB adjustment (potential for safety improvement—PSI)	Yes	Yes

Note: This information is based on Table 4-2 and subsequent discussion from the Highway Safety Manual. (1)

[1] The original LOSS method described in this guide is based on the difference in observed and predicted crashes and did not account for bias due to RTM. In 2015, a new method was proposed for using LOSS in concert with an Empirical Bayes (EB) procedure to correct for RTM bias. (4)

[2] These two measures will not be affected by RTM only if they are based on data from a long time period.

ISSUES RELATED TO REGRESSION-TO-THE-MEAN

Transportation agencies often select sites with high crash counts for further investigation and potential treatment. 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 random nature of crashes. RTM also implies that periods with relatively low crash frequencies are likely followed by periods with relatively high crash frequencies. The intent of network screening is to separate sites with randomly high crashes from sites with high potential for improvement, focusing on the later.

Figure 2 illustrates RTM, comparing the difference between short-term average and long-term average crash history. Due to RTM, the short-term average is not a reliable estimate of the long term crash propensity of a particular site. 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 network screening results may incorrectly identify sites for further investigation and potential treatment.

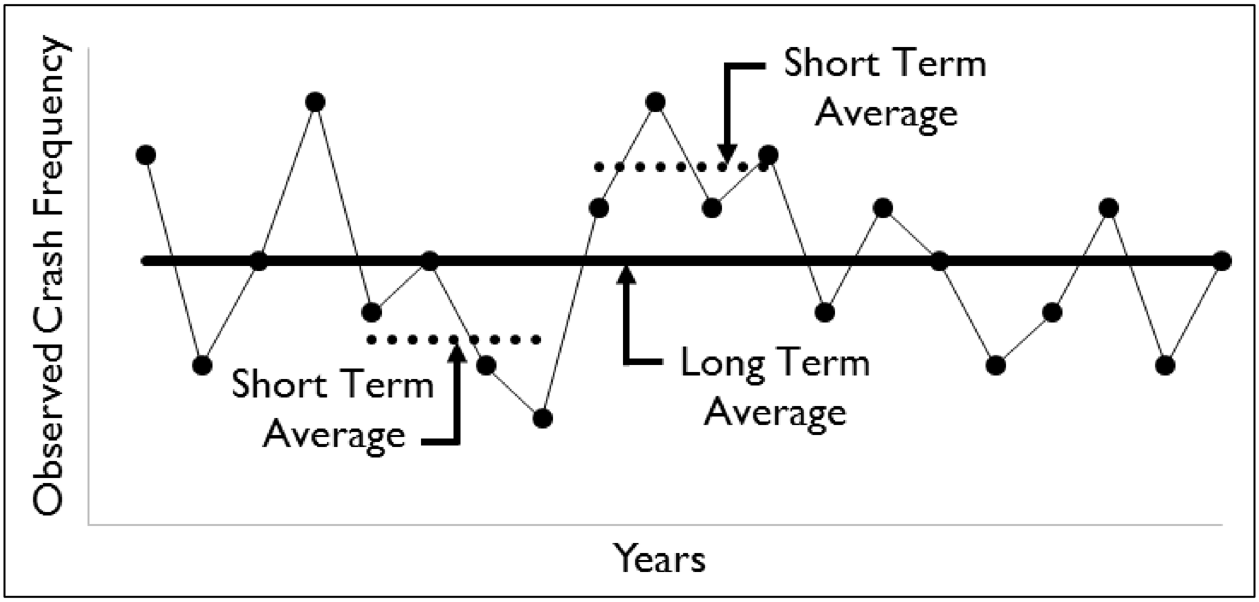


Figure 2. Chart. Illustration of RTM comparing short- and long-term averages.

The first eight measures presented in Table 1 do not account for possible bias due to RTM. Measure 9 (probability of specific crash types exceeding threshold proportion) and measure 10 (excess proportion of specific crash types) are not affected by RTM unless they are based on short-term crash history. Measure 11 (expected average crash frequency with EB adjustments), measure 12 (EPDO average crash frequency with EB adjustment), and measure 13 (excess expected average crash frequency with EB adjustment) account for possible bias due to RTM using the EB adjustments.

ISSUES RELATED TO DIFFERENCES IN TRAFFIC VOLUME

Research studies have established that for most crash types, traffic volume is the most important explanatory variable. This is important because one would expect more crashes at sites with higher traffic volumes than sites with lower traffic volumes. If the analyst does not account for differences in traffic volume among sites, then they may incorrectly identify sites with higher traffic volumes as sites with high potential for improvement. There is further discussion and an example below to illustrate this concept.

Traditionally, analysts have used crash rates to account for differences in traffic volume among sites. Crash rate is the ratio of crash frequency to exposure, which is typically the traffic volume. Crash rates implicitly assume a linear relationship between crash frequency and traffic volume; however, many studies have shown the relationship between crashes and traffic volume is often nonlinear, and this relationship depends on the facility and site type. Nonlinear relationships such as SPFs are more appropriate than linear relationships such as crash rates to account for differences in traffic volume among sites.

SPFs are a more reliable method to account for differences in traffic volume among sites because they reflect the nonlinear relationship between crash frequency and traffic volume. The SPF is an equation, representing a best-fit model that relates annual observed crashes to the annual traffic volume for a group of sites with similar attributes. Figure 3 is a hypothetical SPF where the points represent observed crashes at specific traffic volumes for individual sites, and the solid line represents the best-fit model (i.e., the SPF).

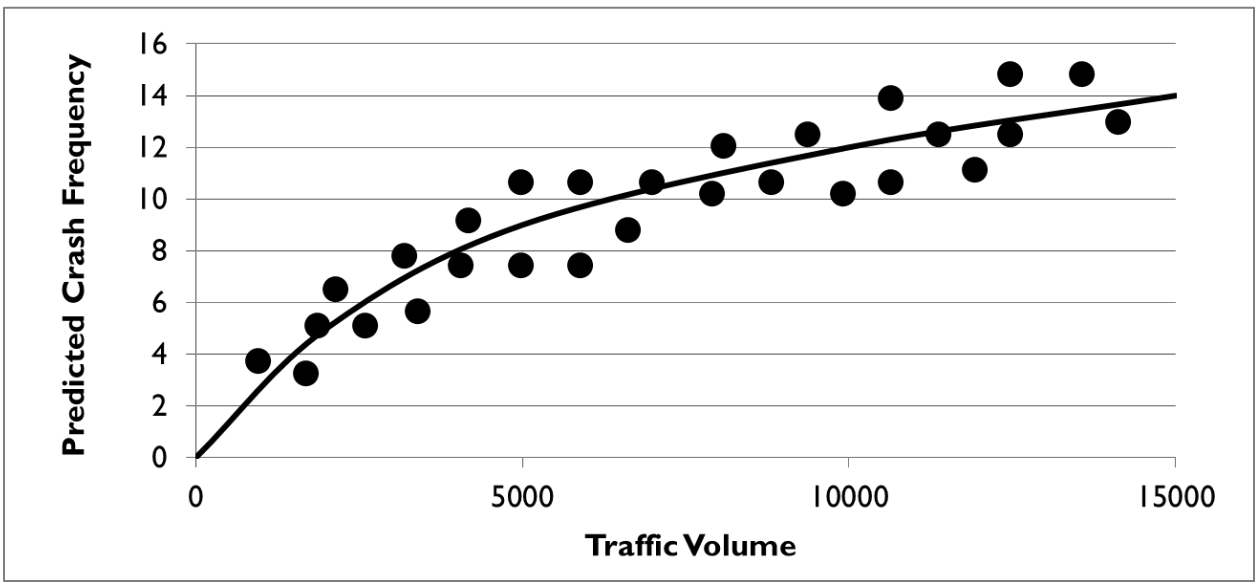


Figure 3. Graph. Example SPF relating crash frequency and traffic volume.

Table 2 provides data for three hypothetical sites. Site 1 has an average of 6 crashes per year with an annual average daily traffic (AADT) volume of 2,000 vehicles per day and corresponding crash rate of 0.003 (crashes per AADT). Site 2 has an average of 8 crashes per year with an AADT of 5,000 vehicles per day and corresponding crash rate of 0.0016 (crashes per AADT). Site 3 has an average of 11 crashes per year with an AADT of 10,000 vehicles per day and corresponding crash rate of 0.0011 (crashes per AADT).

Table 2. Ranking of hypothetical sites by crash frequency and rate.

Site	Crash Frequency	Traffic Volume	Crash Rate	Rank by Frequency	Rank by Rate
1	6	2000	0.0030	3	1
2	8	5000	0.0016	2	2
3	11	10000	0.0011	1	3

Consider the three hypothetical sites from Table 2 and various performance measures to priority rank the sites for further investigation. Using crash frequency (measure 1 from Table 1), the sites are ranked 3, 2, and 1 where site 3 is the highest priority as shown in Table 2. Using crash rate (measure 2 from Table 1), the sites are ranked 1, 2, and 3 where site 1 is the highest priority as shown in Table 2.

Figure 4 plots the hypothetical data from Table 2 and illustrates the difference between a linear and nonlinear trend to define the relationship between crash frequency and traffic volume. In Figure 4, the solid line is a nonlinear relationship between crash frequency and traffic volume, which represents the relationship described by typical SPFs. The dashed line is a linear relationship between crash frequency and traffic volume, which represents the relationship described by crash rate.

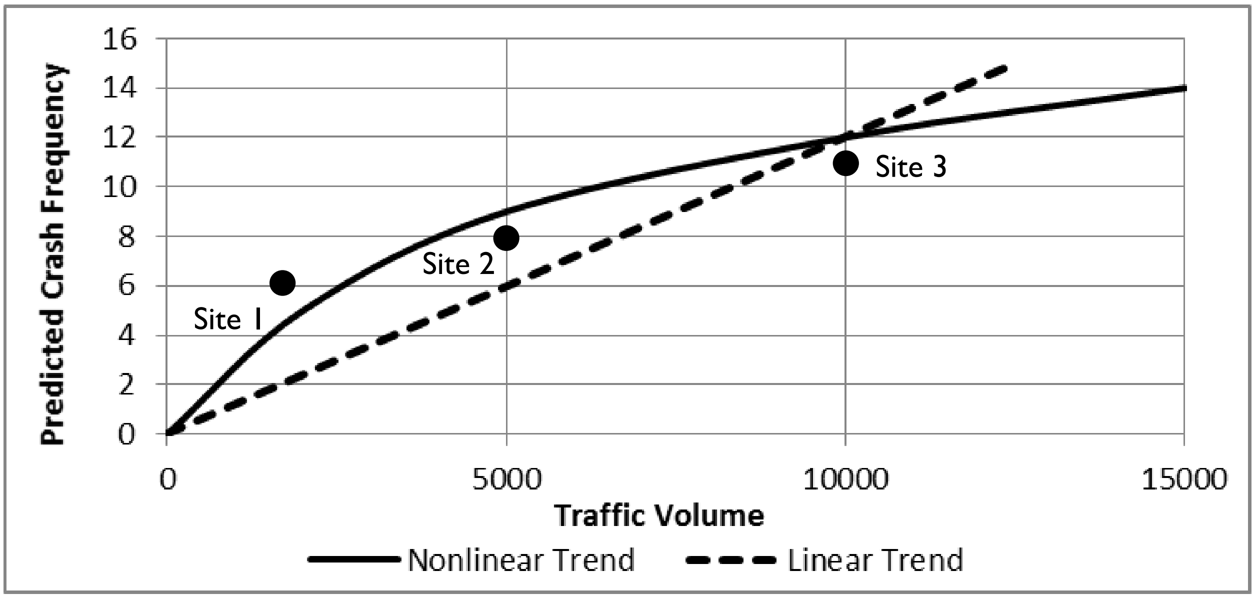


Figure 4. Graph. Relationships between crash frequency and traffic volume.

Figure 4 helps to illustrate the potential limitation of using crash frequency or crash rate as performance measures in network screening compared to SPFs. Consider the performance measure for excess predicted average crash frequency using SPFs (measure 8 from Table 1). Using this measure, the SPF (solid line in Figure 4) is a threshold representing the average crashes for a given traffic volume. Sites above the SPF represent sites with potential for improvement and sites below the SPF represent sites performing well with respect to other similar sites with similar traffic volumes. In Figure 4, site 1 is the only site above the SPF, indicating relative priority over sites 2 and 3. Now, consider the critical rate (measure 5 from Table 1), and assume the dashed linear line in Figure 4 represents the critical rate for this example. If the actual relationship between crashes and traffic volume is nonlinear, and the analyst assumed a linear trend as the threshold to represent the average crashes for a given traffic volume, then they would incorrectly identify any sites between the SPF and dashed line (site 2 in this case) as priority sites for further investigation. It is apparent from this hypothetical example that different performance measures produce different priority rankings.

SPF-based performance measures account for differences in traffic volume among sites and account for the nonlinear relationship between traffic volume and crashes. Measure 7 (level of service of safety), measure 8 (excess predicted average crash frequency using SPFs), measure 11 (expected average crash frequency with EB adjustments), measure 12 (EPDO average crash frequency with EB adjustment), and measure 13 (excess expected average crash frequency with EB adjustment) use SPFs. Hence, these measures properly account for traffic volume.

ISSUES RELATED TO DIFFERENCES IN CRASH SEVERITY

Three of the measures directly account for crash severity:

- Measure 3: EPDO average crash frequency.
- Measure 4: Relative severity index.
- Measure 12: EPDO average crash frequency with EB adjustment.

Measure 3 and measure 12 use the EPDO method, which converts all crashes to a common unit, namely property damage only (PDO) crashes. Using these measures, the analyst assigns points to each crash based on its crash severity level. A PDO crash typically receives one point and the points increase as the severity of the crash increases. While other measures do not explicitly mention severity, analysts can adapt any of the measures to consider any severity level. For example, an analyst could use crash frequency (measure 1), focusing on the frequency of fatal and severe injury crashes to priority rank sites.

The severity distribution of crashes may be a function of site characteristics including AADT. For example, sections with higher AADT values may be associated with lower speeds and consequently fewer severe crashes. A recent study conducted as part of NCHRP Project 17-45 (*Enhanced Safety Prediction Methodology and Analysis Tool for Freeways and Interchanges*) confirmed this relationship. [5] The researchers used data from freeways in Maine, California, and Washington to estimate severity distribution functions. Specifically, they predicted the proportion of crashes in each severity category as a function of geometric and operational characteristics. The results indicated a reduction in the proportion of high-severity crashes with an increase in traffic volume. AASHTO published the results from this study as a supplemental volume (Chapters 18 and 19) to the Highway Safety Manual.

NETWORK SCREENING PERFORMANCE MEASURES

The remainder of this section describes specific network screening performance measures in detail, identifying issues related to specific measures. The following is a description of six potential network screening performance measures discussed throughout this guide:

- Average crash frequency.
- Crash rate.
- Level of service of safety.
- Excess predicted average crash frequency using SPFs.
- Expected average crash frequency with EB adjustments.
- Excess expected average crash frequency with EB adjustment.

Average Crash Frequency

Screening the network by means of the average crash frequency is a comparison of the number of crashes among sites. Analysts rank sites in descending order by the number of crashes, possibly by type or severity. This measure does not account for potential bias due to RTM or differences in traffic volume among sites.

Crash Rate

Screening the network by means of the crash rate is a comparison of the number of crashes per some measure of exposure such as traffic volume. This measure does not account for potential bias due to RTM. While the crash rate does account for differences in traffic volume among sites, it does not account for the nonlinear relationship between crashes and traffic volume.

Level of Service of Safety

The LOSS employs an SPF to compare observed crash frequency for a given site to the predicted crash frequency for the given traffic volume. The analyst defines levels to categorize the sites with respect to the difference between observed and predicted crash frequency. The following is an example categorization:

- Level 1: sites for which the observed crash frequency is greater than the predicted crash frequency and exceeds the upper 95 percent confidence limit of the SPF.
- Level 2: sites for which the observed crash frequency is greater than the predicted crash frequency and does not exceed the upper 95 percent confidence limit of the SPF.
- Level 3: sites for which the observed crash frequency is less than the predicted crash frequency and does not exceed the lower 95 percent confidence limit of the SPF.
- Level 4: sites for which the observed crash frequency is less than the predicted crash frequency and exceeds the lower 95 percent confidence limit of the SPF.

As described, the original version of this measure does not account for potential bias due to RTM, but can be adapted to account for RTM. [(4)] It does account for differences in traffic volume among sites and the nonlinear relationship between crash frequency and traffic volume.

Excess Predicted Average Crash Frequency using SPFs

The excess predicted average crash frequency using SPFs is similar to the LOSS. The excess predicted crash frequency is the difference between the observed and predicted average crash frequency at a site. Analysts rank sites in descending order of excess predicted average crash frequency. This measure does not account for potential bias due to RTM. It does account for differences in traffic volume among sites and the nonlinear relationship between crash frequency and traffic volume.

Expected Average Crash Frequency with EB Adjustments

The EB method combines the observed crash frequency with the predicted crash frequency for a given site to produce an estimate of the expected average crash frequency. The predicted crash frequency is from an SPF. Similar to the average crash frequency measure, analysts rank sites from high to low based on the expected average crash frequency. This measure accounts for potential bias due to RTM, differences in traffic volume among sites, and the nonlinear relationship between crash frequency and traffic volume.

The following is a brief overview of the EB method:

1. **Identify Reference Group**: Identify a group of sites representative of the facility and site type of interest for network screening. The reference group should reflect 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.
3. **Estimate Predicted Crashes**: Use the SPFs and traffic volume data for each site included in the network screening to estimate the predicted number of crashes for each year in the study period.
4. **Estimate Expected Crashes**: Using the EB method, compute the expected crashes for each site-year in the study period as the weighted sum of predicted crashes from the SPF and observed crashes. For details, refer to Hauer or the Highway Safety Manual. [1,6]

The outcome of step 4 is the expected average crash frequency with EB adjustment. Analysts can estimate the EPDO average crash frequency with EB adjustment (measure 12) by repeating the four-step process for each severity category and then combining them based on the weight for each severity category.

Excess Expected Average Crash Frequency with EB Adjustment

The excess expected average crash frequency with EB adjustment is the difference between the expected crashes and the predicted crashes. Analysts rank sites from high to low based on the excess expected average crash frequency. This measure accounts for potential bias due to RTM, differences in traffic volume among sites, and the nonlinear relationship between crash frequency and traffic volume. Further, it establishes a threshold using the SPF to provide an indication of when sites are performing relatively well or not with respect to other similar sites.

Summary of Network Screening Performance Measures

The key to effective network screening is selecting an appropriate performance measure. More reliable performance measures account for potential bias due to RTM, properly account for differences in traffic volume among sites, and consider crash severity. The following measures explicitly account for possible bias due to RTM and the nonlinear relationship between crash frequency and traffic volume.

- Expected average crash frequency with EB adjustments.
- EPDO average crash frequency with EB adjustment.
- Excess expected average crash frequency with EB adjustment.

3. DEMONSTRATING THE VALUE OF MORE RELIABLE METHODS

There is general agreement within the safety research community that EB-based performance measures provide more reliable results than traditional crash-based performance measures such as crash frequency and crash rate for network screening. Specifically, the EB-based measures account for possible bias due to RTM and the nonlinear relationship between crash frequency and traffic volume.

This section demonstrate the value of applying more reliable network screening performance measures to account for RTM bias, changes in traffic volume, and the nonlinear relationship between crash frequency and traffic volume. Empirical examples highlight the shortcomings of less reliable measures, which may lead to less reliable results and conclusions. For interested readers, the appendix presents further details on the empirical examples. The first example presents the results of research performed specifically for this guide. The remaining examples reflect the results of other published studies that used real-world data to compare empirical results for network screenings using different performance measures. Note the examples illustrate general comparative results of the performance measures. Different data and relationships within the data may produce different results. In general, the examples demonstrate the value of applying more reliable performance measures.

EXAMPLE 1: ACCOUNTING FOR RTM BIAS, DIFFERENCES IN TRAFFIC VOLUME, AND THE NONLINEAR RELATIONSHIP BETWEEN CRASH FREQUENCY AND TRAFFIC VOLUME USING EB-BASED MEASURES AND DATA FROM MINNESOTA

The first example presents the results of research performed specifically for this guide. It demonstrates the value of using the EB method to account for RTM bias, differences in traffic volume among sites, and the nonlinear relationship between crash frequency and traffic volume. If RTM is present and not properly accounted for, then the analyst will incorrectly identify sites with randomly high crashes as sites for further investigation and potential treatment. If traffic volumes differ among sites and the analyst does not properly account for these differences, then they may incorrectly identify sites with high traffic volume as sites with high potential for improvement. These issues, left unaccounted, can lead to misallocation of resources. Refer to the appendix for details related to example 1.

This example involves a multiyear network screening based on four performance measures and six years of intersection data for Minnesota. The four network screening performance measures are: crash frequency, crash rate (crash frequency divided by total entering traffic volume), EB expected, and EB expected excess crashes (also called potential for safety improvement—PSI). The dataset included three site types and two study periods for network screening analysis. The site types of interest were three-legged stop-controlled intersections, four-legged stop-controlled intersections, and four-legged signalized intersections. The two study periods were 2007 to 2009 and 2010 to 2012.

The analysts used data in the first period (2007 to 2009) to produce one ranked list of sites, and separately used the data in the second period (2010 to 2012) to produce another ranked list of sites. The analysts compared the network screening results from the two periods to determine the following values for each performance measure.

- **Number of correct positives** = number of top ranked sites from the first period that continue to belong to the list of top ranked sites in the second period.
- **Number of false positives** = number of top ranked sites from the first period that are no longer on the list of top ranked sites in the second period.
- **Number of correct negatives** = number of sites that are not on the top ranked lists in both periods.
- **Number of false negatives** = number of sites that are not on the list of top ranked sites in the first period, but are on the list of top ranked sites in the second period.
- **Total number of positives** = number of correct positives plus false negatives.
- **Total number of negatives** = number of correct negatives plus false positives.

The analysts evaluated results from the various performance measures using the sensitivity and specificity. Figure 5 and Figure 6 provide the equations for sensitivity and specificity, where higher values indicate better measures.

$$\text{Sensitivity} = \frac{\text{number of correct positives}}{\text{total number of positives}}$$

Figure 5. Equation. Sensitivity.

$$\text{Specificity} = \frac{\text{number of correct negatives}}{\text{total number of negatives}}$$

Figure 6. Equation. Specificity.

The evaluation included five different ranked lists: top 10, top 20, top 50, top 100, and top 200 sites. Table 3 to Table 7 present the results of the evaluation, indicating the best results for each evaluation criteria in bold. In general, the results consistently indicate the EB expected measure performs best with respect to the four measures. The crash frequency measure performs well relative to the EB expected measure, and is the second best measure in most cases, particularly as the number of sites increases. Crash rate consistently performs worst with respect to the four measures.

Table 3. Evaluation of network screening performance measures (top 10 sites).

Intersection Type	Evaluation Criteria	Crash Frequency	Crash Rate	EB Expected	EB Expected Excess
3-legged stop-controlled	Sensitivity	0.5	0.3	**0.6**	0.4
	Specificity	0.993	0.99	**0.994**	0.991
4-legged stop-controlled	Sensitivity	0.7	0.2	0.7	**0.8**
	Specificity	0.996	0.991	0.996	**0.998**
4-legged signalized	Sensitivity	0.6	0.5	**0.7**	**0.7**
	Specificity	0.992	0.99	**0.994**	**0.994**

Note: Bold indicates the best result for each evaluation criteria.

Table 4. Evaluation of network screening performance measures (top 20 sites).

Intersection Type	Evaluation Criteria	Crash Frequency	Crash Rate	EB Expected	EB Expected Excess
3-legged stop-controlled	Sensitivity	0.35	0.3	**0.55**	0.45
	Specificity	0.981	0.98	**0.987**	0.984
4-legged stop-controlled	Sensitivity	0.6	0.3	**0.8**	0.5
	Specificity	0.99	0.983	**0.995**	0.988
4-legged signalized	Sensitivity	**0.65**	0.55	**0.65**	0.6
	Specificity	**0.986**	0.982	**0.986**	0.984

Note: Bold indicates the best result for each evaluation criteria.

Table 5. Evaluation of network screening performance measures (top 50 sites).

Intersection Type	Evaluation Criteria	Crash Frequency	Crash Rate	EB Expected	EB Expected Excess
3-legged stop-controlled	Sensitivity	0.5	0.36	**0.56**	0.4
	Specificity	0.962	0.951	**0.966**	0.954
4-legged stop-controlled	Sensitivity	0.64	0.42	**0.7**	0.58
	Specificity	0.978	0.964	**0.981**	0.974
4-legged signalized	Sensitivity	**0.66**	0.58	**0.66**	0.6
	Specificity	**0.963**	0.955	**0.963**	0.957

Note: Bold indicates the best result for each evaluation criteria.

Table 6. Evaluation of network screening performance measures (top 100 sites).

Intersection Type	Evaluation Criteria	Crash Frequency	Crash Rate	EB Expected	EB Expected Excess
3-legged stop-controlled	Sensitivity	0.6	0.38	**0.66**	0.53
	Specificity	0.934	0.898	**0.944**	0.922
4-legged stop-controlled	Sensitivity	0.66	0.39	**0.7**	0.63
	Specificity	0.955	0.919	**0.96**	0.951
4-legged signalized	Sensitivity	0.67	0.61	**0.68**	0.62
	Specificity	0.92	0.906	**0.923**	0.908

Note: Bold indicates the best result for each evaluation criteria.

Table 7. Evaluation of network screening performance measures (top 200 sites).

Intersection Type	Evaluation Criteria	Crash Frequency	Crash Rate	EB Expected	EB Expected Excess
3-legged stop-controlled	Sensitivity	0.63	0.445	**0.775**	0.51
	Specificity	0.854	0.781	**0.911**	0.806
4-legged stop-controlled	Sensitivity	0.635	0.53	**0.725**	0.555
	Specificity	0.889	0.856	**0.916**	0.864
4-legged signalized	Sensitivity	0.79	0.745	**0.795**	0.725
	Specificity	0.866	0.838	**0.869**	0.825

Note: Bold indicates the best result for each evaluation criteria.

EXAMPLE 2: ACCOUNTING FOR RTM BIAS AND THE NONLINEAR RELATIONSHIP BETWEEN CRASH FREQUENCY AND TRAFFIC VOLUME USING EB-BASED MEASURES AND DATA FROM CALIFORNIA

The second example further supports the use of EB-based performance measures in network screening. EB-based performance measures account for potential bias due to RTM, differences in traffic volume among sites, and the nonlinear relationship between crash frequency and traffic volume. These issues, left unaccounted, can lead to misallocation of resources.

This example involves a multiyear network screening based on four performance measures and eight years of data (2000 to 2007) for rural, four-legged, stop-controlled intersections in California. [(7)] The four network screening performance measures are: EB expected, EB expected excess, LOSS, and the Caltrans 'Table C' method. Note the Table C method uses critical crash rates and crash frequency thresholds as described in the appendix. For the EB expected, EB expected excess, and the LOSS measures, the study assessed the performance of three underlying SPFs: SPFs with AADT only (SPF1), SPFs with AADT and additional variables (SPF2), and default SPFs with AADT only from AASHTOWare Safety Analyst™ (SPF SA). Given the similarity in results, the remainder of this section presents only the results using SPF1. Refer to the appendix for details related to example 2, including the Table C method and additional results for the other SPFs.

The evaluation began with an investigation of the potential bias due to RTM. For this, the analysts compared the crash frequency of top ranked sites in 2000 to 2003 with the crash frequency for the same sites in 2004 to 2007. Table 8 presents the results of the analysis, indicating substantial influence due to RTM. Specifically, sites identified as high-crash locations based on data from 2000 to 2003 tend to have much lower average crash counts in 2004 to 2007 due to random variation in crashes. Similarly, sites identified as low-crash locations based on data from 2000 to 2003 tend to have higher average crash counts in 2004 to 2007 due to random variation in crashes. The average number of crashes per intersection in 2000 to 2003 was 3.86 crashes; the example clearly shows the potential for bias due to RTM.

Given the presence of RTM, there is a need for the network screening performance measure to account for RTM. Recall the EB-based measures are able to account for potential bias due to RTM while the other measures are not.

Table 8. Illustration of RTM in rural, four-legged, stop-controlled intersections.

Group (crashes per site 2000 2003)	Number of sites in group	Sum of crashes in group (2000 2003)	Sum of crashes in group (2004 2007)	Average crashes per site (2000 2003)	Average crashes per site (2004 2007)	Percent Change
40+	4	247	195	61.75	48.75	-21.05
30-39	15	494	337	32.93	22.47	-31.78
25-29	9	234	202	26.00	22.44	-13.68
20-24	28	617	545	22.04	19.46	-11.67
15-19	46	781	679	16.98	14.76	-13.06
10-14	112	1298	1213	11.59	10.83	-6.55
9	38	342	300	9.00	7.89	-12.28
8	35	280	310	8.00	8.86	10.71
7	64	448	388	7.00	6.06	-13.39
6	70	420	375	6.00	5.36	-10.71
5	110	550	518	5.00	4.71	-5.82
4	121	484	454	4.00	3.75	-6.20
3	164	492	548	3.00	3.34	11.38
2	242	484	557	2.00	2.30	15.08
1	334	334	513	1.00	1.54	53.59
0	550	0	429	0.00	0.78	Infinite increase

Note: In 2000 to 2003, the mean frequency was 3.86 crashes per site, denoted by the thick line in the table.

The analysts used the following three approaches to compare the network screening measures:

1. **Approach 1 (Table 9):** Compare the ability of each measure to rank locations with high crash frequencies in the future. For a given year, the analyst applies the performance measure of interest to rank the sites. Then, selecting the top 10, 50, 100, and 200 sites, the analyst determines the total crashes in the future years in the study period. The preferred measure identifies sites that remain high-crash sites in future years. As an example, consider applying the measures to rank sites based on data from 2000. First, the analyst ranks all sites using the four performance measures and data for 2000. From this ranking, they select the top 10, 50, 100, and 200 sites for further summary. For each group of sites, the analyst determines the total number of crashes in the 'future' based on data from the remainder of the study period, in this case 2001 to 2007.

2. **Approach 2 (Table 10 and Table 11):** Compare, retrospectively, the performance of each measure in selecting and ranking correct positives and false positives. Correct positives are those locations identified by the Table C method and subsequently investigated and recommended for improvement. False positives are those investigated and not recommended for improvement. For this approach, the first step was to compile the list of sites identified and investigated by Caltrans each year. Note Caltrans used the Table C method to identify sites each year. Then, using data from the previous year, the analyst applied the performance measures and ranked the sites. The preferred measure identifies the most sites for which Caltrans recommended improvement (i.e., Caltrans confirmed a safety concern and identified a targeted mitigation measure). For example, consider the top-ranked sites generated by the Table C method for 2004, and subsequently investigated by Caltrans. The analyst would use data from the previous year (2003) to rank these sites based on the different performance measures. For each measure, the analyst tallies the number of sites for which Caltrans investigated and recommended improvement. It is important to note the starting point in this approach is sites selected for investigation based on results from the Table C method. Thus, the evaluation favors the Table C method with respect to producing an optimal and ranked list of locations. However, the results show the relative performance of other measures when ranking the sites previously recommended for improvement.
3. **Approach 3 (Table 12 and Table 13):** Compare the characteristics of top ranked locations by each measure. The two main characteristics selected for this comparison were the total intersection AADT and the expected number of crashes. The study includes the expected number of crashes (as opposed to the actual number of crashes) because it corrects for possible bias due to RTM and provides a better estimate of the true long-term crash propensity. To implement this approach, the researchers ranked sites based on each performance measure, and computed the average total intersection AADT and the average expected total crashes for the top ranked sites.

Table 9 presents the results of approach 1, which compares the performance measures based on future observed crashes. The table presents the future observed crashes for the various performance measures and four different ranked lists: top 10, top 50, top 100, and top 200 sites. Recall the preferred measure is the one identifying sites with the highest number of future crashes, indicated by bold text in the table.

Results indicate the EB-based measures (EB expected and EB expected excess) performed better than the Caltrans Table C method, which is based on critical crash rates and crash frequency thresholds. Specifically, the top ranked sites based on the EB-based measures had more crashes in the future compared to the top ranked sites from the Table C method. The EB expected measure performed better than the EB expected excess and LOSS measures for this comparison because it ranks sites based on number of expected crashes as opposed to excess crashes. As shown in the appendix, the results based on different SPFs are relatively consistent within a given measure compared to the differences among the measures. This indicates the performance measure is more critical than the type of SPF (AADT only or AADT plus additional variables) used in the process.

Table 9. Future crashes when ranked by performance measures.

Year	Top Ranked Sites	EB Expected	EB Expected Excess	LOSS	Table C Method
2000	10	**648**	468	480	463
	50	**1941**	1733	1567	1354
	100	**3161**	2886	2383	2131
	200	**4989**	4427	3819	2958
2001	10	**562**	451	413	445
	50	**1728**	1570	1419	1271
	100	**2807**	2604	2304	2091
	200	**4418**	3960	3529	2742
2002	10	**394**	374	361	368
	50	**1355**	1322	1255	1169
	100	**2339**	2180	1945	1733
	200	**3677**	3286	3081	2478
2003	10	**338**	336	324	294
	50	1132	**1133**	1054	960
	100	**1915**	1814	1651	1538
	200	**2951**	2695	2477	2050
2004	10	**229**	209	201	217
	50	**806**	759	692	651
	100	**1338**	1267	1182	990
	200	**2119**	1952	1760	1396
2005	10	147	153	156	**159**
	50	**527**	505	470	406
	100	**854**	819	745	687
	200	**1400**	1257	1155	895
2006	10	**71**	64	57	54
	50	**258**	255	215	183
	100	**432**	382	355	307
	200	**664**	609	528	386

Note: The LOSS method presented in this table is the original LOSS method (i.e., difference between observed and predicted crashes) and does not account for bias due to RTM. In 2015, a new method was proposed for using LOSS in concert with the EB method (i.e., difference between expected and predicted) to correct for RTM bias. [4]

The second approach compares performance measures on the identification of correct positives and false positives. Again, correct positives are those locations investigated based on the Table C method and recommended for improvement. False positives are those investigated and not recommended for improvement. The preferred measure is the one producing the most correct positives and fewest false positives.

Table 10 presents the correct positives for rural, four-legged, stop-controlled intersections for 2003 through 2008. The table presents results for the various performance measures and four different ranked lists: top 5, top 10, top 20, and top 50 sites. For each of the six years, bold text indicates the highest value associated with the preferred measure. Caltrans investigated sites after using the Table C method to screen the network. Hence, the Table C method performs well in this comparison because it was the basis for identifying the initial sites. The results indicate the EB-based and LOSS measures performed equally well compared to the Table C method and, in a few cases, performed better than the Table C method.

Table 10. Number of correct positives by performance measure.

Year	Sites Investigated (Sites Recommended for Improvement)	Top Ranked Sites	EB Expected	EB Expected Excess	LOSS	Table C Method
2003	68 (27)	5	3	3	**4**	3
		10	**6**	**6**	4	5
		20	9	10	10	**11**
		50	**23**	21	22	22
2004	81 (18)	5	**1**	**1**	**1**	**1**
		10	**2**	**2**	1	**2**
		20	5	4	4	**6**
		50	10	10	**14**	13
2005	34 (2)	5	**1**	**1**	**1**	**1**
		10	**1**	**1**	**1**	**1**
		20	**2**	**2**	1	**2**
		50	--	--	--	--
2006	76 (30)	5	0	0	1	**2**
		10	0	1	**2**	**2**
		20	4	3	5	**6**
		50	15	15	16	**19**
2007	50 (17)	5	0	**1**	0	0
		10	2	2	**3**	**3**
		20	6	6	6	**7**
		50	**17**	**17**	**17**	**17**
2008	39 (15)	5	**2**	**2**	1	**2**
		10	3	4	3	**5**
		20	**8**	7	7	**8**
		50	--	--	--	--

Note: The second column indicates the total sites identified by the Table C method and investigated by Caltrans. The number in parentheses indicates the number of sites recommended for improvement by Caltrans. The LOSS method presented in this table is the original LOSS method (i.e., difference between observed and predicted crashes) and does not account for bias due to RTM. In 2015, a new method was proposed for using LOSS in concert with the EB method (i.e., difference between expected and predicted) to correct for RTM bias. [4]

Table 11 presents the false positives for rural, four-legged, stop-controlled intersections for 2003 through 2008. The table presents results for the various performance measures and four different ranked lists: top 5, top 10, top 20, and top 50 sites. For each of the six years of analysis, bold text indicates the lowest value corresponding with the preferred measure. The top ranked sites from the EB expected measure generally had fewer false positives compared to the Table C method and the LOSS and EB expected excess measures, particularly for the top 20 and top 50 sites.

Table 11. Number of false positives by performance measure.

Year	Sites Investigated (Sites Rejected for Improvement)	Top Ranked Sites	EB Expected	EB Expected Excess	LOSS	Table C Method
2003	68 (41)	5	2	2	**0**	1
		10	**2**	3	4	3
		20	**6**	8	8	8
		50	**15**	19	22	21
2004	81 (63)	5	**2**	3	3	3
		10	6	**5**	6	7
		20	**8**	11	11	10
		50	**23**	26	29	26
2005	34 (32)	5	2	2	**1**	**1**
		10	**3**	**3**	**3**	4
		20	**4**	7	8	9
		50	**12**	13	13	15
2006	76 (46)	5	5	5	4	**3**
		10	10	9	**8**	**8**
		20	**13**	**13**	**13**	**13**
		50	**19**	25	27	26
2007	50 (33)	5	5	**4**	5	5
		10	6	**5**	7	6
		20	9	8	**7**	**7**
		50	14	16	15	**13**
2008	39 (24)	5	**1**	**1**	2	2
		10	**1**	2	2	3
		20	**3**	**3**	4	4
		50	**8**	**8**	11	10

Note: The second column indicates the total sites identified by the Table C method and investigated by Caltrans. The number in parentheses indicates the number of sites not recommended for improvement by Caltrans. The LOSS method presented in this table is the original LOSS method (i.e., difference between observed and predicted crashes) and does not account for bias due to RTM. In 2015, a new method was proposed for using LOSS in concert with the EB method (i.e., difference between expected and predicted) to correct for RTM bias. [4]

The third approach focused on the characteristics of sites identified by the various measures. Specifically, the characteristics of interest are the average total entering intersection traffic volume and average expected crashes in the last three years of the study period.

Table 12 and Table 13 present the results for traffic volume and expected crashes, respectively. The EB expected measure identifies sites with the highest average traffic volume and the highest number of expected crashes. The Table C method identifies sites with the lowest average traffic volume and the lowest expected number of crashes. The Table C method does not account for potential bias due to RTM or the nonlinear relationship between crash frequency and traffic volume. On average, the LOSS measure identifies sites with higher average traffic volumes and more expected crashes than sites identified by the Table C method. This is likely because the LOSS measure accounts for the nonlinear relationship between crash frequency and traffic volume. On average, the EB expected excess measure identifies sites with higher average traffic volumes and more expected crashes than sites identified by the LOSS measure. This is likely because the EB-based measures explicitly account for potential bias due to RTM.

Table 12. Average traffic volume at top ranked sites.

Year	Top Ranked Sites	EB Expected	EB Expected Excess	LOSS	Table C
2004	10	30,553	23,301	20,474	18,031
	50	23,945	20,301	16,720	11,123
	100	22,109	16,985	15,456	10,308
	200	18,874	15,016	13,199	7,786
2005	10	25,120	24,308	21,535	19,626
	50	23,525	17,711	15,952	12,280
	100	21,356	16,077	13,978	10,532
	200	19,761	14,246	12,468	7,696
2006	10	27,282	25,065	17,223	15,879
	50	22,424	18,347	17,000	11,009
	100	21,538	16,527	14,124	10,422
	200	19,668	14,795	12,602	7,020

Note: The LOSS method presented in this table is the original LOSS method (i.e., difference between observed and predicted crashes) and does not account for bias due to RTM. In 2015, a new method was proposed for using LOSS in concert with the EB method (i.e., difference between expected and predicted) to correct for RTM bias. [4]

Table 13. Average expected crashes at top ranked sites.

Year	Top Ranked Sites	EB Expected	EB Expected Excess	LOSS	Table C
2004	10	11.2	10.7	10.2	9.8
	50	6.6	6.2	5.5	4.9
	100	5.2	4.9	4.5	3.8
	200	3.9	3.6	3.3	2.5
2005	10	10.4	10.1	9.7	9.2
	50	6.3	6.0	5.4	5.0
	100	5.0	4.7	4.3	3.8
	200	3.8	3.4	3.2	2.5
2006	10	10.5	9.3	7.8	8.2
	50	6.2	5.8	5.5	4.5
	100	4.7	4.4	4.0	3.5
	200	3.6	3.3	2.9	2.2

Note: The LOSS method presented in this table is the original LOSS method (i.e., difference between observed and predicted crashes) and does not account for bias due to RTM. In 2015, a new method was proposed for using LOSS in concert with the EB method (i.e., difference between expected and predicted) to correct for RTM bias. [(4)]

The original study included the same comparisons for other site types, including rural, four-legged, signalized intersections; rural, two-lane, undivided road segments; and urban freeway segments. The results for the other site types were consistent with the results for rural, four-legged, stop-controlled intersections. The EB expected measure tends to perform best, and in general, the EB-based measures performed better than the LOSS and Table C method. Again, the LOSS method does not account for potential bias due to RTM and the Table C method does not account for potential bias due to RTM and the nonlinear relationship between crash frequency and traffic volume. This may explain, at least in part, why the LOSS and Table C methods are not as effective in identifying sites with a large number of future crashes.

EXAMPLE 3: ACCOUNTING FOR DIFFERENCES IN CRASH SEVERITY USING EB-BASED MEASURES AND DATA FROM COLORADO

The third example demonstrates the value of using severity-weighted EB-based measures to account for differences in severity among sites. If crash severity differs among sites and the analyst does not properly account for these differences, then they may incorrectly identify sites with higher counts of less severe crashes over sites with lower counts of more severe crashes. The EB-based measures also account for RTM bias, differences in traffic volume among sites, and the nonlinear relationship between crash frequency and traffic volume. If RTM is present and not properly accounted for, then the analyst will incorrectly identify sites with randomly high crashes as sites for further investigation and potential treatment. If traffic volumes differ among sites and the analyst does not properly account for these differences, then they may incorrectly identify sites with high traffic volume as sites with high potential for improvement. These issues, left unaccounted, can lead to misallocation of resources. Refer to the appendix for details related to example 3.

This example involves a network screening exercise based on five performance measures and data for rural, two-lane, undivided roads in Colorado. [8] The objective of this study was to determine the network screening performance measure that is most likely to lead to cost-beneficial projects. The five network screening performance measures are as follows:

1. EB expected total: sites with the most expected crashes.
2. EB expected severity-weighted: sites with the most expected severity-weighted crashes.
3. EB expected excess total: sites with the most excess expected crashes.
4. EB expected excess severity-weighted: sites with the most excess expected severity-weighted crashes.
5. Combination of EB expected and EB expected excess: the product of the expected crashes per mile-year (expressed in crashes per mile-year) and the excess crashes per mile-year (expressed in standard deviations).

The researchers employed a pairwise comparison, comparing two performance measures at a time to identify sites for further investigation and potential treatment. Starting with all sites for the facility type of interest (i.e., rural, two-lane, undivided roads), the researchers generated two ranked lists based on two performance measures. For the top ranked sites *not common* to both lists, the researchers performed a detailed engineering study to diagnose the contributing factors and identify targeted improvements. Finally, the researchers estimated the costs and benefits for the proposed improvements at each location, and determined the performance measure most likely to lead to more cost-beneficial projects. The researchers repeated this process, retaining the superior performance measure from each comparison for comparison against the remaining performance measures. They did not consider the inferior results further.

The researchers performed detailed engineering studies for 22 unique top-ranking sites. These studies involved the following steps:

1. Review detailed crash history of the site.
2. Use geographic information system (GIS) maps to assess horizontal alignment.
3. Review the video log of the site.

The researchers used this information to identify the underlying contributing factors and determine appropriate safety improvements at each site. At the 22 unique sites, the researchers identified 61 actions (projects), and subsequently estimated the costs and safety benefits of each action. They estimated project benefits using crash modification factors (CMFs) applied to raw crash counts (i.e., observed crash history) as well as the EB expected crashes. The researchers present results for both options, but explain the value in using the EB expected crashes as the basis for benefit-cost analyses.

Comparing benefit-cost ratios based on EB expected crashes, the EB expected severity-weighted measure resulted in the most cost-effective projects. The EB expected and EB expected severity-weighted measures outperformed the other measures based on expected excess crashes. This result is intuitive because the benefit of a proposed improvement reflects the change in expected crashes, and not the change in expected excess crashes.

In summary, this study supports the use of the EB expected severity-weighted measure to screen the network. This measure accounts for potential bias due to RTM, changes in traffic volume, and the nonlinear relationship between crash frequency and traffic volume. While it does not establish a threshold to identify sites with high crashes relative to the average expected crashes, it does account for differences in crash severity among sites.

EXAMPLE 4: AN EVALUATION OF FREQUENCY, RATE, EB EXPECTED, AND EB EXPECTED EXCESS MEASURES USING DATA FROM NEW HAMPSHIRE

The fourth example demonstrates the value of using the EB expected measure to account for RTM bias, differences in traffic volume among sites, and the nonlinear relationship between crash frequency and traffic volume. If RTM is present and not properly accounted for, then the analyst will incorrectly identify sites with randomly high crashes as sites for further investigation and potential treatment. If traffic volumes differ among sites and the analyst does not properly account for these differences, then they may incorrectly identify sites with high traffic volume as sites with high potential for improvement. These issues, left unaccounted, can lead to misallocation of resources. Refer to the appendix for details related to example 4.

This example involves a network screening exercise based on four performance measures and data for stop-controlled and signalized intersections in New Hampshire. [9] The objective of this study was to determine the network screening performance measure that is most likely to produce a list of sites with the greatest potential for safety improvement and subsequently result in the greatest safety benefit and most cost-effective safety improvements. The four network screening performance measures are as follows:

1. Fatal and injury crash frequency.
2. Fatal and injury crash rate.
3. EB expected fatal and injury crashes.
4. EB expected excess fatal and injury crashes.

The study simulated the development of projects for a safety program, following the safety management process from network screening through economic analysis. The research team developed ranked lists of intersections based on the four network screening performance measures. For 35 sites selected from network screening, the researchers performed a detailed engineering study to diagnose the contributing factors and identify targeted improvements. The team conducted desktop reviews (i.e., review of all information virtually; no in-field site visits) to identify crash contributing factors and determine appropriate safety improvements at each site. The detailed engineering studies involved the following steps:

1. Review detailed crash history of the site.
2. Develop and review a collision diagram of the site.
3. Review traffic volumes for the major and minor road.
4. Use aerial images and street view images to virtually review site characteristics.

Finally, the team performed an economic analysis to estimate the benefit, cost, and overall benefit-cost ratio (BCR) for each suggested strategy, package of intersection improvements, and the program of projects generated from each network screening measure. The research team estimated project benefits using CMFs applied to raw crash counts (i.e., observed crashes). They estimated project costs based on various sources, including NHDOT cost estimates, State DOT websites, and research reports. The researchers compared the overall economic benefit and overall benefit-cost ratio for each of the four measures.

Table 14 presents the results of the economic analysis by performance measure, including the total estimated benefits, total estimated costs, and overall BCR across all related intersections.

Comparing economic benefits based on observed crash history, the EB excess expected measure and the EB expected measure produced the list of sites with the highest overall economic benefit and the highest return on investment, respectively. The crash frequency measure provided the second highest overall benefit and return on investment. The crash rate method produced the list of sites with the lowest overall benefit and lowest overall return on investment.

Table 14. BCR results by network screening performance measure.

Network Screening Performance Measure	Estimated Benefit	Estimated Cost	BCR
Crash Frequency	$17,942,270	$2,699,700	6.65
Crash Rate	$8,106,398	$3,396,450	2.39
EB Expected	$15,671,311	$2,213,950	**7.08**
EB Expected Excess	**$22,014,117**	$3,891,250	5.66

In summary, this study supports the use of the EB expected measure to screen the network. This measure accounts for potential bias due to RTM, changes in traffic volume, and the nonlinear relationship between crash frequency and traffic volume. While it does not establish a threshold to identify sites with high crashes relative to the average expected crashes, it can account for differences in crash severity among sites. One challenge to using EB measures is the need for an appropriate SPF, which may not be available to some agencies. When the EB expected measure is infeasible, the crash frequency measure may provide a reasonable alternative as it resulted in the second greatest overall benefit and the second greatest BCR.

EXAMPLE 5: SELECTING AN APPROPRIATE PERFORMANCE MEASURE WHEN EB-BASED MEASURES ARE INFEASIBLE

The fifth example provides insight into the selection of an appropriate performance measure when an agency is unable to use EB-based measures due to lack of data or expertise. Again, the EB-based measures account for RTM bias, differences in traffic volume among sites, and the nonlinear relationship between crash frequency and traffic volume. If it is not feasible to employ EB-based performance measures, then it is important to select a measure that performs well with respect to the EB-based measures. If an analyst applies a performance measure that does not perform well, then they may incorrectly identify sites with randomly high crashes and high traffic volumes as sites with high potential for improvement. This can lead to misallocation of resources. Refer to the appendix for details related to example 5.

This example involves a network screening exercise based on five performance measures and data for rural and urban, four-legged, signalized and two-way, stop-controlled intersections in Virginia. [10] The objective of this study was to compare the ranked lists from four traditional network screening performance measures with the ranked list from the excess expected average crash frequency with EB adjustment (i.e., Measure 13 from Table 1). For comparison, the study assumed the results of the EB-based measure as "ground truth." The four traditional performance measures are as follows:

- Crash frequency.
- Crash rate.
- Critical crash rate, also known as rate-quality control.
- Equivalent property damage only (EPDO).

The researchers compared the four traditional performance measures with the results from the EB-based measure based on the following four approaches.

- **Pearson's correlation coefficient:** This is the correlation coefficient between the rankings from the EB-based measure and the rankings from the traditional measure of interest based on all sites. A higher value of correlation is preferred.
- **Correct identification percentage:** This measure indicates how many sites with a PSI *greater than zero* (i.e., locations identified for further investigation by the EB-based measure) coincided with those identified by the traditional measure of interest. The researchers computed this for the top 1, 5, and 10 percent of locations. A higher percentage indicates a better measure.
- **False identification percentage:** This measure indicates how many sites with a PSI *less than zero* (i.e., locations not identified for further investigation by the EB-based measure) coincided with sites identified for further investigation by the traditional measure of interest. The researchers computed this for the top 1, 5, and 10 percent of locations. A lower percentage indicates a better measure.
- **Rank-based mean absolute error (MAE):** This measure is similar to the correct identification percentage, but considers the specific rank of individual sites. The researchers computed the MAE for the top 1, 5, and 10 percent of locations. Lower values of MAE indicate better measures.

Table 15 shows the Pearson correlation coefficient values for the four traditional measures. Bold numbers indicated the preferred measure. Considering all intersections, the critical rate measure performed best with a correlation coefficient of 0.576, and the crash frequency measure was second with a correlation coefficient of 0.494. For intersections identified by the EB measure for further investigation (i.e., PSI > 0), the crash frequency measure performed best with a correlation coefficient of 0.868, and the critical rate measure was second with a correlation coefficient of 0.713. The crash rate and EPDO measures performed poorly in both scenarios.

Table 15. Correlation coefficient values for the traditional measures.

Category	Crash Frequency	Crash Rate	Critical Rate	EPDO
All sites	0.494	0.376	**0.576**	0.252
Sites identified for further investigation (PSI > 0)	**0.868**	0.153	0.713	0.386

Note: This information is the same as Table 3 from Lim and Kweon. [8]

Table 16 through Table 18 show the results from the other three comparisons for the top 1, 5, and 10 percent of locations, respectively. Bold numbers indicated the preferred measure. For the top 1 percent, the crash frequency measure performs best with the highest correct identification percentage, lowest false identification percentage, and lowest MAE rate. For the top 5 and 10 percent, the critical rate measure performs best with respect to correct identification percentage and false identification percentage, whereas the crash frequency measure performs best with respect to MAE rate. While the crash rate measure performs well with respect to false identification percentage, it performs the worst with respect to correct identification percentage and MAE rate in all three tables.

The overall conclusion is the crash frequency and critical rate measures perform well (i.e., are closer to the EB-based measure, which is assumed to represent ground truth) compared to the crash rate and EPDO measures. The crash rate measure does not perform well in two out of the three comparisons and the EPDO measure consistently performs poorly. When the EB expected measure is infeasible, the crash frequency and critical crash rate measures may provide a reasonable alternative.

Table 16. Comparison of traditional measures based on top 1 percent of sites.

Comparison	Crash Frequency	Crash Rate	Critical Rate	EPDO
Correct identification % (count)	**76.5 (13)**	6.9 (1)	52.9 (9)	0.0 (0)
False identification % (count)	**0 (0)**	**0 (0)**	**0 (0)**	29.4 (5)
MAE rate	**3.3**	129.0	9.4	22.9

Note: The EB-based measure identified a total of 17 locations for further investigation. Note: This information is the same as Table 3 from Lim and Kweon. (8)

Table 17. Comparison of traditional measures based on top 5 percent of sites.

Comparison	Crash Frequency	Crash Rate	Critical Rate	EPDO
Correct identification % (count)	67.9 (57)	20.2 (17)	**92.6 (78)**	67.9 (57)
False identification % (count)	8.3 (7)	**0 (0)**	**0 (0)**	9.5 (8)
MAE rate	**26.0**	208.0	33.4	35.9

Note: The EB-based measure identified a total of 84 locations for further investigation. Note: This information is the same as Table 3 from Lim and Kweon. (8)

Table 18. Comparison of traditional measures based on top 10 percent of sites.

Comparison	Crash Frequency	Crash Rate	Critical Rate	EPDO
Correct identification % (count)	65.9 (110)	23.6 (40)	**75.4 (126)**	71.3 (119)
False identification % (count)	15.6 (26)	**0 (0)**	**0 (0)**	13.8 (23)
MAE rate	**61.1**	230.4	66.4	62.5

Note: The EB-based measure identified a total of 167 locations for further investigation. Note: This information is the same as Table 3 from Lim and Kweon. (8)

SUMMARY OF NETWORK SCREENING PERFORMANCE MEASURES

The examples presented in this information guide demonstrate the value of applying more reliable performance measures such as the EB expected measure in network screening. More reliable measures account for potential bias due to RTM, changes in traffic volume, the nonlinear relationship between crash frequency and traffic volume, and differences in crash severity. Table 19 presents the 10 general measures compared in this guide, and indicates the ability of each to account for the potential sources of bias.

If RTM is present and not properly accounted for, then the analyst may incorrectly identify sites with randomly high crashes as sites for further investigation and potential treatment. If traffic volumes differ among sites and the analyst does not properly account for these differences, then they may incorrectly identify sites with high traffic volume as sites with high potential for improvement. Similarly, if crash severity differs among sites and the analyst does not properly account for these differences, then they may incorrectly identify sites with higher counts of less severe crashes over sites with lower counts of more severe crashes. These issues (RTM, differences in traffic volume, and differences in crash severity), left unaccounted, can lead to misallocation of resources.

Crash frequency alone does not account for any of the above issues. Performance measures such as crash rate, critical rate, and the Caltrans Table C method account for differences in traffic volume, but do not account for possible bias due to RTM or the nonlinear relationship between crash frequency and traffic volume. The EPDO measure accounts for differences in crash severity, but does not account for possible bias due to RTM or differences in traffic volume. The original LOSS measure accounts for the nonlinear relationship between crash frequency and traffic volume, but does not account for potential bias due to RTM. The LOSS measure described in this guide (i.e., the difference in observed and predicted crashes) has been revised in recent research to incorporate the EB method to account for potential bias due to RTM. The EB-based measures can be used to account for all sources of potential bias listed in Table 19.

The examples presented throughout this guide consistently demonstrate the value of applying the EB expected performance measure in network screening. The examples also provide an indication of the potential magnitude of differences in results obtained from various measures. In these examples, the EB expected measure outperforms traditional measures such as crash frequency, crash rate, critical rate, and LOSS. Within the EB-based measures, the EB expected measures tend to outperform the EB expected excess measures and the severity-weighted measures tend to outperform the non-severity-weighted measures. Note the examples reflect specific datasets, facility types, and site types, and results may vary for other datasets.

While the examples demonstrate the general reliability of the EB expected measure, there are situations when an agency may choose alternate performance measures for network screening. For example, an agency may select the EB expected excess measure over the EB expected measure when comparing sites among different facility or site types. The EB expected measure favors sites with the greatest expected crashes such as high-volume, signalized intersections or multilane facilities. The EB expected excess measure establishes a threshold and demonstrates the relative need, helping to normalize for general differences in safety performance among groups of sites. As such, the EB expected excess measure may help to combine ranked lists from different groups such as stop-controlled intersections, signalized intersections, and segments in both rural and urban areas. This hypothesis requires further investigation. In other cases, an agency may not have the required data or resources to employ EB-based performance measures. Several examples showed the crash frequency measure performs nearly as well as the EB expected measure, and often better than the EB expected excess measure. Further, the examples consistently showed the crash rate measure does not perform well. As such, the crash frequency measure may serve as a suitable alternative when the EB expected measure is impractical.

The results from measures 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 reinforce the need to apply more reliable measures such as the EB expected measure when conducting network screening. Otherwise, agencies may misallocate time and resources to sites incorrectly identified for further investigation and possible treatment, while locations with a truly high potential for cost-effective safety improvement remain untreated.

Table 19. Summary of sources of bias accounted for by performance measures.

Performance Measure	RTM	Changes in Traffic Volume	Nonlinear Relationship	Crash Severity
Crash frequency				
Crash rate		●		
Critical rate		●		
EPDO				●
Caltrans Table C method		●		
LOSS		●	●	
EB expected	●	●	●	
EB expected severity-weighted	●	●	●	●
EB expected excess	●	●	●	
EB expected excess severity-weighted	●	●	●	●

Note: The EPDO and severity-weighted measures directly account for differences in crash severity; however, analysts can adapt any of the measures to account for differences in severity. The LOSS method presented in this table is the original LOSS method (i.e., difference between observed and predicted crashes) and does not account for bias due to RTM. In 2015, a new method was proposed for using LOSS in concert with the EB method (i.e., difference between expected and predicted) to correct for RTM bias. [(8)]

4. DATA REQUIREMENTS FOR NETWORK SCREENING

Table 20 summarizes the data requirements for network screening performance measures. The following is a description of each data element.

- **Crash Data**: Summary of crashes by site for the study period, crash type, and crash severity of interest.
- **Roadway Data**: Characteristics to define the facility and site type of interest for network screening. Typical segment-level characteristics include area type (rural or urban), number of lanes, and median type. Typical intersection-level characteristics include area type (rural or urban), number of approaches, and type of traffic control.
- **Traffic Volume Data**: Summary of traffic volume by site for each year in the analysis. For years where traffic volumes are not available, consider estimating the value based on linear interpolation or extrapolation.
- **Other**: Additional data are required for specific performance measures such as factors to weight crashes by severity level and calibrated SPFs to predict average crashes.

Table 20. Data requirements for network screening performance measures.

Performance Measure	Crash	Roadway	Traffic	Other
Average Crash Frequency	•			
Crash Rate	•		•	
Equivalent Property Damage Only (EPDO) Average Crash Frequency	•			Factors
Relative Severity Index	•	•		Indices
Critical Rate (rate quality control)	•	•	•	
Excess Predicted Average Crash Frequency Using Method of Moments	•	•	•	
Level of Service of Safety	•	•	•	SPF
Excess Expected Average Crash Frequency Using SPFs	•	•	•	SPF
Probability of Specific Crash Types Exceeding Threshold Proportion	•	•		
Excess Proportions of Specific Crash Types	•	•		
Expected Average Crash Frequency with EB Adjustments	•	•	•	SPF
Equivalent Property Damage Only (EPDO) Average Crash Frequency with EB Adjustment	•	•	•	SPF
Excess Expected Average Crash Frequency with EB Adjustments	•	•	•	SPF

The results summarized in this guide indicate the EB expected measure and the crash frequency measure provide relatively high and comparable benefits and return on investment, particularly when compared to the crash rate measure. The question remains whether it is worthwhile to employ the EB expected measure over crash frequency for network screening. The answer depends on the cost to implement the EB expected measure and the difference in overall benefits given a fixed budget. If an agency can implement the EB expected measure for less than the difference in benefits between the EB expected and crash frequency measure, then it is worthwhile to pursue the more reliable EB expected measure.

Based on the results from the fourth example in this guide, the EB expected measure provides the greatest return on investment (BCR = 7.08) and the crash frequency measure provides the second greatest return on investment (BCR = 6.65). Assuming a $10M safety program budget, and assuming the BCRs hold for the entire program, the difference in annual benefits between the EB expected and crash frequency measure is $4.3M ($70.8M - $66.5M).

An agency could likely implement the EB expected network screening measure for less than $4.3M per year. The associated costs would include a basic roadway inventory of all centerline miles (i.e., the Model Inventory Roadway Elements—Fundamental Data Elements), an intersection inventory, development or calibration of intersection SPFs, and programming to integrate the EB expected measure in the existing network screening process. Any software upgrades to conduct network screening would be similar regardless of the performance measure.

This guide can assist highway safety practitioners in selecting network screening measures based on their ability to produce effective projects. The results provide insight into the effectiveness of the various screening measures. Specifically, the EB expected measure appears to provide a high return on investment, and the crash frequency measure provides a reasonable alternative when the EB expected measure is infeasible or in the interim while an agency prepares for more reliable methods. Based on these results, it is likely that agencies can implement the EB expected measure for less than the difference in benefits between the EB expected and crash frequency measures. The FHWA report, *Benefit-Cost Analysis of Investing in Data Systems and Processes for Data-Driven Safety Programs*, provides a methodology to quantify the economic returns from investing in safety data improvements. [11]

5. TOOLS AND RESOURCES FOR NETWORK SCREENING

Tools and resources are available to support network screening, including guides, databases, and software. Some guides provide a discussion of the network screening process or specific performance measures, while other guides relate to specific components of the process. For example, EB-based measures require calibrated SPFs, and guides are available to explain how to develop and calibrate SPFs for the facility and site type of interest.

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 network screening 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.

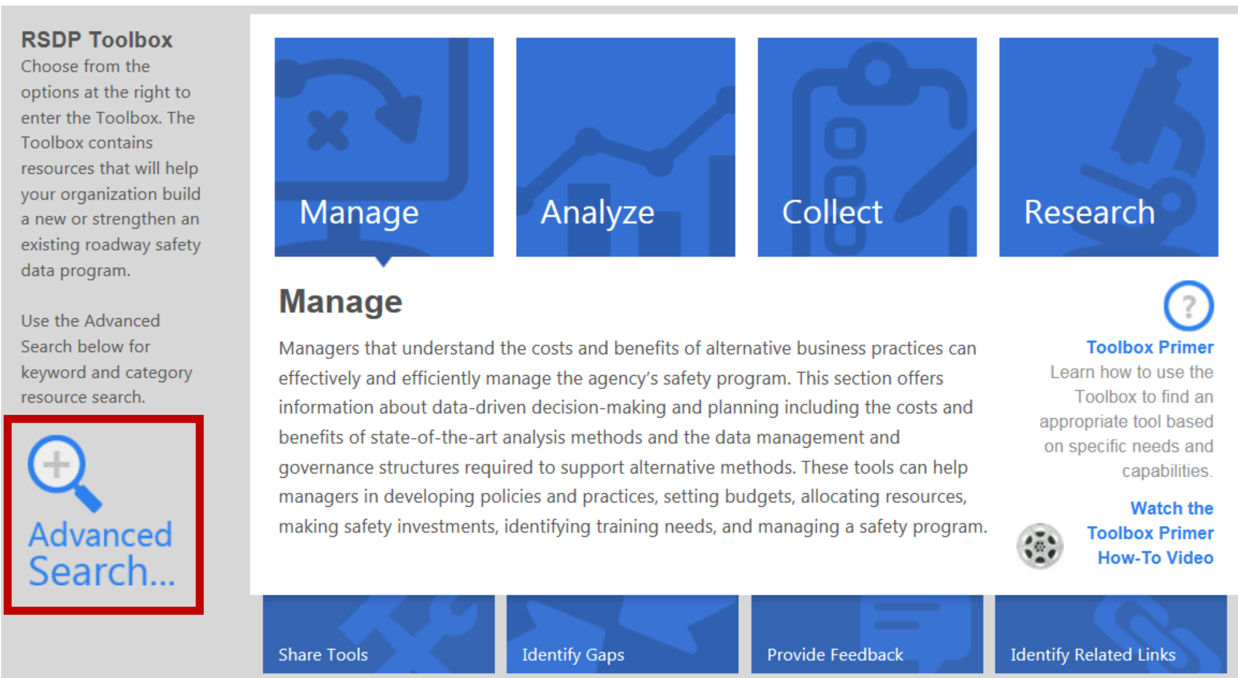


Figure 7. Image. Screenshot of Roadway Safety Data and Analysis Toolbox.

The following is a brief demonstration of the stepwise process to identify an appropriate tool to support network screening.

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.

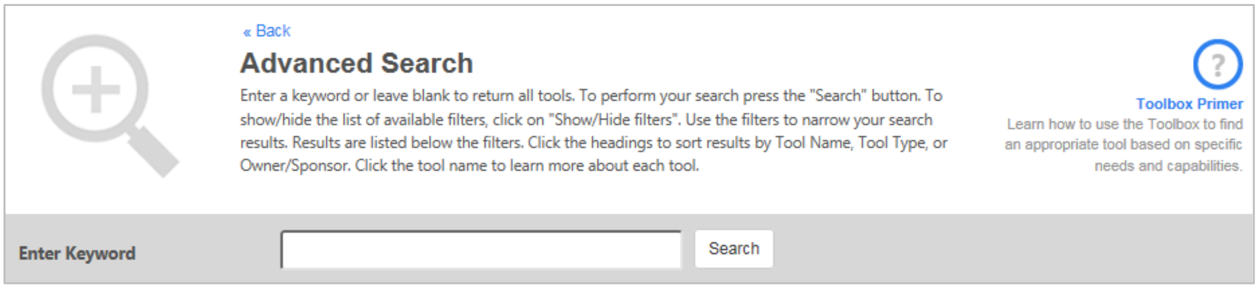


Figure 8. Image. Screenshot of advanced search feature.

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 'Network Screening' 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.

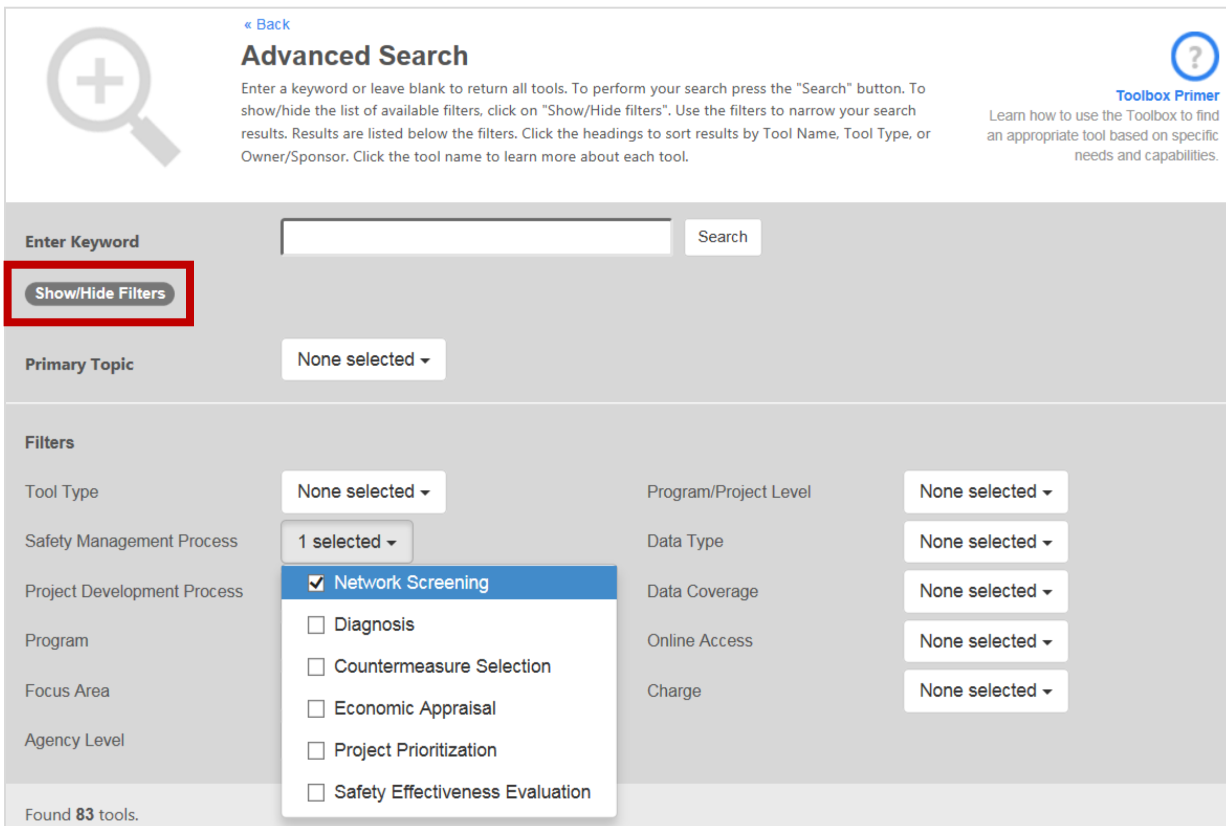


Figure 9. Image. Screenshot of filter options from advanced search page.

Using the stepwise process described in this section, the Toolbox returns guides such as FHWA's Highway Safety Improvement Program (HSIP) Manual and Systemic Safety Project Selection Tool. 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. Fixing America's Surface Transportation Act, Highway Safety Improvement Program [23 U.S.C. 148(b)].
3. Gross, F., T. Harmon, K. Peach, and G. Bahar. Reliability of Safety Management Methods: Systemic Safety Programs, Report No. FHWA-SA-16-041, Federal Highway Administration, Washington, DC, September 2016.
4. Kononov, J., Durso, C., Lyon, C., and Allery, B. (2015), Level of Service of Safety Revisited, Transportation Research Record: Journal of the Transportation Research Board 2514, pp. 10-20.
5. Srinivas G., Bonneson, J., Pratt, M., and Lord, D. (2012), Safety Prediction Methodology and Analysis Tool for Freeways and Interchanges, Project 17-45, National Cooperative Highway Research Program, Transportation Research Board, Washington, DC.
6. Hauer, E. (1997), Observational Before-After Studies in Road Safety, Pergamon, Elsevier Science Ltd.
7. Srinivasan, R., Lyon, C., Persaud, B., Martell, C., and Baek, J. (2011), Methods for Identifying High Collision Concentration Locations (HCCL) for Potential Safety Improvements – Phase II: Evaluation of Alternative Methods for Identifying HCCL, Prepared for California Department of Transportation, Sacramento, California.
8. Hauer, E., Allery, B.K., Kononov, J., and Griffith, M.S., (2004), How Best to Rank Sites with Promise, Journal of Transportation Research Board 1897, pp. 48-54.
9. Gross, F., T. Harmon, M. Albee, S. Himes, R. Srinivasan, D. Carter, and M. Dugas. Evaluation of Four Network Screening Performance Measures, Report No. FHWA-SA-16-103, Federal Highway Administration, Washington, DC, October 2016.
10. Lim, I.K. and Kweon, Y.J., (2013) Identifying High-Crash-Risk Intersections: Comparison of Traditional Methods with the Empirical Bayes-Safety Performance Function Method, Journal of Transportation Research Board 2364, pp. 44-50.
11. Lawrence, M., D. Cartwright-Smith, J. Mans, P. Nguyen, and N. Lefler. Benefit-Cost Analysis of Investing in Data Systems and Processes for Data-Driven Safety Programs: Project Report, Report No. FHWA-SA-12-029, Federal Highway Administration, Washington, DC, August 2012.

APPENDIX: EXAMPLE DETAILS

EXAMPLE 1: COMPARISON OF EB EXPECTED, EB EXPECTED EXCESS, CRASH RATE, AND CRASH FREQUENCY MEASURES USING DATA FROM MINNESOTA

Objective

The objective of this study was to compare and evaluate the results of four network screening performance measures. The results will help readers to understand the differences in the measures. The example includes the following four performance measures:

- Crash frequency.
- Crash rate (number of crashes divided by the total entering vehicles).
- EB expected.
- EB expected excess crashes (also called potential for safety improvement—PSI).

Description of Performance Measures

The earlier section titled, *Overview of Network Screening*, describes the performance measures in detail. In summary, the crash frequency measure does not account for potential bias due to RTM or changes in traffic volume. The crash rate measure accounts for differences in traffic volume among sites, but does not account for possible bias due to RTM or the nonlinear relationship between crash frequency and traffic volume. The EB expected and EB expected excess measures are able to properly account for all of these issues, including potential bias due to RTM, changes in traffic volume, and the nonlinear relationship between crash frequency and traffic volume.

Approach for Comparing Measures

The analysts developed a multiyear database, including two study periods: 2007 to 2009 and 2010 to 2012. The analysts used data in the first period (2007 to 2009) to produce one ranked list of sites, and separately used the data in the second period (2010 to 2012) to produce another ranked list of sites. The two EB-based measures incorporate SPFs developed from the multiyear dataset. The analysts compared the network screening results from the two periods to determine the following values for each performance measure.

- **Number of correct positives** = number of top ranked sites from the first period that continue to belong to the list of top ranked sites in the second period.
- **Number of false positives** = number of top ranked sites from the first period that are no longer on the list of top ranked sites in the second period.
- **Number of correct negatives** = number of sites that are not on the top ranked lists in both periods.
- **Number of false negatives** = number of sites that are not on the list of top ranked sites in the first period, but are on the list of top ranked sites in the second period.
- **Total number of positives** = number of correct positives plus false negatives.
- **Total number of negatives** = number of correct negatives plus false positives.

The analysts evaluated results from the various performance measures using the sensitivity specificity. Figure 10 and Figure 11 provide the equations for sensitivity and specificity, where higher values indicate better measures.

$$\text{Sensitivity} = \frac{\text{number of correct positives}}{\text{total number of positives}}$$

Figure 10. Equation. Sensitivity.

$$\text{Specificity} = \frac{\text{number of correct negatives}}{\text{total number of negatives}}$$

Figure 11. Equation. Specificity.

Data Description

The analysts used the Highway Safety Information System (HSIS) to obtain roadway geometry, traffic volume, and crash data from Minnesota for six years and three site types. Again, the study period included 2007 through 2012. The site types included three-legged stop-controlled, four-legged stop-controlled, and four-legged signalized intersections. The analysts removed intersections from the dataset if the number of legs, traffic control, or the number of through lanes on major/minor roads changed during the study period. The study focused on total crashes for the analysis. To overcome issues related to excessive zeros in the first period, and ensure the quality of the SPFs, the analysts removed sites with no crashes in the first period from the dataset. The final dataset included 706 three-legged stop-controlled, 855 four-legged stop-controlled, and 514 four-legged signalized intersections

Discussion of Results

The evaluation included five different ranked lists: top 10, top 20, top 50, top 100, and top 200 sites. Table 21 to Table 25 present the results of the evaluation, indicating the best results for each evaluation criteria in bold. Figure 12 to Figure 20 further illustrate the difference in results among the various performance measures. In general, the results consistently indicate the EB expected crashes performs best with respect to the four measures. The crash frequency measure performs well relative to the EB expected measure, and is the second best measure in most cases, particularly as the number of sites increases. Crash rate consistently performs worst with respect to the four measures.

Table 21. Evaluation of network screening performance measures (top 10 sites).

Intersection Type	Evaluation Criteria	Crash Frequency	Crash Rate	EB Expected	EB Expected Excess
3-legged stop-controlled	Sensitivity	0.5	0.3	**0.6**	0.4
	Specificity	0.993	0.99	**0.994**	0.991
4-legged stop-controlled	Sensitivity	0.7	0.2	0.7	**0.8**
	Specificity	0.996	0.991	0.996	**0.998**
4-legged signalized	Sensitivity	0.6	0.5	**0.7**	**0.7**
	Specificity	0.992	0.99	**0.994**	**0.994**

Note: Bold indicates the best result for each evaluation criteria.

Table 22. Evaluation of network screening performance measures (top 20 sites).

Intersection Type	Evaluation Criteria	Crash Frequency	Crash Rate	EB Expected	EB Expected Excess
3-legged stop-controlled	Sensitivity	0.35	0.3	**0.55**	0.45
	Specificity	0.981	0.98	**0.987**	0.984
4-legged stop-controlled	Sensitivity	0.6	0.3	**0.8**	0.5
	Specificity	0.99	0.983	**0.995**	0.988
4-legged signalized	Sensitivity	**0.65**	0.55	**0.65**	0.6
	Specificity	**0.986**	0.982	**0.986**	0.984

Note: Bold indicates the best result for each evaluation criteria.

Table 23. Evaluation of network screening performance measures (top 50 sites).

Intersection Type	Evaluation Criteria	Crash Frequency	Crash Rate	EB Expected	EB Expected Excess
3-legged stop-controlled	Sensitivity	0.5	0.36	**0.56**	0.4
	Specificity	0.962	0.951	**0.966**	0.954
4-legged stop-controlled	Sensitivity	0.64	0.42	**0.7**	0.58
	Specificity	0.978	0.964	**0.981**	0.974
4-legged signalized	Sensitivity	**0.66**	0.58	**0.66**	0.6
	Specificity	**0.963**	0.955	**0.963**	0.957

Note: Bold indicates the best result for each evaluation criteria.

Table 24. Evaluation of network screening performance measures (top 100 sites).

Intersection Type	Evaluation Criteria	Crash Frequency	Crash Rate	EB Expected	EB Expected Excess
3-legged stop-controlled	Sensitivity	0.6	0.38	**0.66**	0.53
	Specificity	0.934	0.898	**0.944**	0.922
4-legged stop-controlled	Sensitivity	0.66	0.39	**0.7**	0.63
	Specificity	0.955	0.919	**0.96**	0.951
4-legged signalized	Sensitivity	0.67	0.61	**0.68**	0.62
	Specificity	0.92	0.906	**0.923**	0.908

Note: Bold indicates the best result for each evaluation criteria.

Table 25. Evaluation of network screening performance measures (top 200 sites).

Intersection Type	Evaluation Criteria	Crash Frequency	Crash Rate	EB Expected	EB Expected Excess
3-legged stop-controlled	Sensitivity	0.63	0.445	**0.775**	0.51
	Specificity	0.854	0.781	**0.911**	0.806
4-legged stop-controlled	Sensitivity	0.635	0.53	**0.725**	0.555
	Specificity	0.889	0.856	**0.916**	0.864
4-legged signalized	Sensitivity	0.79	0.745	**0.795**	0.725
	Specificity	0.866	0.838	**0.869**	0.825

Note: Bold indicates the best result for each evaluation criteria.

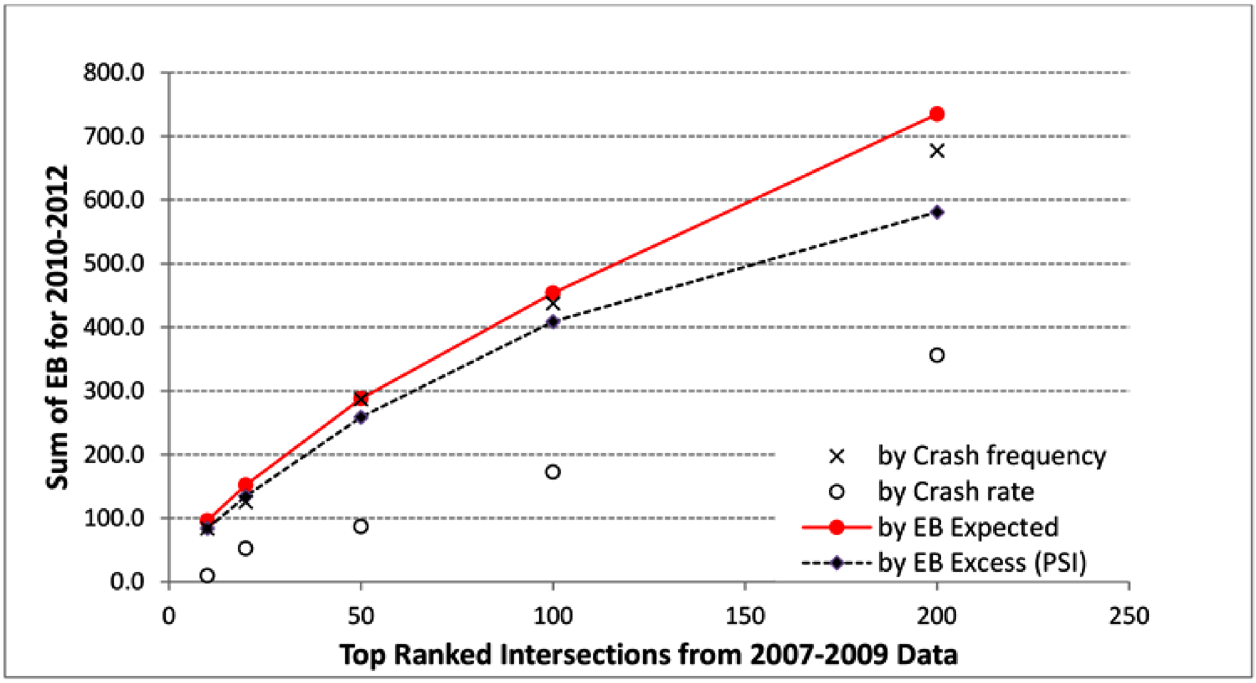


Figure 12. Graph. Sum of EB expected crashes (2010-2012) by various performance measures for MN 3-legged stop-controlled intersections (706 sites).

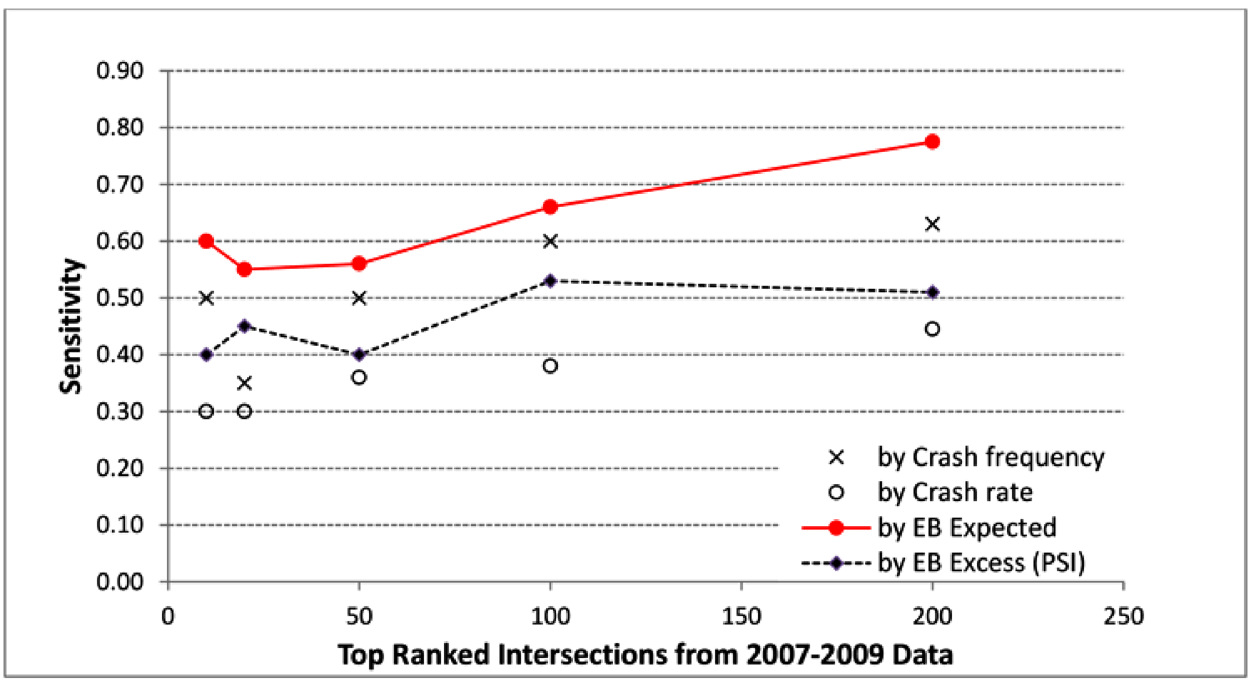


Figure 13. Graph. Sensitivity (2010-2012) by various performance measures for MN 3-legged stop-controlled intersections (706 sites).

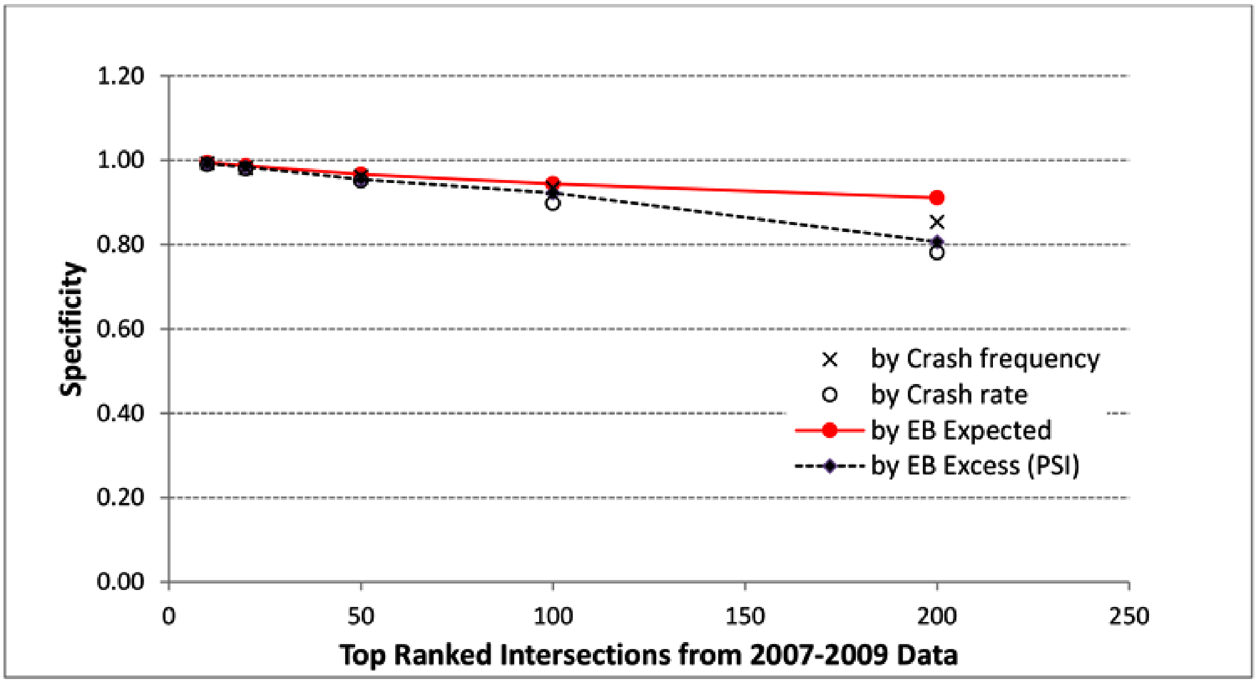


Figure 14. Graph. Specificity (2010-2012) by various performance measures for MN 3-legged stop-controlled intersections (706 sites).

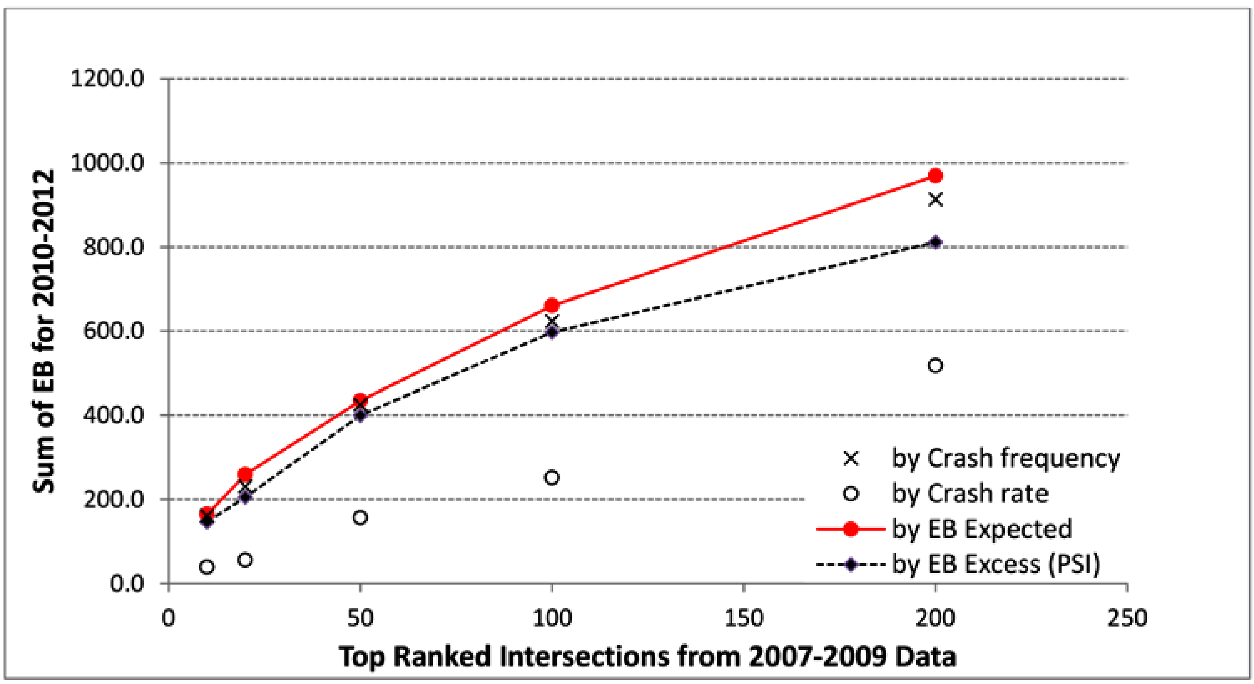


Figure 15. Graph. Sum of EB expected crashes (2010-2012) by various performance measures for MN 4-legged stop-controlled intersections (855 sites).

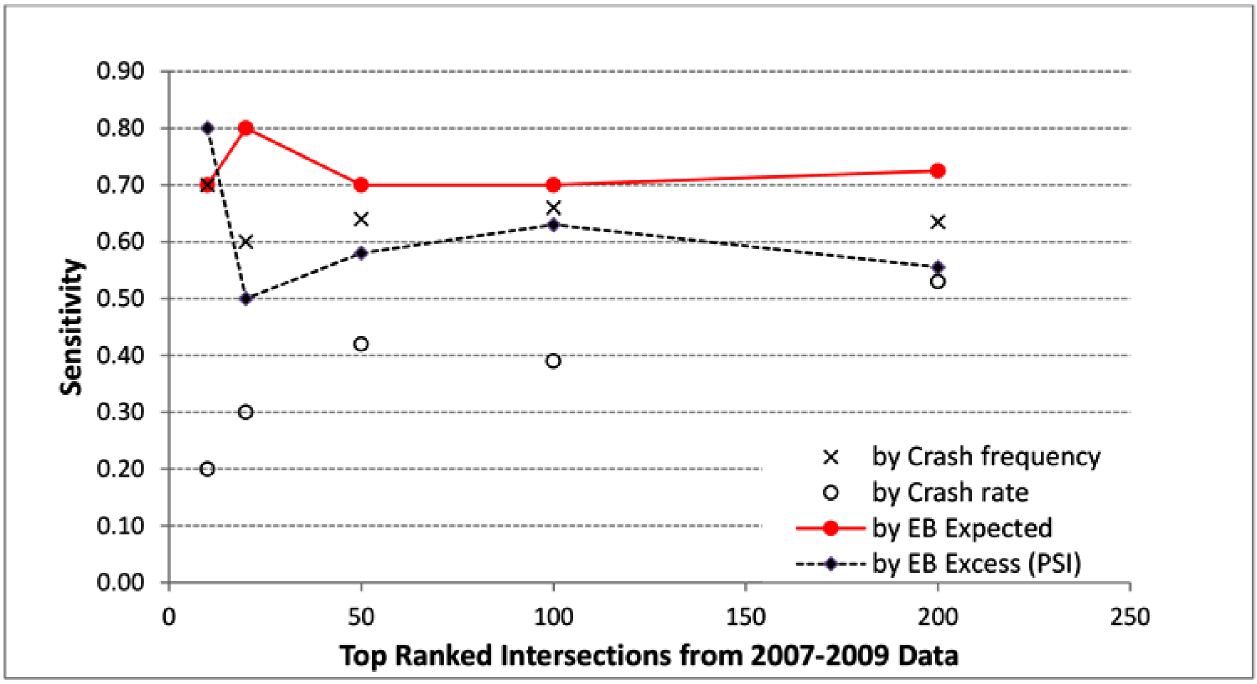


Figure 16. Graph. Sensitivity (2010-2012) by various performance measures for MN 4-legged stop-controlled intersections (855 sites).

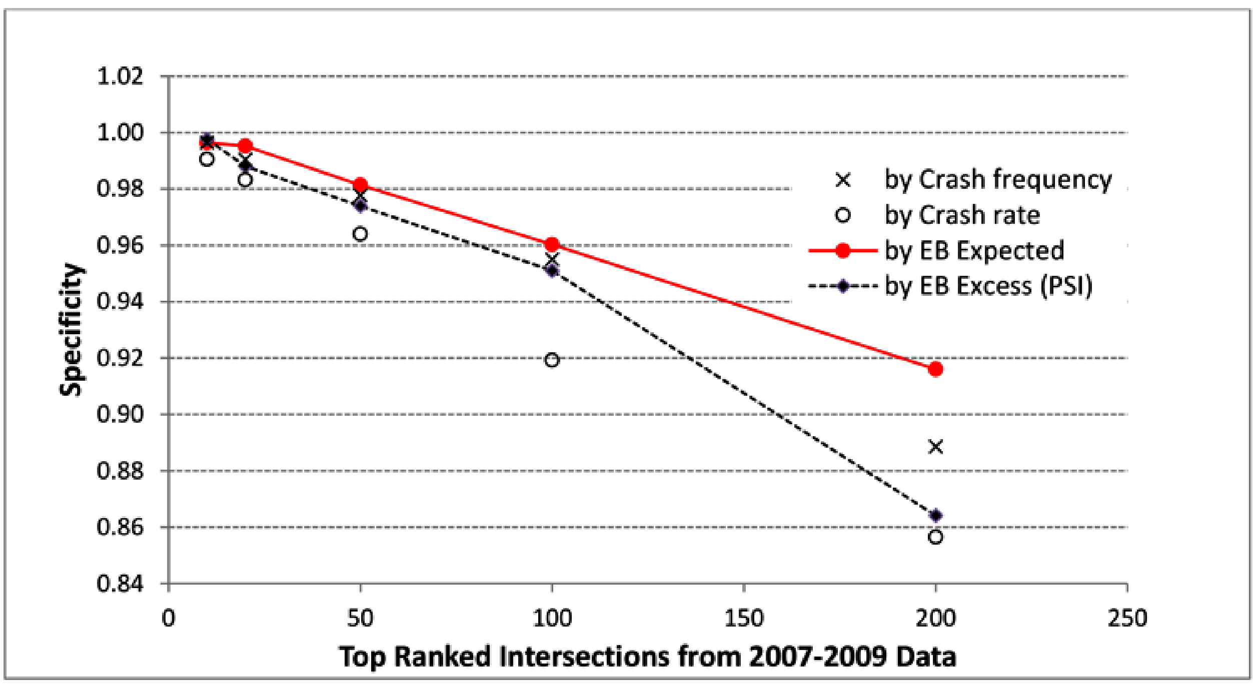


Figure 17. Graph. Specificity (2010-2012) by various performance measures for MN 4-legged stop-controlled intersections (855 sites).

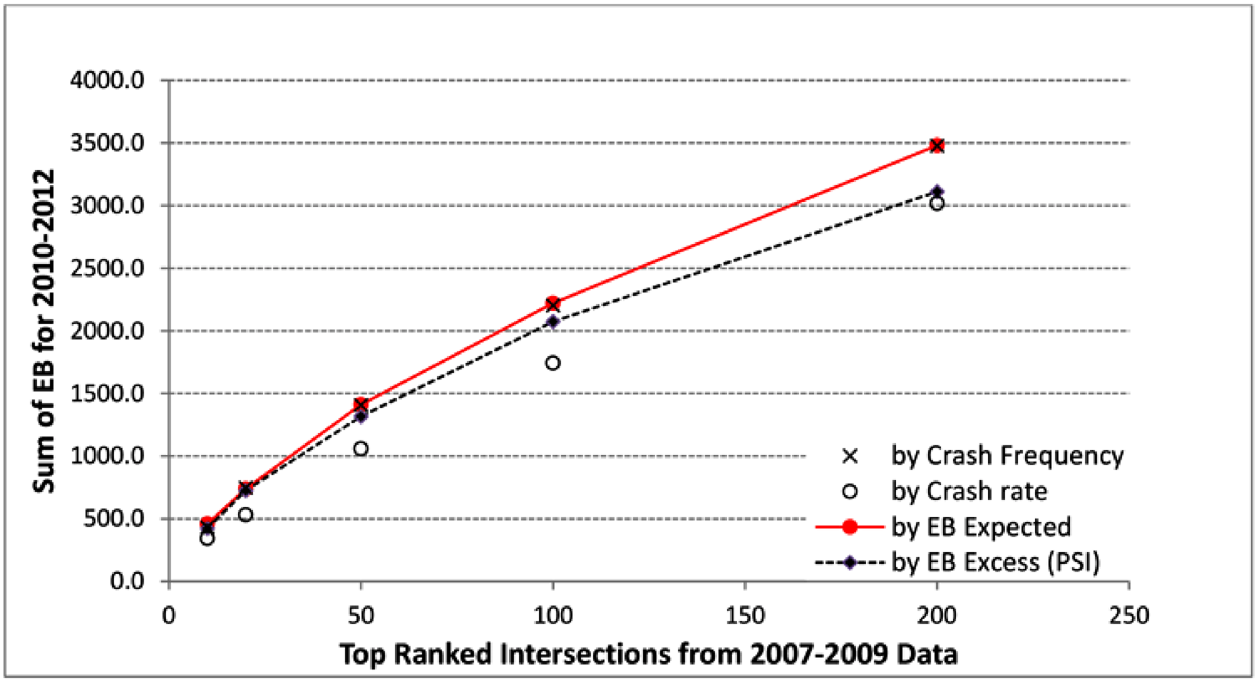


Figure 18. Graph. Sum of EB expected crashes (2010-2012) by various performance measures for MN 4-legged signalized intersections (514 sites).

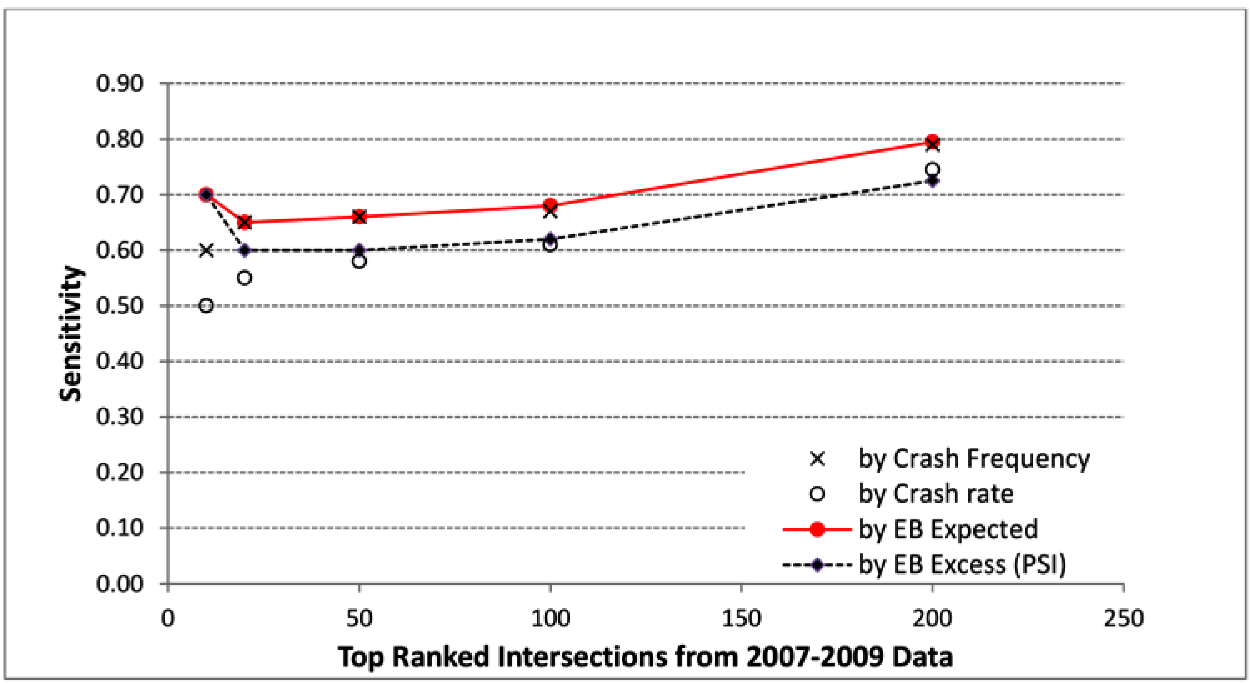


Figure 19. Graph. Sensitivity (2010-2012) by various performance measures for MN 4-legged signalized intersections (514 sites).

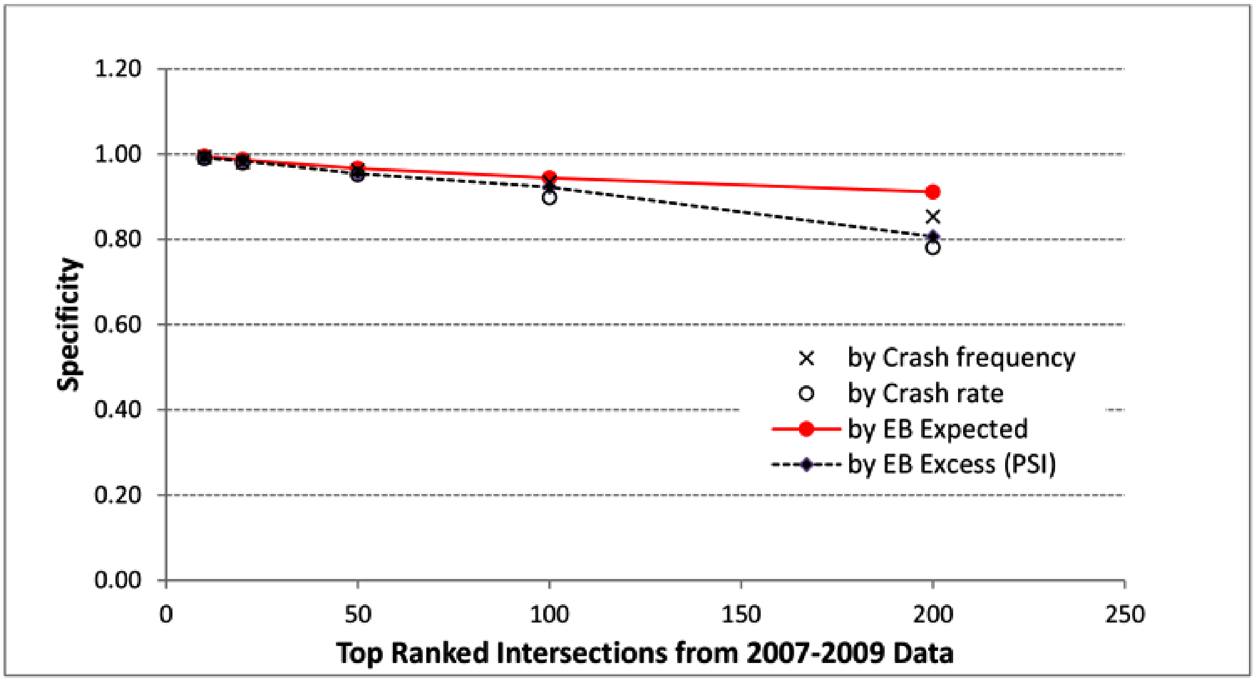


Figure 20. Graph. Specificity (2010-2012) by various performance measures for MN 4-legged signalized intersections (514 sites).

Example 1 References

Srinivasan, R., F. Gross, B. Lan, and G. Bahar. *Reliability of Safety Management Methods: Network Screening*, Report No. FHWA-SA-16-037, Federal Highway Administration, Washington, D.C., October 2016.

The authors developed this example as part of the research to support this guide.

EXAMPLE 2: COMPARISON OF EB EXPECTED, EB EXPECTED EXCESS, LOSS, AND CALTRANS 'TABLE C' MEASURES USING DATA FROM CALIFORNIA

Objective

The objective of this study was to compare EB-based performance measures (EB expected and EB expected excess) with the LOSS and Caltrans 'Table C' method for network screening (Srinivasan et al., 2011). The results will help readers to understand the differences in the measures.

Description of Performance Measures

The earlier section titled, *Overview of Network Screening*, describes the EB expected, EB expected excess, and LOSS performance measures in detail. These measures incorporate information from SPFs. For this study, the researchers developed two SPFs, one based on traffic volume as the only predictor and one based on traffic volume and other site characteristics as predictor variables. Additionally, the researchers applied the related default SPF from AASHTOWare Safety Analyst™, which includes traffic volume as the only predictor variable.

The Table C method identifies, in a given time period, sites that have experienced significantly more crashes per unit of traffic volume than the statewide average. Caltrans screens sites within groups of similar sites based on facility and site type. At the time of this study, there were 30 rate groups for intersections. For intersections, Caltrans screens sites based on all crashes within a predetermined influence area, which is usually 250 ft. The following are the two criteria that must be met for flagging a site for investigation based on the Table C method:

1. The observed crash frequency is greater than the average for the rate group with 99.5 percent confidence.
2. There are four or more crashes in the time period.

Figure 21 defines the minimum number of observed crashes required for significance (N_R):

$$N_R = N_E + 2.576 \times \sqrt{N_E} + 1.329$$

Figure 21. Equation. Minimum number of observed crashes required for significance (N_R).

Figure 22 defines the average number of crashes for the rate group (N_E):

$$N_E = \frac{ADT \times t \times L \times R_E}{10^6}$$

Figure 22. Equation. Average number of crashes for the rate group (N_E).

Where:

N_R = minimum number of observed crashes required for significance.
N_E = average number of crashes for rate group.
ADT = average daily traffic, in vehicles per day.
t = time, in days.
L = length, in miles (= 1 for Ramps and Intersections).
R_E = average crash rate for group, in crashes per million vehicles (intersections) or crashes per million vehicle miles (segments). This is the base rate plus an ADT factor if applicable.

Each rate group has a base rate that is determined by looking at all crashes in a three-year period. Some highway segment rate groups also include an ADT factor, which adjusts the base rate given a site's ADT. For those rate groups that do not include an ADT factor, the assumed relationship between crash frequency and ADT is linear. In addition, the Table C method does not account for possible bias due to RTM.

In summary, the EB-based performance measures account for potential bias due to RTM, changes in traffic volume, and the nonlinear relationship between crash frequency and traffic volume. While the LOSS performance measure accounts for changes in traffic volume and the nonlinear relationship between crash frequency and traffic volume, the application in this study does not account for potential bias due to RTM. The Caltrans Table C method accounts for changes in traffic volume, but does not account for potential bias due to RTM or the nonlinear relationship between crash frequency and traffic volume.

Approach for Comparing Measures

The analysts used the following three approaches to compare the network screening measures:

1. Compare the ability of each measure to rank locations more likely to have high crash frequencies in the future. For a given year, the analyst applies the performance measure of interest to rank the sites. Then, selecting the top 10, 50, 100, and 200 sites, the analyst determines the total crashes in the future years in the study period. The preferred measures are those for which top ranked sites remain as high-crash locations in future years. As an example, consider the application of the LOSS measure to rank sites based on data from 2000 and a study period of 2000 to 2007. First, the analyst ranks all sites using the LOSS measure and data for 2000. From this ranking, the top 10, 50, 100, and 200 sites are selected for further summary. For each group of sites, the analyst determines the total number of crashes in the 'future' based on data from 2001 to 2007. The researchers repeated this process for all four performance measures, for four ranked lists (10, 50, 100, and 200), and for four baseline years (2000, 2001, 2002, and 2003).
2. Compare, retrospectively, the performance of each measure in selecting and ranking correct positives and false positives. Correct positives are those locations identified by the Table C method and subsequently investigated and recommended for improvement. False positives are those investigated and not recommended for improvement. For this approach, the first step was to compile the list of sites identified and investigated by

Caltrans each year. Caltrans used the Table C method to identify sites each year. Then, the analyst used data from the previous year to rank these sites based on the different performance measures. The preferred measures are those for which Caltrans recommended improvement to highly-ranked sites (i.e., Caltrans confirmed a safety concern and identified a targeted mitigation measure). For example, consider the top-ranked sites generated by the Table C method for 2004, and subsequently investigated by Caltrans. The analyst would use data from the previous year (2003) to rank these sites based on the different performance measures. For each measure, the analyst tallies the number of sites for which Caltrans investigated and recommended improvement. The preferred measures should give a high ranking to the investigated sites found to be deserving of treatment, and a low ranking to the investigated sites subsequently found to be undeserving of treatment. It is important to note the starting point in this approach is the sites selected for investigation based on results from the Table C method. Thus, the evaluation favors the Table C method with respect to producing an optimal and ranked list of locations. However, the results show how the other measures performed when ranking the sites recommended for improvement.

3. Compare the characteristics of top ranked locations by each measure. The two main characteristics selected for this comparison were the total intersection AADT and the expected number of crashes. The study includes the expected number of crashes (as opposed to the actual number of crashes) because it corrects for possible bias due to RTM and provides a better estimate of the true long-term crash propensity. To implement this approach, the researchers ranked sites based on each performance measure, and computed the average total intersection AADT and the average expected total crashes for the top ranked sites.

In addition to comparing the measures using the three approaches, the researchers also investigated and demonstrated the significance of the RTM issue by comparing the crash frequency of top ranked sites in 2000 to 2003 with the crash frequency for the same sites in 2004 to 2007.

Data Description

The analysts used the Highway Safety Information System (HSIS) to obtain roadway geometry, traffic volume, and crash data from California for intersections over an eight-year study period. The study period included 2000 through 2007. The intersections included rural, four-legged stop-controlled intersections. The study focused on total crashes within an influence area of 250 feet from the intersection. Caltrans also provided information on investigations triggered by the Table C method and the recommendations for improvement for each site.

Discussion of Results

This section begins with an illustration of the potential for bias related to RTM and then includes a discussion of results from the following three comparisons.

1. Comparison of measures based on future observed crashes.
2. Comparison of measures based on correct positives and false positives.
3. Comparison of measures based on AADT and future expected crashes.

Table 26 shows the potential for bias due to RTM. The first column shows the intersection groups based on the number of crashes per intersection in 2000 to 2003. The second column shows the number of intersections per group. The third and fourth columns show the sum of crashes for each intersection group from 2000 to 2003 and from 2004 to 2007, respectively. The fifth and sixth columns show the average number of crashes per intersection for each group from 2000 to 2003 and from 2004 to 2007, respectively. The final column shows the percentage change in the number of crashes, comparing the average from 2004 to 2007 to the average from 2000 to 2003.

Table 26. Illustration of RTM in rural, four-legged, stop-controlled intersections.

Group (crashes per site 2000 2003)	Number of sites in group	Sum of crashes in group (2000 2003)	Sum of crashes in group (2004 2007)	Average crashes per site (2000 2003)	Average crashes per site (2004 2007)	Percent Change
40+	4	247	195	61.75	48.75	-21.05
30-39	15	494	337	32.93	22.47	-31.78
25-29	9	234	202	26.00	22.44	-13.68
20-24	28	617	545	22.04	19.46	-11.67
15-19	46	781	679	16.98	14.76	-13.06
10-14	112	1298	1213	11.59	10.83	-6.55
9	38	342	300	9.00	7.89	-12.28
8	35	280	310	8.00	8.86	10.71
7	64	448	388	7.00	6.06	-13.39
6	70	420	375	6.00	5.36	-10.71
5	110	550	518	5.00	4.71	-5.82
4	121	484	454	4.00	3.75	-6.20
3	164	492	548	3.00	3.34	11.38
2	242	484	557	2.00	2.30	15.08
1	334	334	513	1.00	1.54	53.59
0	550	0	429	0.00	0.78	Infinite increase

Note: In 2000 to 2003, the mean frequency was 3.86 crashes per site, denoted by the thick line in the table.

The results show that sites with high average crash counts in the first half of the study period (2000 – 2003) tend to have much lower average crash counts in the second half of the study period (2004 – 2007) due to random variation in crashes. For example, intersections in the first group of 40+ crashes experienced an average of 61.75 crashes in 2000 to 2003 and 48.75 in 2004 to 2007. This represents a reduction of 21 percent. The RTM bias is notable even when averaging four years of crash data.

The results also show sites with low average crash counts in the first half of the study period (2000 – 2003) tend to have higher average crash counts in the second half of the study period (2004 – 2007) due to random variation in crashes. The average number of crashes per intersection in 2000 to 2003 was 3.86. It is clear the intersection groups with an average crash frequency less than 3.86 in 2000 to 2003 show a notable increase in crashes in 2004 to 2007.

Given the presence of RTM, there is a need for the network screening performance measure to account for RTM. Recall the EB-based measures are able to account for potential bias due to RTM while the other measures are not.

The basis of the first comparison of performance measures is future observed crashes. Table 27 presents the results for the various performance measures and four different ranked lists: top 10, top 50, top 100, and top 200 sites. For the EB expected, EB expected excess, and the LOSS measures, results are shown for three underlying SPFs: SPFs with AADT only (SPF1), SPFs with AADT and additional variables (SPF2), and default SPFs with AADT only from AASHTOWare Safety Analyst™ (SPF SA). Recall the preferred measure is the one identifying sites with the highest number of future crashes, indicated by bold text in the table. Each analysis includes only one year for this comparison because the Caltrans Table C method includes only one year of data.

Results indicate the EB-based measures (EB expected and EB expected excess) performed better than the Caltrans Table C method, which is based on critical crash rates and crash frequency thresholds. Specifically, the top ranked sites based on the EB-based measures had more crashes in the future compared to the top ranked sites from the Table C method. The EB expected measure performed better than the EB expected excess and LOSS measures for this comparison because it ranks sites based on number of expected crashes as opposed to excess crashes. For the majority of cases, the results based on different SPFs are relatively consistent within a given measure compared to the differences among the measures. This indicates the performance measure is more critical than the type of SPF (AADT only or AADT plus additional variables) used in the process.

Table 27. Future crashes when ranked by performance measures.

Year	Top Ranked Sites	EB Expected			EB Expected Excess			LOSS			Table C Method
		SPF1	SPF2	SPF SA	SPF1	SPF2	SPF SA	SPF1	SPF2	SPF SA	
2000	10	**648**	**648**	642	468	455	483	480	480	480	463
	50	1941	**1990**	1937	1733	1852	1744	1567	1523	1417	1354
	100	**3161**	3110	3147	2886	3052	2812	2383	2527	2279	2131
	200	4989	5000	**5064**	4427	4424	4295	3819	4009	3679	2958
2001	10	**562**	**562**	561	451	451	431	413	398	398	445
	50	1728	**1731**	1728	1570	1648	1547	1419	1478	1402	1271
	100	2807	2771	**2825**	2604	2631	2452	2304	2312	2186	2091
	200	**4418**	4408	4378	3960	3956	3791	3529	3597	3350	2742
2002	10	394	**408**	**408**	374	374	374	361	343	295	368
	50	1355	1363	**1411**	1322	1322	1339	1255	1195	1181	1169
	100	2339	2345	**2403**	2180	2144	2121	1945	1946	1920	1733
	200	3677	3692	**3700**	3286	3265	3250	3081	3057	3027	2478
2003	10	**338**	**338**	**338**	336	336	336	324	281	272	294
	50	1132	**1148**	1141	1133	1116	1120	1054	1057	1026	960
	100	1915	1896	**1922**	1814	1804	1800	1651	1669	1624	1538
	200	**2951**	2921	2944	2695	2711	2664	2477	2504	2442	2050
2004	10	229	**233**	229	209	209	197	201	201	191	217
	50	**806**	**806**	773	759	746	760	692	684	695	651
	100	**1338**	1304	1316	1267	1267	1249	1182	1173	1152	990
	200	2119	2110	**2153**	1952	1944	1907	1760	1771	1725	1396
2005	10	147	147	147	153	154	154	156	156	141	**159**
	50	527	528	**536**	505	508	491	470	462	478	406
	100	854	**862**	854	819	819	797	745	750	734	687
	200	**1400**	1388	1395	1257	1267	1231	1155	1147	1126	895
2006	10	71	71	71	64	**74**	59	57	57	59	54
	50	258	256	**259**	255	258	234	215	208	200	183
	100	**432**	429	**432**	382	387	368	355	354	336	307
	200	**664**	663	660	609	615	579	528	515	497	386

The second comparison of performance measures focused on the identification of correct positives and false positives. Again, correct positives are those locations investigated based on the Table C method and recommended for improvement. False positives are those investigated and not recommended for improvement. The preferred measure is the one producing the most correct positives and fewest false positives.

Table 28 through Table 33 present the results of this comparison of rural, four-legged, stop-controlled intersections for 2003 through 2008, respectively. The tables present results for the various performance measures and four different ranked lists: top 5, top 10, top 20, and top 50 sites. The tables show the number of sites recommended for improvement by Caltrans. For each of the six years of analysis, bold text indicates the highest value corresponding with the preferred measure.

It is important to note Caltrans investigated sites after using the Table C method to screen the network. Hence, the Table C method performs well in this comparison because it was the basis for identifying the initial sites. The results indicate the EB-based and LOSS measures performed equally well as Table C in many cases and, in a few cases, performed better than the Table C method. The tables do not present the number of false positives, but the top ranked sites from the EB expected measure generally had fewer false positives compared to the Table C method and the LOSS and EB expected excess measures, particularly for the top 20 and top 50 sites.

Again, the EB-based and LOSS measures include results for three different SPFs. Overall, the results are similar regardless of the underlying SPF (AADT only versus AADT plus additional variables). Based on further examination of the cumulative residual plots, the jurisdiction-specific SPFs (i.e., those *developed* from California data) performed better than the default AASHTOWare Safety Analyst™ SPFs (i.e., those developed from a multistate dataset and *calibrated* with the same California data).

Table 28. Number of intersections selected as 'improvement recommended' in 2003.

Performance Measure	Top 5	Top 10	Top 20	Top 50
Table C	3	5	**11**	22
EB expected (SPF1)	3	**6**	9	**23**
EB expected (SPF2)	3	**6**	9	**23**
EB expected (SPF SA)	3	**6**	9	**23**
EB expected excess (SPF1)	3	**6**	10	21
EB expected excess (SPF2)	3	**6**	10	22
EB expected excess (SPF SA)	**4**	**6**	10	21
LOSS (SPF1)	**4**	4	10	22
LOSS (SPF2)	3	4	10	22
LOSS (SPF SA)	**4**	**6**	9	21

Note: In 2003, Caltrans investigated 68 stop-controlled intersections and recommended 27 for improvement.

Table 29. Number of intersections selected as 'improvement recommended' in 2004.

Performance Measure	Top 5	Top 10	Top 20	Top 50
Table C	**1**	**2**	**6**	13
EB expected (SPF1)	**1**	**2**	5	10
EB expected (SPF2)	**1**	**2**	5	11
EB expected (SPF SA)	**1**	**2**	5	10
EB expected excess (SPF1)	**1**	**2**	4	10
EB expected excess (SPF2)	**1**	**2**	4	10
EB expected excess (SPF SA)	**1**	**2**	4	11
LOSS (SPF1)	**1**	1	4	**14**
LOSS (SPF2)	**1**	1	4	13
LOSS (SPF SA)	0	0	3	**14**

Note: In 2004, Caltrans investigated 81 stop-controlled intersections and recommended 18 for improvement.

Table 30. Number of intersections selected as 'improvement recommended' in 2005.

Performance Measure	Top 5	Top 10	Top 20	Top 50
Table C	**1**	1	**2**	--
EB expected (SPF1)	**1**	1	**2**	--
EB expected (SPF2)	**1**	**2**	**2**	--
EB expected (SPF SA)	**1**	**2**	**2**	--
EB expected excess (SPF1)	**1**	1	**2**	--
EB expected excess (SPF2)	**1**	1	**2**	--
EB expected excess (SPF SA)	**1**	1	**2**	--
LOSS (SPF1)	**1**	1	1	--
LOSS (SPF2)	**1**	1	1	--
LOSS (SPF SA)	**1**	1	1	--

Note: In 2005, Caltrans investigated 34 stop-controlled intersections and recommended 2 for improvement.

Table 31. Number of intersections selected as 'improvement recommended' in 2006.

Performance Measure	Top 5	Top 10	Top 20	Top 50
Table C	**2**	**2**	**6**	**19**
EB expected (SPF1)	0	0	4	15
EB expected (SPF2)	0	0	4	14
EB expected (SPF SA)	0	0	4	15
EB expected excess (SPF1)	0	1	3	15
EB expected excess (SPF2)	0	1	3	15
EB expected excess (SPF SA)	0	1	4	15
LOSS (SPF1)	1	**2**	5	16
LOSS (SPF2)	1	**2**	5	16
LOSS (SPF SA)	**2**	**2**	5	15

Note: In 2006, Caltrans investigated 76 stop-controlled intersections and recommended 30 for improvement.

Table 32. Number of intersections selected as 'improvement recommended' in 2007.

Performance Measure	Top 5	Top 10	Top 20	Top 50
Table C	0	**3**	**7**	**17**
EB expected (SPF1)	0	2	6	**17**
EB expected (SPF2)	0	2	6	**17**
EB expected (SPF SA)	0	2	6	**17**
EB expected excess (SPF1)	**1**	2	6	**17**
EB expected excess (SPF2)	**1**	2	6	**17**
EB expected excess (SPF SA)	**1**	2	6	**17**
LOSS (SPF1)	0	**3**	6	**17**
LOSS (SPF2)	0	**3**	6	**17**
LOSS (SPF SA)	0	**3**	6	**17**

Note: In 2007, Caltrans investigated 50 stop-controlled intersections and recommended 17 for improvement.

Table 33. Number of intersections selected as 'improvement recommended' in 2008.

Performance Measure	Top 5	Top 10	Top 20	Top 50
Table C	**2**	**5**	**8**	--
EB expected (SPF1)	**2**	3	**8**	--
EB expected (SPF2)	**2**	3	7	--
EB expected (SPF SA)	**2**	3	**8**	--
EB expected excess (SPF1)	**2**	4	7	--
EB expected excess (SPF2)	**2**	4	7	--
EB expected excess (SPF SA)	**2**	4	**8**	--
LOSS (SPF1)	1	3	7	--
LOSS (SPF2)	1	4	**8**	--
LOSS (SPF SA)	1	4	**8**	--

Note: In 2008, Caltrans investigated 39 stop-controlled intersections and recommended 15 for improvement.

The final comparison focused on the characteristics of sites identified by the various measures. Specifically, the characteristics of interest are the average total entering intersection traffic volume and average expected crashes in the last three years of the study period. This comparison includes only the type 1 SPFs (AADT only) for the EB-based and LOSS measures.

Table 34 and Table 35 show the results for average total volume and average expected crashes, respectively. The top ranked sites from the EB expected measure have the highest average traffic volume and the highest number of expected crashes. The top ranked sites from the Table C method have the lowest average traffic volume and the lowest expected number of crashes. It is important to note the Table C method does not account for the nonlinear relationship between crashes frequency and traffic volume, which may be a reason why the top ranked sites in the Table C method have fewer expected crashes and tend to have lower average traffic volumes compared to the top ranked sites in the LOSS measure. On average, the top ranked sites from the EB expected excess measure have more expected crashes compared to the LOSS measure because the EB-based measures explicitly account for potential bias due to RTM.

Table 34. Average traffic volume at top ranked sites.

Year	Top Ranked Sites	EB Expected	EB Expected Excess	LOSS	Table C
2004	10	30,553	23,301	20,474	18,031
	50	23,945	20,301	16,720	11,123
	100	22,109	16,985	15,456	10,308
	200	18,874	15,016	13,199	7,786
2005	10	25,120	24,308	21,535	19,626
	50	23,525	17,711	15,952	12,280
	100	21,356	16,077	13,978	10,532
	200	19,761	14,246	12,468	7,696
2006	10	27,282	25,065	17,223	15,879
	50	22,424	18,347	17,000	11,009
	100	21,538	16,527	14,124	10,422
	200	19,668	14,795	12,602	7,020

Table 35. Average expected crashes at top ranked sites.

Year	Top Ranked Sites	EB Expected	EB Expected Excess	LOSS	Table C
2004	10	11.2	10.7	10.2	9.8
	50	6.6	6.2	5.5	4.9
	100	5.2	4.9	4.5	3.8
	200	3.9	3.6	3.3	2.5
2005	10	10.4	10.1	9.7	9.2
	50	6.3	6.0	5.4	5.0
	100	5.0	4.7	4.3	3.8
	200	3.8	3.4	3.2	2.5
2006	10	10.5	9.3	7.8	8.2
	50	6.2	5.8	5.5	4.5
	100	4.7	4.4	4.0	3.5
	200	3.6	3.3	2.9	2.2

The original study included the same comparisons for other site types, including rural, four-legged, signalized intersections; rural, two-lane, undivided roadway segments; and urban freeway segments. The results for the other site types were consistent with the results for rural, four-legged, stop-controlled intersections. The EB expected measure tends to perform best with respect to the specific comparisons, and in general, the EB-based measures performed better than the LOSS and Table C method. Again, the LOSS method does not account for potential bias due to RTM and the Table C method does not account for potential bias due to RTM and the nonlinear relationship between crash frequency and traffic volume. This may explain, at least in part, why the LOSS and Table C methods are not as effective in identifying sites with a large number of future crashes.

Example 2 References

Srinivasan, R., Lyon, C., Persaud, B., Martell, C., and Baek, J., (2011), Methods for Identifying High Collision Concentration Locations (HCCL) for Potential Safety Improvements – Phase II: Evaluation of Alternative Methods for Identifying HCCL, Prepared for California Department of Transportation, Sacramento, California.

EXAMPLE 3: COMPARING THE PERFORMANCE OF FREQUENCY-BASED AND SEVERITY-WEIGHTED EB MEASURES USING DATA FROM COLORADO

Objective

The objective of this study was to determine the network screening performance measure that is most likely to lead to cost-beneficial projects (Hauer et al., 2004). The scope of this study included rural, two-lane, undivided roads in Colorado. The study included the following five performance measures.

1. EB expected total.
2. EB expected severity-weighted.
3. EB expected excess total.
4. EB expected excess severity-weighted.
5. Combination of EB expected and EB expected excess.

Description of Performance Measures

The earlier section titled, *Overview of Network Screening*, describes the EB expected and EB expected excess performance measures in detail. In this study, the researchers developed two separate measures of the EB expected and EB expected excess crashes as follows.

1. EB expected total: sites where the most crashes are expected.
2. EB expected severity-weighted: sites where the most severity-weighted crashes are expected.
3. EB expected excess total: sites where the most excess crashes are expected.
4. EB expected excess severity-weighted: sites where the most excess severity-weighted crashes are expected.

The severity-weighted measures are similar to the EPDO measure described in the section titled, *Overview of Network Screening*. The severity-weighted measures convert all crashes to a common unit, assigning points to each crash based on the severity level. A PDO crash typically receives one point and the points increase as the severity of the crash increases.

The researchers developed SPFs and incorporated them in the EB-based measures to estimate the expected and expected excess crashes using the peak search method. The peak search method is a screening method applied in step four of the network screening process, and is applicable to roadway segments. The peak search method ranks sites based on the window within the site with the maximum value of the performance measure. Refer to the Highway Safety Manual for further discussion of the peak search method (AASHTO, 2010).

The fifth performance measure is a combination of the EB expected and EB expected excess measures. Specifically, the performance measure is the product of the expected crashes per mile-year (expressed in crashes per mile-year) and the excess crashes per mile-year (expressed in standard deviations). For example, if the EB expected value is 5.0 crashes per mile-year and the EB expected excess value is 1.5 crashes per mile-year with a standard deviation of 0.5, then the EB expected excess is expressed as 3.0 standard deviations above the mean, and the value of the fifth performance measure is 15.0 (5.0 * 3.0). Sites with larger values rank higher than sites with smaller values.

In summary, the EB-based performance measures account for potential bias due to RTM, changes in traffic volume, and the nonlinear relationship between crash frequency and traffic volume. The EB expected measures identify sites with the highest number of expected crashes while the EB expected excess measures employ a threshold to identify sites with relatively high crashes compared to the average expected value. The severity-weighted measures account for the difference in crash severity among sites.

Approach for Comparing Measures

The study employed a pairwise comparison approach, comparing two performance measures at a time to identify sites for further investigation and potential treatment. Starting with all sites for the facility type of interest (i.e., rural, two-lane, undivided roads), the researchers generated two ranked lists based on two performance measures. For the top ranked sites *not common* to both lists, the researchers performed a detailed engineering study to diagnose the contributing factors and identify targeted improvements. Finally, the researchers estimated the costs and benefits for the proposed improvements at each location, and determined the performance measure most likely to lead to more cost-beneficial projects. The researchers repeated this process, retaining the superior performance measure from each comparison for comparison against the remaining performance measures. They did not consider the inferior results further.

The detailed engineering study involved the following steps:

1. Review detailed crash history of the site.
2. Use GIS maps to assess horizontal alignment.
3. Review the video log of the site.

The researchers used this information to identify the underlying contributing factors and determine appropriate safety improvements at each site. They estimated project benefits using crash modification factors (CMFs) applied to raw crash counts (i.e., observed crash history) as well as the EB expected crashes. The study presents results for both options, but explains the value in using the EB expected crashes as the basis for benefit-cost analyses.

Data Description

The scope of this study included rural, two-lane, undivided roads in Colorado. The researchers used the five performance measures to rank sites for further investigation. For those sites not included on multiple lists, the researchers performed detailed engineering studies. This included 22 of the top-ranking sites. At these 22 sites, the researchers identified 61 actions (projects), and subsequently estimated the costs and safety benefits of each action.

Discussion of Results

Comparing benefit-cost ratios based on observed crash history, the EB expected measure resulted in the most cost-effective projects. Comparing benefit-cost ratios based on EB expected crashes, the EB expected severity-weighted measure resulted in the most cost-effective projects. Given the benefit of a proposed improvement is based on the change in expected crashes, and not the change in expected excess crashes, the researchers expected the EB expected and EB expected severity-weighted measures to outperform the other measures based on the number of expected excess crashes. The unexpected result was the superior performance of the EB expected measure over the EB expected severity-weighted measure when observed crash history represents the basis to estimate project benefits. This is further evidence supporting the use of EB expected crashes as the basis for estimating project benefits.

In summary, this study supports the use of the EB expected severity-weighted measure to screen the network. This measure accounts for potential bias due to RTM, changes in traffic volume, and the nonlinear relationship between crash frequency and traffic volume. While it does not establish a threshold to identify sites with high crashes relative to the average expected crashes, it does account for differences in crash severity among sites.

Example 3 References

American Association of State Highway and Transportation Officials (AASHTO), Highway Safety Manual, First Edition, Washington, DC, 2010.

Hauer, E., Allery, B.K., Kononov, J., and Griffith, M.S., (2004), How Best to Rank Sites with Promise, Journal of Transportation Research Board 1897, pp. 48-54.

EXAMPLE 4: AN EVALUATION OF FREQUENCY, RATE, EB EXPECTED, AND EB EXPECTED EXCESS MEASURES USING DATA FROM NEW HAMPSHIRE

Objective

The objective of this study was to determine the network screening performance measure that is most likely to produce a list of sites with the greatest potential for safety improvement and subsequently result in the greatest safety benefit and most cost-effective safety improvements (Gross et al., 2016). The scope of this study included intersections in New Hampshire. The study included the following four performance measures.

1. Fatal and injury crash frequency.
2. Fatal and injury crash rate.
3. EB expected fatal and injury crashes.
4. EB expected excess fatal and injury crashes.

Description of Performance Measures

The earlier section titled, *Overview of Network Screening*, describes the performance measures in detail. The EB-based measures incorporate information from SPFs. The researchers developed SPFs for various facility types to predict fatal and injury crashes for urban and rural, three-legged and four-legged, and stop-controlled and signalized intersections.

In summary, the crash frequency measure does not account for potential bias due to RTM or changes in traffic volume. The crash rate measure accounts for differences in traffic volume among sites, but does not account for possible bias due to RTM or the nonlinear relationship between crash frequency and traffic volume. The EB expected and EB expected excess measures are able to properly account for all of these issues, including potential bias due to RTM, changes in traffic volume, and the nonlinear relationship between crash frequency and traffic volume.

Approach for Comparing Measures

The study simulated the development of projects for a safety program, following the safety management process from network screening through economic analysis. The study included statewide intersection data from New Hampshire. The research team developed ranked lists of sites based on each of the four network screening performance measures. The research team identified the top 20 sites from each screening. Of the 80 potential sites given the top 20 sites from each of the four lists, there were 39 unique sites due to overlap among the lists. Of the 39 unique sites, four sites appeared on all four network screening lists. The research team subsequently removed those sites from the analysis because they would not provide any comparative difference between the screening measures, leaving 35 sites for further consideration.

For the 35 sites selected from network screening, the researchers performed a detailed engineering study to diagnose the contributing factors and identify targeted improvements. The team conducted desktop reviews (i.e., review of all information virtually; no in-field site visits). The detailed engineering studies involved the following steps:

1. Review detailed crash history of the site.
2. Develop and review a collision diagram of the site.
3. Review traffic volumes for the major and minor road.
4. Use aerial images and street view images to virtually review the site.

The researchers used this information to identify the underlying crash contributing factors and determine appropriate safety improvements at each site. Following the engineering studies, the research team provided the intersection summary packages and suggested improvements to the New Hampshire DOT (NHDOT) for review and confirmation. NHDOT provided comments for each site and provided additional insight into the field conditions. In most cases, NHDOT confirmed the appropriateness of suggested strategies, and offered additional feedback on those that might not be appropriate. The research team excluded strategies deemed inappropriate by NHDOT from further analysis.

Finally, the team performed an economic analysis to estimate the benefit, cost, and overall benefit-cost ratio (BCR) for each suggested strategy, package of intersection improvements, and the program of projects generated from each network screening measure. The research team estimated project benefits using CMFs applied to raw crash counts (i.e., observed crashes). They estimated project costs based on various sources, including NHDOT cost estimates, State DOT websites, and research reports. The researchers compared the overall economic benefit and overall benefit-cost ratio for each of the four measures.

Data Description

The scope of this study included intersections in New Hampshire. The dataset included fatal and injury crash, traffic, and roadway data for years 2010 through 2014. The researchers limited the scope of the study to at-grade intersections between two-way major and minor roads with available major and minor road traffic volumes. The team excluded intersections with one or more of the following characteristics from the study:

- Ramp terminals (functional class of major or minor road coded as 'Principal Arterial - Interstate' or 'Principal Arterial - Other Freeway/Expressway').
- Major or minor road AADT was not available.
- Intersection type changed during the study period.
- Split tee intersections (i.e., leg offset was over 20 feet).
- Number of legs was less than three (not an intersection) or greater than four (multi-leg intersection).
- Traffic control was something other than stop control or traffic signal.

The economic analysis required information on the related benefit (i.e., CMF) and cost for each strategy. The research team identified CMFs from the CMF Clearinghouse and other recent research reports. When selecting CMFs, the research team used the most applicable CMF, considering the number of intersection legs, area type, and existing traffic control type. For example, the team identified and applied different CMFs for installing a traffic signal at an urban, four-leg, all-way stop-control intersection and installing a traffic signal at a rural, four-leg, two-way stop-control intersection. The research team estimated project benefits by applying CMFs for total crashes, individual crash severities, and individual crash types as appropriate. The research team used crash costs by crash type and severity level as appropriate based on the CMF. For costs by injury level (K, A, B, C, and O on the KABCO scale), the research team used the NHDOT *2013 HSIP Guidelines* (NHDOT, 2013). For costs by crash type, the research team used the FHWA *Crash Cost Estimates by Maximum Police-Reported Injury Severity within Selected Crash Geometries*. (Council et al., 2005). The team estimated project costs based on various sources, including NHDOT cost estimates, State DOT websites, and research reports.

Discussion of Results

Table 36 presents the results of the economic analysis by performance measure, including the total estimated benefits, total estimated costs, and overall BCR across all related intersections. The following are key observations from the results:

- **Highest overall benefit:** The **EB expected excess** measure produced the list of sites with the highest overall benefit ($22,014,117 in total estimated benefits).
- **Highest return on investment:** The **EB expected** measure produced the list of sites with the highest return on investment (7.08 BCR).
- **Lowest overall benefit:** The **crash rate** measure produced the list of sites with the lowest (by a large margin) overall benefit ($8,106,398 in total estimated benefits).
- **Lowest return on investment:** The **crash rate** measure produced the list of sites with the lowest (by a large margin) return on investment (2.39 BCR).

While the EB expected excess measure produced the list of sites with the greatest overall benefit and the EB expected measure produced the list of sites with the greatest return on investment, all four measures produced a list of sites that could be improved cost-effectively (i.e., BCR greater than 1.0). Further, the EB measures require appropriate SPFs, which may not be available to some agencies. When the EB expected measure is infeasible, it appears the crash frequency measure provides a reasonable alternative. Specifically, the crash frequency measure resulted in the second greatest overall benefit and the second greatest BCR.

Table 36. BCR results by network screening performance measure.

Network Screening Performance Measure	Estimated Benefit	Estimated Cost	BCR
Crash Frequency	$17,942,270	$2,699,700	6.65
Crash Rate	$8,106,398	$3,396,450	2.39
EB Expected	$15,671,311	$2,213,950	**7.08**
EB Expected Excess	**$22,014,117**	$3,891,250	5.66

The economic analysis included 35 intersection improvement packages, and some intersections appeared in multiple performance measure lists. Table 37 provides a summary of the number of sites that overlap among performance measures. For example, there were two sites on the KABC crash frequency list that also appeared on the KABC crash rate list. There was generally limited overlap between the crash rate and other measures, and no overlap in sites between the crash rate measure and EB expected measure. The greatest overlap is between the EB expected and crash frequency measures (14 of 18 sites). There is also substantial overlap among the EB expected excess, EB expected, and crash frequency measures. This is further evidence supporting the use of the crash frequency measure as a reasonable alternative when the EB expected measure is infeasible.

Table 37. Number of sites identified by multiple screening performance measures.

Network Screening Performance Measure	Crash Frequency	Crash Rate	EB Expected	EB Expected Excess
Crash Frequency	--	--	--	--
Crash Rate	2	--	--	--
EB Expected	14	0	--	--
EB Expected Excess	10	5	8	--

In summary, this study supports the use of the EB expected measure to screen the network. This measure accounts for potential bias due to RTM, changes in traffic volume, and the nonlinear relationship between crash frequency and traffic volume. While it does not establish a threshold to identify sites with high crashes relative to the average expected crashes, it can account for differences in crash severity among sites.

Example 4 References

Council, F., Zaloshnja, E., Miller, T., and Persaud, B. Crash Cost Estimates by Maximum Police-Reported Injury Severity within Selected Crash Geometries. Report No. FHWA-HRT-05- 051, Federal Highway Administration, McLean, VA, 2005. Available online at: http://www.fhwa.dot.gov/publications/ research/safety/05051/.

Gross, F., T. Harmon, M. Albee, S. Himes, R. Srinivasan, D. Carter, and M. Dugas. Evaluation of Four Network Screening Performance Measures, Report No. FHWA-SA-16-103, Federal Highway Administration, Washington, D.C., October 2016.

New Hampshire Department of Transportation (NHDOT). Highway Safety Improvement Program Guidelines, 2013.

EXAMPLE 5: COMPARISON OF CRASH FREQUENCY, CRASH RATE, CRITICAL CRASH RATE, EPDO, AND EB-BASED MEASURES

Objective

The objective of this study was to compare four traditional network screening performance measures with the EB-based measure (Lim and Kweon, 2013). The results will help readers understand the differences in the measures, and determine which measure(s) to use if they cannot use EB-based measures due to lack of data or expertise. For comparison, the study assumed the results of the EB-based measure as "ground truth" and compared the following four traditional performance measures:

- Crash frequency.
- Crash rate.
- Critical rate, also known as rate-quality control.
- Equivalent property damage only (EPDO).

Description of Performance Measures

The earlier section titled, *Overview of Network Screening*, describes the four traditional network screening performance measures and the EB-based measure in detail. The following is a brief discussion of how this study employed these performance measures:

- **Crash frequency:** Sites ranked by the number of reported crashes during the study period.
- **Crash rate:** Sites ranked by the crash rate during the study period. The crash rate is the total number of crashes divided by million entering vehicles (MEV) for each intersection.
- **Critical rate:** Sites ranked by the crash rate during the study period. This measure uses the crash rate (crashes per MEV) with the addition of a minimum threshold. The study established a minimum threshold based on a statistical test to determine whether the crash rate is abnormally high at a particular location compared to the average crash rate for locations with similar characteristics.
- **EPDO:** Sites ranked by the EPDO score during the study period. This measure assigns weights to crashes based on the severity level. This study assigned the following weights to the severity categories: fatal crashes assigned a weight of 542, A-injury crashes assigned a weight of 29, B-injury crashes assigned a weight of 11, C-injury crashes assigned a weight of 6, and PDO crashes assigned a weight of 1.
- **EB:** Sites ranked by potential for safety improvement (PSI) during the study period. PSI is also known as excess expected average crash frequency with EB adjustment (i.e., Measure 13 from Table 1). In this study, the researchers estimated separate SPFs for stop-controlled and signalized intersections for rural and urban areas using major and minor road traffic volumes from all intersections in the sample. In order to compare the EPDO measure with the EB-based measure, the researchers used the severity proportions in the sample data to partition the SPF and then applied severity weights to produce an EPDO-equivalent EB estimate.

Approach for Comparing Measures

The researchers compared the four traditional network screening performance measures with the results from the EB-based measure. They applied the following four approaches to compare the various measures.

- **Pearson's correlation coefficient:** This is the correlation coefficient between the rankings from the EB-based measure and the rankings from the traditional measure of interest based on all sites. A higher value of correlation is preferred.
- **Correct identification percentage:** This measure indicates how many sites with a PSI *greater than zero* (i.e., locations identified for further investigation by the EB-based measure) coincided with those identified by the traditional measure of interest. The researchers computed this for the top 1, 5, and 10 percent of locations. A higher percentage indicates a better measure.
- **False identification percentage:** This measure indicates how many sites with a PSI *less than zero* (i.e., locations not identified for further investigation by the EB-based measure) coincided with sites identified for further investigation by the traditional measure of interest. The researchers computed this for the top 1, 5, and 10 percent of locations. A lower percentage indicates a better measure.
- **Rank-based mean absolute error (MAE):** This measure is similar to the correct identification percentage, but considers the ranks. The researchers computed this for the top 1, 5, and 10 percent of locations using the equation in Figure 23. Lower values of MAE indicate better measures.

$$MAE(rank) = \frac{1}{n}\sum_{i=1}^{n}\left|rank(EB - SPF)_i - rank(T)_i\right|$$

Figure 23. Equation. Rank-based MAE.

Where:

MAE(rank) = mean absolute error in ranks.

$rank(EB - SPF)_i$ = rank of location *i* based on the PSI from the EB-based measure.

$rank(T)_i$ = rank of location *i* based on performance measure from traditional measure *T*.

n = number of locations, varying with the specified top percentage (1, 5, or 10 percent).

Data Description

The scope of the study included rural and urban, four-legged, signalized and two-way, stop-controlled intersections in Virginia. The dataset included data for a total of 1670 intersections from 2004 to 2008.

Discussion of Results

Table 38 shows the Pearson correlation coefficient values for the four traditional measures. Bold numbers indicated the preferred measure. Considering all intersections, the critical rate measure performed best with a correlation coefficient of 0.576, and the crash frequency measure was second with a correlation coefficient of 0.494. For intersections identified by the EB measure for further investigation (i.e., PSI > 0), the crash frequency measure performed best with a correlation coefficient of 0.868, and the critical rate measure was second with a correlation coefficient of 0.713. The crash rate and EPDO measures performed poorly in both scenarios.

Table 38. Correlation coefficient values for the traditional measures.

Category	Crash Frequency	Crash Rate	Critical Rate	EPDO
All sites	0.494	0.376	**0.576**	0.252
Sites identified for further investigation (PSI > 0)	**0.868**	0.153	0.713	0.386

Note: This information is the same as Table 3 from Lim and Kweon. [8]

Table 39 through Table 41 show the results from the other three comparisons for the top 1, 5, and 10 percent of locations, respectively. Bold numbers indicated the preferred measure. For the top 1 percent, the crash frequency measure performs best with the highest correct identification percentage, lowest false identification percentage, and lowest MAE rate. For the top 5 and 10 percent, the critical rate measure performs best with respect to correct identification percentage and false identification percentage, whereas the crash frequency measure performs best with respect to MAE rate. While the crash rate measure performs well with respect to false identification percentage, it performs the worst with respect to correct identification percentage and MAE rate in all three tables. The overall conclusion is the crash frequency and critical rate measures perform well with respect to the EB-based measure, while the crash rate measure does not perform well in two out of the three comparisons.

Table 39. Comparison of traditional measures based on top 1 percent of sites.

Comparison	Crash Frequency	Crash Rate	Critical Rate	EPDO
Correct identification % (count)	**76.5 (13)**	6.9 (1)	52.9 (9)	0.0 (0)
False identification % (count)	**0 (0)**	**0 (0)**	**0 (0)**	29.4 (5)
MAE rate	**3.3**	129	9.4	22.9

Note: The EB-based measure identified a total of 17 locations for further investigation. This information is from Table 4 from Lim and Kweon. [8]

Table 40. Comparison of traditional measures based on top 5 percent of sites.

Comparison	Crash Frequency	Crash Rate	Critical Rate	EPDO
Correct identification % (count)	67.9 (57)	20.2 (17)	**92.6 (78)**	67.9 (57)
False identification % (count)	8.3 (7)	**0 (0)**	**0 (0)**	9.5 (8)
MAE rate	**26**	208	33.4	35.9

Note: The EB-based measure identified a total of 84 locations for further investigation. This information is from Table 4 from Lim and Kweon. [8]

Table 41. Comparison of traditional measures based on top 10 percent of sites.

Comparison	Crash Frequency	Crash Rate	Critical Rate	EPDO
Correct identification % (count)	65.9 (110)	23.6 (40)	**75.4 (126)**	71.3 (119)
False identification % (count)	15.6 (26)	**0 (0)**	**0 (0)**	13.8 (23)
MAE rate	**61.1**	230.4	66.4	62.5

Note: The EB-based measure identified a total of 167 locations for further investigation. This information is from Table 4 from Lim and Kweon. [8]

Example 5 References

Lim, I.K. and Kweon, Y.J., (2013) Identifying High-Crash-Risk Intersections: Comparison of Traditional Methods with the Empirical Bayes-Safety Performance Function Method, Journal of Transportation Research Board 2364, pp. 44-50.

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-16-037