VALIDATION OF NETWORK SCREENING METHODS ON RURAL HIGHWAYS

by

Bishal Dhakal

A thesis submitted in partial fulfillment
of the requirements for the degree

of

Master of Science

in

Civil Engineering

MONTANA STATE UNIVERSITY
Bozeman, Montana

May 2024

DEDICATION

To my family, whose unwavering support has been a constant source of strength throughout this academic journey.

ACKNOWLEDGEMENTS

I would like to thank Dr. Ahmed Al-Kaisy, P.E., for his guidance and encouragement in the completion of this thesis. I would also like to thank my committee members: Mr. Michael Sanderson, P.E., and Dr. Kelvin Wang, for their insightful feedback and support.

I would also like to thank Dr. Craig Woolard, the head of the department of civil engineering.

TABLE OF CONTENTS

TABLE OF CONTENTS CONTINUED

TABLE OF CONTENTS CONTINUED

LIST OF TABLES

Table Page

LIST OF TABLES CONTINUED

Table Page

LIST OF FIGURES

Figure Page

LIST OF FIGURES CONTINUED

Figure Page

ABSTRACT

Road safety in rural areas poses challenges due to lower traffic volume, random crash occurrences, and limited access to data and technical expertise for roads managed by local agencies. Consequently, current network screening methodologies and related safety analysis tools become less effective in addressing these features. While numerous network screening techniques have been developed and proposed in literature, each with its own set of strengths and limitations, there has been relatively less research dedicated to evaluating their performance. This thesis evaluates the efficacy of different network screening methods for identifying safety improvement sites on rural highways. It comprehensively assesses three methods: the Global Risk Scoring Method (GRS), the Empirical Bayes (EB) Prediction Method, and the Crash Risk Index (CRI) method. These methods are specifically designed for rural highways, considering the challenges faced by local agencies. Their performance is compared against well-established methods using rural highways in Oregon as a case study. The research employs a robust methodological framework, including spearman's rank correlation coefficient, true positive identification, and root mean square error, among other metrics, to assess the effectiveness of each method in network screening. The validation analysis was conducted separately for all methods, revealing that although the GRS method is less predictive, it outperforms the well-established empirical Bayes (EB) method in network screening. This makes it a viable option for local agencies with limited data resources. Similarly, the CRI method either outperformed or showed comparable results to the EB method in network screening across various evaluation metrics considered. The EB prediction method demonstrates comparable performance to the EB method assessed using crash frequency. The method consistency test within the network screening methods for two periods suggests that the EB prediction method excels in identifying the same hotspots across different time frames.

CHAPTER ONE

INTRODUCTION

Context

Every year, road traffic crashes result in a significant number of injuries and fatalities worldwide. According to the latest global road safety report by the World Health Organization (WHO) ("World Health Organization," 2023), traffic crashes account for 1.19 million deaths globally per year and traffic injuries representing the leading cause of death for people aged 5-29 years. Although several efforts have been made to improve road safety and are having an impact, more actions are needed to achieve the goal of reducing traffic deaths and injuries by at least 50% by 2030 as released by WHO and United Nations regional commissions ("World Health Organization," 2021). Moreover, the road crash costs range from 0.4% to 4% of the national Gross Domestic Product (GDP) across different countries (Wijnen et al., 2019). These statistics suggest that road unsafety still represents a huge health, social, and economic burden globally. Thus, the problem is clear: traffic crashes must be reduced in terms of both severity and frequency. Understanding the factors related to road crashes is crucial to provide road safety experts and practitioners with tools, resources and strategies that should be employed more effectively to tackle such a silent and long-lasting pandemic.

Rural Highway Safety

Highway safety research has made significant advancement over the past two decades. Particularly, development of methodologies in identifying safety improvement locations that have experienced exceptionally high crash frequencies, also known as hotspot or blackspots,

have seen notable progress. However, safety in rural areas presents challenges compared to urban areas due to the lower traffic volume and random crash occurrence typically observed on rural highways. The current methodologies to identify safety improvement locations and related safety analysis tools become less effective when coping with random and widely distributed crash events.

Rural roads constitute an integral part of the roadway network providing vital access to rural towns and communities (including farms and ranches). According to the fatality facts 2021, around 40% of traffic fatalities took place in rural areas, although only 20% of people in the U.S. live there, and 32% of all vehicle miles traveled (VMT) occurred in rural areas ("Fatality Facts," 2021). These statistics demonstrate the need for improving rural road safety and the requirement to effectively include it in the ongoing highway safety improvement programs (HSIPs). A major issue with safety improvement for rural highways is that the crashes on rural and local roads are typically spread over hundreds of miles and are not as densely clustered as crashes in urban areas. The current hot spot identification methodologies and related safety analysis tools become less effective when coping with random and widely distributed crash events. Thus, the need to identify and develop new approaches to address these issues is imperative and apparent.

HSIPs and Network Screening

Highway safety improvement programs (HSIPs) are the primary means for managing safety on the highway system, thus promoting sustainable mobility on the urban and rural highway networks. The HSIP is a core Federal-aid program for states to achieve a significant reduction in fatalities and serious injuries on all public roads. The HSIP requires a data-driven, strategic approach to improving highway safety on all roads, including non-state-owned roads

and roads on tribal lands with a focus on performance (Herbel, 2010). The HSIP is a six-step process including network screening, crash analysis and diagnosis, countermeasure selection, economic appraisal, project prioritization, and safety effectiveness evaluation as shown in figure 1. HSIP history dates back to the Highway Safety Act 1966, enacted by Congress on September 9, 1966, which was the first major effort at the Federal level to reduce the number and severity of highway-related crashes ("Highway Safety Improvement Program (HSIP) - Safety | Federal Highway Administration,"). The Highway Safety Manual (HSM), published by the American Association of State Highway Transportation Officials (AASHTO) is the recognized source of information and methods for quantitatively evaluating traffic safety performance on existing or proposed roadways (AASHTO, 2010).

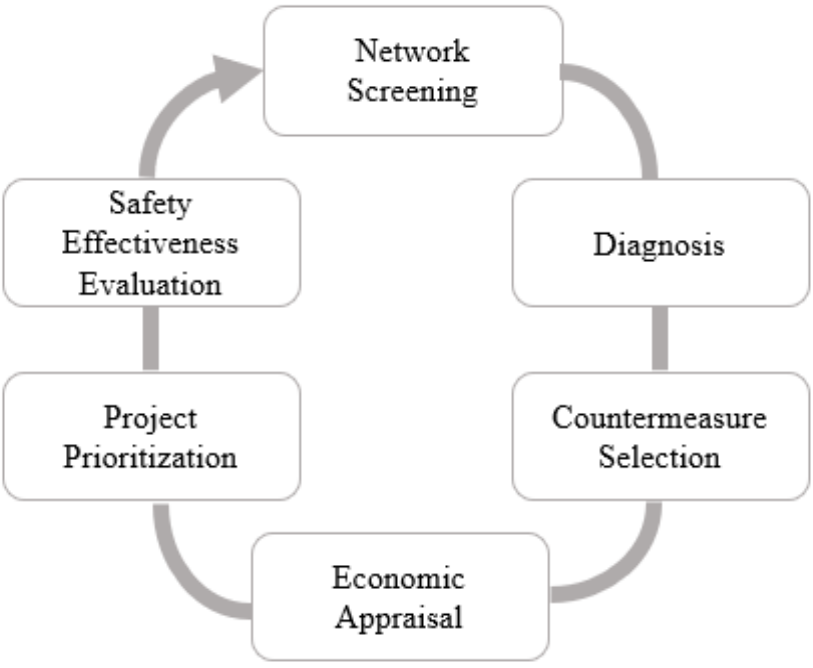


Figure 1. HSM six-step highway safety improvement programs (AASHTO, 2010)

Transportation agencies ideally would like to improve all the locations that have safety problems. However, this is not feasible due to budget constraints. Transportation engineers are, therefore, tasked with selecting the potential locations at the network level that are likely to provide maximum benefits for safety improvements. This is commonly termed network screening. Network screening is the first step of highway safety improvement programs (HSIPs) which seek to identify sites at the network level that are good candidates for further consideration and investigation for safety improvement.

In network screening, the resulting list of identified sites are given priority for detailed engineering studies to identify crash patterns, contributing factors, and potential countermeasures (Hauer et al., 2002; Schmelkin, 1991). Errors in ranking the sites may produce a large number of false negatives (i.e., sites needing safety improvement mistakenly considered as safe) and a large number of false positives (i.e., truly safe sites identified as needing safety improvement). These errors result in an inefficient use of the resources dedicated to safety improvements and eventually reduce the overall effectiveness of the safety management programs. Therefore, the correct identification of safety improvement sites is essential for the successful implementation of any highway safety program.

Most traditional network screening methods rely on the crash frequency, crash rate, crash severity, or any combination of these measures (Anastasopoulos and Mannering, 2011; Park and Sahaji, 2013). Most of the existing ranking techniques for identifying safety improvement sites are reactive in nature (Agerholm and Lahrmann, 2012), as they rely on crash data only (Huang et al., 2009). Thus, these techniques require crashes to occur before identifying the potential sites for safety improvement. On the other hand, several proactive approaches have been proposed

and used that require detailed data on roadway geometry and roadside characteristics, besides crash and traffic data (Lee et al., 2006). A major challenge for network screening on rural roads is the lack of access to the required crash, roadway, and traffic data at the network level especially for roadways that are owned and operated by local agencies.

Study Motivation

The objective of this study is to evaluate the performance of new proposed network screening methods that were tailored for use on rural roadways including those owned and operated by local agencies. Rural highways present unique challenges regarding their physical characteristics, travel patterns, and safety risks, often exacerbated by limited access to data and technical expertise for roads managed by local agencies. Specifically, local agencies often lack access to detailed databases and do not have safety engineers among their staff. Simple methods for network screening were developed and proposed for use to address these challenges. However, no evaluations were performed to have a better understanding of the strengths and limitations of the proposed methods. Such evaluation is important for making informed decisions about the implementation of the proposed network screening methods. Therefore, the major motivation behind this research was to examine the performance of the proposed network screening methods that may prove valuable for network screening applications on rural road networks. Effective network screening methods should contribute to successful safety management programs and help promote sustainable mobility on the highway system by reducing the number of crashes, fatalities, and injuries.

The objective of this study is to compare various network screening methods to validate their effectiveness in identifying safety improvement sites on rural highways.

Thesis Organization

This thesis is comprised of nine chapters, as follows:

Chapter 1: An introduction with a brief overview of the research background, rural highway safety, Highway Safety Improvement Programs (HSIPs) and network screening, study motivation, and thesis organization.

Chapter 2: The literature is reviewed in this chapter. The review is provided mainly for existing network screening methodologies (traditional and non-traditional methods), empirical bayes (EB) method for network screening, and evaluation of different network screening methods for rural and urban road networks. It also provides information to the readers about regression-to-mean and safety performance functions.

Chapter 3: This chapter discusses the study data. Specifically, the study area from where the sample for this validation analysis is extracted is discussed in detail. It also discusses how the data is collected for use in the analyses including data cleaning and roadway segmentation process.

Chapter 4: This chapter details the methodology for this study. It explains the study design, and different evaluation analyses and performance metrics like precision, accuracy, correlation.

Chapter 5: Validation analysis for method I: Global Risk Scoring Scheme is presented in this chapter. The overview of the proposed methodology and study results are discussed in brief in this section. Also, the comparison of the performance of this method with well-established methods are shown in this part of the thesis.

Chapter 6: This chapter presents the validation of method II: EB Prediction Model for Network Screening. The chapter discusses the development of the regression model using the proposed approach and the validation analysis of this method for network screening applications.

Chapter 7: Validation analysis of method III: Crash Risk Index is evaluated and checked against existing widely used network screening methods like EB method in this chapter.

Chapter 8: The criteria for comparative analysis of different network screening methods on rural highways is discussed in this chapter. Specifically, three new proposed methods are compared with each other for identifying safety improvement sites.

Chapter 9: The summary of major findings and future recommendations are presented in this chapter.

CHAPTER TWO

LITERATURE REVIEW

The process of examining the road network (roadway segments or intersections) to locate the sites that demonstrate a need for safety improvement is commonly known as network screening. The result of network screening is a rank list of sites, with the most unsafe sites at the top. Some or all of these sites are considered as candidates for more detailed investigation, sometimes called detailed engineering study (DES). The aim of this study is to suggest appropriate crash-reduction countermeasures for those candidate sites. Over time, numerous network screening techniques have been used in practice or proposed in the literature with each having its strengths and limitations. Some of these techniques are well established in practice while others have seen limited or no practical implementations to learn about their effectiveness. This chapter provides the reader a background into traditional crash observational analysis and modern predictive network screening methods that are proposed and developed. Also, in this section, the major studies that evaluated the use of the Highway Safety Manual (HSM) safety performance function (SPF) predictive models, traditional network screening methods and Empirical Bayes (EB) method for network screening are summarized and presented.

Existing Network Screening Methods

Traditional Descriptive Analysis Method

The traditional descriptive analysis method uses historical crash data alone. These methods tend to have very simple data requirements and, more importantly, do not require any assumptions. The observed number can reflect the safety performance of a facility if the

observation is not fluctuating considerably over a long period of observation. The methods focus on network screening using crash frequency, crash density, crash severities, crash types, and more advanced crash rate (uses traffic exposure). Some variations or combinations are also available. Various methods using these parameters are discussed in this section.

Spot Map Method: The spot map method includes generating a map where clusters of symbols represent crash locations along road segments ("SEMCOG," 1997). By examining the map for geographical concentrations of crashes, areas with the highest total crash counts (or specific types of crashes) are flagged as high crash locations. While the spot map method is straightforward and easy to use, it offers only a basic approximation of high crash areas and doesn't furnish a definitive list of such locations. This method is best suited for smaller areas with fewer crashes, as it becomes less effective in larger areas or with higher crash volumes. In such scenarios, alternative methods for identifying high-crash areas would be more advisable.

Crash Frequency Method: In the HSM, crash frequency is defined as the number of crashes occurring at a particular site, facility, or network in a one-year period. It can be used as a preliminary screening tool, to identify locations in need of more detailed analysis. This method involves analyzing the number of crashes occurring at specific sites or segments within a transportation network, such as intersections, bridges, or roadway segments. These sites are ranked based on the frequency of crashes, which can be calculated for overall crashes or specific crash types (e.g., run-of-the-road crashes, pedestrian crashes). Sites with crash frequencies exceeding a predetermined critical value are identified as high crash locations. Critical frequency values are determined either through statistical analysis of similar sites or by selecting a

significantly high threshold for a specific type of location (Fitzpatrick et al., 2000; Pawlovich, 2002; “SEMCOG,” 1997; Zegeer, 1986).

Crash Density Method: When comparing locations with different lengths, crash density is often used. The crash density method involves calculating the number of crashes per mile for specific roadway segments, with a segment defined as the minimum length of roadway exhibiting similar characteristics. These segments are then ranked in descending order, and those with a crash density exceeding a predetermined critical threshold are identified as high crash locations. The critical density is calculated in a similar way as described in the frequency method (Pawlovich, 2002; “SEMCOG,” 1997; Zegeer, 1986).

The strengths of the crash frequency and crash density methods lie in their simplicity and effectiveness in identifying locations with a high number of crashes. However, these methods overlook the factor of exposure, such as traffic volumes, in their network screening process. This can lead to potentially misleading results, particularly in road systems where traffic volumes vary significantly. The crash frequency and crash density methods tend to classify high-volume locations as high-crash areas, even if the actual number of crashes relative to their traffic volume is relatively low. Likewise, one drawback of crash density method arises from extremely short segments, where the crash density may appear disproportionately large due to the division of crash counts by the short length of the roadway segment.

Crash Severity Method: Various methods within the network screening process incorporate severity, whether it pertains to the crashes themselves or the injuries resulting from them. A method commonly employed to assess crash injury severity involves calculating the ratio of fatal crashes to the total number of crashes, which aids in identifying sites warranting

further examination. Additionally, fatal crash rates, fatal plus injury crash rates, and total crash rates are alternative metrics that can be utilized for this purpose. These approaches employ diverse techniques to integrate severity measures, including assessing the frequency/density of more severe crashes, the rate of such crashes, and the ratio of more severe incidents. Essentially, crashes or injuries deemed more severe are assigned greater relative importance compared to those considered less severe. Often, the results for each site are then benchmarked against systemwide averages for similar roadways.

A standard definition of severity levels has been defined by the National Safety Council (NSC) and is an American National Standards Institute (ANSI) standard (Ogden, 1996) to distinguish severity of crashes and injuries:

Fatal: one or more deaths (commonly signified by K)

A-level injury: incapacitating injury preventing victim from functioning normally (e.g., paralysis, broken/distorted limbs, etc.)

B-level injury: non-incapacitating but visible injury (e.g., abrasions, bruising, swelling, limping, etc.)

C-level injury: probable but not visible injury (e.g., sore/stiff neck)

PDO: property-damage only (commonly signified by O)

This standard is known as the KABCO injury scale and is used commonly in police reporting of crashes. Many of the crash severity methods utilize this scale.

The two types of crash severity methods used for network screening methods are Equivalent Property-Damage-Only (EPDO) method, and Relative Severity Index (RSI) method.

In the equivalent property damage only (EPDO) method, fatal and injury crashes are assigned weights relative to property damage only crashes, which serve as a baseline (Fitzpatrick et al., 2000; “SEMCOG,” 1997; Zegeer, 1986). Each injury level (KABC) is assigned a specific weight compared to property damage only crashes, which are given a weight of 1. These weight coefficients are determined based on the average crash costs by severity. Notably, K-type and A-type crashes often share the same weight. These weights are integrated into the SICL process either by computing an EPDO index or an EPDO rate.

The relative severity index (RSI) method incorporates the weighted average cost of crashes at sites (Fitzpatrick et al., 2000; Zegeer, 1986). This method is best-suited for the further evaluation of sites already identified by other methods as high-crash sites. In the RSI method, crash frequency at each severity level is multiplied by the average "comprehensive cost" for crashes at that severity level. The subtotals for each of these severity-specific costs are summed and the sum is divided by the total crash frequency.

Crash Rate Method: The crash rate method involves normalizing the crash data by the volume of traffic on a roadway. The crash rate is typically expressed as crashes per million entering vehicles for intersections or crashes per million vehicle miles traveled for segments. A critical crash rate is determined as that in crash frequency method and locations with higher than predetermined rate is flagged as high-crash locations and is considered for further investigations. The crash rate method often uses total crashes in calculating rates, however, rates for specific crash type (e.g., single-vehicle crashes) and severity levels (e.g., fatal crashes) are also used.

The primary rationale behind employing the crash rate method is its incorporation of exposure through traffic volume. It acknowledges that a road location or section might

experience a high number of crashes solely due to its extensive use, rather than being hazardous. By utilizing crash rates, this issue is addressed and mitigated. Generally, the crash rate method yields superior outcomes compared to the crash frequency or crash density methods. However, it is more complex than either of those methods, especially as it adds the further complication of requiring non-crash data.

The crash rate method has the advantage of considering the traffic exposure, a major contributing factor to crashes. The data needed for application of crash rate is easily available. While the AADT is needed, most roadways studied will have an AADT measured, or an estimate readily available. Similar to crash density, crash rate measurement is undermined by extremely short roadway segments or segments carrying very low traffic volume.

Frequency Rate Method: The frequency-rate method combines elements from both the crash frequency/crash density methods and the crash rate method (Fitzpatrick et al., 2000; Pawlovich, 2002; Zegeer, 1986). It identifies locations as high-crash areas if they exceed the prescribed minimum crash frequency or crash density and also surpass the minimum crash rate threshold. While the individual methods—crash frequency/crash density and crash rate—each have their limitations, the frequency-rate method aims to mitigate or eliminate these deficiencies. By integrating aspects of both approaches, it offers the potential to enhance effectiveness in identifying high-crash locations.

Though all of the above methods generate usable lists for hazardous site ranking, none of them include any measure of statistical significance or any statistical control. However, a couple currently utilized methods exist that incorporate some simple statistics, one based on crash frequency/density, and one based on crash rate.

Quality Control Methods: Similar to the frequency-rate method, quality control methods also account for different highway categories. These methods ensure the quality of analysis by employing statistical tests to identify any unusual crash frequencies/densities or rates (Fitzpatrick et al., 2000; Pawlovich, 2002; "SEMCOG," 1997). These statistical tests are grounded on the widely accepted notion that crashes adhere to the Poisson distribution (Abdulhafedh, 2016; MacLean and Teale, 1982; Nicholson and Wong, 1993).

The number quality control method applies a statistical test to determine the significance of a site's crash frequency/density when compared to the mean crash frequency/density for similar sites (Fitzpatrick et al., 2000). The statistical test applied is based on the Poisson distribution, the commonly accepted distribution for crashes. Use of the number quality control method effectively addresses sites with high crash frequencies/densities but low exposures. Inputs for the number quality control method, for identification of hazardous sites, include average crash frequency/density for site category, crash frequency/density at the site, and level of statistical significance.

The rate quality control method applies a statistical test to determine the significance of a site's crash rate when compared to the mean crash rate for similar sites (Fitzpatrick et al., 2000). The statistical test applied is based on the Poisson distribution, the commonly accepted distribution for crashes. Use of the rate quality control method effectively eliminates sites with high crash rates but low exposures. Inputs for the rate quality control method, for identification of hazardous sites, include average crash rate (per 100 million vehicle miles) for site category, crash rate at the site, and level of statistical significance.

Predictive Crash Analysis Methods

With the advancement in road safety research, the network screening method was also transformed from traditional analysis to quantitative predictive analysis. In contrast to traditional observational crash analysis, predictive methods predict the expected number of crashes for a site of interest. This method uses the mathematical regression model to find the expected number of crashes based on which network screening is performed. The models are developed based on the relationship between crash numbers and risk elements (roadway, traffic and geometric factors).

The fundamental components of the predictive models outlined in Part C of the HSM may vary depending on the facility and site type, but they share common elements.

Safety Performance Functions (SPFs), which are statistical models designed to estimate the average crash frequency for a specific element under defined conditions.

Crash Modification Factors (CMFs) represent the ratio of effectiveness between one specific condition and another.

And, Calibration Factor (C), which involves multiplying the predicted crash frequency by the SPF. This factor aims to accommodate variations across jurisdictions and time periods for which the predictive models have been developed (AASHTO, 2010).

Safety Performance Functions (SPFs) serve as a predictive tool for estimating the number of crashes on roadways. They are statistical functions designed to assess and predict the average crash frequency for specific site types under defined base conditions, taking into account both traffic volume and roadway segment length. Part B of the American Association of State Highway and Transportation Officials (AASHTO) Highway Safety Manual (HSM) outlines the utilization of SPFs for Network Screening, while Part C illustrates their application in predicting crash numbers for particular sites, incorporating various SPFs into the six-step roadway safety

management process. SPFs are developed through statistical multiple regression techniques using observed crash data collected over a number of years at sites with similar characteristics and covering a wide range of AADTs (AASHTO, 2010). However, the SPFs outlined in Part C of the HSM rely on multistate data and require local calibration using specific factors to ensure accurate application to individual locations.

Accident Modification Factors (AMFs) or CMFs represent the relative change in crash frequency due to a change in one specific condition (when all other conditions and site characteristics remain constant). AMFs are the ratio of the crash frequency of a site under two different conditions Therefore, an AMF may serve as an estimate of the effect of a particular geometric design or traffic control feature or the effectiveness of a particular treatment or condition.

Crash frequencies, even for nominally similar roadway segments or intersections, can vary widely from one jurisdiction to another. Calibration is the process of adjusting the SPFs to reflect the differing crash frequencies between different jurisdictions. Calibration can be undertaken for a single state, or where appropriate, for a specific geographic region within a state.

This section will discuss the studies that develop models for predicting crash frequency using relationships between the crash occurrence and risk elements.

Early crash prediction models often relied on simple multiple linear regression methods, with an underlying assumption of normally distributed errors. The use of multivariate or non-multivariate linear regression models for crash prediction have been developed and proposed as seen on some literatures (Golob et al., 2004; Golob and Recker, 2003; Persaud et al., 2000).

Over the years, multiple linear regression models have been widely used in crash analysis and prediction. However, it has been observed that crash data often exhibit characteristics that align more closely with a Poisson distribution rather than a normal distribution. One common mistake is to inappropriately treat crash data as continuous and apply ordinary least squares regression (Abdulhafedh, 2017). Consequently, generalized linear modeling variates, specifically tailored to accommodate the characteristics of Poisson regression, have been proposed and developed to investigate the correlation between risk factors and the modeling of road crashes (Abdulhafedh, 2016; El-Basyouny and Sayed, 2009; Ma et al., 2008; Park and Lord, 2007). However, it has been identified that Poisson regression methods are subject to a crucial restriction: the requirement for the mean to be equal to the variance. Deviation from this condition can lead to biased standard error estimates through maximum likelihood estimation, consequently producing inaccurate test statistics from the model. Recent research has indicated that crash data typically exhibit over-dispersion, wherein the variance surpasses the mean. This discrepancy highlights the potential for misestimation in the likelihood of crash occurrence when employing the Poisson regression model (Lord and Mannering, 2010).

To address the issue of over-dispersion, researchers have turned to utilizing the Negative Binomial (NB) distribution, also known as the Poisson-Gamma distribution, as an alternative to the Poisson distribution. Unlike the Poisson distribution, the NB distribution does not impose the constraint of equal mean and variance, thus allowing for the accommodation of over-dispersion in counts of crash data (Lord and Mannering, 2010). As a result, different NB models have been extensively proposed and used in crash frequency prediction and modeling (Abdulhafedh, 2016; Daniels et al., 2010; El-Basyouny and Sayed, 2009; Geedipally et al., 2012; Kim et al., 2007;

Malyshkina and Mannering, 2009). Nonetheless, the Negative Binomial (NB) model is not without its limitations, as it cannot effectively address situations of under-dispersion in data counts, wherein the mean of crash counts exceeds the variance (Amoros et al., 2003; Lord, 2006; Oh et al., 2006).

In response to the constraints of Negative Binomial (NB) models, the Poisson-lognormal model was introduced (Aguero-Valverde and Jovanis, 2008; Daniels et al., 2010; Lord and Miranda-Moreno, 2008). In this model, the error term follows a Poisson-lognormal distribution instead of a gamma distribution, which enhances its capability to manage under-dispersed data counts. While offering greater flexibility compared to the negative binomial model, the Poisson-lognormal model is not without drawbacks, notably in its complex parameter estimation process due to the absence of a closed form for the Poisson-lognormal distribution (Lord and Miranda-Moreno, 2008).

Another prevalent approach in crash frequency modeling involves the use of zero-inflated Poisson and zero-inflated negative binomial models. These models were primarily developed to address the issue of over-dispersion caused by an excessive number of zeros, typically observed in traffic data counts. The zero-inflated methodology allows for the modeling of crash frequencies in two states: the zero-crash state and the non-zero crash state, where crash frequencies follow either a Poisson or negative binomial distribution (Shankar et al., 1997). The probability of a section being in either the zero or non-zero state can be determined using a binary logit or probit model.

In crash data, the presence of numerous zero observations are often attributed to several factors, including underreporting of minor crashes, the existence of dangerous crash sites nearby

(which render neighboring sites safer by comparison), and the possibility that certain sites may be free from specific types of crashes while still experiencing others. Zero-inflated models aim to address these excess zeros by assuming a dual-state crash system. In this system, one state represents the zero-crash state, which is essentially considered safe during the observation period, while the other state represents the non-zero crash state, where crash occurrences are expected and follow a count distribution (Lambert, 1992).

Other Network Screening Methods

(Tarko and Kanodia, 2004) proposed two crash screening methods for ranking hazardous locations. One of the methods is based on the difference between expected and true crash numbers and the other is based on crash cost. The two proposed methods are index of crash frequency and index of crash cost.

(Al-Kaisy et al., 2019) developed a risk index to identify locations along Oregon's low volume rural roads that deserve further consideration. The crash risk index was developed using three major elements: geometric features, crash history and traffic exposure. Weights, which show the contribution of geometric and roadside features, crash history, and traffic exposure elements in the overall crash risk index, are assigned to each of these elements.

A Canadian study (Leur and Sayed, 2002)developed a Road Safety Risk Index (RSRI) utilizing concepts related to the traffic conflict observation technique and drive through safety reviews. Well-defined and quantifiable characteristics of road features are studied and scored while completing a drive-through review. These scores are then combined to produce an overall road safety risk, by combining three components of risk: the exposure of road users to road

features, the probability of becoming involved in a collision, and the resulting consequences should a collision occur.

A safety index (SI) that quantitatively measures the relative safety performance of a road segment is calculated from the procedure (Cafiso et al., 2007). The SI is formulated by combining three components of risk: the exposure of road users to road hazards, the probability of a vehicle's being involved in an accident, and the resulting consequences should an accident occur. The Safety Index (SI) was evaluated in 30 segments of two-lane rural highways in Italy using the empirical Bayes (EB) method to obtain actual accident data. The analysis revealed a strong agreement between SI scores and EB estimates, with a correlation coefficient of 0.87, significant at the 99.9% level. Some other methods that uses safety index and scoring system for network screening on highway are developed and proposed in the literatures (Al-Kaisy and Raza, 2023; Bonera et al., 2022; Himes and Donnell, 2020; Tamburri and Smith, 1970).

Empirical Bayes Method (EB)

The statistical randomness can be demonstrated by observing the number of crashes occurring at a site for a certain time period, which usually changes from year to year. A closer examination will reveal some trends in this fluctuation. A year with a low number of crashes usually follows a year with a high number of crashes and vice versa. This phenomenon is called regression-to-the-mean (RTM), when unusually large or small measurements tend to be followed by measurements that are closer to the mean (Bland and Altman, 1994). RTM complicates a safety analysis because most “unsafe” locations will, if nothing is done, experience a lower crash rate after time elapses. More importantly, the safety decision is sometimes based on a short-term average that may be different from the actual long-term average (Hauer et al., 2002; Maher and

Mountain, 2009; Persaud, 2001; Sharma and Datta, 2007). This tendency frequently misleads the analyst to draw incorrect conclusions as to high-risk locations, resulting in applying countermeasures in locations where the crash rate will likely decrease regardless of what countermeasures may be employed. More critically, the areas with potentially high crash risk may be overlooked. Figure 2 illustrates a hypothetical example of RTM. To address this issue, a more refined approach is necessary for identifying sites with genuine safety improvement potential.

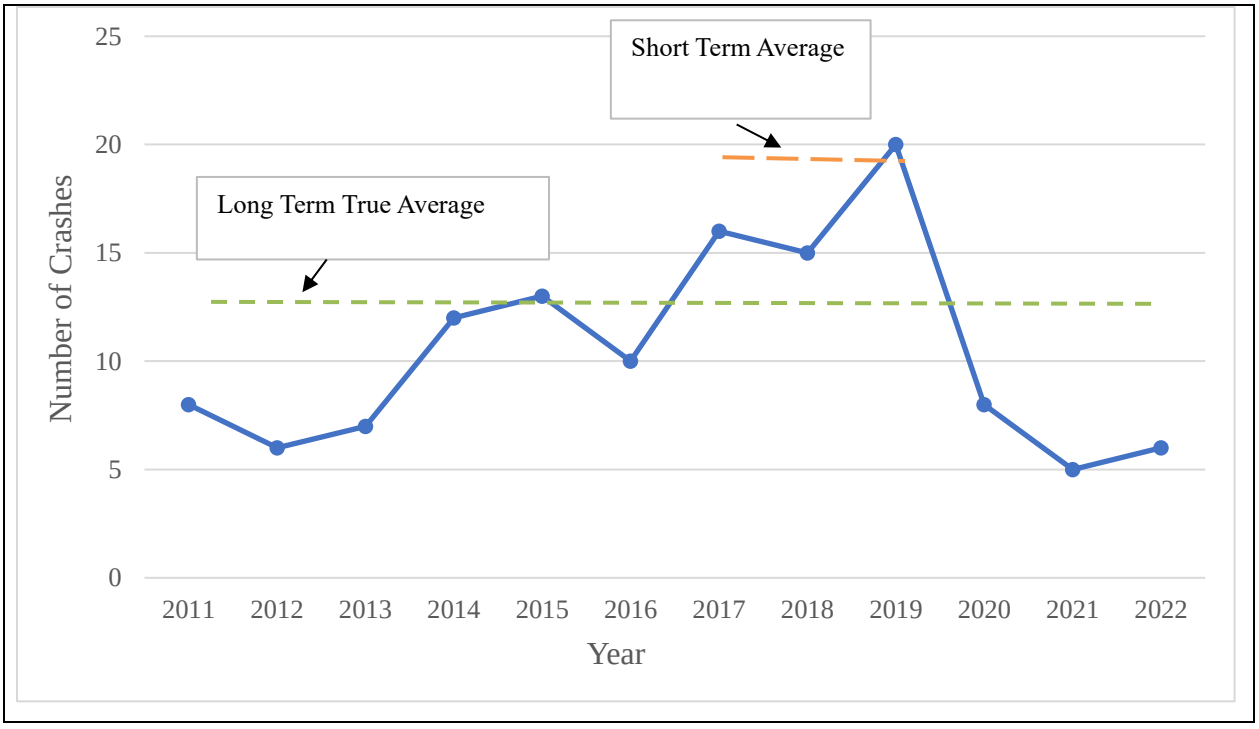


Figure 2. Hypothetical example of regression-to-mean

The Empirical Bayes estimate obtained from the EB method is the weighted sum of the expected value and the actual value, where the weights are determined based on the number of actual measurements for a certain variable. In other words, the Empirical Bayes method estimates the expected number of crashes using the HSM predicted number of crashes and the

observed number of crashes from crash history. This estimation is developed using the following equation:

$$N_{expected} = w * N_{predicted} + (1 - w) * N_{observed} \quad (1)$$

Where, $N_{expected}$ is the estimated expected number of crashes, $N_{predicted}$ is the predicted number of crashes using SPFs and CMFs, $N_{observed}$ is the observed crash count over the study period, and w is the weight for predicted number of crashes.

This equation assigns different weights (contributions) to the predicted and observed numbers of crashes in estimating the expected number of crashes. The predicted number of crashes is found using a predictive mathematical model, also known as safety performance function (SPF), and a set of crash modification factors (CMFs). The predictive HSM method predicts the number of crashes at a specific site in a two-step process. First, the method utilizes the SPF mathematical model in predicting the number of crashes under base conditions (e.g., for two-lane highway segments, inputs are the AADT and segment length). equation (2) shows the SPF for rural two-lane highway segments.

$$N_{spf} = AADT * L * 3655 * e^{-0.312} \quad (2)$$

Where, N_{spf} is the estimate of predicted average crash frequency for SPF base conditions, $AADT$ is the annual average daily traffic volume, L is the length of roadway segment.

Then, the method accounts for roadway and roadside characteristics that deviate the base conditions and the predicted crash numbers using crash modification factors and safety performance function is shown in equation (3) below:

$$N_{predicted} = N_{spf} * CMF_1 * CMF_2 * \ldots\ldots * CMF_n \quad (3)$$

Where, $N_{predicted}$ is the predicted crash numbers using SPFs and CMFs, $CMF_1 \ldots CMF_n$ are the crash modification factors for roadway and roadside characteristics that deviate from base conditions.

The weight for predicted number of crashes is calculated as:

$$w = \frac{1}{1+k*N_{predicted}} \quad (4)$$

Where, w is the weight of predicted number of crashes, k is the overdispersion parameter, and $N_{predicted}$ is the predicted number of crashes.

Once predicted number of crashes are obtained, then equation (1) is used to find the EB expected number of crashes.

There has been an extensive discussion of the regression methods used for developing SPFs in the EB (e.g. (Dixon et al., 2012; Shin and Washington, 2012; Wood et al., 2015a, 2015b).

EB methods have been widely used and discussed in traffic safety studies over the last two decades, especially in before–after evaluations (Aguero-Valverde and Jovanis, 2008; El-Basyouny and Sayed, 2009; Elvik et al., 2017; Hauer, 1992; Hauer et al., 2002; Li et al., 2008; Miaou and Lord, 2003; Park and Lord, 2007; Persaud and Lyon, 2007; Quddus, 2008; Wood and Donnell, 2017).

The potential for safety improvements (PSI) is estimated as difference between the EB expected number of crashes and predicted crashes what is normal for similar sites (Persaud et al., 1998). A segment is said to have potential for improvement if the observed number of crashes is greater than the predicted number of crashes for that segment. The PSI is calculated using equation 5 as below:

$$PSI = N_{expected} - N_{predicted} \quad (5)$$

Where, PSI = potential for safety improvement value, $N_{expected}$ is the EB estimated expected number of crashes, and $N_{predicted}$ is the predicted number of crashes.

The basis of this index was first suggested by McGuigan (McGuigan, 1981; O, 1972) and was called the potential accident reduction. However, the authors suggested using the accident count instead of EB expected number of crashes in equation 4. This could be problematic because accident counts can vary randomly, especially when based on a short history of accidents, which is often the case. More recently, (Tarko et al., 1996) sought to overcome this difficulty by using the concept of the confidence level for which the above-norm number of crashes is larger than zero. A validation of the potential-for-safety-improvement concept (Persaud et al., 1998) showed that the method is not only conceptually sound but is also comparatively efficient. However, some concerns arose, and it was decided to continue the research into the concept and to undertake a more extensive validation. One area of concern is in the definition of what is normal for similar sites—that is, the value of predicted number of crashes to be used in Equation 4. If treatable variables, such as shoulder width, that may contribute to a site being unsafe are part of the predicted number of crashes, then both $N_{expected}$ and $N_{predicted}$ will reflect this contribution. Consequently, even if a site is unsafe, Equation 4 may suggest it has little or no potential for safety improvement.

Evaluation of Network Screening Methods

Numerous network screening techniques have been used in practice or proposed in the literature with each having its own strengths and limitations. Compared with the large number of

studies focused on the development of various hotspot identifications methods, considerably less research has been dedicated to evaluating the performance of various methods (Cheng and Washington, 2005). In this section, the major studies that evaluated the performance of different network screening techniques are summarized and presented.

A study by (Ambros et al., 2015) investigated the difference between network screening results based on multivariate and simple crash prediction models. The study was performed to answer the research question: “What is the difference between network screening results based on multivariate and simple crash prediction models?” For the purpose of investigation multivariate and simple crash prediction models were developed for regional road network of South Moravia (Czech Republic) and used in network screening. This study first compared the list of segments using Spearman’s rank correlation coefficient between the two models. Secondly, this study assessed the equality of statistical distributions of potential for safety improvement (PSI) value, which was obtained as a difference between predicted crash frequency and EB estimate. The third comparison was the percentage of segments identified in both lists. The findings suggested that the results from both methods were comparable.

The effectiveness of several network screening techniques was examined in an Italian study (Montella, 2010) using a set of criteria. These included reliability in detecting hazardous sites over a period, efficiency in finding sites with poor safety performance, and consistency in ranking. In this research, seven commonly applied hotspot identification (HSID) methods were compared against four robust and informative quantitative evaluation criteria. The potential for improvements (PFI) was examined along with the crash frequency, equivalent property damage only (EPDO), crash rate, proportion method, and EB estimates for total and severe crashes. The

HSID methods were compared using the site consistency test, the method consistency test, the total rank differences test, and the total score test. This analysis used crash data for five years from an Italian highway. The quantitative evaluation tests showed that the EB method performs better than the other HSID methods investigated in this study. Test results highlight that the EB method is the most consistent and reliable method for identifying priority investigation locations. Also, study results suggest that the crash frequency method outperformed the crash rate method.

A different Italian study (Cafiso and Di Silvestro, 2011) used simulated data to evaluate the effectiveness of various hotspot identification techniques. A Monte Carlo simulation was used to produce theoretical crash data similar to empirical data, and the data were used to define a priori hazardous sites and, therefore, to assess whether a method could correctly identify such site. This study assessed the estimated EB, the potential for safety improvement (PSI), observed crash frequency, and crash rate. The EB and PSI methods outperformed the crash frequency and crash rate methods in this study's identification of high crash locations.

(Cheng and Washington, 2008) proposed four new quantitative HSID evaluation tests to address the gap in systematic assessments. These tests assess various aspects of HSID method performance, including result reliability, ranking consistency, and false identification consistency and reliability. The aim is for practitioners to employ these evaluation tests alongside existing ones to compare HSID method performances effectively and select the most suitable method for identifying sites requiring further analysis. This study illustrates these new criteria using three years of Arizona road section accident data and four commonly used HSID methods: accident frequency ranking, accident rate ranking, accident reduction potential, and empirical Bayes (EB). The results demonstrate that the EB HSID method generally outperforms others in the evaluation

tests, while accident rate ranking performs the least favorably. Accident frequency and accident reduction potential methods show similar performances, with subtle distinctions elucidated.

Another study by (Cheng and Washington, 2005) evaluated the network screening methods using experimentally derived simulated data. By manipulating various factors to simulate real-world conditions, the study assesses the performance of each method across different levels of confidence while considering the occurrence of false positives (identifying safe sites as high risk) and false negatives (identifying high-risk sites as safe). Additionally, the influence of crash history duration on the effectiveness of the HSID approaches was examined. The findings highlight the superiority of the Empirical Bayes technique over ranking and confidence interval methods, albeit with certain considerations. Consistent with prior research, there exists an inverse relationship between false positives and negatives. Moreover, the study suggests that a three-year crash history duration generally yields optimal results.

In a recent study (Ambros et al., 2016b) focusing on a rural-road network spanning approximately 1,000 km in South Moravia, Czech Republic, data spanning 8 years of annual crash frequencies, alongside exposure and geometrical variables, were utilized to develop various prediction models and the performance of these models for network screening were evaluated. Through a rigorous assessment involving consistency tests comparing different model variants and their diagnostic performance, it was established that simpler models, incorporating traffic volume, segment length, and curvature change rate, suffice for effective network screening. Notably, the study suggests that given the relative stability of segment length and curvature over time, periodic updates may primarily focus on traffic volume data. Furthermore,

consistency analyses indicated that a four-year updating cycle for these models aligns with optimal performance.

Another study in Czech Republic (Ambros et al., 2016a) evaluate three distinct approaches for identifying hazardous road locations within a regional rural road network spanning approximately 1000 kilometers in South Moravia, Czech Republic. These approaches include: (1) the conventional reactive accident-based method, focusing on pinpointing accident black spots; (2) the state-of-the-art empirical Bayes technique utilizing accident prediction models to identify critical locations, encompassing both existing and potential black spots; and (3) a proactive approach involving preliminary road safety inspections to identify risk factors that could contribute to increased accident occurrence and severity. The study finds that the traditional black spot approach is unsuitable, particularly within low-volume road networks characterized by sporadic accident occurrences. Conversely, the risk index derived from road safety inspections emerges as a viable alternative, demonstrating comparable ranking performance to the empirical Bayes method. Furthermore, the study compares the data requirements of both the empirical Bayes and risk index methods and recommends their combined application as a robust replacement for traditional black spot management.

(Deacon et al., 1974) performed a study in Kentucky to determine the hazardous rural highway locations. Study findings suggest that to identify hazardous highway locations, it is necessary to distinguish between short highway segments (spots) and large segments (sections) and to further classify spots as intersection and non-intersection locations. Intersection spots should include a distance of 0.15 mile (0.24 km) along all approaches; non-intersection spots should be 0.3-mile (0.48-km), floating segments; and sections should be 3-mile (4.8-km),

floating segments. Both spots and sections should be classified by highway type and location. The use of dual time intervals of 1 and 2 years for accumulating and evaluating accident statistics was found to be desirable.

The study by (Elvik, 2008a) presents a comparative analysis of five methodologies aimed at identifying hazardous road locations, each characterized by varying degrees of control for randomness in accident counts. Utilizing data from Norwegian roads, the study establishes a definition of hazardous road locations as those with an expected higher number of accidents compared to similar sites due to local risk factors. The five techniques under scrutiny include: recording the number of accidents within a specific period, observing the accident rate (accidents per million vehicle kilometers) during a specific period, combining a critical count of accidents with an above-normal accident rate during a specific period, utilizing the empirical Bayes estimate of expected accidents at each location, and determining the contribution of local risk factors to the empirical Bayes estimate. Each technique is applied to subsets representing the upper 1%, 2.5%, and 5% of the distribution of sites and evaluated against various criteria, such as accident count and rate. The diagnostic performance of the techniques is assessed using epidemiological criteria, including sensitivity and specificity. The empirical Bayes technique emerges as the most effective method according to epidemiological criteria, suggesting that it offers the most reliable means of identifying hazardous road locations.

(Huang et al., 2009) introduced a novel model-based hot spot identification framework utilizing the full Bayes (FB) technique. Compared to the empirical Bayes method (EB), FB offers the advantage of seamlessly integrating prior information and available data into posterior distributions for ranking criteria. Using intersection crash data from Singapore, six approaches

were assessed, including naive ranking, standard EB ranking, FB ranking using a Poisson-gamma model, FB ranking using a Poisson-lognormal model, FB ranking using a hierarchical Poisson model, and FB ranking using a hierarchical Poisson (AR-1) model. The study findings demonstrated that all model-based approaches outperform naive ranking method in safety ranking when using the expected crash rate-related decision parameters. Also, the FB approach using hierarchical models significantly performs better than the standard EB approach in correctly identifying hazardous sites.

A study conducted in Belgium (Moons et al., 2009) adopts a novel approach by leveraging Moran's I to account for the underlying structure of the road network, thereby identifying hot spots indicative of heightened risk. Taking a step beyond mere identification, the study delves into a comprehensive hot zone analysis, offering deeper insights into these identified hot spots. This extended analysis holds significance both theoretically, enriching our understanding of conceptualizing hazardous locations, and practically, providing valuable insights for addressing traffic safety issues. The study demonstrates that integrating the hot zone methodology within hot spot analysis yields a clearer depiction of hazardous road locations.

(Kwon et al., 2013) conducted a comprehensive evaluation of three distinct methods for segmenting freeway sites aimed at identifying locations with a high concentration of collisions. The methods under evaluation—Sliding Moving Window (SMW), Peak Searching (PS), and Continuous Risk Profile (CRP)—were assessed based on their performances in segmenting sites using traffic collision data under two different roadway definitions. The study used Empirical Bayes adjustment to estimate excess expected average crash frequency, considering two sets of Safety Performance Functions (SPFs). The results were used to prioritize identified sites for

safety investigation, with subsequent comparisons made against confirmed high collision concentration locations. Notably, the CRP method exhibited superior performance, demonstrating the lowest false positive rate. Furthermore, the study revealed the robustness of the CRP method, with its performance being less affected by variations in the sets of SPFs compared to the SMW and PS methods.

A Federal Highway Administration (FHWA) study (Gross et al., 2016) performed a comprehensive analysis using intersection data from New Hampshire to evaluate the performance of four network screening method on identifying high-priority sites for safety investigations. The four network screening measures – crash frequency, crash rate, EB expected crashes, and EB excess expected crashes – were evaluated by analyzing the overall economic benefit and benefit-cost ratio for each of the four techniques. Study findings suggested that 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.

(Ghadi and Török, 2019) compared four common accident black spot identification methods (empirical Bayes, excess empirical Bayes, accident frequency and accident rate) and four segmentation methods (spatial clustering, constant length, constant traffic volume, and the standard Highway Safety Manual segmentation method) to study the effect of methodological diversity of road network segmentation on the performance of different black spot identification methods. Initially, the approach evaluated the segmentation methods based on the accuracy of the developed safety performance function (SPF). Then, a consistency test was applied to compare the joint performances of the black spot methods and segmentation methods. Study

results suggested that the black spot methods showed a significant change in their performance depending on the different segmentation methods applied.

A study conducted in Brazil (Mesquita Xavier and Craveiro Cunto, 2017) compared different safety performance measures and their practical limitations using a sample of signalized intersections located in Fortaleza City, Brazil. The safety performance measures evaluated in this study average crash frequency (ACF), crash rate (CR), Equivalent Property Damage Only (EPDO), Level of Service of Safety (LOSS), Excess Predicted Average Crash Frequency using Safety Performance Functions (SPFs), Expected Average Crash Frequency with EB Adjustments (EB), and Excess Expected Average Crash Frequency with EB Adjustment (EEB). The difference in rank between each safety performance measure assessed, and the Excess Expected Average Crash Frequency with EB Adjustment (EEB) was used to evaluate the performance of the subjected method. The findings suggested that the most comprehensive measure EEB, and fundamental measures like crash frequency and crash rate displayed reasonably similar rankings.

In a study (Lee et al., 2020), a novel network screening method is proposed for identifying hotspots, defined as locations characterized by high collision concentration and significant potential for safety enhancement. Notably, the proposed approach allows for the dynamic determination of hotspot lengths based on user-defined constraints. The calculation of the Dynamic Site Length (DSL) method, facilitated by Dynamic Programming, is demonstrated to yield close-to-optimal solutions with manageable computational complexity. To validate the efficacy of the screening method, historical crash data from extended freeway routes in San Francisco, California, is utilized. Employing the Empirical Bayesian (EB) estimate as a safety measure, the performance of the DSL method is compared against conventional screening

methods such as Sliding Window (SW) and Continuous Risk Profile (CRP). Evaluation criteria include the optimal objective value, assessing the ability to detect sites with the highest risk, as well as spatiotemporal consistency, analyzed through site and method consistency tests. Results indicate that the DSL method exhibits superior performance compared to SW and CRP by identifying more hotspots with greater accuracy and demonstrating improved spatiotemporal consistency under equivalent resource allocations for DES.

Literature Review Summary

The literature review in the document extensively covers various network screening methods used for road safety analysis, from traditional descriptive and crash frequency methods to advanced predictive crash analysis models. It discusses the evolution and evaluation of different techniques, highlighting their strengths, limitations, and applicability in identifying high-risk locations on road networks. Traditional methods, such as crash frequency and density analyses, offer simplicity but may not account for traffic volume, leading to potential inaccuracies. In contrast, predictive models, including the Empirical Bayes method, incorporate traffic and roadway characteristics for a more nuanced understanding of crash risks. These methodologies reflect a significant shift towards a more analytical approach, focusing on the underlying factors contributing to road incidents, rather than merely depending on crash occurrences. However, the literature review also reveals a difficulty in the implementation of these advanced models across different regions, particularly in rural areas and local agencies where data may be limited.

This study addresses the notable gap in literature regarding the comprehensive evaluation and adaptation of predictive road safety models in regions with limited or inconsistent data

quality. By exploring these under-researched areas, this study aims to bridge the gap, by developing and validating different methodologies that are easily applicable in rural areas overcoming the difficulties posed by limited data availability.

CHAPTER THREE

STUDY DATA

This chapter primarily focuses on describing the case study area, from which the data required for validation are extracted. It presents a map showing the study sample highway within the road network of the state of Oregon. Extensive and detailed road characteristics, traffic data, and crash data have been collected and analyzed for a large representative sample for this validation study. The following sections provide detailed information about the study area and the data collection process.

Study Area

The study area for this research included the state-owned two-lane rural roadway network in the state of Oregon. The state of Oregon was selected in this study for its diverse terrain and area character and for considerations related to data availability and accessibility. A total roadway segment sample of around 1495 miles was used in this investigation to ensure adequate geographic coverage and mixed traffic exposure. The study sample comprises of around 850 miles roadway segment of volume less or equal to 1000 vehicles per day (vpd) or low-volume roads and remaining 645 miles of volume greater than 1000 vpd but less than 4000 vpd. In this study, low-volume roads are defined as rural two-lane highways with Average Annual Daily Traffic (AADT) equal to or less than 1000 vehicles per day. This is consistent with a few other studies that used the same definition in the literature (Ewan et al., 2016; Gross et al., 2011). All state-owned roads with AADT less than or equal to 1,000 vpd were queried using online GIS data, and then random selections were made from that query to arrive at the target length of road

sample. Also, to get the target sample of AADT greater than 1000 vpd and less than 4000 vpd, similar procedure was performed in GIS. The study sample included roadways from different parts of the state, as shown in Figure 3. In general, roadways in the eastern part of the state run in flatter terrain with less restrictive alignment, while those in the western part of the state run in mountainous terrain with more restrictive alignment (winding routes, sharp radii, etc.).

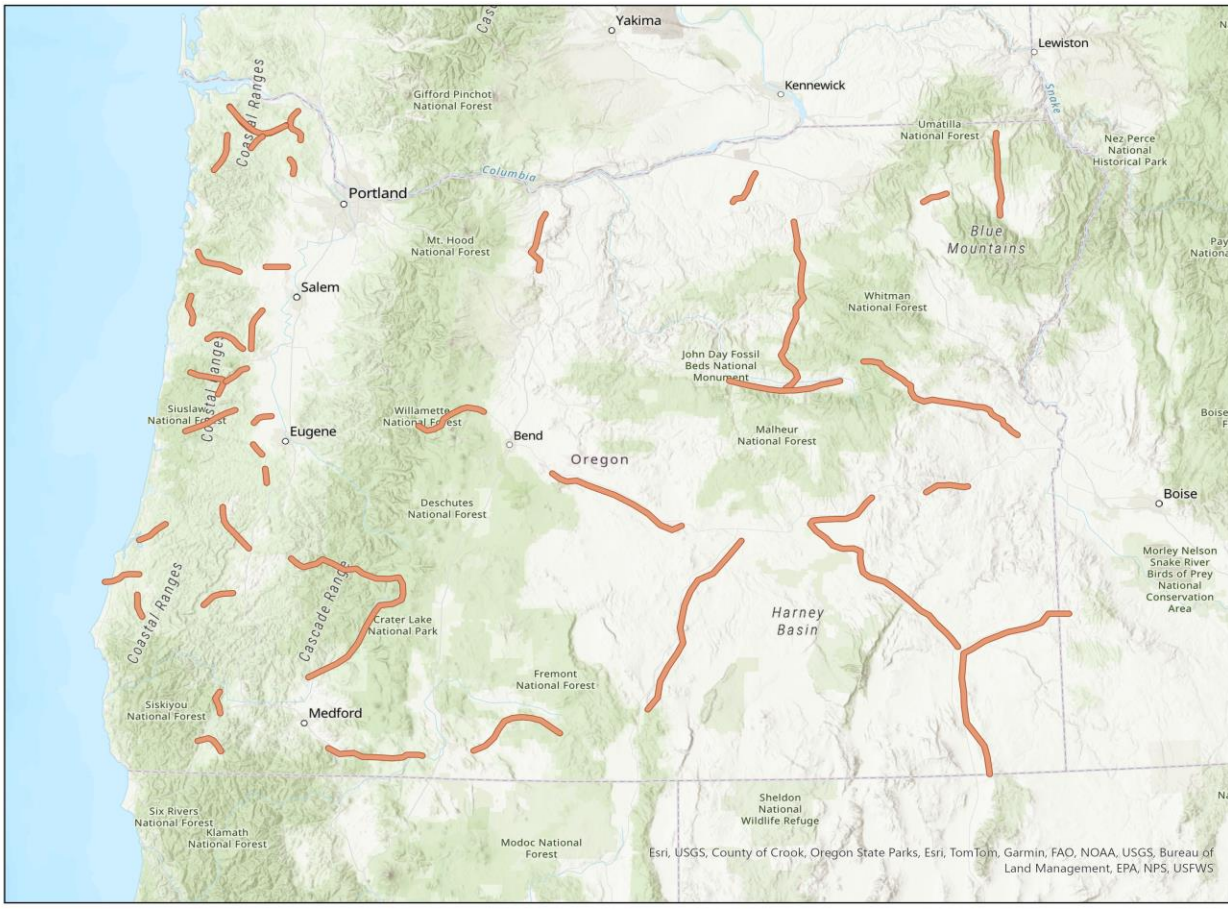


Figure 3. Study area showing sampled highway sections.

Data Collection

All state-owned rural two-lane roads were identified using online geographic information system (GIS) data ("ODOT TransGIS"). The focus was on state-owned roadways for considerations related to data availability and accessibility. Then, selections were made from the results to arrive at the target sample considering geographical coverage. Data were collected for roadway segments using 0.05-mile increments to ensure that data would capture all changes in the physical characteristics of the roadway, thus eliminating the possibility of missing significant differences between consecutive observation points. All segments in the study sample had a posted speed limit of 55 mph. Intersections were not included in the study sample. Intersections and 0.05-mile segments along upstream and downstream approaches were excluded from the dataset. Three different categories of data collected for this study including roadway geometry, traffic data, and crash data are discussed below:

Roadway Geometry

The Oregon Department of Transportation (ODOT) online databases and video logs were used to identify and compile roadway geometry data, and roadside characteristics ("Oregon Department of Transportation : Road Assets and Mileage : Data & Maps : State of Oregon," n.d.; "Oregon Department of Transportation," 2000). The roadway geometric data consisted of lane type and width, shoulder type and width, horizontal curve presence and direction, degree of curvature, length of the horizontal curve, spiral curve presence and length, vertical curve presence and type, grade, and length of vertical curve taken from the online database. Roadside characteristics comprised driveway density, side slope ratings, and fixed object ratings taken from ODOT video logs at 0.05-mile resolution. Side slope rating is provided from 1 through 3

for steep, moderate, and flat roadside terrain, respectively. Similarly, fixed object rating is provided from 1 through 3 for many, some, and few fixed objects/obstacles within 15 feet of the road, respectively.

Crash Data

Ten years of crash data from 2011 to 2020 were collected for the study sample from the ODOT online database ("TDS - Crash Reports"). As mentioned earlier, for the study sample, intersection, or intersection- related crashes were not collected in the data set. Data on more than 20 individual crash characteristics—including crash location, road character, impact location, traffic control, crash type, crash severity, vehicle type, and weather condition—were collected and combined with the geometric and roadside database, forming an integrated dataset for analysis. The reported injury severities were fatal, non-fatal injury, and property damage only.

Traffic Data

Traffic volumes were reported as average annual daily traffic (AADT). Traffic data for ten years (2011-2020) were collected separately from the ODOT online database ("ODOT: Road Assets and Mileage"). The traffic data also consist of percentage of the vehicles according to their functional classification. The percentage of heavy vehicles in the traffic stream (AASHTO class 4 or larger) was compiled separately and combined with traffic data into a unified database for analysis.

The resulting comprehensive data set is very rich and includes all road characteristics described combined with all crash data and traffic data for the 1495 miles of roadway segment (850 miles low-volume, and 645 miles higher-volume state-owned roads in Oregon) at 0.05-mile resolution. The database prepared for this study consists of over 30,000 records, each

representing a 0.05-mile sub-segment with 62 individual road and crash characteristics for each sub-segment.

Data Processing

Most of the road characteristics gathered from ODOT databases required no interpretation and could be used as recorded by ODOT including lane type and width, shoulder type and width, AADT, grade, horizontal curve data (degree of curvature, length, spiral presence), and vertical curve data (type, length). Other data necessary for analysis required manual video log review and data entry. Driveway density, side slopes, number of fixed objects present in the clear zone and guardrail presence were all recorded manually while reviewing video log images at 0.05-mile increments. Side slope ratings and fixed object ratings are somewhat subjective, requiring the data collector to assign values from video log images. Side slope ratings were characterized as 1 (flat), 2 (moderate), or 3 (steep). Fixed object ratings were 1 (few fixed objects in the clear zone), 2 (some), or 3 (many). Side slope and fixed object ratings were collected for each side of the roadway independently then averaged. The same data collector was used for recording both ratings for the entire sample to limit any variability that may have been introduced had multiple persons been used. All crash data gathered was used as recorded in the ODOT database, requiring no subjective interpretation.

As mentioned earlier, the data were collected using 0.05-mile increments. Afterward, the total sample was compiled into homogeneous segments concerning the different variables, such as annual average daily traffic (AADT), speed limit, lane plus shoulder (total) width, side slope, and lane type. Any change in one or more variables marked the end of a segment and the beginning of another segment. For example, let us consider the 1-mile stretch of a highway

segment shown in Figure 4. This stretch can be divided into five segments based on variations in AADT, total lane and shoulder width, lane type, and speed limit. Figure 4 illustrates the segmentation process, also reported in the work of (Ambros et al., 2016b).

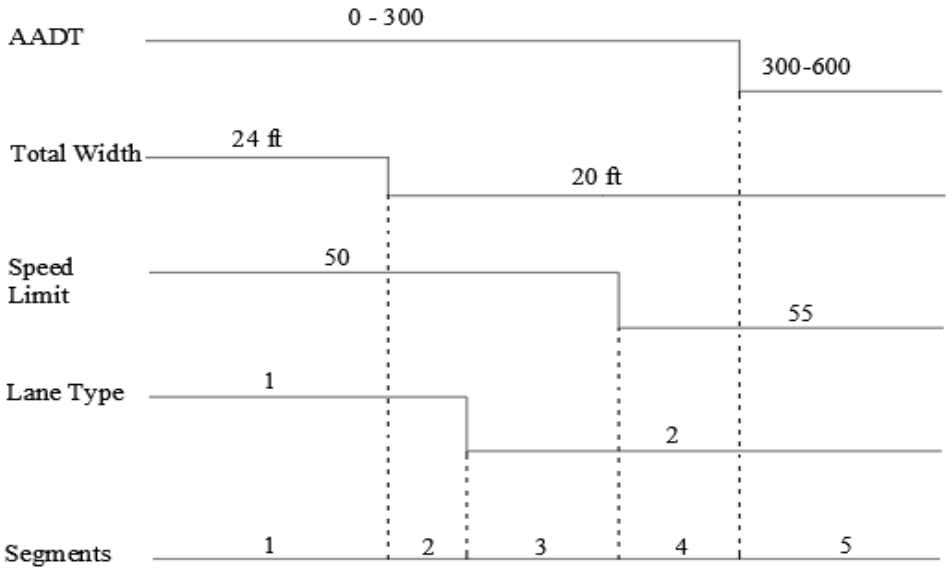


Figure 4. Principle of division into homogeneous segments

Upon completing the segmentation, a total of 377 segments including 1495 miles of roadway segments were compiled. Out of 377 segments, 200 segments were low-volume roads and remaining 177 segments were higher-volume roads with AADT greater than 1000 vpd.

CHAPTER FOUR

METHODOLOGY

The objective of the validation of the network screening methods is to examine the effectiveness of these methods in identifying sites in need of safety improvement. This chapter presents the methodology for performing the validation study of three different proposed network screening methods (the GRS method, the CRI method, and the EB prediction method). The common methodological procedure and metrics used for the evaluation of all three proposed network screening methods are discussed in this chapter. However, any specific evaluation criteria for different methods are presented in the respective chapters, later in this thesis, that presents the validation analysis of those methods.

The approach that was used in evaluating the network screening methods is to compare site rankings using the different proposed methods (the GRS method, the CRI method, and the EB prediction method) with the respective rankings using crash history data. Furthermore, rankings from the Empirical Bayes method were included in the analyses to see how the performance of the proposed method compares to that of the well-established Empirical Bayes technique. In the context of this study, “ranking” refers to the process of arranging roadway segments in accordance with their respective crash risk levels, with the intent of establishing a hierarchical order from segments exhibiting a high crash risk to those displaying a low crash risk. The Empirical Bayes method estimates the expected number of crashes at a site using the observed number of crashes and the predicted number of crashes found from the respective HSM safety performance function. Its main advantage is to alleviate the effect of randomness by not relying solely on crash history. The use of the Empirical Bayes method in this study is based on

the fact that this technique has become the gold standard in performing quantitative safety analyses, with merits and effectiveness well established in the literature (Elvik, 2008b; Hauer, 2002; Hauer et al., 1991; Manepalli and Bham, 2016, 2011; Persaud and Lyon, 2007).

Evaluation Metrics

Spearman's Rank Correlation Coefficient

The Spearman rank correlation coefficient was calculated between the proposed method's ranking and the ranking using observed crash data. Spearman's rank correlation coefficient is a nonparametric or distribution-free rank statistical measure of the strength and the direction of the arbitrary monotonic association between two ranked variables or one ranked variable and one measurement variable (Hauke and Kossowski, 2011). Spearman's rank correlation value varies from −1 to +1, where ±1 represents perfect correlation, and 0 means no correlation. The simple expression for Spearman coefficient "P" based on the difference between the two ranked variables is as shown in Equation (6):

$$P = 1 - \frac{6\Sigma\, d_i^2}{n^3 - n} \tag{6}$$

where P = Spearman's rank correlation coefficient; di = R(Xi) − R(Yi) is the difference between the two ranks of segment "i" by two compared methods; and n = number of observations.

In this study, the evaluation was carried out for five subsets of the total dataset (increasing 20 upper tails in the ranking list with each subset successively) and for the whole

dataset, using the three different analysis periods. If the value is close to 1, the correlation is stronger, and the performance of the method is deemed more favorable.

Root Mean Square Error

Root Mean Square Error (RMSE) is a widely used metric in statistics and machine learning to evaluate the accuracy of a predictive model. It measures the average magnitude of the errors between predicted and actual values, providing a single value that summarizes the model's performance. RMSE is calculated by taking the square root of the average of squared differences between predicted and actual values. A lower RMSE indicates better model performance, with values closer to zero representing higher accuracy (i.e., less disparity between rankings). RMSE is particularly useful for comparing different models or assessing improvements in predictive performance over time.

The three different proposed methods are analyzed according to the average difference between its ranking and the ranking of the observed crash data using the root mean square error as in Equation (7):

$$RMSE = \sqrt{\frac{1}{n}\Sigma d_i^2} \tag{7}$$

where RMSE = root mean square error; di= R(Xi) − R(Yi) is the difference between the two ranks of segment i by two compared methods; n is the number of observations.

True Positive Identification

True positive identification refers to correctly identifying accidents that actually occurred. In other words, a true positive occurs when the predictive model correctly predicts the presence of an accident on a given road segment or within a specified time frame. This is a crucial aspect of model validation because it indicates the model's ability to accurately detect and anticipate real accidents, which is essential for effective accident prevention and intervention strategies. The number of sites identified as critical is estimated. This evaluation listed the number and percentage of common segments between two lists: the one compiled using ranks from the proposed method and the other using ranks from observed crash data. The higher the number and percentage of the common sites, the better the performance of the proposed method because this indicates a stronger match between the prediction of the proposed method and the real-world crash data.

Average Rank Difference

When validating the network screening performance of proposed methods using observed crashes as a reference, calculating the average difference in rank involves comparing the rankings of road segments based on predicted crash risk against their actual observed crash frequencies. This process quantifies how well the proposed methods prioritize high-risk segments compared to the observed crash data. The difference in rank is calculated for a specified number of upper tails, indicating segments with the highest predicted crash risk. A smaller average difference suggests better alignment between predicted and observed crash frequencies, indicating the effectiveness of the proposed methods in identifying high-risk segments. This approach provides a nuanced evaluation of network screening performance,

considering both the accuracy of risk predictions and their relative ranking compared to actual crash data. The average difference in rank is calculated by using equation (8) as follow:

$$Average\ Rank\ Difference = [\sum_{k=1}^{N}\{ R(\ k_{j,i}\ ) - R(\ k_{j,i+1}\ )\}]/N \quad (8)$$

Where, R is the rank for site k in period i for method j,

$i + 1$ is the subsequent time period,

n is the number of upper tails considered.

CHAPTER FIVE

VALIDATION OF METHOD I: GLOBAL RISK SCORING METHOD

Overview of the Global Risk Scoring Method (GRS)

The Global Risk Scoring (GRS) network screening method uses two heuristic scoring schemes for intersections and roadway segments on rural two-lane highways independently (Al-Kaisy and Raza, 2023). The Highway Safety Manual (HSM) (AASHTO, 2010), the nation's preeminent manual for conducting quantitative safety analyses, was largely the basis for developing the proposed scoring schemes. Using these schemes, an individual site in a network obtains a score that is a function of roadway and roadside characteristics, crash history, and traffic exposure during the analysis period. The proposed scoring scheme for highway segments is shown in Table 1. Scores for roadway and roadside characteristics are determined depending on the existence of specific road elements (including horizontal curvature, gradient, roadside fixed objects, road-side slope, driveway density, and the total width of the road). Horizontal curves are those curved sections of highways associated with changes in the direction of the road. The radius of the horizontal curve (an indicator of curve sharpness) is used to provide the score for this road characteristic. A roadside fixed object refers to the presence of non-breakaway fixed objects like trees, utility poles, traffic signs, or any hazardous object within 15 feet from the edge of travel lanes. The side slope is defined as the slope of the cut or fill ex-pressed as the ratio of vertical distance to horizontal distance. The scoring scheme assigns a score for side slopes steeper than 1 V:3 H, also known as non-traversable side slopes. The total width of the road is the

sum of lane width and shoulder width. The driveway density refers to the concentration or frequency of driveways (access points) along a specific roadway segment expressed as the number of driveways per mile. The scores in the proposed scheme were developed using the crash modification factors (CMFs) for rural two-lane roadways that are supplied in the HSM, published in the CMFs clearinghouse maintained by the Federal Highway Administration ("The CMF Clearinghouse: A Handy Safety Tool | FHWA"), or other CMFs reported in published research. The scores assigned to observed crashes were mainly selected to ensure that sites with one or more fatal or serious injury crashes receive further consideration/review for potential safety improvements regardless of the risk factors present. The score assigned to fatal and serious injury crashes is large enough to cause these sites to make it to the priority list of sites. The property-damage-only (PDO) crashes were assigned a weight that is in proportion to the weight assigned to fatal and serious injury crashes (fatal to PDO proportion is 1:16 versus the typical range of 1:15 to 1:25). The proposed screening method also accounts for traffic exposure. The method assigns a multiplier (multiplicative factor) to adjust the relative risk score based on traffic level. These multipliers were derived using the HSM safety performance functions (SPFs) for rural two-lane highways. Specifically, traffic volumes representing volume ranges in the scoring scheme were used in the safety performance function, and the proportions of the resulting number of crashes were used to derive the multipliers. The proposed scoring scheme for this method is as shown in table 1.

Table 1. Proposed scoring scheme for roadway segments (Al-Kaisy and Raza, 2023)

Safety Related Questions	If Yes, Add:
Risk Factors	
Total Width (TD)	
TD ≤ 20 ft?	7
20 ft < TD ≤ 24 ft?	4
Horizontal curve [Radius (R)]	
Flatter curve (R ≥ 300 ft)	30
Sharper curve (R < 300 ft)	60
Grade steeper than 4%?	3
Six or more driveways per mile?	5
Side slope steeper than 1 V:3 H?	4
Fixed objects within 15 ft of travel lane?	4
Unpaved road?	14
Poor pavement condition? (Rutting, potholes, etc.)	7
Crash History Available?	
Number of fatal or serious injury crashes (N1)	N1 * 80
Number of other crashes (N2)	N2 * 5
Relative risk Compound Scores (RRCS)	
Speed ≥ 50 mph?	RRCS * 1.25
Got Annual Daily Traffic (ADT)?	
ADT ≤ 300	RRCS * 1.0
300 < ADT ≤ 600	RRCS * 3.0
600 < ADT ≤ 1000	RRCS * 5.0
ADT ≥ 1000	RRCS * 7.0
Global Risk Score (GRS)	

Methodology

Three different analysis periods were used to check the consistency of the performance of the proposed method for network screening: three years, five years, and ten years. The use of three- and five-year study periods is consistent with the current practice in safety analyses, as found in many studies in the literature (Montella, 2010; Persaud and Lyon, 2007). The rationale for using a ten-year analysis period is that, while crashes are rare events that tend to be randomly distributed over space and time, observing crash occurrence over an extended period of time is believed to largely alleviate the effect of randomness in crash data. Therefore, it was of interest

to see how the analysis results may differ from the shorter study periods used in this study. Figure 5 shows the data flowchart of the study design used in this research.

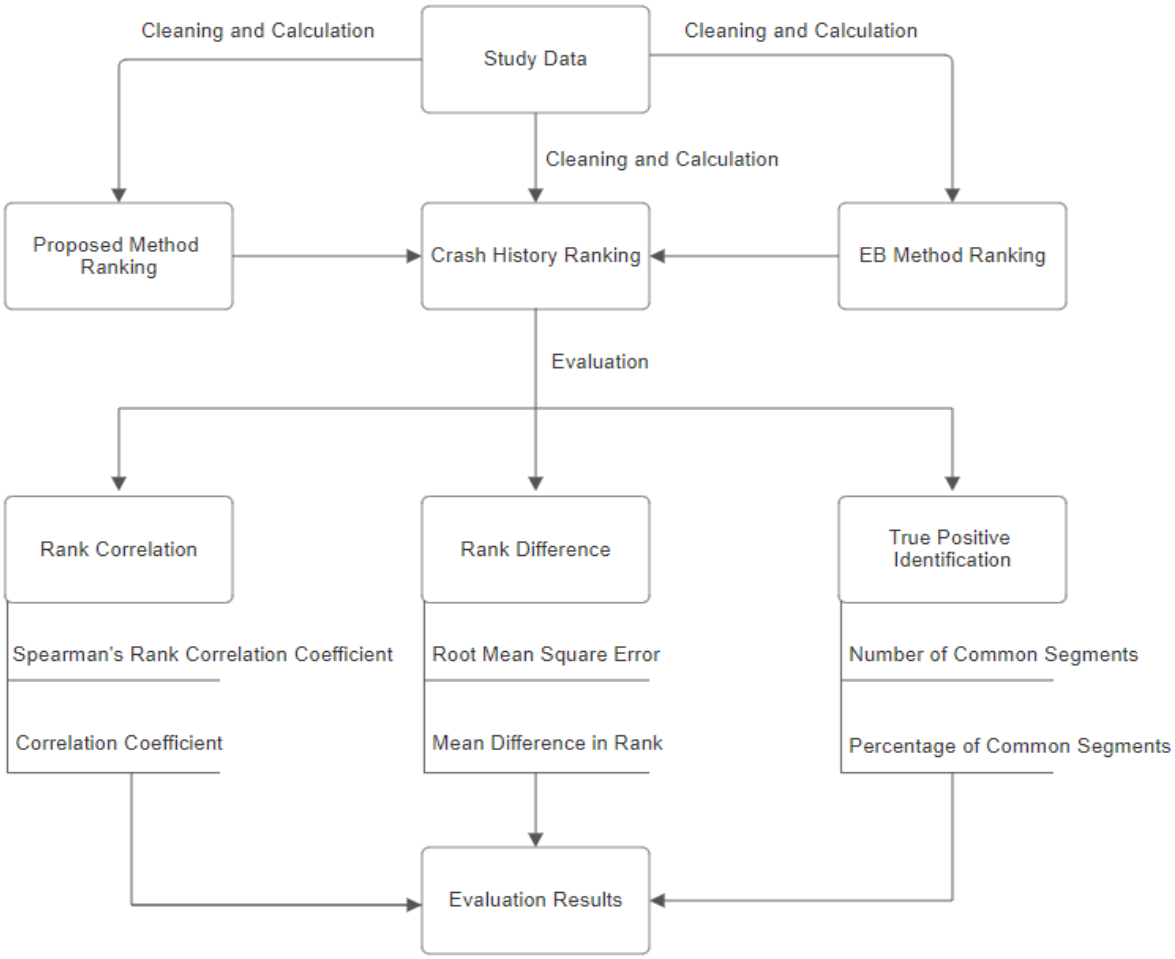


Figure 5. Data flowchart for evaluation of the proposed method.

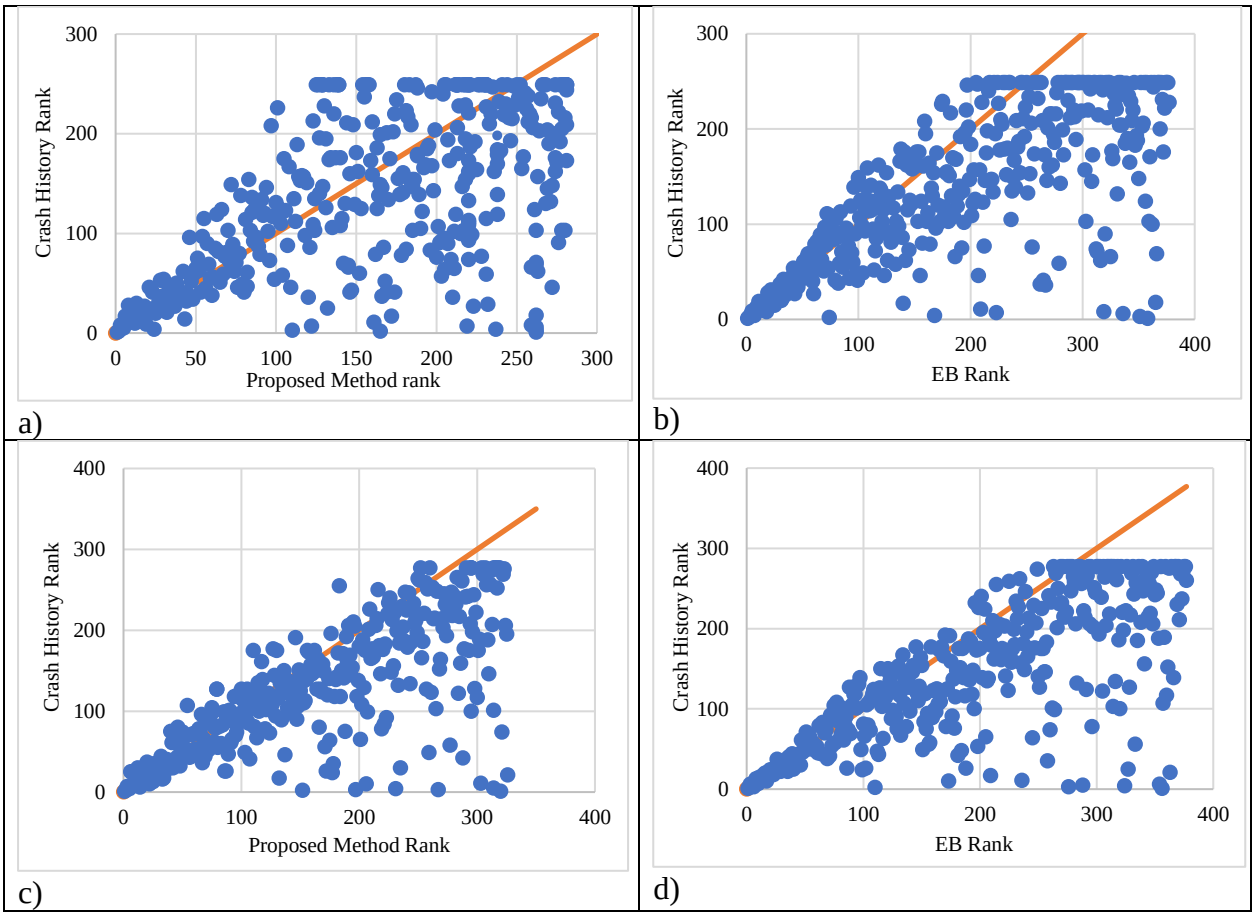
Validation Analysis and Results

Figure 6 shows a scatterplot with site rankings using the proposed method and the EB method versus crash history for the three different analysis periods. A careful examination of all the scatterplots for the proposed method versus crash history (Figure 6a, c, e) reveals a few

important observations. First, for most roadway segments, the two rankings correspond well to each other, as exhibited by the clustered observations around the diagonal line. This is particularly evident in sites that ranked high on the list (first seventy ranks for Figure 6a, and around a hundred ranks for Figure 6c, e). The figures also show a slight downward deviation of clustered observations below the diagonal line, which increases with the increase in rank. This is mostly the result of having two or more sites sharing the same rank using crash history (e.g., segments that had seen no crashes during the analysis period). This is confirmed by the fact that the total number of ranks using crash history is smaller than that using the proposed method. It is important to note that for observations that are scattered outside the diagonal cluster, the rank using crash his-tory almost always resulted in sites being higher on the list (lower ranks) compared to the ranks using the proposed method (observations below the diagonal cluster). The tightness of the data points around the diagonal line in the scatterplot indicates a strong correlation between the two rankings, which is supported by a correlation coefficient of 0.821 for ten years of analysis. However, for the shorter study period of three years (Figure 6a), the data points are much more scattered compared to the longer analysis periods. This is due to the fact that for the shorter time period, more segments in the study sample had no crashes compared to those for longer analysis periods. A perfect correlation is visually represented as a straight diagonal line.

The EB method ranking was also compared with the ranking using the observed number of crashes, and the results are shown in Figure (6b, d, f) for different analysis periods. In these figures, the proximity of data points to the line passing through the origin (diagonal line) indicates that it is somewhat scattered compared to that of the proposed method. This observation

is confirmed by a slightly lower correlation coefficient (r = 0.781) between the two rankings for the ten-year study period. Also, for the three-year study period, the EB method exhibits somewhat similar observations to that of the proposed method. This is also the result of a large number of segments having no crashes. However, it is worth noting that the EB method shows better performance than the proposed method for the three-year period.



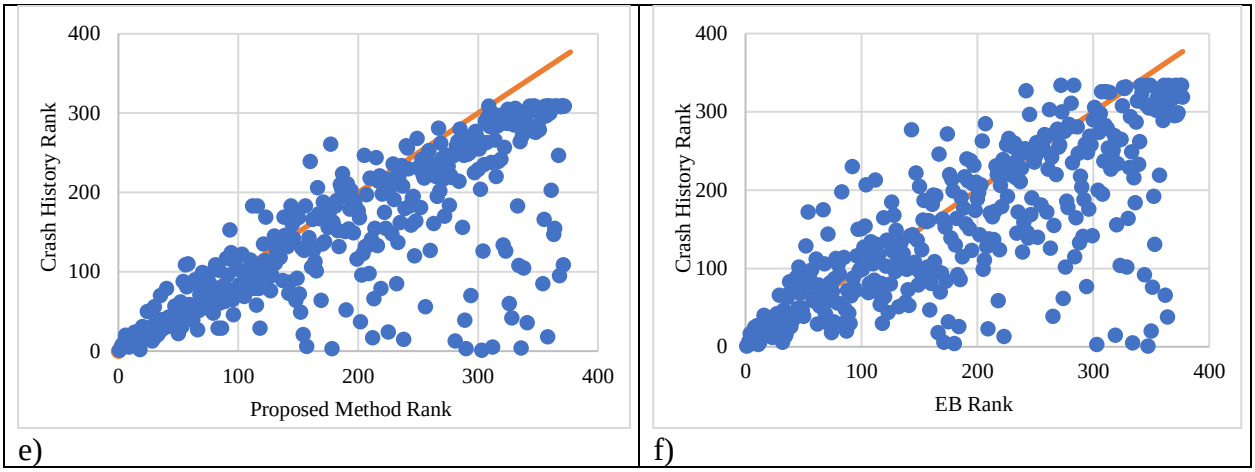


Figure 6. Rank comparison using proposed method and EB method versus crash history for different analysis periods. (a) Three-year rank comparison using proposed method versus crash history. (b) Three-year rank comparison using EB method versus crash history. (c) Five-year rank comparison using proposed method versus crash history. (d) Five-year rank comparison using EB method versus crash history. (e) Ten-year rank comparison using proposed method versus crash history. (f) Ten-year rank comparison using EB method versus crash history.

Further analyses were conducted to better understand the correlation between rankings using the proposed and the EB methods on one hand and crash history on the other hand, and the results are reported in Table 2. The table shows the values of Spearman's rank correlation coefficient along with the correlation coefficient for the total study sample as well as for subsets of sample data using upper tail 20, 40, 60, 80, and 100 sites (top ranking sites on the list). Also, the results show that the consistency of the proposed method and crash history increases with the increase in the analysis period. This is both logical and expected, given that longer analysis periods are believed to alleviate the randomness in crash data.

Table 2. Correlation coefficients for study sample for different analysis periods

(# of Segments) % of Sample Segments		Spearman's Rank Correlation Coefficient		Correlation Coefficient	
		Proposed Methodology	EB Method	Proposed Methodology	EB Method
Three-Year Analysis Period					
Upper Tail (20)	5.31%	0.820	0.881	0.812	0.876
Upper tail (40)	10.61%	0.814	0.930	0.806	0.917
Upper Tail (60)	15.92%	0.834	0.921	0.828	0.902
Upper Tail (80)	21.22%	0.825	0.871	0.807	0.852
Upper Tail (100)	26.53%	0.835	0.837	0.813	0.824
Total (377)	100%	0.787	0.794	0.760	0.766
Five-Year Analysis Period					
Upper Tail (20)	5.31%	0.801	0.792	0.784	0.787
Upper tail (40)	10.61%	0.876	0.830	0.841	0.776
Upper Tail (60)	15.92%	0.891	0.854	0.872	0.840
Upper Tail (80)	21.22%	0.884	0.814	0.875	0.792
Upper Tail (100)	26.53%	0.864	0.767	0.851	0.751
Total (377)	100%	0.821	0.801	0.811	0.786
Ten-Year Analysis Period					
Upper Tail (20)	5.31%	0.952	0.832	0.949	0.834
Upper tail (40)	10.61%	0.966	0.930	0.969	0.935
Upper Tail (60)	15.92%	0.975	0.935	0.975	0.942
Upper Tail (80)	21.22%	0.951	0.915	0.957	0.934
Upper Tail (100)	26.53%	0.916	0.867	0.924	0.884
Total (377)	100%	0.821	0.794	0.854	0.824

The results clearly demonstrate that the proposed method consistently outperforms the Empirical Bayes method across upper tail segments and the total sample, as shown in Table 2 for the five-year and ten-year analysis periods. Higher correlation coefficients are observed for the proposed methodology in all segment categories, indicating a stronger association and better alignment with the observed crash data. For example, considering upper tail (20) segments for the ten-year period, which accounts for 5.31% of sample segments, the proposed methodology yields a Spearman's rank correlation coefficient of 0.952 versus 0.832 for the Empirical Bayes method. Similarly, in the upper tail (100) segments, representing 26.53% of sample segments, the

proposed methodology yields a correlation coefficient of 0.916 compared to 0.867 for the Empirical Bayes method. Using the total study sample, Spearman's rank correlation coefficients for the proposed methodology (0.821) and the Empirical Bayes method (0.794) indicate the overall favorable performance of the two methods across all segments. The higher correlation coefficients signify a higher association between the rankings generated by the proposed method and the observed crash data.

However, the three-year study period shows that the EB method outperformed the proposed method in every upper tail group and the total sample when comparing the correlation coefficients and Spearman's rank correlation coefficients.

Table 3 presents the rank root mean square error (RMSE) and mean rank difference for the proposed and the EB methods, in reference to the observed crash history over the three different study periods. The rank RMSE and mean rank difference were estimated using upper tail segments for a total of 20, 40, 60, 80, and 100 segments. A quick examination of Table 3 indicates that, overall, as the percentage of upper tail segments increases, the mean difference between ranks and RMSE tends to increase for both methods. This trend indicates more discrepancies in rankings as the number of segments increases. For the three-year analysis period, the EB method consistently outperformed the proposed method with lower mean difference and RMSE values. The five-year analysis period still shows a better performance by the EB method compared with the proposed method; however, the discrepancies in mean difference and RMSE are not as significant and are relatively close, especially for upper tail groups 60, 80, and 100. Unexpectedly, the ten-year analysis period results showed a different trend than that of the shorter analysis periods. Specifically, the proposed method consistently

outperformed the EB method for all upper tail groups using ten years of crash and traffic data. It is worth noting that the discrepancies in mean difference and RMSE for the ten-year analysis period are notable, generally in the order of 30 to 40 percent.

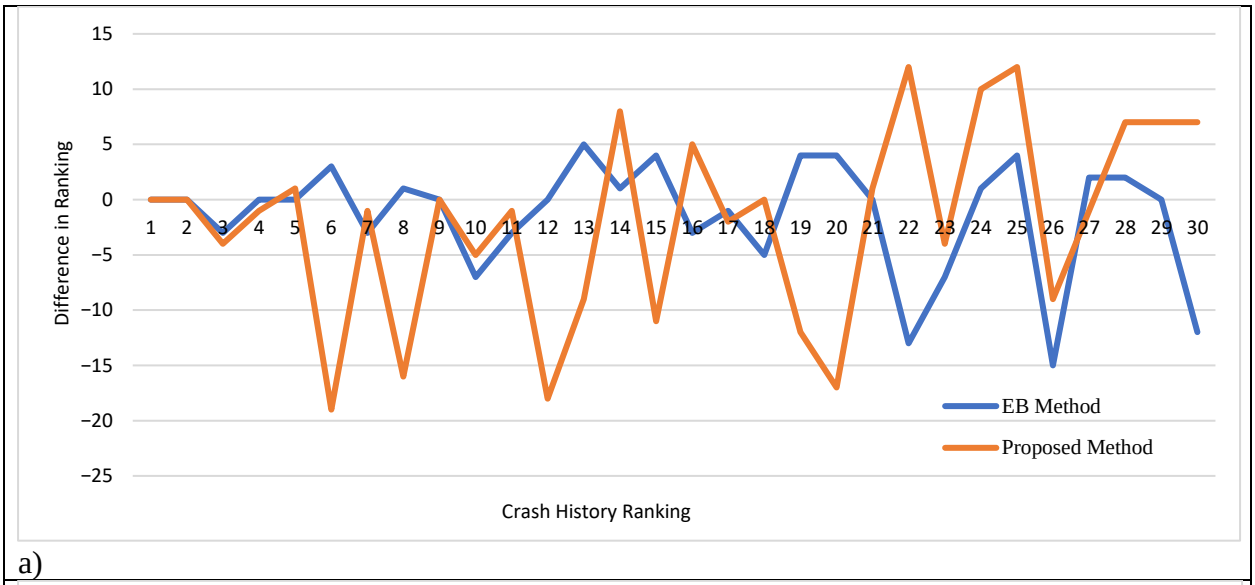
Table 3. Mean difference in ranks and root mean square error for different analysis periods

(# of Segments)	% of Sample Segments	Mean Difference in Rank		Root Mean Square Error	
		Proposed Methodology	EB Method	Proposed Methodology	EB Method
Three-Year Analysis Period					
Upper Tail (20)	5.31%	5.500	2.915	6.624	3.521
Upper tail (40)	10.61%	7.800	3.875	8.152	5.102
Upper Tail (60)	15.92%	11.750	5.617	13.245	7.655
Upper Tail (80)	21.22%	15.162	9.687	17.183	12.202
Upper Tail (100)	26.53%	18.520	13.31	20.224	18.323
Five-Year Analysis Period					
Upper Tail (20)	5.31%	4.200	2.300	5.464	3.521
Upper tail (40)	10.61%	5.175	3.225	5.875	4.102
Upper Tail (60)	15.92%	5.933	4.667	7.229	5.655
Upper Tail (80)	21.22%	9.275	8.125	9.687	9.202
Upper Tail (100)	26.53%	12.190	11.970	13.187	12.323
Ten-Year Analysis Period					
Upper Tail (20)	5.31%	5.250	7.300	6.124	9.124
Upper tail (40)	10.61%	7.475	11.000	8.412	12.142
Upper Tail (60)	15.92%	10.533	15.450	11.224	17.452
Upper Tail (80)	21.22%	12.037	19.237	13.450	21.230
Upper Tail (100)	26.53%	14.910	23.400	15.112	25.345

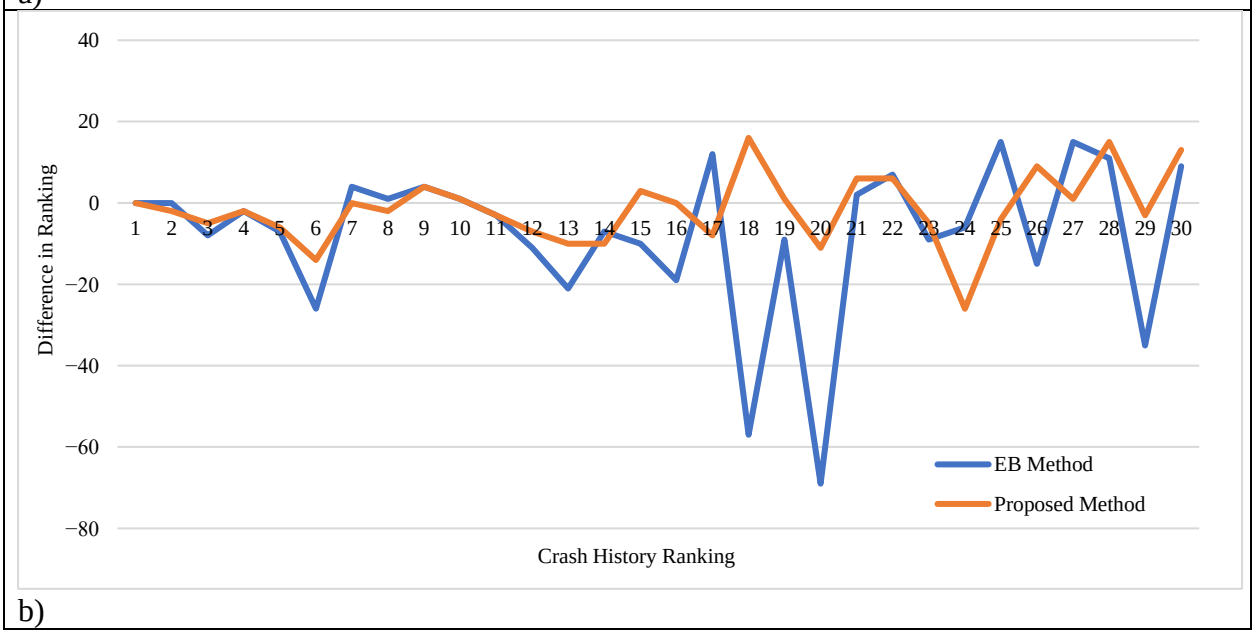
Figure 7 shows the difference in ranking between the crash history and that of the proposed and the EB methods for the first thirty highest-ranking segments in the study sample and for the three different analysis periods. The y-axis represents the difference between the ranks, while the x-axis represents the rank of the subject segment using the observed crash history.

This figure clearly shows a higher level of consistency between the proposed method ranks and the crash history ranks compared with that of the EB method for the 5-year and 10-

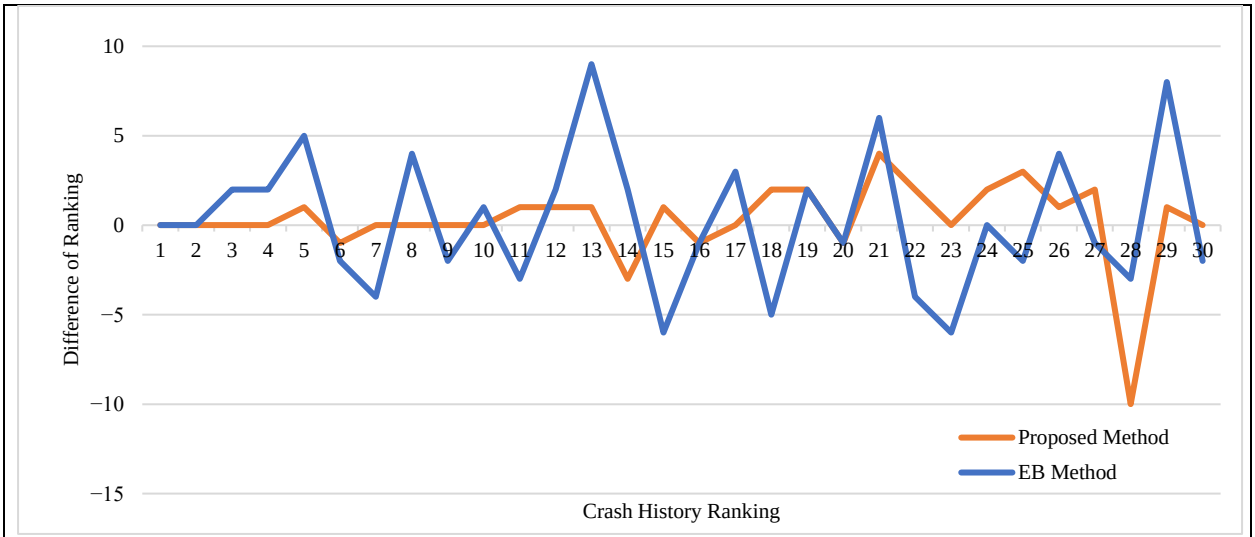
year analysis periods. For the ten-year analysis period, the difference in ranks for the proposed method was between 0 and 2 in all observations except for ranks 21, 25, and 28, as shown in Figure 7c. However, when the analysis was performed using a three-year period, the EB method outperformed the proposed method with ranks showing higher consistency with the observed crash data.



a)



b)



c)

Figure 7. Rank difference between observed crash data versus the proposed and EB methods for different analysis periods: (a) three years, (b) five years, and (c) ten years.

Table 4 presents the results of identifying true positive segments with a crash history for the proposed method and the EB method across different upper tail segment groups and analysis periods. The results are presented in terms of the number of common segments and the percentage of common segments identified by both methods.

For every analysis period, the proposed method identified more true positive sites in all upper tail segments except for the upper tail (20), where the data show less favorable performance of the proposed method. Specifically, in considering five-year and ten-year study periods, the proposed method identified 100% of the segments as identified by the crash history ranking for upper tails 60 or greater. For the upper tail (40), the proposed method identified 97.5% of true positive segments under both study periods. While the percent common segments for the EB method was 80% or higher for all upper tail segments, the proposed method consistently outperformed the EB method for being more consistent with the crash history rankings for all upper tail segment groups except for the upper tail (20).

Table 4. Number and percent of common segments with crash history for the proposed and the EB methods by various upper tail segment groups.

(# of Segments)	% of Sample Segments	# of Common Segments		% of Common Segments	
		Proposed Method	EB Method	Proposed Method	EB Method
Three-Year Analysis Period					
Upper Tail (20)	5.31%	13	19	65	95
Upper Tail (40)	10.61%	38	36	95	90
Upper Tail (60)	15.92%	54	53	90	88.33
Upper Tail (80)	21.22%	70	67	87.5	83.75
Upper Tail (100)	26.53%	88	85	88	85
Five-Year Analysis Period					
Upper Tail (20)	5.31%	13	18	65	90
Upper Tail (40)	10.61%	39	36	97.5	90
Upper Tail (60)	15.92%	60	53	100	88.33
Upper Tail (80)	21.22%	80	67	100	83.75
Upper Tail (100)	26.53%	100	85	100	85
Ten-Year Analysis Period					
Upper Tail (20)	5.31%	14	18	70	90
Upper Tail (40)	10.61%	39	38	97.5	95
Upper Tail (60)	15.92%	60	56	100	93.33
Upper Tail (80)	21.22%	80	74	100	92.5
Upper Tail (100)	26.53%	100	92	100	92

Traffic Volume Investigation

As discussed earlier, the development of the proposed method was largely based on the HSM safety performance functions and crash modification factors for rural two-lane highway segments and intersections. Therefore, the proposed method should be applicable to rural two-lane highways regardless of traffic volume. However, as the method was specifically proposed for use on low-volume roads, it is deemed valuable to evaluate the effectiveness of the proposed method for lower as well as higher traffic volumes. The results of this investigation are presented in this section.

Low-Volume Roads Analysis: In this study, low-volume roads are defined as rural two-lane highways with Average Annual Daily Traffic (AADT) equal to or less than 1000 vehicles per day. This is consistent with a few other studies that used the same definition in the literature (Ewan et al., 2016; Gross et al., 2011). Approximately a total of 850 miles representing 200 segments of state-owned and operated low-volume roads were used in this analysis.

Figure 8 shows scatterplots with rankings by the proposed method and the EB method versus the observed crash history for different analysis periods. As shown in Figure 8a, c, e for the proposed method and observed crash history, the consistency between the two rankings increases steadily with the increase in the analysis period. This is evident in the density of data points clustered around the diagonal line (the correlation coefficient is 0.742, 0.817, and 0.869 for three, five, and ten years, respectively). The general pattern shown here is consistent with that shown in Figure 6 for the whole dataset. This is due to the fact that a larger number of segments do not have crashes when shorter analysis periods are used. This is also evident in the total number of ranks for different analysis periods.

Figure (8b, d, f), which compares rankings by the EB method and observed crash data, show similar patterns with slightly more scattered around the diagonal line. Specifically, the coefficients of correlation are only 0.647, 0.734, and 0.854 for the three analysis periods, which indicate good yet less consistent relationships between the two rankings compared to those of the proposed method. The increase in correlation coefficients for longer analysis periods is similar to the case of the proposed method.

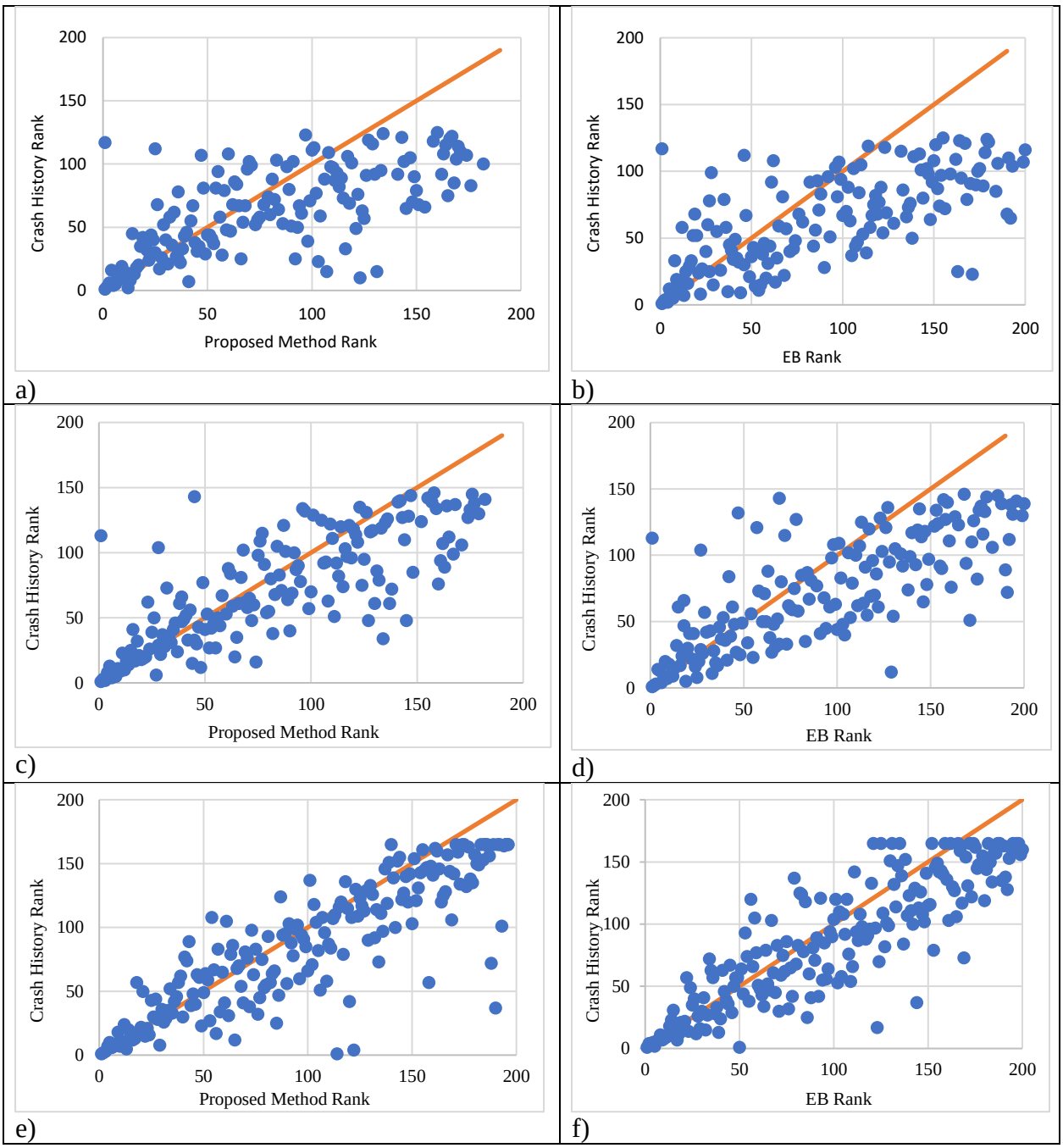


Figure 8. Scatterplots showing rankings by the proposed method and the EB method versus observed crash history—low-volume roads. (a) Three-year rank comparison using the proposed method versus crash history. (b) Three-year rank comparison using the EB method versus crash history. (c) Five-year rank comparison using the proposed method versus crash history. (d) Five-year rank comparison using the EB method versus crash history. (e) Ten-year rank comparison using the proposed method versus crash history. (f) Ten-year rank comparison using the EB method versus crash history.

Table 5 provides the Spearman's rank correlation coefficient, the correlation coefficient, the average rank difference, and the rank RMSE for the proposed and the EB methods. As shown clearly in this table, the proposed method consistently outperforms the EB method in every metric for all low-volume road segments (200 segments) and for each study period. The average difference in rankings and root mean square errors decreases with the increase in the analysis period, whereas the correlation coefficients increase with the increase in the analysis period for both methods. This was discussed earlier in discussing results for the whole dataset.

Table 5. Ranking consistency metrics for the proposed and the EB method—low-volume roads.

Metrics	Proposed Method	Empirical Bayes Method
Three-Year Analysis Period		
Spearman's Rank Correlation Coefficient	0.760	0.617
Correlation Coefficient	0.742	0.637
Average Difference in Ranking	28.415	38.912
Root Mean Square Error	29.753	40.201
Five-Year Analysis Period		
Spearman's Rank Correlation Coefficient	0.842	0.724
Correlation Coefficient	0.817	0.734
Average Difference in Ranking	21.415	31.645
Root Mean Square Error	23.753	36.311
Ten-Year Analysis Period		
Spearman's Rank Correlation Coefficient	0.853	0.847
Correlation Coefficient	0.868	0.857
Average Difference in Ranking	19.415	22.885
Root Mean Square Error	21.753	25.311

Table 6 shows the number and percentage of common segments for various upper tail segment groups for the proposed and the EB methods separately. Overall, the results for the two

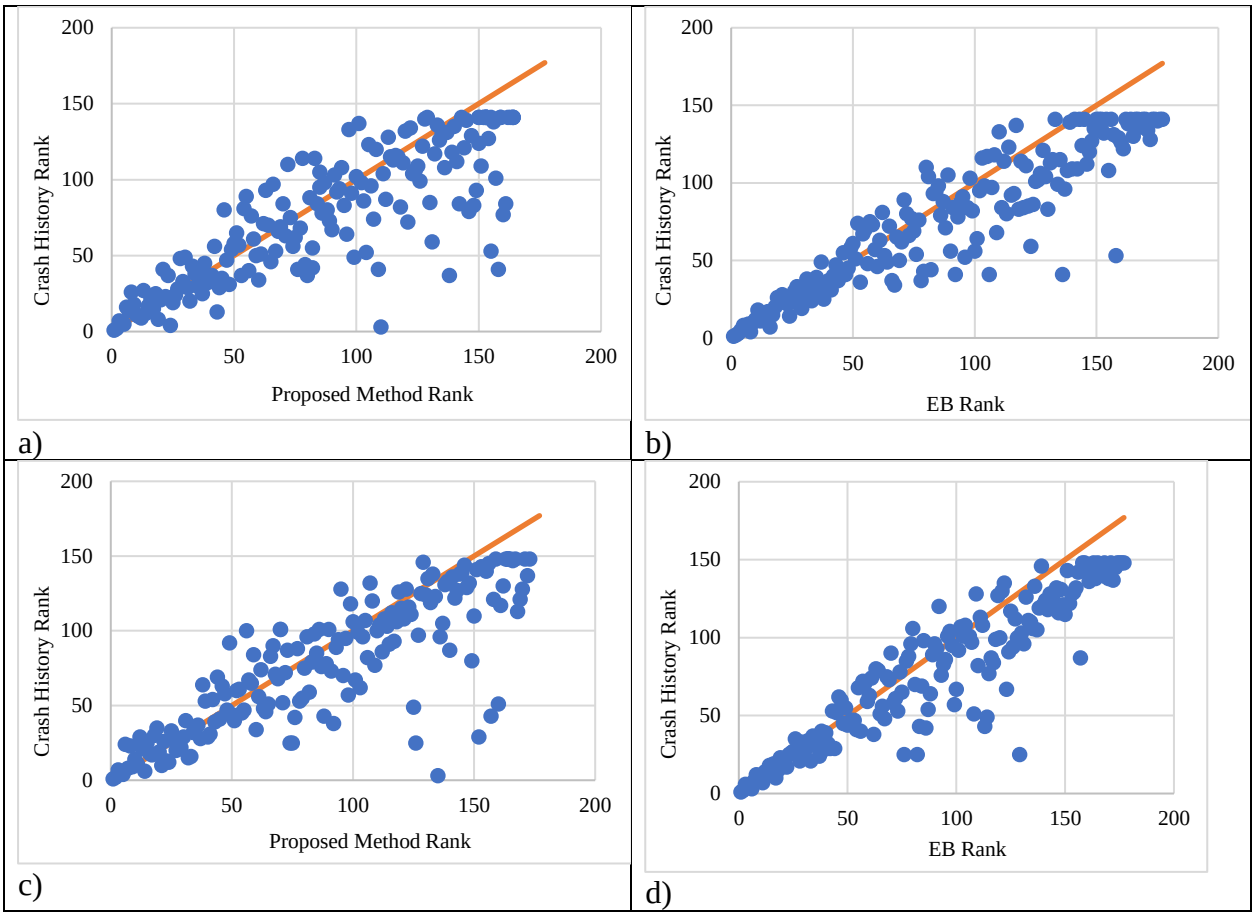
methods are comparable, with slightly better performance exhibited by the proposed method for the three analysis periods.

Table 6. Number and percent of common segments with crash history for the proposed and the EB methods for various upper tail segment groups—low-volume roads.

(# of Segments)	% of Sample Segments	# of Common Segments		% of Common Segments	
		Proposed Method	EB Method	Proposed Method	EB Method
Three-Year Analysis Period					
Upper Tail (10)	5.00%	6	6	60	60
Upper Tail (20)	10.00%	15	12	75	60
Upper Tail (30)	15.00%	19	25	63.33	60
Upper Tail (40)	20.00%	28	24	70	60
Upper Tail (50)	25.00%	39	33	78	66
Five-Year Analysis Period					
Upper Tail (10)	5.00%	8	6	80	60
Upper Tail (20)	10.00%	15	14	75	70
Upper Tail (30)	15.00%	23	21	76.67	70
Upper Tail (40)	20.00%	29	29	72.5	72.5
Upper Tail (50)	25.00%	40	38	80	76
Ten-Year Analysis Period					
Upper Tail (10)	5.00%	7	7	70	70
Upper Tail (20)	10.00%	15	16	75	80
Upper Tail (30)	15.00%	26	25	86.67	83.33
Upper Tail (40)	20.00%	33	32	82.5	80
Upper Tail (50)	25.00%	44	39	88	78

Higher Volume Roads: In this analysis, higher volume roads represent the rest of the dataset, i.e., roads with AADT greater than 1000 vehicles per day. Around 645 miles of two-lane roadways were extracted from the dataset, accounting for 177 total segments for the analysis.

The consistency of rankings between the proposed method, the EB method, and the observed crash history is shown in Figure 9. While the data points in Figure 9a, c, e show a relatively good relationship between the rankings of crash history and the proposed method, it is not as strong as that for the crash history and EB method shown in Figure 9b, d, f. This is contradictory to the trend shown in Figure 6 for low-volume roads.



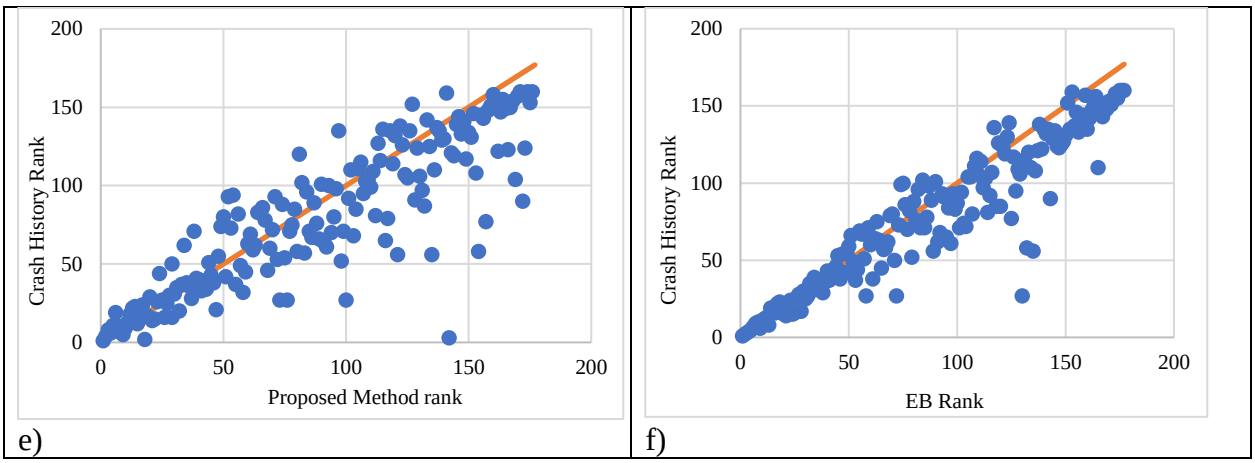


Figure 9. Scatterplots showing rankings by the proposed method and the EB method versus observed crash history—higher-volume roads. (a) Three-year rank comparison using the proposed method versus crash history. (b) Three-year rank comparison using the EB method versus crash history. (c) Five-year rank comparison using the proposed method versus crash history. (d) Five-year rank comparison using the EB method versus crash history. (e) Ten-year rank comparison using the proposed method versus crash history. (f) Ten-year rank comparison using the EB method versus crash history.

Table 7 provides Spearman's rank correlation coefficient, correlation coefficient, average rank difference, and rank RMSE for the proposed and the EB methods using rankings by crash history as a reference for three analysis periods. This table clearly shows that for the higher-volume roads, the EB method rankings are more consistent with those of the observed crash history than the rankings of the proposed method. This is contrary to the respective trend discussed in the low-volume road analysis. However, it should be mentioned that the correlation coefficients still suggest a relatively strong relationship between the proposed method ranks and those of the observed crash history.

Table 7. Ranking consistency metrics for proposed and the EB methods—higher volume roads.

Metrics	Proposed Method	Empirical Bayes Method
Three-Year Analysis Period		
Spearman's Rank Correlation Coefficient	0.844	0.889
Correlation Coefficient	0.840	0.917
Average Difference in Rank	19.186	17.446
Root Mean Square Error	22.745	20.355
Five-Year Analysis Period		
Spearman's Rank Correlation Coefficient	0.841	0.903
Correlation Coefficient	0.848	0.927
Average Difference in Rank	18.723	16.969
Root Mean Square Error	21.745	18.169
Ten-Year Analysis Period		
Spearman's Rank Correlation Coefficient	0.866	0.924
Correlation Coefficient	0.871	0.941
Average Difference in Rank	17.875	13.536
Root Mean Square Error	19.745	15.355

The number and percentage of common segments for the proposed and the EB methods with those of the observed crash history for the three different analysis periods are provided in Table 8. As shown in this table, the EB method outperforms the proposed method in all upper tail segment groups. While this may initially seem unexpected, it can easily be explained with a proper understanding of the EB method formulation. Specifically, the contribution of crash history in the EB expected number of crashes increases with traffic volume, and thus the correlation between crash history and the EB expected number of crashes. This explains why results for the low-volume roads and higher-volume roads are somewhat different.

Table 8. Number and percent of common segments with crash history for the proposed and the EB methods for various upper tail segment groups -higher volume roads.

(# of Segments)	% of Sample Segments	# of Common Segments		% of Common Segments	
		Proposed Method	EB Method	Proposed Method	EB Method
Three-Year Analysis Period					
Upper Tail (10)	5.60%	5	9	50	90
Upper Tail (20)	11.20%	15	18	75	90
Upper Tail (30)	16.90%	25	27	83.33	90
Upper Tail (40)	22.60%	33	36	82.5	90
Upper Tail (50)	28.40%	42	44	84	88
Five-Year Analysis Period					
Upper Tail (10)	5.60%	7	8	70	80
Upper Tail (20)	11.20%	13	18	65	90
Upper Tail (30)	16.90%	26	26	86.67	86.67
Upper Tail (40)	22.60%	36	37	90	92.5
Upper Tail (50)	28.40%	42	45	84	90
Ten-Year Analysis Period					
Upper Tail (10)	5.60%	8	9	80	90
Upper Tail (20)	11.20%	15	16	75	80
Upper Tail (30)	16.90%	26	29	86.67	96.67
Upper Tail (40)	22.60%	34	38	85	95
Upper Tail (50)	28.40%	44	46	88	92

Summary

Study results showed an overall strong performance of the proposed method in identifying candidate sites for safety improvements. Given the simplicity and limited data requirements of the proposed method, it can provide a viable option for network screening in cases where the use of the well-established EB method is deemed impractical for lack of accessible data or technical expertise (e.g., roads owned and operated by local agencies such as counties, townships, and tribal governments).

The proposed method outperformed the well-established EB method when the whole study sample was used in the analysis. However, when the study sample was split into two datasets based on traffic volume, the proposed method outperformed the EB method for low-volume roads, while the EB method was found more effective for higher-volume roads. This can be explained by the fact that the contribution of crash history into the EB expected number of crashes increases with traffic volume, thus the more consistency between the two. Therefore, the use of the proposed method, while overall effective, can be particularly valuable for use on low-volume roads for its better performance and ease of implementation, particularly by local agencies.

The proposed screening method was originally developed for the application on low-volume roads. Nevertheless, the performance has been assessed for low and higher traffic volumes, extending up to 4000 ADT (Annual Daily Traffic). It is worth noting that for exceptionally high traffic volumes, the effectiveness of the proposed method may be limited.

CHAPTER SIX

VALIDATION OF METHOD II: EMPIRICAL BAYES PREDICTION METHOD

Overview of the Empirical Bayes (EB) Prediction Method

A new network screening approach was recently proposed for use on rural two-lane highways (Huda and Al-Kaisy, 2024). The main merit of the proposed approach is that it can easily be implemented by local agencies lacking access to detailed databases and technical expertise. Specifically, the more sophisticated network screening methods (e.g., Empirical Bayes method) use exact values for various roadway characteristics (sometimes referred to as risk factors). While the use of exact values may improve the accuracy of the screening process, it requires access to extensive databases or on-site detailed measurements which are typically beyond the re-sources available for small local agencies. Hence, implementing such sophisticated methods might not be feasible for most local agencies. To minimize data requirements, the proposed approach employs classified variables that can easily be compiled by local agencies without the need to access detailed and extensive databases. Using the classified variables, the proposed method consists of regression models that are developed using the EB expected number of crashes to predict the level of risk (or safety) of roadway segments that are part of the roadway network. Two models were proposed, one with and one without traffic data. The response (dependent) variable in both models was the EB expected number of crashes, which is a function of the HSM predicted number of crashes and the observed number of crashes. The explanatory (independent) variables included roadway and roadside characteristics

besides traffic exposure (AADT). Table 9 shows the explanatory variables for the proposed models. The development of the mathematical model for this study using the proposed approach is discussed later in the methodology section.

Table 9. Explanatory variables of proposed model (Huda and Al-Kaisy, 2024)

Risk Factors	Approximate Ranges of Variables	Categories	Terms
Segment Length (SL)	Exact Length		
Lane Width (LW)	LW < 11	1	Narrower
	LW ≥ 11	2	Wider
Shoulder Width (SW)	SW < 1.8	1	Narrower
	SW ≥ 1.8	2	Wider
Degree of Horizontal Curvature (DC)	DC = 0	0	Straight
	DC < 10	1	Mild
	10 ≤ DC < 27	2	Moderate
	DC ≥ 27	3	Sharp
Grade (G)	G < 4	0	Mild
	G ≥ 4	1	Steep
Driveway Density (DD) (driveways per mile)	Exact Number		
Side Slope (SS)	Steep	1	Steep
	Moderate	2	Moderate
	Flat	3	Flat
Fixed Objects (FO)	Many	1	Many
	Some	2	Some
	Few	3	Few
Volume (V)	Exact Volume		

Methodology

The proposed approach was implemented using the study data and evaluated using observed crash history collected for the study sample. Specifically, the crash data collected over ten years were divided into two datasets: a training dataset spanning five years (2011–2015) to develop the regression model, and a testing dataset covering the subsequent five-year period (2016–2020). The approach that was used in evaluating the proposed network screening method

is to compare site rankings using the proposed method with the respective rankings using crash history data. Both crash frequency and crash density (number of crashes per mile) were used in ranking sites. Further, rankings from the Empirical Bayes (EB) method and the potential for safety improvement (PSI) method were included in the analyses to see how the performance of the proposed method compares to that of the two well-established techniques. Using the EB and PSI in this study is based on the fact that these techniques have become the gold standard in performing quantitative safety analyses, with merits and effectiveness well established in the literature (Cafiso et al., 2010; Elvik, 2008b; Hauer, 2002; Hauer et al., 1991; Manepalli and Bham, 2016, 2011; Persaud and Lyon, 2007).

Model Development

Multivariate linear regression analysis was used in developing the model for road network screening. The open-source statistical software R 4.3.1 was used for running the analysis. The model with different explanatory variables is shown in Equation (9).

$$Ln\,Exp = -8.2709 + 0.7292 * \frac{1}{FO} + 0.02717 * DD + 0.98309 * Ln\,(V) + 0.10126 * DC + 0.94290 * Ln\,(SL) \quad (9)$$

Where exp = EB expected number of crashes; SL = segment length in miles; V = traffic volume (AADT); FO = fixed object; DD = driveway density; and DC = degree of curvature.

This model has a coefficient of determination (adjusted R-squared value) of 0.932. This indicates that the model can explain about 93 percent of the variability of the EB expected number of crashes, which is relatively high given the classified format used for most of the variables in this model. All variable coefficients are found significant at the 95% confidence level. The regression output using R 4.3.1 statistical software is shown in Figure 10.

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -8.270493   0.195408 -42.324  < 2e-16 ***
`1/FO`       0.729629   0.186246   3.918 0.000106 ***
DD           0.027170   0.003241   8.384 1.07e-15 ***
`Ln (V)`     0.983096   0.021492  45.743  < 2e-16 ***
DC           0.101260   0.019409   5.217 3.03e-07 ***
`Ln (SL)`    0.942908   0.017038  55.340  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3178 on 371 degrees of freedom
Multiple R-squared:  0.9325,    Adjusted R-squared:  0.9316
F-statistic:  1025 on 5 and 371 DF,  p-value: < 2.2e-16

> coef(regression_models)
(Intercept)      `1/FO`          DD    `Ln (V)`          DC   `Ln (SL)`
-8.27049313  0.72962895  0.02717009  0.98309595  0.10126046  0.94290773
```

Figure10. R-output of regression model

Validation Analysis and Results

This section presents the discussion of the metrics used for the evaluation of the performance of the proposed method. The performance of the EB method and the PSI method is also calculated and compared with the proposed method.

Spearman Correlation Coefficient

Figure 11 shows a scatterplot with site rankings using the proposed method, EB method, and PSI method versus crash history. An examination of Figure 11a, which shows the proposed method versus the crash history rank, reveals that the data points are spread around the diagonal line. The discrepancy between the ranks increases with the increase in rank. The tightness of the data points around the line and the correlation coefficient of 0.686 indicates a moderate correlation between the ranks using the two methods. A perfect correlation is a diagonal line passing through the origin.

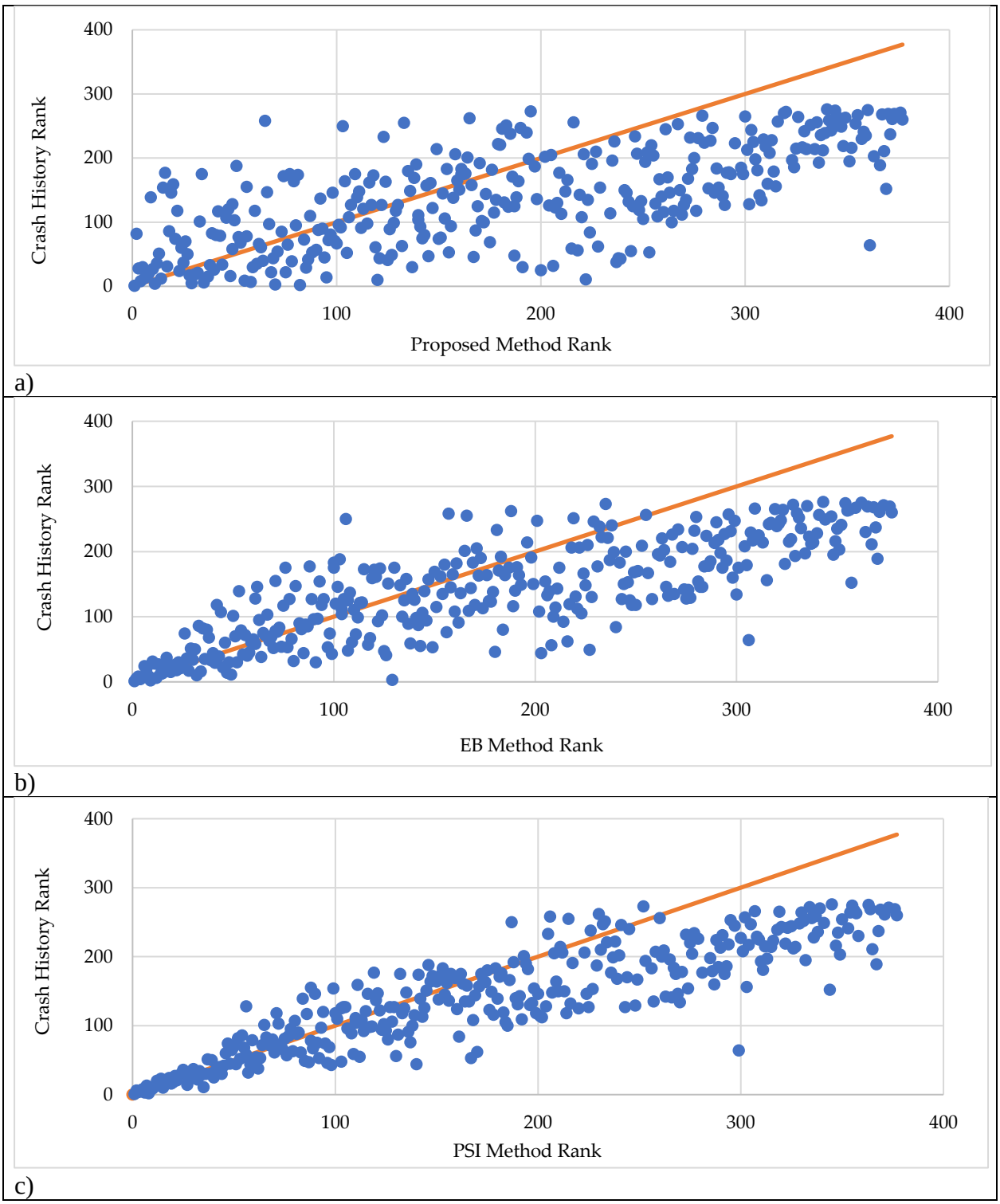


Figure 11. Rank comparison using different network screening methods versus crash history. (a) proposed method versus crash history rank; (b) EB method versus crash history rank; and (c) PSI method versus crash history rank.

The EB method ranking was also compared with the ranking using the observed number of crashes and the results are shown in Figure 11b. In this figure, the proximity of data points to the line passing through the origin (diagonal line) indicates that it is more compact compared to that of the proposed method. This observation is confirmed by a higher correlation coefficient (r = 0.816) between the two rankings.

Additionally, figure 11c shows the comparison of ranking using PSI method versus the ranking using the observed number of crashes. The data points clustered around the diagonal line and the correlation coefficient of 0.879 indicate the very strong association between the ranks.

Table 10 presents the Spearman's rank correlation coefficient values for three different methods (proposed method, Empirical Bayes method, and PSI method) based on crash density and crash frequency in different upper tail segments and for the total sample size of 377 segments. In analyzing the crash density for the upper tail segments (20, 40, 60, 80, 100), the proposed method shows moderate to strong positive correlations with crash history, with values ranging from 0.587 to 0.744. The Empirical Bayes method and PSI method consistently exhibit higher correlations, ranging from 0.786 to 0.828 and 0.954 to 0.959, respectively. For the total sample size of 377 segments, the proposed method exhibits a moderate positive correlation (0.683), while the Empirical Bayes method (0.816) and PSI method (0.879) display stronger positive correlations. This shows that, in analyzing crash densities, the Empirical Bayes and PSI methods outperform the proposed method at identifying safety improvement sites.

Table 10. Comparative analysis of rank correlation for proposed method versus the EB method versus the PSI method.

Group/Method	Spearman's Rank Correlation Coefficient		
	Proposed Method	EB Method	PSI Method
Crash Density-Based Analysis			
Upper Tail (20)	0.744	0.815	0.958
Upper Tail (40)	0.587	0.786	0.954
Upper Tail (60)	0.638	0.813	0.959
Upper Tail (80)	0.655	0.828	0.959
Upper Tail (100)	0.605	0.825	0.958
Total Sample (377)	0.683	0.816	0.879
Crash Frequency-Based Analysis			
Upper Tail (20)	0.502	0.406	0.738
Upper Tail (40)	0.562	0.768	0.878
Upper Tail (60)	0.65	0.84	0.893
Upper Tail (80)	0.703	0.823	0.87
Upper Tail (100)	0.702	0.829	0.867
Total Sample (377)	0.815	0.928	0.942

In the case of crash frequency, for the upper tail segments (20, 40, 60, 80, 100), the proposed method shows weak to moderate positive correlations, ranging from 0.502 to 0.703. The Empirical Bayes method exhibits overall stronger positive correlations, varying in a broad range from 0.406 to 0.840, while the PSI method consistently demonstrates the highest correlations, ranging from 0.738 to 0.893. In the total sample, the proposed method shows a relatively strong positive correlation (0.815), whereas the Empirical Bayes method (0.928) and PSI method (0.942) display even stronger positive correlations with crash history. This indicates a higher association between all three methods and the crash history when the total sample is analyzed. The difference in performance between the proposed method and the other existing methods can be explained by properly understanding the formulation of the EB and the PSI methods. Specifically, observed crash history is a major contributor to the EB and the PSI

methods, which explains the higher correlation between crash history and these methods. On the other hand, observed crash history is not an in-put to the proposed method.

True Positive Identification

Table 11 shows a comparison of the identification of true positive segments by three different methods—the proposed method, the Empirical Bayes (EB) method, and the PSI method—while using observed crash data as a reference in the context of crash density and crash frequency analysis for various upper tail segments.

When evaluating the identification of common segments using crash density, for upper tail segments (20), the proposed method was least consistent with crash history (30%), followed by the EB method (65%) and the PSI method (90%), respectively. With the increase in upper tail segments, the discrepancy between the proposed method and the EB and PSI methods decreases. However, the performance of the proposed method is overall less favorable compared to that of the other two methods. For example, in upper tail segments (100), the proposed method identified 68 sites versus 78 sites for the EB method and 87 sites for the PSI method.

When using crash frequency in the analysis, show better performance for the proposed method and lower discrepancy between the three methods. For instance, the pro-posed method identified 78% of true positive segments compared to 86% by the EB method and 88% by the PSI method when considering 100 upper tail segments. Further, for every other upper tail considered, the result from the proposed method shows lower discrepancy with those using the other two methods.

Table 11. True positive segment identification comparison for proposed method versus the EB method versus the PSI method.

Groups	% of segments	# of Common Segments			% of Common Segments		
		Proposed Method	EB Method	PSI Method	Proposed Method	EB Method	PSI Method
Crash Density Analysis							
Upper Tail (20)	5.31%	6	13	18	30	65	90
Upper Tail (40)	10.61%	22	31	37	55	77.5	92.5
Upper Tail (60)	15.92%	30	45	51	50	75	85
Upper Tail (80)	21.22%	46	60	64	57.5	75	80
Upper Tail (100)	26.53%	68	78	87	68	78	87
Crash Frequency Analysis							
Upper Tail (20)	5.31%	12	17	17	60	85	85
Upper Tail (40)	10.61%	30	34	34	75	85	85
Upper Tail (60)	15.92%	41	49	50	68.33	81.67	83.33
Upper Tail (80)	21.22%	61	68	70	76.25	85	87.5
Upper Tail (100)	26.53%	78	86	88	78	86	88

Precision and Percent Deviation

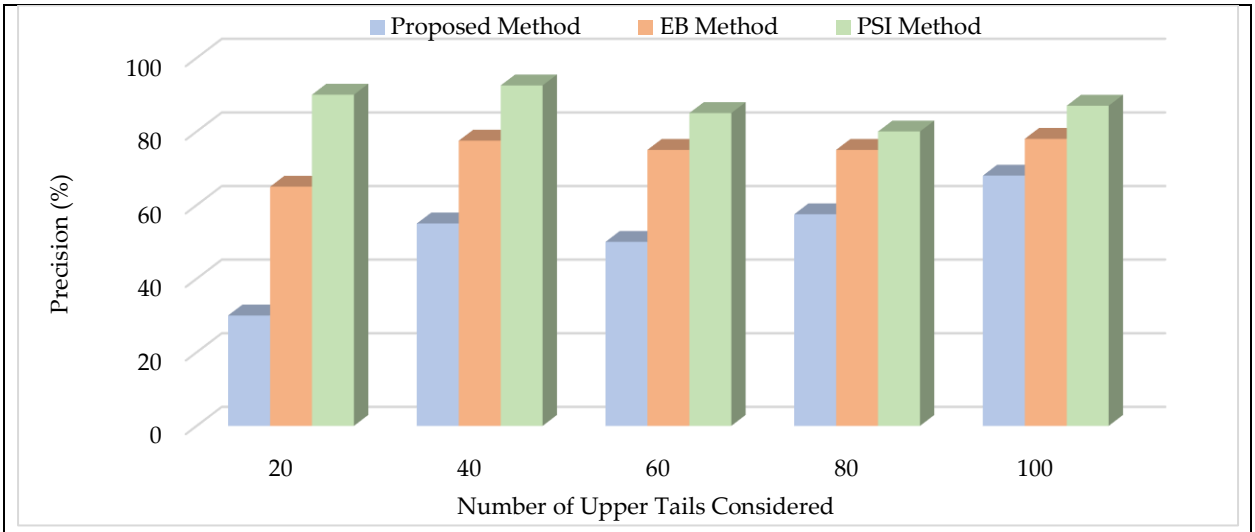
The precision and percent deviation from crash history are calculated for various upper tail segment groups (20, 40, 60, 80, and 100) for the proposed method and the performance is compared with that of the EB and the PSI methods. A higher percentage of precision and lower value of percent deviation indicates better performance.

$$Precision = \frac{TP}{TP+FP} * 100\% \quad (10)$$

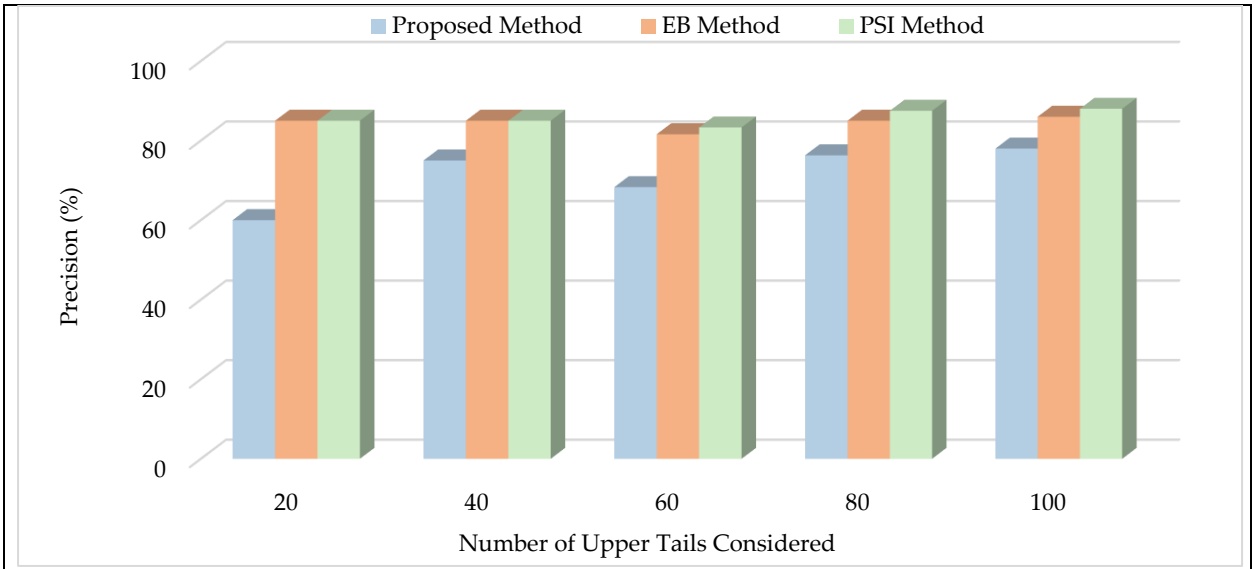
$$Percent\ Deviation = \left(1 - \frac{TP}{M}\right) * 100\% \quad (11)$$

Where TP, true positive, is the number of sites correctly identified as safety improvement sites by the evaluating method; FP, false positive, is the number of sites falsely identified as safety improvement sites by the evaluating method (not listed in crash history); and M is the number of upper tail segments that are considered.

Figure 12 shows precision across various upper tail groups (20, 40, 60, 80, and 100). Precision is also calculated for the EB method, and the PSI method to see how the performance of the proposed method compares to that of the other two methods. The precision percentage based on crash density analysis is shown in Figure 12a. The figure shows that the precision value is less for the proposed method when compared with the EB and the PSI methods. The value for the proposed method varies from 30% to 68%, versus 62% to 76% for the EB method and 85% to 90% for the PSI methods. On the other hand, the precision percentage based on crash frequency shown in Figure 12b indicates that the precision value varies from 60% to 78% for the proposed method versus 81.6% to 86% for the EB method and 83.3% to 88% for the PSI method. This confirms that both the EB and PSI methods were more consistent with the crash history compared with the proposed method. Further, the proposed method exhibited better performance and higher consistency with crash history when analyzed using crash frequency over crash density.



a)



b)

Figure 12. Precision comparison using proposed method, EB method, and PSI method versus crash history: (a) crash density-based analysis; (b) crash frequency-based analysis.

Figure 13 shows the percent deviation from the crash history for the proposed method, the EB method, and the PSI method. A quick examination of Figure 13 clearly shows higher deviation for the proposed method compared to the EB and the PSI methods both for crash density and crash frequency analyses. The other trend clearly exhibited in this figure is that the percent deviation for the proposed method decreases with the increase in the number of upper tail segments. Specifically, the percent deviation of 70% for density-based analysis and 40% for frequency-based analysis were observed for the 20 upper tail segments. However, the corresponding percent deviation for the 100 upper tails is 32% and 22% for density-based and frequency-based analyses, respectively. The EB and the PSI methods, while generally showing less deviation and better performance, exhibited less variation in percent deviation with the increase in upper tail segments.

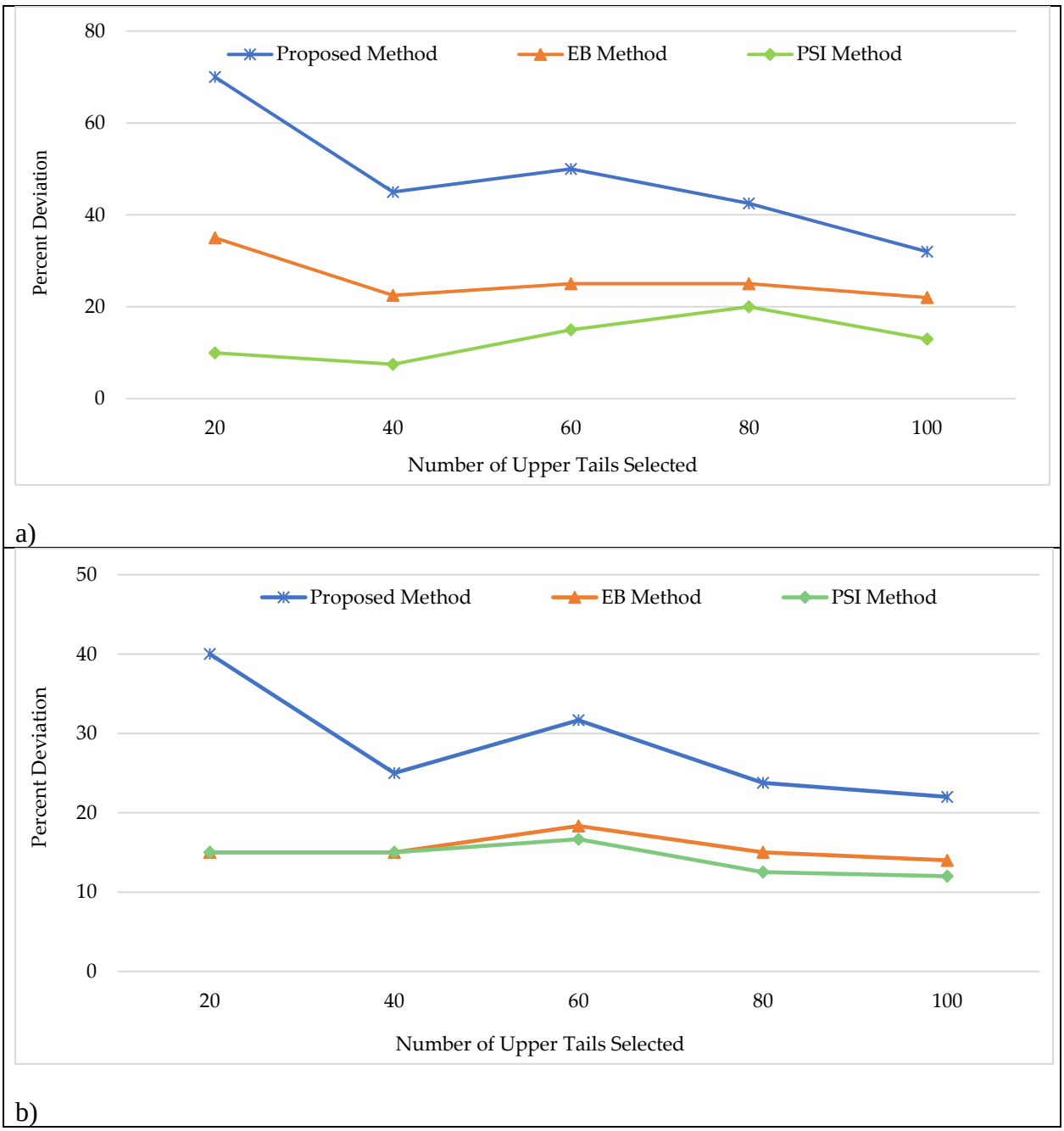


Figure 13. Percent deviation comparison using proposed method, EB method, and PSI method versus crash history: (a) crash density-based analysis; (b) crash frequency-based analysis.

Treatable Crashes

The last analysis in the evaluation of the proposed method involved the use of the number of treatable crashes (calculated for five years: 2016–2020) as the performance measure. Treatable

crashes are defined as those crashes that may be treated by engineering countermeasures such as lane widening. The remaining crashes constitute what might be expected due to traffic exposure and may not be treatable by engineering countermeasures. A reasonable estimate for the number of treatable crashes at a location is the number of crashes in excess of what would normally be expected at locations with similar characteristics and traffic exposure (Persaud et al., 1999). To assess the degree of similarity between the segments identified by a method and those identified by the crash history, a segment overlap metric was used. The higher the percentage of segment overlap and the closer the estimate of treatable crashes with crash history, the better the performance of the compared method.

Table 12 shows the summary of the estimate of treatable crashes for the proposed method the EB method, and crash history. The table also provides the ratio and overlap for treatable crashes in the top 100 segments defined earlier using both crash density and crash frequency.

Table 12. Treatable crashes for the top 100 segments identified.

Methods	Total Length (mi)	Total Crashes (2016–2020)	Treatable Crashes (2016–2020)	
			Estimate	Overlap
Crash Density-Based Analysis				
Proposed Method	260.55	565	469	66%
EB	265	614	522	81%
Crash History	241.1	805	726	-
Crash Frequency-Based Analysis				
Proposed Method	807.45	2220	1147	79%
EB	787.05	2144	1152	85%
Crash History	805.75	2439	1299	-

The table shows that, based on crash density, the proposed method identified a fewer number of treatable crashes for the top 100 segments compared to crash history (469 versus 726, respectively). The segment overlap was 66%. Using crash frequency, the segment overlap increased to 79% indicating a significant improvement in performance. Regarding the EB method, the number of treatable crashes and overlap indicate a higher level of consistency with crash history. However, the discrepancy in performance between the pro-posed and the EB methods is less evident when crash frequency is used in the analysis. Specifically, there were only 11.6% fewer treatable crashes on the segments identified with the proposed method than those identified with the crash history, compared to 11.3% fewer treatable crashes using the EB method. Further, 79% of all segments identified by the proposed method were also on the top 100 segments list of the observed crash.

In addition to evaluating the top 100 segments identified by the methods, similar calculations were performed for other possible levels of selection up to upper tails having positive treatable crashes. Plots of the treatable crashes identified at various levels are shown in Figure 14. For the range of top segments studied, the chart shows relatively close lines for the proposed and the EB methods, with the EB method consistently identifying a slightly higher number of treatable crashes. The figure also suggests that the larger the total segments selected, the closer the proposed method is in terms of selection efficiency compared to that of EB method. These patterns are similar for the crash density and crash frequency analyses.

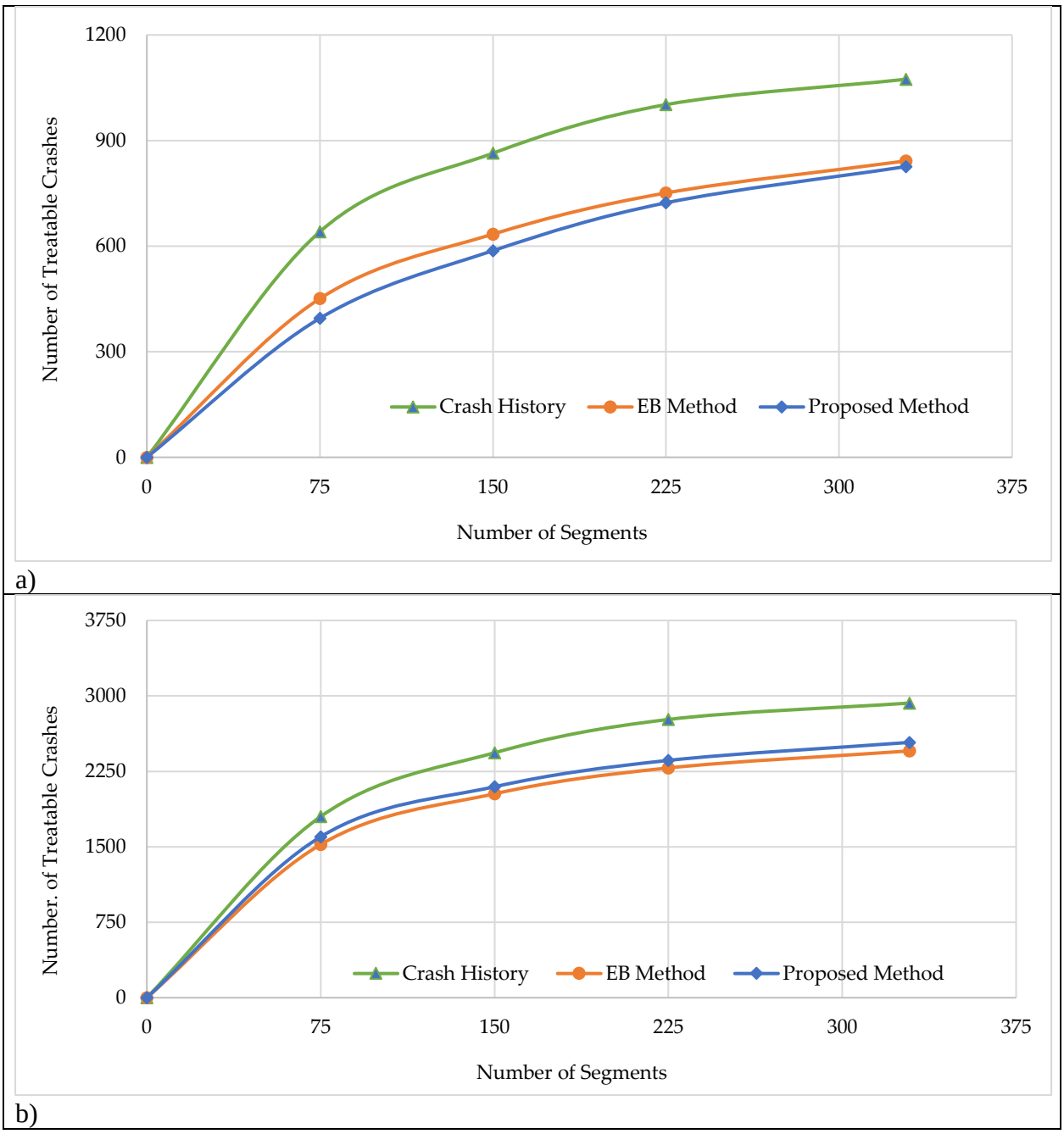


Figure 14. Treatable crash count by different methods and total number of segments selected: (a) crash density-based analysis; (b) crash frequency-based analysis.

Summary

The research presented in this chapter examined a new proposed approach for identifying safety improvement sites on rural highways. The proposed approach involves the use of EB prediction models using traffic and roadway variables only. Specifically, the multiple linear regression was used to develop a mathematical model to predict the EB expected number of crashes. The proposed model utilizes classified variables for roadway and roadside characteristics that do not require detailed databases or extensive technical expertise. The evaluation was performed using a dataset comprising 1495 miles of rural two-lane highway segments in the state of Oregon. The data used in the evaluation included roadway geometry, roadside features, traffic conditions, and ten years of crash records. A training dataset spanning five years (2011–2015) was employed to develop the model, while the subsequent five-year period (2016–2020) served as the basis for evaluating the model's performance in network screening.

The study findings suggest that using crash density for highway segments, the performance of the proposed method was fair and not as effective as the well-established EB and PSI methods. This is despite the high R-square value of the predictive model used by the proposed approach. However, when using crash frequencies for highway segments, the performance of the proposed method was found to be comparable to the EB and PSI methods. Moreover, the results suggest that, in identifying segments most likely to have treatable crashes, the proposed method outperformed the EB method for higher upper tail segments considered in the analysis. Despite the overall lower performance of the proposed method, it provides a

valuable tool to local agencies where the application of the more sophisticated methods is deemed impractical for lack of resources.

To promote the use of the proposed method in practice, further evaluation using highway networks from other regions and states may reinforce the evidence about the effectiveness of the proposed method in identifying safety improvement sites. Further, evaluating the proposed approach at rural highway intersections would make the use of the proposed method more attractive to local agencies by applying the same approach to various highway network components (i.e., segments and intersections).

CHAPTER SEVEN

VALIDATION OF METHOD III: THE CRASH RISK INDEX (CRI) METHOD

Overview of the Crash Risk Index Method

A Crash Risk Index (CRI) was developed and proposed to serve as a multi-criteria risk assessment tool, integrating three key elements: geometric and roadside features, crash history, and traffic exposure (Al-Kaisy et al., 2019). The CRI is defined by the equation below.

$$CRI = W_G * X_G + W_C * X_C + W_T * X_T \quad (12)$$

where, CRI is crash risk index, $W_G \; and \; X_G$ are the weight and numerical score of geometric and roadside feature, $W_C \; and \; X_C$ are the weight and score for the contribution of the crash history, and $W_T \; and \; X_T$ are the weight and score to the contribution of traffic exposure. The weights used in the CRI formula should reflect agency priorities and preferences and can be calibrated as per the need.

The geometric and roadside features score (X_G) , which contributes to increased crash risk for the CRI calculation is estimated using the formula shown below:

$$X_G = W_{dc}(y_{dc}) + W_{lvc}(y_{lvc}) + W_{lw}(y_{lw}) + W_g(y_g) + W_{sw}(y_{sw}) + W_{dd}(y_{dd}) + W_{ss}(y_{ss}) + W_{fo}(y_{fo}) \quad (13)$$

Where, dc is degree of curvature, lvc is length of vertical curve, lw is lane width, g is the vertical curve grade, sw is shoulder width, dd is driveway density, ss is the side slope rating, fo is the fixed object rating, W is the weight, and y is the numerical scores.

The equations for calculation of numerical scores and weight of geometric and roadside features are not provided in this section, and interested readers are encouraged to check reference for details (Al-Kaisy et al., 2019).

The crash history risk index values (X_C) are based on the average crash rate and the critical crash rate for the sample rural roads as defined in the Highway Safety Manual (AASHTO, 2010). Table 13 provides (X_C) equations using the crash rates.

Table 13. Crash history risk index score equations

Equation	**Condition**
$\boldsymbol{X_C}$ = 0.502 * crash rate – 0.266	0.53 ≤ crash rate ≤ 2.52
$\boldsymbol{X_C}$ = 0.00	crash rate < 0.53
$\boldsymbol{X_C}$ = 1.00	Crash rate > 2.52

The proposed CRI considers two aspects of traffic exposure in estimating the exposure risk index values: total vehicular traffic volume expressed as ADT and the percentage of heavy vehicles in the traffic stream (AASHTO class 4 or larger). The traffic exposure score X_T is provided in table 14.

Table 14. Traffic exposure numerical scores

AADT (vpd)	**Percent heavy vehicles**		
	< 29%	**29 - 39%**	**>39%**
<300	0.20	0.30	0.40
300 - 499	0.40	0.50	0.60
500 - 699	0.60	0.70	0.80
700 - 900	0.80	0.90	1.00
>900	1.00	1.00	1.00

Note: AADT = average annual daily traffic, vpd = vehicles per day

Methodology

The determination of weights within the Crash Risk Index (CRI) formula can be calibrated with agency priorities and preferences. The weights calibrated in this study utilizing a five-year dataset spanning from 2011 to 2015, is summarized below:

Crash History Weight Calculation

The weight attributed to crash history was found through the calculation of the Empirical Bayes (EB) average weight for the predicted number of crashes within the sample segments. This weight (denoted as 'w') is computed using the formula:

$$w = \frac{1}{1+k*(N_{predicted})} \quad (13)$$

Where, w = Weight for predicted number of crashes, k = Overdispersion parameter associated with the specific SPF, and $N_{predicted}$ = HSM predicted number of crashes.

After calculation, the weight assigned to the crash history was found to be 50%.

Traffic Exposure and Roadway Geometry Weight Calculation

The weights assigned to traffic exposure and roadway geometry were determined by analyzing the correlations between their respective scores and the crash rate for every roadway segment considered. With 50% of the weight allocated to the crash rate score, the remaining 50% weight distribution was calculated as 24% for roadway geometry and 26% for traffic exposure.

The performance of the EB method was also calculated and compared with the proposed method. Two different analysis periods were used to check the consistency of the performance of the proposed method for network screening: three years and five years, as these study periods are

consistent with the current practice in safety analyses as found in many studies in the literature (Montella, 2010; Persaud and Lyon, 2007). The crash rate was determined using both the estimated number of crashes from EB data and historical crash data. Subsequently, a ranking was assigned based on this crash rate. This standardized approach ensured that all considered methods (proposed, EB, and crash history) were evaluated using the same parameters. The crash rate on a roadway was determined using following equation:

$$R = \frac{C*10^6}{L*V*365*N} \quad (14)$$

Where, R = crash rate expressed as crashes per million vehicle miles travelled.

C = total number of observed crashes or EB expected number of crashes in study period.

L = length of roadway segments in miles.

V = traffic exposure or Annual Average Daily Traffic (AADT).

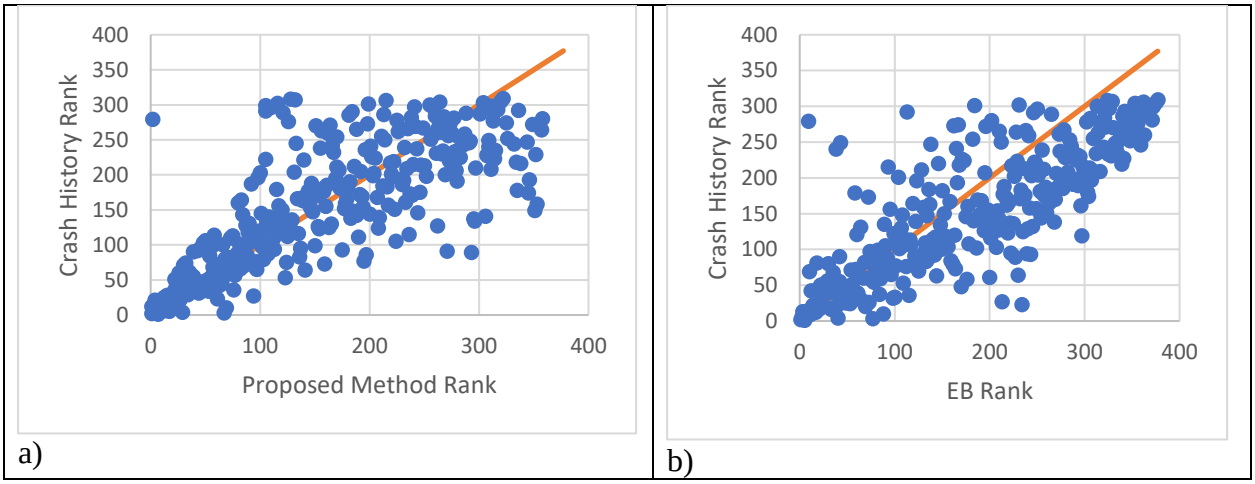
N = number of years of data.

Validation Analysis and Results

Figure 15 shows scatterplot with site rankings using the proposed method, and the EB method versus crash history. A careful examination of figures 15a and 15c, which is proposed method versus crash history rank for 3-years analysis and 5-years analysis respectively, reveals that for most roadway segments the data points correspond well with each other. This is particularly evident in sites that ranked high on the list (first hundred ranks). The discrepancy between the ranks increases with the increase in ranks. There are many segments which are expected to have a high crash based on the proposed method. But the crash history data shows

that they have zero (or very few) observed crashes. This is evident as these segments are ranked much higher by the proposed method than by the crash history. However, this does not necessarily mean that the model is performing poorly. These segments have the features of high-risk segments as determined by the method, even though they have no or less observed crashes. Therefore, these sites would be of interest because of their potential for future crashes. The tightness of the data points around the line and the correlation coefficient of 0.785 and 0.776 for the three- and five-years analysis period respectively indicates a strong correlation between the ranks by the two methods.

The EB method ranking was also compared with the ranking using the observed number of crashes and the results are shown in figure 15b and 15d. In this figure, the proximity of data points to the line passing through the origin (diagonal line) indicates that it is more compact compared to that of the proposed method. This observation is confirmed by a higher correlation coefficient (r=0.806, and r=0.832 for three- and five-years analysis) between the two rankings.



a) b)

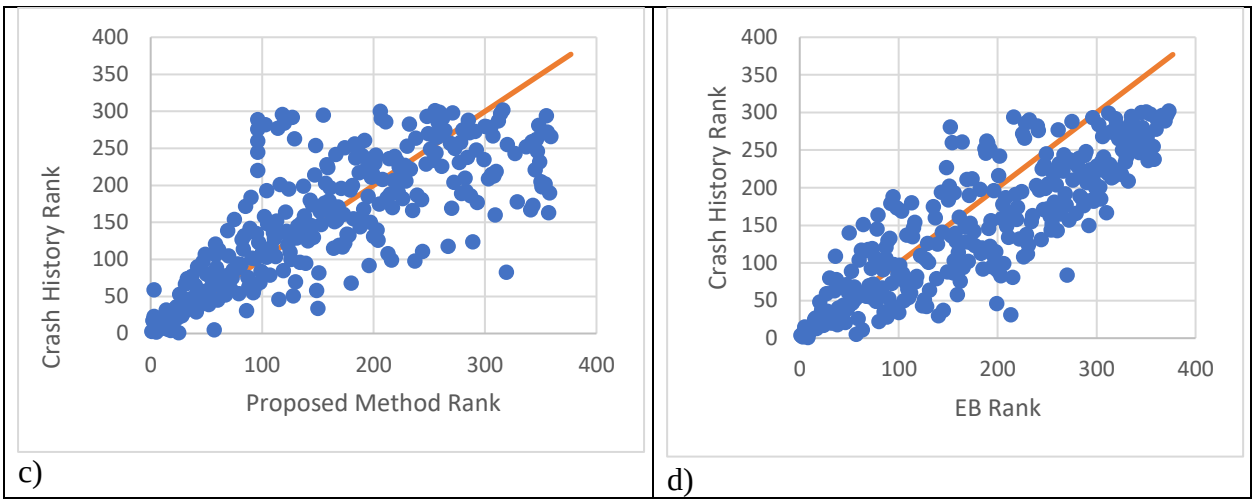


Figure 15. Rank comparison using proposed method, and EB method versus crash history. (a) three-year analysis for proposed method versus crash history (b) three -year analysis for EB method versus crash history (c) five-year analysis for proposed method versus crash history (d) five-year analysis for proposed method versus crash history.

Table 15 presents the spearman's rank correlation coefficient values for two different methods (proposed method, and EB method) based on three- and five-years analysis periods in different upper tail segments and for the total sample size of 377 segments. The results clearly demonstrate that the proposed method consistently outperforms the Empirical Bayes method across upper tail segments, while for the total sample EB method shows slightly higher correlation value for both analysis periods. Higher correlation coefficients are observed for the proposed methodology in all segment categories, indicating a stronger association and better alignment with the observed crash data. For example, considering upper tail (100) segments for the three-year period, which accounts for 26.53% of sample segments, the proposed methodology yields a Spearman's rank correlation coefficient of 0.712 versus 0.551 for the Empirical Bayes method. Similarly, in the upper tail (100) segments for a five-year study period, the proposed method yields a correlation coefficient of 0.736 compared to 0.646 for the Empirical Bayes method. However, it is worth notetaking that for upper tails (20), both methods

show very weak correlation with the crash history. Using the total study sample, the Spearman's rank correlation coefficients for the proposed methodology (0.786 in three years, and 0.780 in five years) and the Empirical Bayes method (0.807 in three years, and 0.832 in five years) indicate the overall favorable performance of the two methods across all segments. The higher correlation coefficients signify a higher association between the rankings generated by the proposed method and the observed crash data.

Table 15. Spearman's ranking correlation coefficients for study sample for different analysis periods

Groups/Methods	**% of segments**	**Spearman's Rank Correlation Coefficient**	
		Proposed Method	**EB Method**
Three-Years Analysis			
Upper Tail (20)	5.31%	0.246	0.154
Upper Tail (40)	10.61%	0.728	0.415
Upper Tail (60)	15.92%	0.762	0.327
Upper Tail (80)	21.22%	0.642	0.445
Upper Tail (100)	26.53%	0.712	0.551
Total Sample (377)	100%	0.786	0.807
Five-Years Analysis			
Upper Tail (20)	5.31%	0.079	0.300
Upper Tail (40)	10.61%	0.718	0.528
Upper Tail (60)	15.92%	0.722	0.575
Upper Tail (80)	21.22%	0.780	0.528
Upper Tail (100)	26.53%	0.736	0.646
Total Sample (377)	100%	0.780	0.832

Table 16 presents the rank root mean square error (RMSE) and mean rank difference for the proposed and the EB methods, in reference to the observed crash history over the two different study periods (three years and five years). The rank RMSE and mean rank difference were estimated using upper tail segments for a total of 20, 40, 60, 80, and 100 segments respectively. A quick examination of (Table 16) indicates that the mean rank difference and RMSE tends to increase for both methods in both study periods as the percentage of upper tail segments increases. This trend suggests a greater degree of discrepancy in rankings as the number of segments increases. Throughout both analysis periods, the proposed method consistently demonstrates superior performance than the EB method in terms of both mean rank difference and RMSE. This is evident as the proposed method consistently yields lower values for these metrics compared to the EB method across all upper tail segment groups. For example, in three-year analysis period, the mean rank difference value is 8.500 for the proposed method, while it is 15.900 for the EB method when 20 upper tail segment is considered. Likewise, the RMSE value is 30.266 for the proposed method compared to 59.070 for the EB method when examining 100 upper tail segments. Similarly, in the five-year study period, the RMSE value is 10.266 versus 17.655 for the proposed method and the EB method, respectively, when 20 upper tail segment is analyzed. Notably, the lower disparities in mean rank difference and RMSE for both analysis periods suggest a closer correlation between the proposed method's ranking and crash history compared to the EB method.

TABLE 16. Mean rank difference and root mean square error for different analysis periods.

Groups/Methods	% of segments	Average Difference in Rank		Root Mean Square Error	
		Proposed Method	EB Method	Proposed Method	EB Method
Three-Years Analysis					
Upper Tail (20)	5.31%	8.500	15.900	9.607	28.335
Upper Tail (40)	10.61%	13.100	31.575	16.648	54.791
Upper Tail (60)	15.92%	15.967	30.400	21.864	51.308
Upper Tail (80)	21.22%	17.813	35.225	23.840	55.139
Upper Tail (100)	26.53%	21.820	39.380	30.266	59.070
Five-Years Analysis					
Upper Tail (20)	5.31%	8.600	9.400	10.266	17.655
Upper Tail (40)	10.61%	10.850	24.150	14.866	44.604
Upper Tail (60)	15.92%	13.933	30.933	20.210	49.859
Upper Tail (80)	21.22%	16.363	31.200	23.687	47.299
Upper Tail (100)	26.53%	24.650	37.970	42.079	56.093

Figure 16 shows the difference in ranking between the crash history and that of the proposed and the EB methods for the first fifty highest-ranking segments in the study sample for two different analysis periods (three-years, and five-years). The y-axis represents the difference between the ranks while the x-axis represents the rank of the subject segment using the observed crash history. This figure clearly shows a higher level of consistency between the proposed method ranks and the crash history ranks compared to the EB method for three-year and five-year analysis periods. For both analysis periods, the difference in ranks for the proposed method was less than that of EB method for most of the high-ranked segments considered. For example, for three years study period considered as shown in figure 16(a), maximum rank difference for

the compared segment is 53 between the proposed method and crash history which is 207 for the EB method. Similarly, the maximum rank difference between the segment compared is 110 for the proposed method while it is 190 for the EB method when five years of analysis period is considered. This result shows higher consistency of the proposed method with the observed crash data.

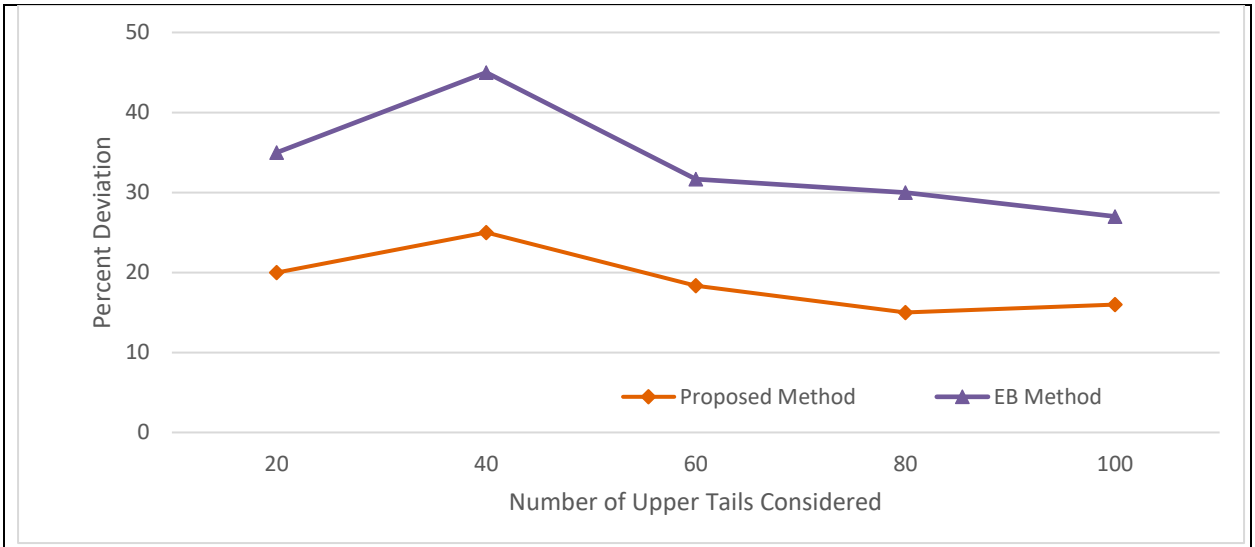
Figure 16. Rank difference between observed crash data versus the proposed and EB methods for different analysis periods; a) three years analysis; and b) five years analysis.

Table 17 presents the number and percentage of true positive segments identified by the proposed as well as the EB methods and those of the observed crash history for different upper tail segments and analysis periods. For both study periods, the proposed method consistently identifies more true positive segments than the EB method in all upper tails considered. For example, while considering the three-year analysis period, the proposed method identifies 80% of true positive segments which is 65% for the EB method in upper tail (20). Similarly, for upper tail (100), which is 26.53% of the total sample, the proposed method identifies 81% of true positive segments versus 67% for the EB method when five-year analysis period is considered.

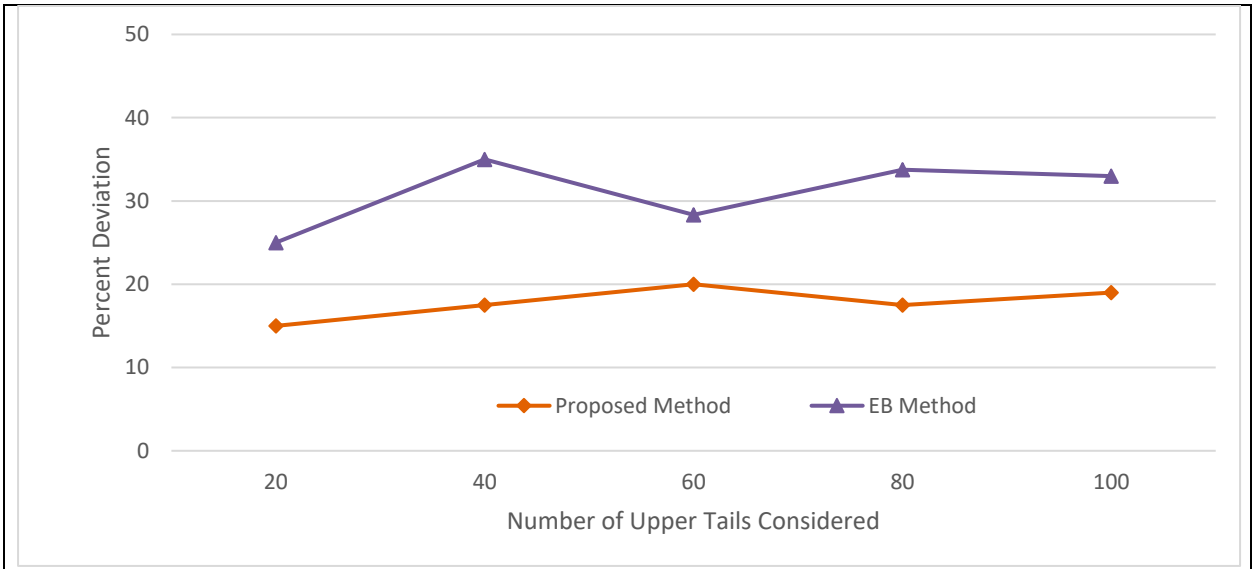
Table 17. Number and percent of common segments with crash history for the proposed and the EB methods by various upper tail segment groups.

Groups/Methods	**% of segments**	**# of Common Segments**		**% of Common Segments**	
		Proposed Method	**Empirical Bayes Method**	**Proposed Method**	**Empirical Bayes Method**
Three-Years Analysis					
Upper Tail (20)	5.31%	16	13	80	65
Upper Tail (40)	10.61%	30	22	75	55
Upper Tail (60)	15.92%	49	41	81.67	68.33
Upper Tail (80)	21.22%	68	56	85	70
Upper Tail (100)	26.53%	84	73	84	73
Five-Years Analysis					
Upper Tail (20)	5.31%	17	15	85	75
Upper Tail (40)	10.61%	33	26	82.5	65
Upper Tail (60)	15.92%	48	43	80	63.33
Upper Tail (80)	21.22%	66	53	82.5	66.25
Upper Tail (100)	26.53%	81	67	81	67

Figure 17 shows percent deviation from the crash history for the proposed method, and the EB method. A quick examination of figure clearly shows lower deviation and hence better performance for the proposed method compared to the EB method both for three-years and five-years analyses consistently throughout the upper tails considered. Specifically, the percentage deviations of 20% for three-years analysis and 15% for five-years analysis are observed for the 20 upper tail segments for the proposed method. Similarly, the corresponding percent deviations for the 100 upper tails are 16% and 19% for three-year and five-year analyses respectively. The EB method, on the other hand, continuously shows a higher percent deviation than the proposed method for every upper tail and study periods considered.



a)



b)

Figure 17. Percent deviation comparison using proposed method, and EB method versus crash history; a) three-years analysis; b) five-years analysis.

Summary

The study findings consistently demonstrate the superior performance of the proposed method over the EB method in identifying sites requiring safety enhancements, particularly within the upper tail segments of the study sample. Notably, the proposed method demonstrated a strong correlation with observed crash data, underlining the effectiveness in identifying sites needing safety improvements. The lower value of mean rank difference and RMSE further reinforce the robustness of the proposed method in network screening compared to the EB method. Additionally, the proposed method proved to be more consistent than the EB method in identifying true positive segments, highlighting the reliability in network screening. In terms of deviation from crash history, the proposed method showed a lower percent deviation compared to the EB method, which further validated the accuracy.

Overall, given the fact about the practical merits of the proposed method, the findings from this study suggest that the proposed method is a promising tool for identifying safety improvement locations on rural highways especially for the local agencies.

CHAPTER EIGHT

COMPARATIVE ANALYSIS OF PERFORMANCE OF NETWORK SCREENING METHODS ON RURAL HIGHWAYS

Methodology

Once the dataset is cleaned and prepared, the roadway segments undergo ranking using three distinct network screening methods: GRS method, CRI method, and EB prediction method. Each method produces a different ranking for the segments.

In the GRS method, segments are assigned scores based on a variety of factors, including roadway and roadside characteristics, crash history, and traffic exposure. These scores contribute to determining a global risk score. Subsequently, segments are ranked based on crash density, derived by dividing the score by the total length of the segment considered.

In the CRI method, weights are assigned to different risk factors such as roadway geometry, traffic exposure, and crash history, calibrated using sample data. Segments are then ranked based on the crash rate obtained by dividing the risk index by the total segment length and traffic exposure.

In the EB prediction method, a multivariate regression approach is employed to develop a model equation for predicting the expected number of crashes (EB) for each segment. Segments are ranked based on crash density, calculated by dividing the expected number of crashes by the total segment length.

The study design used in this comparative analysis research is shown in a data flowchart in figure 18 below.

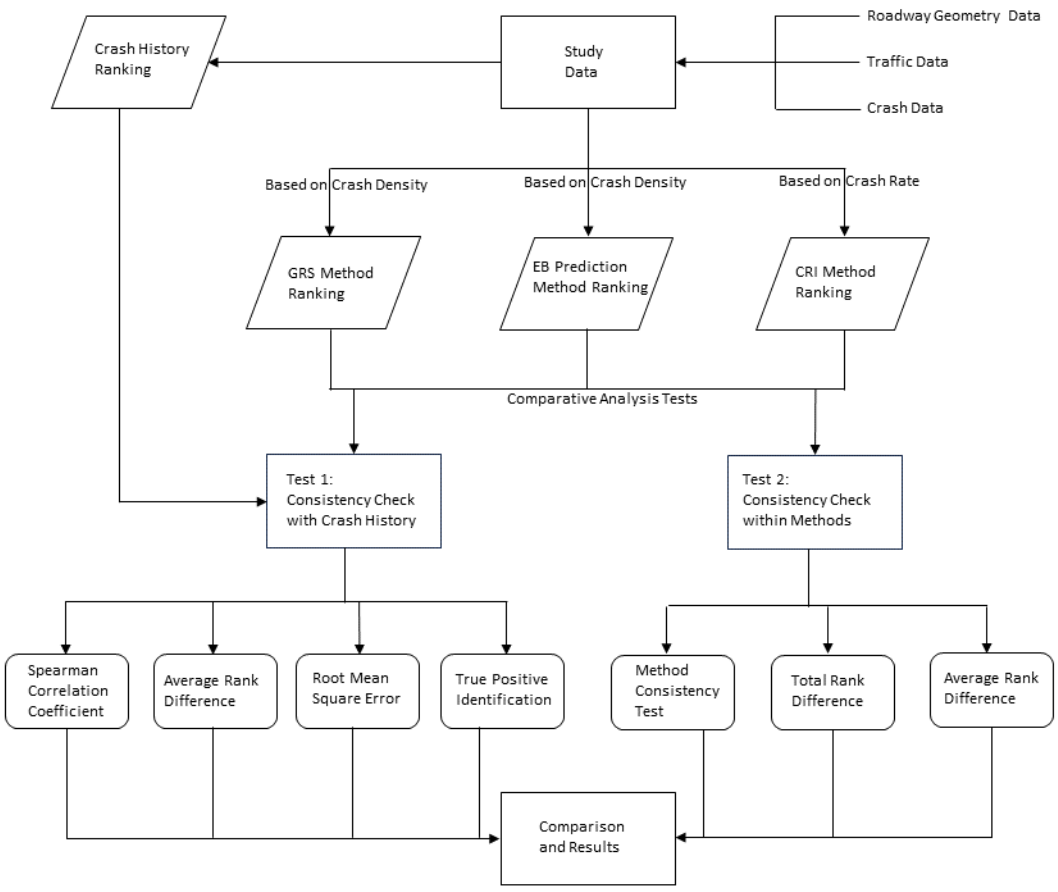


Figure 18. Data flowchart for study design

Study Analysis and Results

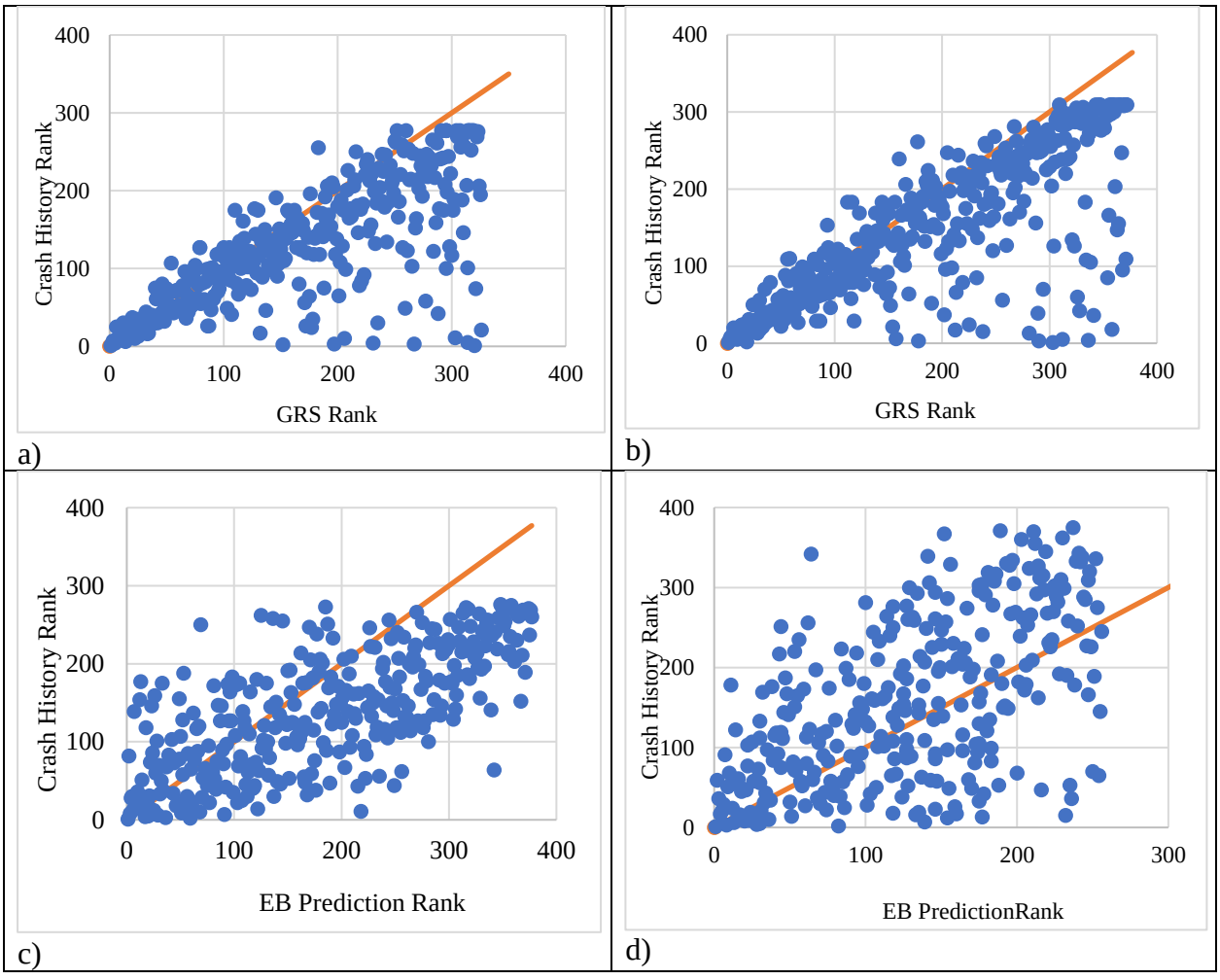
This section presents the study results for the comparative analysis of the performance of three different methods (CRI, GRS, and EB Prediction methods) for network screening on rural highways. The metrics used for evaluating the methods are discussed in detail in this section. Particularly, two different types of analysis are done to evaluate and compare the performance of these methods: consistency check of different methods with crash history, and consistency check within methods.

Consistency Check of Methods with Crash History

Figure 19 shows a scatterplot with site rankings using the three different methods versus crash history for the two different analysis periods: three-year and five-year. A careful examination of all the scatterplots for different methods versus crash history for three-year analysis period (Figure 19a, b, c) reveal a few important observations. First, for most roadway segments, the two rankings correspond well to each other, as exhibited by the clustered observations around the diagonal line. This is particularly evident in sites that ranked high on the list (first hundred fifty ranks for Figure 19a, and around seventy ranks for Figure 19c that are for GRS and CRI methods, respectively). However, figure 19b, which shows scatterplot for EB prediction method versus crash history, suggests that the rankings do not correspond well to each other. The figures also show a slight downward deviation of clustered observations below the diagonal line, which increases with the increase in rank. This is mostly the result of having two or more sites sharing the same rank using crash history (e.g., segments that had seen no crashes during the analysis period). This is confirmed by the fact that the total number of ranks using crash history is smaller than using the other methods. Overall, the tightness of the data points around the diagonal line in the figure 19a and 19c, which is for GRS method and CRI methods respectively, indicate most strong correlation between the two rankings, which is supported by a correlation coefficient of 0.787 for GRS method and 0.786 for CRI method. On the other hand, the EB prediction method performs worst as seen with scattered data points and correlation coefficient value of 0.654. A perfect correlation is visually represented as a straight diagonal line and a correlation coefficient of ±1.

Figures (19d, 19e, and 19f) present the scatterplot for the GRS, EB prediction, and CRI methods, respectively versus crash history for five-year study period. In these figures, the results

seen are like that for a three-year analysis period. However, it is worth noting that the strongest correlation between the ranks is shown by GRS method with a correlation coefficient of 0.821 which is followed by CRI method with correlation coefficient value of 0.776. Again, the less favorable performance is shown by the EB prediction method with low correlation coefficient (r = 0.683) between two rankings.



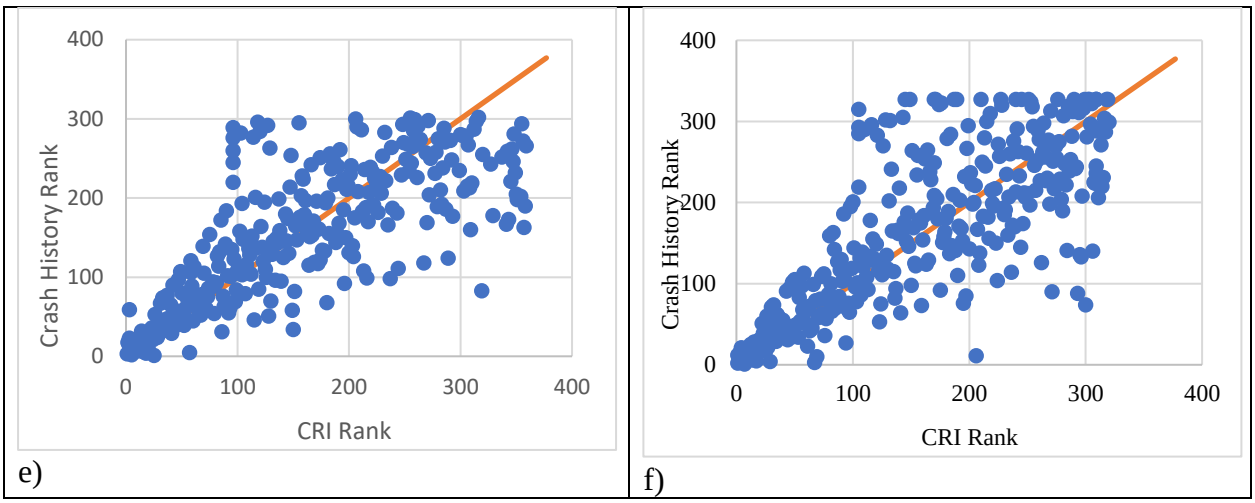


Figure 19: Rank comparison using three different methods versus crash history for different analysis periods. (a) Three-year rank comparison using GRS method versus crash history. (b) Three-year rank comparison using EB prediction method versus crash history. (c) Three-year rank comparison using CRI method versus crash history. (d) Five-year rank comparison using GRS method versus crash history. (e) Five-year rank comparison using EB prediction method versus crash history. (f) Five-year rank comparison using CRI method versus crash history.

The Spearman rank's correlation coefficient is also calculated for three different methods versus observed crash rankings as is shown in table 18. The table shows the values of Spearman's rank correlation coefficient for the total study sample as well as for subsets of sample data using upper tail 20, 40, 60, 80, and 100 sites (top ranking sites on the list). The results clearly demonstrate that the GRS method consistently outperforms the other two methods across upper tail segments, as shown in Table 18. Higher correlation coefficients are observed for the GRS method in all upper tail segment categories, indicating a stronger association and better alignment with the observed crash data. For example, considering upper tail (20) segments for the three-year period, which accounts for 5.31% of sample segments, the GRS method yields a Spearman's rank correlation coefficient of 0.820 versus 0.246 for the CRI method and 0.133 for the EB prediction method. Similarly, in the upper tail (100) segments, representing 26.53% of sample segments, the GRS method yields a value of 0.735 compared to 0.712 and 0.612 for the

CRI method, and EB prediction methods, respectively. However, using the total study sample, the Spearman's rank correlation coefficients for the GRS method (0.787) and the CRI method (0.786) indicate the overall favorable performance of the two methods across all segments. While the coefficient value for the EB prediction method is 0.654 using total study sample. The higher correlation coefficients signify a higher association between the rankings generated by the GRS method and the observed crash data.

When a five-year study period is considered, the findings are not surprising and replicates three-year analysis period. Instead, the GRS method outperforms the CRI method using the total study sample as well, where the coefficient value for these two methods is 0.821 and 0.776 respectively. Likewise, for every upper tail considered, GRS method outperforms other two methods in terms of spearman's rank correlation coefficient.

Table 18. Spearman's rank correlation coefficient for three different methods versus crash history for different analysis period

(# of segments)	% of segments	Spearman Rank Correlation Coefficient		
		CRI Method	GRS Method	EB Prediction Method
Three-Year Analysis				
Upper Tail (20)	5.31%	0.246	0.82	0.133
Upper Tail (40)	10.61%	0.728	0.814	0.587
Upper Tail (60)	15.92%	0.762	0.834	0.612
Upper Tail (80)	21.22%	0.642	0.825	0.635
Upper Tail (100)	26.53%	0.712	0.835	0.612
Total Sample (377)	100.00%	0.786	0.787	0.654
Five-year Analysis				
Upper Tail (20)	5.31%	0.079	0.801	0.744
Upper Tail (40)	10.61%	0.718	0.876	0.587
Upper Tail (60)	15.92%	0.722	0.891	0.638
Upper Tail (80)	21.22%	0.780	0.884	0.655
Upper Tail (100)	26.53%	0.736	0.864	0.605
Total Sample (377)	100.00%	0.776	0.821	0.683

Table 19 presents the rank root mean square error (RMSE) and average rank difference for three methods in reference to the observed crash history over the two different study periods. The rank RMSE and mean rank difference were estimated using upper tail segments for a total of 20, 40, 60, 80, and 100 segments and for the total study sample as well. A quick examination of Table 19 indicates that, overall, as the percentage of upper tail segments increases, the mean difference between ranks and RMSE tends to increase for all methods. This trend indicates more discrepancies in rankings as the number of segments increases. For the three-year analysis period, the GRS method consistently outperformed the other two methods with lower average

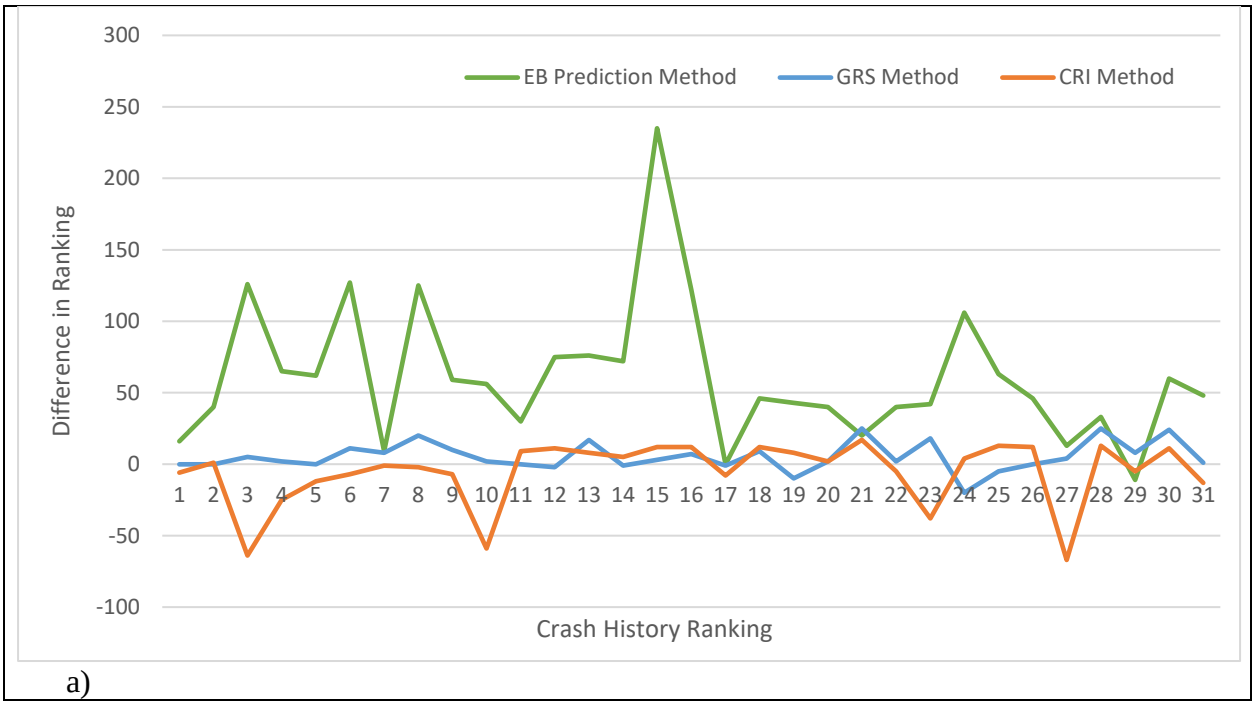
difference and RMSE values. Specifically, the average difference in rank for GRS method is 5.500 compared to 8.500 for CRI method and 71.200 for EB prediction method, when upper tail 20 is considered.

The five-year analysis period still shows a better performance by the GRS method compared with other methods in every upper tail segment considered. For example, the RMSE value for upper tail 100, is 20.224 for the GRS method followed by CRI method with a value of 30.226 and 64.176 for the EB prediction method. However, it is worth noting that the value for average difference in rank and RMSE for total study sample is least for the CRI method compared to other two methods.

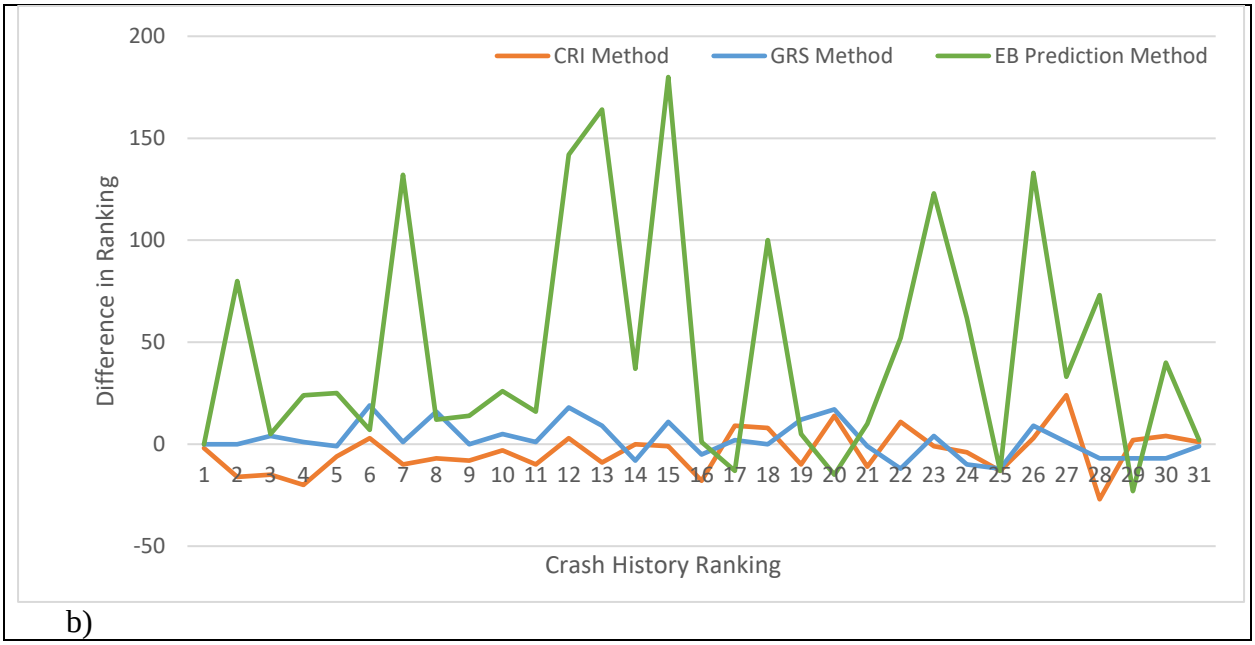
Table 19. Average difference in ranks and root mean square error for different analysis periods.

(# of segments)	Average Difference in Rank			Root Mean Square Error		
	CRI Method	GRS Method	EB Prediction Method	CRI Method	GRS Method	EB Prediction Method
Three-Year Analysis						
Upper Tail (20)	8.5	5.5	71.2	9.607	6.624	88.716
Upper Tail (40)	13.1	7.8	58.3	16.648	8.152	77.99
Upper Tail (60)	15.967	11.75	54.55	21.864	13.245	71.2
Upper Tail (80)	17.813	15.162	51.85	23.84	17.183	67.945
Upper Tail (100)	21.82	18.52	49.74	30.266	20.224	64.716
Total Sample (377)	47.976	44.405	81.743	68.077	49.384	105.768
Five-Year Analysis						
Upper Tail (20)	8.6	6.5	54	10.266	7.053	87.711
Upper Tail (40)	10.85	7.475	60.875	14.866	10.652	92.701
Upper Tail (60)	13.933	9.933	55.817	20.21	15.965	85.393
Upper Tail (80)	16.363	12.275	60.875	23.687	16.862	90.826
Upper Tail (100)	24.65	14.19	60.8	42.079	20.575	88.354
Total Sample (377)	48.971	54.594	75.252	70.412	81.408	92.564

Figure 20 shows the difference in ranking between the crash history and that of the three different methods for the first thirty highest-ranking segments in the study sample and for the two different analysis periods. The y-axis represents the difference between the ranks, while the x-axis represents the rank of the subject segment using the observed crash history. Figure clearly shows a higher consistency between the GRS method and CRI method with observed crash compared to that of the EB prediction method. Specifically, higher consistency is shown by the GRS method followed by the CRI method. For example, the difference in ranking of a subject segment is less than 25 for the GRS method which is up to 67 for the CRI method and 235 for the EB prediction method when a three-year study period is considered.



a)



b)

Figure 20. Difference in ranking for three methods for different analysis periods. (a) Three-year analysis. (b) Five-Year analysis

Likewise, when a five-year study period is considered, difference in ranking is less than 19 for the GRS method compared to 27 for the CRI method and 164 for the EB prediction method. Overall, study results suggest that the consistency of the GRS method with crash history is higher followed by the CRI method, while the EB prediction method performs worst when thirty high rank segments are analyzed.

Table 20 presents the comparison of three different network screening methods—CRI Method, GRS Method, and EB Prediction Method— across different upper tail segment groups and analysis periods in terms of identifying true positive with crash history. The results are presented in terms of the number of common segments and the percentage of common segments identified by all methods.

For every analysis period, the GRS method identified more true positive sites in all upper tail segments except for the upper tail (20), where the data show less favorable performance of

this method. Specifically, in considering the three-year study period, the CRI method outperforms the GRS method as these two methods identified 80% and 65% of true positive using upper tail 20 segment. The EB prediction method consistently shows poor performance in every upper tail segment group and study periods.

When considering five- study periods, the GRS method identified 100% of the segments as identified by the crash history ranking for upper tails 60 or greater. For the upper tail (40), this method identified 97.5% of true positive segments. While the percent common segments for the CRI method was 80% or higher for all upper tail segments, the GRS method consistently outperformed the CRI method for being more consistent with the crash history rankings for all upper tail segment groups except for the upper tail (20).

Table 20. True positive identification for three different methods compared.

(# of segments)	# of Common Segments			% of Common Segments		
	CRI Method	GRS Method	EB Prediction Method	CRI Method	GRS Method	EB Prediction Method
Three-Year Analysis						
Upper Tail (20)	16	13	7	80	65	35
Upper Tail (40)	30	38	18	75	95	45
Upper Tail (60)	49	54	56	81.67	90	50
Upper Tail (80)	68	70	46	85	87.5	57.5
Upper Tail (100)	84	88	65	84	88	65
Five-Year Analysis						
Upper Tail (20)	17	13	6	85	65	30
Upper Tail (40)	33	39	22	82.5	97.5	55
Upper Tail (60)	48	60	56	80	100	50
Upper Tail (80)	66	80	46	82.5	100	57.5
Upper Tail (100)	81	100	68	81	100	68

Consistency Check Within Network Screening Methods

A test was carried out to assess the consistency of rankings for identified safety improvement locations between two distinct time periods (2016-2017 and 2018-2020). This evaluation test is done for all three network screening methods. The term "within-method consistency" was employed to differentiate this assessment from the comparison of two ranked lists using different screening methods, which is known as "consistency across methods" as shown in previous sub-section of study analysis and results section.

A method consistency test proposed in literature (Cheng and Washington, 2008; Montella, 2010) was used in this study. Since no safety improvement or upgrade was carried out during the study period, road sections are in the same or similar underlying operational states (similar traffic volumes, geometric designs) and that their expected safety performance remains virtually unaltered over the two periods. A good network screening method will identify the same set of safety improvement locations across two periods. The greater the number of hot spots that are identified in both periods, the more reliable and consistent is the performance of the network screening method. This within-method consistency test involves comparing and evaluating two ranked lists of the same network screening method in consecutive periods (i and i + 1) using the following evaluation criteria:

$$MCT = \{k_1, k_2, \dots\dots, k_n\}_{j,i} \cap \{k_1, k_2, \dots\dots, k_n\}_{j,i+1} \quad (15)$$

Where, MCT is the number of segments identified in both time periods,

k_i is the i^{th} ranked site identified as a safety improvement location,

n is the number of upper tails considered,

j is the network screening method being compared.

In this test, the intersection of segments identified as high risk in two subsequent periods is compared over methods $j = 1$ to J, and the method yielding the largest intersection of sites is said to be the most consistent. The method consistency test is carried out for various upper tail segment groups (20, 40, 60, 80, and 100).

Table 21. Method consistency test results of compared methods for different upper tails.

(# of segments)	Number of Segments Match			Percentage of Segments Match		
	CRI Method	GRS Method	EB Prediction Method	CRI Method	GRS Method	EB Prediction Method
Upper Tail (20)	13	14	20	65	70	100
Upper Tail (40)	25	32	38	62.50	80	95
Upper Tail (60)	41	43	59	68.33	71.67	98.33
Upper Tail (80)	62	62	76	77.50	77.50	95
Upper Tail (100)	78	81	96	78.00	81	96

Table 21 shows the number of similarly identified poor performing sites identified by alternate network screening methods over the two periods. The EB prediction method is superior in this test by identifying the largest number of the same sites in every upper tail considered, with 20, 38, 59, 76, and 96 sites, respectively. In other words, the EB prediction method identified 96 sites in 2016-2017 that were also identified in 2018 to 2020 using upper tail 100. The GRS method, which performs slightly better than the CRI method, places 2nd with identifying 14, 32, 43, 62, and 81 consistent sites (in the case of upper tail 20, 40, 60, 80, and 100, respectively). The CRI performs last, with the lowest number of consistent sites identified in the two periods. However, it is worth mentioning that the results obtained from the CRI method and GRS method are highly comparable.

Also shown in Table 21 are the percentages of segments match for a method over two study periods. The results obtained from this test may seem initially unexpected to what was obtained when comparing the performance of network screening with crash history as reference where the EB prediction method shows the worst performance. However, this can be easily explained by proper understanding the nature of crash occurrence which is sporadic and random. Also, this is due to regression-to-mean effect which means a year with a low number of crashes usually follows a year with a high number of crashes and vice versa regardless of the safety improvement done in the sites. This is explained in detail in a later section.

Total rank difference test was also done to compare the performance of three different network screening methods. The test is conducted by calculating the sum of total rank differences of the risk segments identified across the two periods. The smaller the total rank difference, the more consistent is the network screening method on this test, reflecting consistent ranking of sites across periods. The total rank difference for a method being compared is calculated as:

$$TRD = \sum_{k=1}^{n}\{ R\left(k_{j,i} \right) - R(k_{j,i+1})\} \quad (16)$$

Where, TRD is the total rank difference,

R is the rank for site k in period i for method j,

$i + 1$ is the subsequent time period,

n is the number of upper tails considered.

Table 22. Total and average rank difference results for various compared methods

(# of segments)	Average Rank Difference			Total Rank Difference		
	CRI Method	GRS Method	EB Prediction Method	CRI Method	GRS Method	EB Prediction Method
Upper Tail (20)	18.250	7.550	1.000	365	151	20
Upper Tail (40)	48.950	11.225	2.000	1958	449	80
Upper Tail (60)	43.650	14.867	2.800	2619	892	168
Upper Tail (80)	39.375	18.350	3.738	3150	1468	299
Upper Tail (100)	38.680	20.680	5.240	3868	2068	524

The total rank difference and average difference in rank of the same sites for three different methods —CRI Method, GRS Method, and EB Prediction Method— over two periods are calculated and presented in table 22. This table illustrates that the EB prediction method is vastly superior in this test. In every upper tail used in analysis, the EB prediction method has significantly smaller-ranked difference. For example, for upper tail 20, the total rank difference for the EB prediction is 86.75% lower than the GRS method and 94.52% lower than CRI method. Similarly, for upper tail 100, the average rank difference for EB prediction is 74.66% less compared to GRS, and 86.45% less compared to CRI method. This result suggests that the EB prediction method is the best network screening method (of the three evaluated here) for ranking sites consistently from period to period.

Discussion

In understanding the fluctuation of crash occurrences and the concept of regression to the mean, it is essential to recognize the phenomenon where extreme outcomes tend to move closer to the average over time. In the context of road safety, this means that areas experiencing

unusually high crash rates in one period are likely to see a reduction in crash rates in subsequent periods due to chance alone. This regression to the mean effect suggests that even without any improvement in road safety, the crash rates in these areas are likely to stabilize or decrease over time simply due to random variation.

To assess the effectiveness of different network screening methods, two distinct tests were conducted. The first test involved evaluating the performance of the methods based on historical crash data, that is, using observed crash as a reference to identify hotspots. The second test, however, focused on assessing the consistency of the methods in identifying hotspots across two distinct periods, irrespective of crash history. Since no safety improvements were implemented during the study period, a reliable network screening method should consistently identify the same sites as hotspots in both periods.

A method consistency test was conducted to understand the nature of crash occurrence, focusing solely on crash history. Sites were ranked based on crash density for two consecutive periods (2015-2017 and 2018-2020), assessing whether sites identified as hotspots in the first period were also identified in the second period. The number and percentages of the same sites identified in both periods for various upper tails were calculated and presented in table 23.

Table 23. Method consistency test for observed crash ranking in two periods.

Group of Segments	% of segments	Number of Common Segments	Percentage of Common Segments
Upper Tail (20)	5.31%	7	35
Upper Tail (40)	10.61%	25	62.5
Upper Tail (60)	15.92%	38	63.33
Upper Tail (80)	21.22%	49	61.25
Upper Tail (100)	26.53%	65	65

Surprisingly, the results of the consistency test revealed the random occurrence of crashes. Sites identified as hotspots in one period did not consistently retain their hotspot status in the subsequent period. For example, using 20 upper tail segments, only 7 seven segments identified as safety improvement locations in period (2015-2017) is identified in (2018-2020), which is 35%. Similarly, when 100 upper tails are considered, 65% of the sites identified as hotspots in period 1 is identified as hotspots in period 2. This inconsistency highlights the transient nature of crash occurrences and emphasizes the importance of considering factors beyond historical crash data alone when evaluating performance of network screening methods. It underscores the necessity of robust network screening methods capable of effectively identifying potential high-risk areas, even in the absence of significant changes in crash history. Such methods are crucial for facilitating proactive safety interventions and mitigating road traffic accidents.

Further analysis was done to know about the rank difference of a subject segment in two consecutive periods and results are presented in table 24. The results from this table are consistent with what was observed in method consistency test in table 23.

Table 24. Average rank difference and total rank difference for crash history ranking in two periods.

Group of Segments	% of segments	Average Rank Difference	Total Rank Difference
Upper Tail (20)	5.31%	40.650	813
Upper Tail (40)	10.61%	44.975	1799
Upper Tail (60)	15.92%	47.633	2858
Upper Tail (80)	21.22%	50.250	4020
Upper Tail (100)	26.53%	54.170	5417

Specifically, the average rank difference for 20 upper tail segments that were identified as hotspots in period 1 is 40.650 which is relatively high knowing the fact that the road sections are in the same or similar underlying operational states (similar traffic volumes, geometric designs) and that their expected safety performance remains virtually unaltered over the two periods. Likewise, the total rank difference for upper tail 100 is 5417. It underscores the necessity of robust network screening methods capable of effectively identifying potential high-risk areas, even in the absence of significant changes in crash history.

Summary

The study presented in this paper has examined and documented a comparative analysis for the performance of three network screening methods (Global Risk Scoring, Crash Risk Index, and EB prediction methods) The analysis was performed using a dataset of 1495 miles of rural two-lane highway segments in the state of Oregon. The data used in the evaluation included roadway geometry, roadside features, traffic conditions, and ten years of crash records (2011-2020). The research first assessed the performance of the three network screening methods using consistency check with crash history for two different study periods. In this test, the ranking from the three network screening methods is compared with each other with observed crash as reference. Then, it evaluated the performance based on the temporal consistency test. In other words, ranking consistency check within the network screening method is performed for two different periods and the results are compared.

After evaluation of these three methods, following conclusions are drawn:

1. Consistency with observed crash

- The GRS method outperforms the other methods on spearman's rank correlation test, followed closely by the CRI method, on both the study periods. That is, the GRS and CRI methods show higher spearman's correlation coefficient value when comparing the ranking of the subject segments with the observed crash.
- The GRS method is superior to the other two methods for the rank root mean square error and average rank difference of the subject segments compared with crash history. That is, the GRS method shows lower RMSE, and average rank difference value followed by the CRI method. The EB prediction method is the worst to perform with relatively high value of these metrics.
- In the true positive identification, the GRS method revealed a higher percentage of hotspots identified in both study periods, except for upper tail 20, where the CRI method outperforms the GRS method. Again, the EB prediction method is the worst to perform in this criterion.
- Overall, in this first test – consistency check with crash history – the GRS method is the most consistent and reliable method for identifying safety improvement locations, followed by the CRI method, which outperforms the GRS method under some metrics as well. In contrast, the EB prediction method performed consistently worst in all the criteria and study periods.

2. Ranking consistency within the method

- In contrast to what was observed in the first test, the EB prediction method is superior to the other two methods on the method consistency test. That is, the EB prediction method consistently identified a larger number of intersections across

observation periods. The GRS and CRI methods are highly comparable with GRS method showing relatively better performance in this criterion by a little margin.

- Based on the rank difference criteria, the EB prediction method outperformed all competing methods by a wide margin of more than 70%, showing great consistency in ranking sites across observation periods. Again, the EB prediction method is followed by GRS and CRI method, respectively.

The analysis reveals distinct strengths among the screening methods. The GRS method demonstrates high effectiveness in consistency with historical crash data, indicating its utility in immediate safety prioritization. Conversely, the EB Prediction method exhibits superior consistency across different time periods, suggesting its value for long-term safety planning. However, better understanding the nature of crash occurrence and regression-to-mean effect, the result obtained from the later test is more reliable. This is reinforced by a test performed to know the consistency in identifying common hotspots across two observation periods when ranking was solely based on observed crash. Sites identified as hotspots in one period did not consistently retain their hotspot status in the subsequent period even though safety improvements were not implemented during the study period which is due to the random occurrence of crash and the regression-to-mean effect.

The study concludes that while no single method outperforms the others in all scenarios, each has unique advantages that can guide agencies in selecting the most appropriate approach for their specific needs. Given the simplicity and limited data requirements of all three methods tailored for use on rural highways, it can provide a viable option for network screening in cases where the use of the more sophisticated and data extensive method is deemed impractical for

lack of accessible data or technical expertise (e.g., roads owned and operated by local agencies such as counties, townships, and tribal governments).

CHAPTER NINE

CONCLUSIONS

Network screening is a critical initial step of the highway safety improvement programs which aims to identify sites with potential for safety improvement. Ineffective network screening can lead to a significant number of false positives and/or false negatives resulting in an inefficient allocation of agency resources and consequently impacting the overall efficacy of safety management programs. Hence, the accurate identification of sites that are in need of safety improvements is crucial for the success of any highway safety program.

This study aimed to evaluate the performance of different network screening methods for identifying candidate sites for safety improvements on rural highways. The evaluation of network screening methods in this thesis utilized a comprehensive dataset encompassing crash data, traffic volume information, and detailed road characteristics from rural highways in Oregon. The data were acquired for the 1495-mile study sample using Oregon DOT online databases and archived video logs. It comprehensively assesses three different network screening methods: the Global Risk Scoring Method (GRS), the Empirical Bayes (EB) Prediction Method, and the Crash Risk Index (CRI) method, using robust methodological framework.

The validation of the GRS method conducted in this thesis presents a promising approach for identifying candidate sites for safety improvements, particularly on low-volume roads where traditional methods may be impractical due to data limitations or technical constraints. The simplicity and minimal data requirements of this method make it a viable option for network screening by local agencies, such as counties, townships, and tribal governments. Its strong correlation with observed crash data and lower mean rank difference and Root Mean Square

Error (RMSE) demonstrate its effectiveness and reliability in identifying sites requiring safety enhancements.

The second validation study, which is of EB prediction method, reveals that when crash density is utilized for highway segments, the performance of the proposed method was deemed fair but not as effective as the well-established EB and PSI methods, despite the high R-square value of the predictive model used by the proposed approach. However, upon using crash frequencies for highway segments, the performance of the proposed method was found to be comparable to that of the EB and PSI methods. Additionally, the findings indicate that the proposed method excels in identifying segments most likely to have treatable crashes, particularly within higher upper tail segments considered in the analysis, outperforming the EB method.

The results from validation study of the CRI method consistently demonstrates the superior performance of the proposed method compared to the EB method in identifying sites necessitating safety enhancements, particularly within the upper tail segments of the study sample. Notably, the proposed method exhibits a strong correlation with observed crash data, emphasizing its effectiveness in identifying sites requiring safety improvements. Furthermore, the lower mean rank difference and Root Mean Square Error (RMSE) strengthen the robustness of the proposed method in network screening when compared to the EB method. Additionally, the proposed method proves to be more consistent than the EB method in identifying true positive segments, thereby highlighting its reliability in network screening. Moreover, regarding deviation from crash history, the proposed method demonstrates a lower percent deviation compared to the EB method, further validating its accuracy.

The fourth study compares the performance of these three network screening methods. The evaluation considers consistency with observed crash data and ranking consistency within each method across two study periods.

Results show that the GRS method outperforms others in spearman's rank correlation and rank root mean square error, indicating higher consistency with observed crash data. It also identifies a higher percentage of hotspots, except for upper tail 20, where CRI method performs better. In contrast, the EB prediction method performs consistently worse in all criteria.

However, in terms of ranking consistency within each method, the EB prediction method excels, consistently identifying a larger number of intersections of hotspots identified across observation periods and showing great consistency in ranking sites.

The GRS method demonstrates effectiveness in immediate safety prioritization, whereas the EB prediction method holds value for long-term safety planning. However, the results from the EB prediction method are deemed more reliable due to a deeper comprehension of crash occurrence and the regression-to-mean effect. This is underscored by the inconsistency observed in identifying common hotspots across observation periods solely based on observed crash data.

Overall, the validation studies of the Global Risk Scoring (GRS), Empirical Bayes (EB) prediction, and Crash Risk Index (CRI) methods highlight their respective strengths and applications in network screening for safety improvements. While the GRS method proves effective for immediate safety prioritization, particularly on low-volume roads, the EB prediction method offers value for long-term safety planning despite its inferior performance in certain aspects. The CRI method consistently outperforms the EB method in identifying sites necessitating safety enhancements, demonstrating strong correlation with observed crash data

and greater reliability in network screening. Ultimately, the choice of method should align with the specific needs and contexts of safety assessment and improvement initiatives.

The performance has been assessed for low and higher traffic volumes, extending up to 4000 ADT (Annual Daily Traffic). It is important to acknowledge that exceptionally high traffic volumes (well above 4000 vpd) were not within the scope of this study. Consequently, there is lack of evidence regarding the performance of the methods under such extreme traffic conditions. Furthermore, the evaluation was exclusively conducted using data from the state of Oregon. To gain a more comprehensive understanding of the strengths and limitations of the proposed methods, it is imperative to conduct further evaluations using highway networks from diverse regions and states. Specifically, incorporating traffic volumes greater than 4,000 ADT in these evaluations could provide valuable insights into the effectiveness of the methods in identifying safety improvement sites.

REFERENCES CITED

AASHTO, 2010. Highway Safety Manual, First Edition. American Association of State Highway and Transportation, Washington, D.C.

Abdulhafedh, A., 2017. Road Traffic Crash Data: An Overview on Sources, Problems, and Collection Methods. J. Transp. Technol. 7, 206–219. https://doi.org/10.4236/jtts.2017.72015

Abdulhafedh, A., 2016. Crash Frequency Analysis. J. Transp. Technol. 6, 169–180. https://doi.org/10.4236/jtts.2016.64017

Agerholm, N., Lahrmann, H., 2012. Identification of Hazardous Road Locations on the basis of Floating Car Data: Method and first results.

Aguero-Valverde, J., Jovanis, P.P., 2008. Analysis of Road Crash Frequency with Spatial Models. Transp. Res. Rec. 2061, 55–63. https://doi.org/10.3141/2061-07

Al-Kaisy, A., Ewan, L., Hossain, F., 2019. Identifying Candidate Locations for Safety Improvements on Low-Volume Rural Roads: The Oregon Experience. Transp. Res. Rec. 2673, 690–698. https://doi.org/10.1177/0361198119853549

Al-Kaisy, A., Raza, S., 2023. A Novel Network Screening Methodology for Rural Low-Volume Roads. J. Transp. Technol. 13, 599–614. https://doi.org/10.4236/jtts.2023.134026

Ambros, J., Havránek, P., Valentová, V., Křivánková, Z., Striegler, R., 2016a. Identification of Hazardous Locations in Regional Road Network – Comparison of Reactive and Proactive Approaches. Transp. Res. Procedia, Transport Research Arena TRA2016 14, 4209–4217. https://doi.org/10.1016/j.trpro.2016.05.392

Ambros, J., Valentová, V., Janoška, Z., 2015. Investigation of difference between network screening results based on multivariate and simple crash prediction models. Presented at the Transportation Research Board 94th Annual MeetingTransportation Research Board, Transportation Research Board.

Ambros, J., Valentová, V., Sedoník, J., 2016b. Developing Updatable Crash Prediction Model for Network Screening: Case Study of Czech Two-Lane Rural Road Segments. Transp. Res. Rec. 2583, 1–7. https://doi.org/10.3141/2583-01

Amoros, E., Martin, J.L., Laumon, B., 2003. Comparison of road crashes incidence and severity between some French counties. Accid. Anal. Prev. 35, 537–547. https://doi.org/10.1016/S0001-4575(02)00031-3

Anastasopoulos, P.Ch., Mannering, F.L., 2011. An empirical assessment of fixed and random parameter logit models using crash- and non-crash-specific injury data. Accid. Anal. Prev. 43, 1140–1147. https://doi.org/10.1016/j.aap.2010.12.024

Bland, J.M., Altman, D.G., 1994. Statistic Notes: Regression towards the mean. BMJ 308, 1499. https://doi.org/10.1136/bmj.308.6942.1499

Bonera, M., Barabino, B., Maternini, G., 2022. A Straightforward Framework for Road Network Screening to Lombardy Region (Italy). Sustainability 14, 12424. https://doi.org/10.3390/su141912424

Cafiso, S., Di Silvestro, G., 2011. Performance of Safety Indicators in Identification of Black Spots on Two-Lane Rural Roads. Transp. Res. Rec. 2237, 78–87. https://doi.org/10.3141/2237-09

Cafiso, S., Di Silvestro, G., Persaud, B., Begum, M.A., 2010. Revisiting Variability of Dispersion Parameter of Safety Performance for Two-Lane Rural Roads. Transp. Res. Rec. 2148, 38–46. https://doi.org/10.3141/2148-05

Cafiso, S., La Cava, G., Montella, A., 2007. Safety Index for Evaluation of Two-Lane Rural Highways. Transp. Res. Rec. 2019, 136–145. https://doi.org/10.3141/2019-17

Cheng, W., Washington, S., 2008. New Criteria for Evaluating Methods of Identifying Hot Spots. Transp. Res. Rec. 2083, 76–85. https://doi.org/10.3141/2083-09

Cheng, W., Washington, S.P., 2005. Experimental evaluation of hotspot identification methods. Accid. Anal. Prev. 37, 870–881. https://doi.org/10.1016/j.aap.2005.04.015

Daniels, S., Brijs, T., Nuyts, E., Wets, G., 2010. Explaining variation in safety performance of roundabouts. Accid. Anal. Prev. 42, 393–402. https://doi.org/10.1016/j.aap.2009.08.019

Deacon, J., Zegeer, C., Deen, R., 1974. Identification of Hazardous Rural Highway Locations. Ky. Transp. Cent. Res. Rep. http://dx.doi.org/10.13023/KTC.RR.1974.410

Dixon, K., Monsere, C., Xie, F., Gladhill, K., 2012. Calibrating the HSM Predictive Methods for Oregon Highways. TREC Final Rep. https://doi.org/10.15760/trec.24

El-Basyouny, K., Sayed, T., 2009. Collision prediction models using multivariate Poisson-lognormal regression. Accid. Anal. Prev. 41, 820–828. https://doi.org/10.1016/j.aap.2009.04.005

Elvik, R., 2008a. Comparative Analysis of Techniques for Identifying Locations of Hazardous Roads. Transp. Res. Rec. 2083, 72–75. https://doi.org/10.3141/2083-08

Elvik, R., 2008b. The predictive validity of empirical Bayes estimates of road safety. Accid. Anal. Prev. 40, 1964–1969. https://doi.org/10.1016/j.aap.2008.07.007

Elvik, R., Ulstein, H., Wifstad, K., Syrstad, R.S., Seeberg, A.R., Gulbrandsen, M.U., Welde, M., 2017. An Empirical Bayes before-after evaluation of road safety effects of a new motorway in Norway. Accid. Anal. Prev. 108, 285–296. https://doi.org/10.1016/j.aap.2017.09.014

Ewan, L., Al-Kaisy, A., Hossain, F., 2016. Safety Effects of Road Geometry and Roadside Features on Low-Volume Roads in Oregon. Transp. Res. Rec. J. Transp. Res. Board 2580, 47–55. https://doi.org/10.3141/2580-06

Fatality Facts [WWW Document], 2021. . IIHS-HLDI Crash Test. Highw. Saf. URL https://www.iihs.org/topics/fatality-statistics/detail/urban-rural-comparison (accessed 7.20.23).

Fitzpatrick, K., Balke, K., Harwood, D.W., Anderson, I.B., 2000. ACCIDENT MITIGATION GUIDE FOR CONGESTED RURAL TWO-LANE HIGHWAYS. NCHRP Rep.

Geedipally, S.R., Lord, D., Dhavala, S.S., 2012. The negative binomial-Lindley generalized linear model: Characteristics and application using crash data. Accid. Anal. Prev. 45, 258–265. https://doi.org/10.1016/j.aap.2011.07.012

Ghadi, M., Török, Á., 2019. A comparative analysis of black spot identification methods and road accident segmentation methods. Accid. Anal. Prev. 128, 1–7. https://doi.org/10.1016/j.aap.2019.03.002

Golob, T.F., Recker, W.W., 2003. Relationships Among Urban Freeway Accidents, Traffic Flow, Weather, and Lighting Conditions. J. Transp. Eng. 129, 342–353. https://doi.org/10.1061/(ASCE)0733-947X(2003)129:4(342)

Golob, T.F., Recker, W.W., Alvarez, V.M., 2004. Tool to Evaluate Safety Effects of Changes in Freeway Traffic Flow. J. Transp. Eng. 130, 222–230. https://doi.org/10.1061/(ASCE)0733-947X(2004)130:2(222)

Gross, F., Eccles, K., Nabors, D., 2011. Low-Volume Roads and Road Safety Audits: Lessons Learned. Transp. Res. Rec. 2213, 37–45. https://doi.org/10.3141/2213-06

Gross, F.B., Harmon, T., Albee, M., Himes, S., Srinivasan, R., Carter, D., Dugas, M., 2016. Evaluation of Four Network Screening Performance Measures (Tech Report No. FHWA-SA-16-103). United States. Federal Highway Administration. Office of Safety.

Hauer, E., 2002. Observational before--after studies in road safety: estimating the effect of highway and traffic engineering measures on road safety. Pergamon, Amsterdam.

Hauer, E., 1992. Empirical bayes approach to the estimation of “unsafety”: The multivariate regression method. Accid. Anal. Prev. 24, 457–477. https://doi.org/10.1016/0001-4575(92)90056-O

Hauer, E., Kononov, J., Allery, B., Griffith, M.S., 2002. Screening the Road Network for Sites with Promise. Transp. Res. Rec. 1784, 27–32. https://doi.org/10.3141/1784-04

Hauer, E., Persaud, B.N., Smiley, A., Duncan, D., 1991. Estimating the accident potential of an Ontario driver. Accid. Anal. Prev. 23, 133–152. https://doi.org/10.1016/0001-4575(91)90044-6

Hauke, J., Kossowski, T., 2011. Comparison of Values of Pearson's and Spearman's Correlation Coefficients on the Same Sets of Data. QUAGEO 30, 87–93. https://doi.org/10.2478/v10117-011-0021-1

Herbel, S.L., Lorrie; McGovern, Colleen, 2010. Highway Safety Improvement Program Manual: The Focus Is Results (No. FHWA-SA-09-029).

Highway Safety Improvement Program (HSIP) - Safety | Federal Highway Administration [WWW Document], n.d. URL https://safety.fhwa.dot.gov/HSIP/about.cfm (accessed 3.15.24).

Himes, S., Donnell, E., 2020. Safety Effects of Horizontal Curve Reliability Index. Transp. Res. Rec. 2674, 627–636. https://doi.org/10.1177/0361198120930715

Huang, H., Chin, H.C., Haque, Md.M., 2009. Empirical Evaluation of Alternative Approaches in Identifying Crash Hot Spots: Naive Ranking, Empirical Bayes, Full Bayes Methods. Transp. Res. Rec. 2103, 32–41. https://doi.org/10.3141/2103-05

Huda, K.T., Al-Kaisy, A., 2024. Network Screening on Low-Volume Roads Using Risk Factors. Future Transp. 4, 257–269. https://doi.org/10.3390/futuretransp4010013

Kim, D.-G., Lee, Y., Washington, S., Choi, K., 2007. Modeling crash outcome probabilities at rural intersections: application of hierarchical binomial logistic models. Accid. Anal. Prev. 39, 125–134. https://doi.org/10.1016/j.aap.2006.06.011

Kwon, O.H., Park, M.J., Yeo, H., Chung, K., 2013. Evaluating the performance of network screening methods for detecting high collision concentration locations on highways. Accid. Anal. Prev. 51, 141–149. https://doi.org/10.1016/j.aap.2012.10.019

Lambert, D., 1992. Zero-Inflated Poisson Regression, with an Application to Defects in Manufacturing. Technometrics 34, 1–14. https://doi.org/10.2307/1269547

Lee, C., Hellinga, B., Ozbay, K., 2006. Quantifying effects of ramp metering on freeway safety. Accid. Anal. Prev. 38, 279–288. https://doi.org/10.1016/j.aap.2005.09.011

Lee, J., Chung, K., Papakonstantinou, I., Kang, S., Kim, D.-K., 2020. An optimal network screening method of hotspot identification for highway crashes with dynamic site length. Accid. Anal. Prev. 135, 105358. https://doi.org/10.1016/j.aap.2019.105358

Leur, P. de, Sayed, T., 2002. Development of a Road Safety Risk Index. Transp. Res. Rec. 1784, 33–42. https://doi.org/10.3141/1784-05

Li, W., Carriquiry, A., Pawlovich, M., Welch, T., 2008. The choice of statistical models in road safety countermeasure effectiveness studies in Iowa. Accid. Anal. Prev. 40, 1531–1542. https://doi.org/10.1016/j.aap.2008.03.015

Lord, D., 2006. Modeling motor vehicle crashes using Poisson-gamma models: Examining the effects of low sample mean values and small sample size on the estimation of the fixed dispersion parameter. Accid. Anal. Prev. 38, 751–766. https://doi.org/10.1016/j.aap.2006.02.001

Lord, D., Mannering, F., 2010. The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives. Transp. Res. Part Policy Pract. 44, 291–305. https://doi.org/10.1016/j.tra.2010.02.001

Lord, D., Miranda-Moreno, L.F., 2008. Effects of low sample mean values and small sample size on the estimation of the fixed dispersion parameter of Poisson-gamma models for modeling motor vehicle crashes: A Bayesian perspective. Saf. Sci. 46, 751–770. https://doi.org/10.1016/j.ssci.2007.03.005

Ma, J., Kockelman, K.M., Damien, P., 2008. A multivariate Poisson-lognormal regression model for prediction of crash counts by severity, using Bayesian methods. Accid. Anal. Prev. 40, 964–975. https://doi.org/10.1016/j.aap.2007.11.002

MacLean, A.S., Teale, G., 1982. PROBABILITY DISTRIBUTIONS FOR TRAFFIC ACCIDENTS, INJURIES AND DEATHS. Aust. Road Res. 12.

Maher, M., Mountain, L., 2009. The sensitivity of estimates of regression to the mean. Accid. Anal. Prev. 41, 861–868. https://doi.org/10.1016/j.aap.2009.04.020

Malyshkina, N.V., Mannering, F.L., 2009. Markov switching multinomial logit model: An application to accident-injury severities. Accid. Anal. Prev. 41, 829–838. https://doi.org/10.1016/j.aap.2009.04.006

Manepalli, U.R.R., Bham, G.H., 2016. An evaluation of performance measures for hotspot identification. J. Transp. Saf. Secur. 8, 327–345. https://doi.org/10.1080/19439962.2015.1048015

Manepalli, U.R.R., Bham, G.H., 2011. Crash Prediction: Evaluation of Empirical Bayes and Kriging Methods. Presented at the 3rd International Conference on Road Safety and Simulaion, Transportation Research Board, Indianapolis, p. 13p.

McGuigan, D.R.D., 1981. THE USE OF RELATIONSHIPS BETWEEN ROAD ACCIDENTS AND TRAFFIC FLOW IN “BLACK-SPOT” IDENTIFICATION. Traffic Eng. Control 22.

Mesquita Xavier, V.J., Craveiro Cunto, F.J., 2017. Comparative Analysis of Performance Measures for Network Screening: A Case Study of Brazilian Urban Areas. J. Adv. Transp. 2017, e4360414. https://doi.org/10.1155/2017/4360414

Miaou, S.-P., Lord, D., 2003. Modeling Traffic Crash-Flow Relationships for Intersections: Dispersion Parameter, Functional Form, and Bayes Versus Empirical Bayes Methods. Transp. Res. Rec. 1840, 31–40. https://doi.org/10.3141/1840-04

Montella, A., 2010. A comparative analysis of hotspot identification methods. Accid. Anal. Prev. 42, 571–581. https://doi.org/10.1016/j.aap.2009.09.025

Moons, E., Brijs, T., Wets, G., 2009. Identifying Hazardous Road Locations: Hot Spots versus Hot Zones, in: Gavrilova, M.L., Tan, C.J.K. (Eds.), Transactions on Computational Science VI, Lecture Notes in Computer Science. Springer, Berlin, Heidelberg, pp. 288–300. https://doi.org/10.1007/978-3-642-10649-1_17

Nicholson, A., Wong, Y.-D., 1993. Are accidents poisson distributed? A statistical test. Accid. Anal. Prev. 25, 91–97. https://doi.org/10.1016/0001-4575(93)90100-B

O, J.N., 1972. Statistical Detection of Accident Blackspots. OTA-PIARC 11th Study Week Transp. Saf.

ODOT TransGIS [WWW Document], n.d. URL https://gis.odot.state.or.us/transgis (accessed 6.5.23).

Ogden, K.W., 1996. SAFER ROADS: A GUIDE TO ROAD SAFETY ENGINEERING.

Oh, J., Washington, S.P., Nam, D., 2006. Accident prediction model for railway-highway interfaces. Accid. Anal. Prev. 38, 346–356. https://doi.org/10.1016/j.aap.2005.10.004

Oregon Department of Transportation : Road Assets and Mileage : Data & Maps : State of Oregon [WWW Document], n.d. URL https://www.oregon.gov/odot/Data/pages/road-assets-mileage.aspx (accessed 6.5.23).

Oregon Department of Transportation [WWW Document], 2000. URL https://dvlprod-ordot.msappproxy.net/cf/dvl/ (accessed 3.10.24).

Park, E.S., Lord, D., 2007. Multivariate Poisson–Lognormal Models for Jointly Modeling Crash Frequency by Severity. Transp. Res. Rec. J. Transp. Res. Board.

Park, P.Y., Sahaji, R., 2013. Safety network screening for municipalities with incomplete traffic volume data. Accid. Anal. Prev. 50, 1062–1072. https://doi.org/10.1016/j.aap.2012.08.012

Pawlovich, M., 2002. Safety Improvement Candidate Location (SICL) Methods.

Persaud, B., Cook, W., Kazakov, A., 1998. DEMONSTRATION OF NEW APPROACHES FOR IDENTIFYING HAZARDOUS LOCATIONS AND PRIORITIZING SAFETY TREATMENT, in: Publication of: Swedish National Road and Transport Research Institute. Presented at the Traffic Safety on Two ContinentsPTRC Education and Research Services Limited.

Persaud, B., Lyon, C., 2007. Empirical Bayes before-after safety studies: lessons learned from two decades of experience and future directions. Accid. Anal. Prev. 39, 546–555. https://doi.org/10.1016/j.aap.2006.09.009

Persaud, B., Lyon, C., Nguyen, T., 1999. Empirical Bayes Procedure for Ranking Sites for Safety Investigation by Potential for Safety Improvement. Transp. Res. Rec. 1665, 7–12. https://doi.org/10.3141/1665-02

Persaud, B., Retting, R.A., Lyon, C., 2000. Guidelines for Identification of Hazardous Highway Curves. Transp. Res. Rec. 1717, 14–18. https://doi.org/10.3141/1717-03

Persaud, B.N., 2001. STATISTICAL METHODS IN HIGHWAY SAFETY ANALYSIS. NCHRP Synth. Highw. Pract.

Quddus, M.A., 2008. Modelling area-wide count outcomes with spatial correlation and heterogeneity: An analysis of London crash data. Accid. Anal. Prev. 40, 1486–1497. https://doi.org/10.1016/j.aap.2008.03.009

Schmelkin, E.J.P., Liora Pedhazur, 1991. Measurement, Design, and Analysis: An Integrated Approach. Psychology Press, New York. https://doi.org/10.4324/9780203726389

SEMCOG, 1997.

Shankar, V., Milton, J., Mannering, F., 1997. Modeling accident frequencies as zero-altered probability processes: an empirical inquiry. Accid. Anal. Prev. 29, 829–837. https://doi.org/10.1016/s0001-4575(97)00052-3

Sharma, S.L., Datta, T.K., 2007. Investigation of Regression-to-Mean Effect in Traffic Safety Evaluation Methodologies. Transp. Res. Rec. 2019, 32–39. https://doi.org/10.3141/2019-05

Shin, K., Washington, S.P., 2012. Empirical Bayes method in the study of traffic safety via heterogeneous negative multinomial model. Transportmetrica 8, 131–147. https://doi.org/10.1080/18128601003680976

Tamburri, T.N., Smith, R.N., 1970. The Safety Index: A Method of Evaluating and Rating Safety Benefits, in: Highway Research Record. Presented at the 49th Annual Meeting of the Highway Research Board.

Tarko, A.P., Kanodia, M., 2004. Effective and Fair Identification of Hazardous Locations. Transp. Res. Rec. 1897, 64–70. https://doi.org/10.3141/1897-09

Tarko, A.P., Sinha, K.C., Farooq, O., 1996. Methodology for Identifying Highway Safety Problem Areas. Transp. Res. Rec. 1542, 49–53. https://doi.org/10.1177/0361198196154200108

TDS - Crash Reports [WWW Document], n.d. URL https://tvc.odot.state.or.us/tvc/ (accessed 6.5.23).

The CMF Clearinghouse: A Handy Safety Tool | FHWA [WWW Document], n.d. URL https://highways.dot.gov/public-roads/novemberdecember-2010/cmf-clearinghouse-handy-safety-tool (accessed 5.22.23).

Wijnen, W., Weijermars, W., Schoeters, A., van den Berghe, W., Bauer, R., Carnis, L., Elvik, R., Martensen, H., 2019. An analysis of official road crash cost estimates in European countries. Saf. Sci. 113, 318–327. https://doi.org/10.1016/j.ssci.2018.12.004

Wood, J.S., Donnell, E.T., 2017. Causal inference framework for generalizable safety effect estimates. Accid. Anal. Prev. 104, 74–87. https://doi.org/10.1016/j.aap.2017.05.001

Wood, J.S., Donnell, E.T., Porter, R.J., 2015a. Comparison of safety effect estimates obtained from empirical Bayes before–after study, propensity scores-potential outcomes framework, and regression model with cross-sectional data. Accid. Anal. Prev. 75, 144–154. https://doi.org/10.1016/j.aap.2014.11.019

Wood, J.S., Gooch, J.P., Donnell, E.T., 2015b. Estimating the safety effects of lane widths on urban streets in Nebraska using the propensity scores-potential outcomes framework. Accid. Anal. Prev. 82, 180–191. https://doi.org/10.1016/j.aap.2015.06.002

World Health Organization [WWW Document], 2023. URL https://www.who.int/teams/social-determinants-of-health/safety-and-mobility/global-status-report-on-road-safety-2023 (accessed 2.12.24).

World Health Organization [WWW Document], 2021. URL https://www.who.int/teams/social-determinants-of-health/safety-and-mobility/decade-of-action-for-road-safety-2021-2030 (accessed 2.12.24).

Zegeer, C.V., 1986. METHODS FOR IDENTIFYING HAZARDOUS HIGHWAY ELEMENTS. NCHRP Synth. Highw. Pract.