CHAPTER

8

Identification of hazardous sites

8.1 Introduction

Hazardous sites, also called hotspots, blackspots, sites with promise, priority investment locations, and crash-prone locations, refer to transport system locations that experience more crashes than would be expected compared to other sites with similar characteristics. The process to identify potentially hazardous sites across the network is called network screening or blackspot identification. Network screening is conducted to identify a smaller subgroup of sites from the entire road network for detailed investigation or site diagnostics (Hauer et al., 2002). The network screening analysis is primarily used to assess the conditions of an existing infrastructure to assist transportation agencies in their long-range plan or corridor planning analysis and for selecting appropriate countermeasures. Network screening ensures an efficient identification of hazardous sites where limited agency resources are devoted by implementing safety improvements with the objective of reducing the number and severity of crashes in a most efficient manner.

Network screening can be conducted using either a reactive or a proactive approach. The reactive approach relies on analyses of historic crash data, whereas the proactive approach relies on analyses and identification of geometric and operational characteristics that are highly associated with crash risk but not necessarily with crashes themselves. Although proactive approaches have gained increasing attention, reactive approaches are still the most popular methods used in the hazardous site selection.

Different methods are available to identify hazardous sites, and each method has its own advantages and disadvantages. If sites are improperly identified, the "true" high-risk sites may not be treated or the relatively "safe" sites may be identified for further investigation. In the former case, safety is not improved while, in the latter case, safety funds are wrongly

https://doi.org/10.1016/B978-0-12-816818-9.00001-9

invested or wasted. For this reason, it is recommended to use robust methods if hazardous sites are to be detected in the most successful way.

The selection of a method depends on two key factors:

- *Data availability*—determination of a method depends on the availability of data such as crash frequency, crash severity, crash types, crash costs, traffic volume, and crash-frequency models. Simple methods often rely on historic crashes, whereas more sophisticated methods need additional information such as predicted or expected crashes.
- *Random Fluctuations in Crashes*—crashes fluctuate from year to year randomly even when nothing is changed. If the random and large fluctuation is not accounted for, then it could result in inaccurate identification of sites. In this case, the identification method needs to be adjusted to account for this temporal variation. Some of the methods described in the next section account for random fluctuations in crash data.

The other important aspect of network screening is the threshold determination with which hazardous sites are detected. A comparison of a site performance relative to how it is expected to perform reveals the safety performance of that site. The sites are selected for a treatment based on the budget available or a specified performance threshold. In some instances, sites are ranked in descending order by a specific safety performance metric and a certain number of sites (e.g., top 5% or until the allocated funds are depleted) are selected for treatment. The threshold or reference point is used to decide which sites can be selected for further investigation. This threshold value can be either an assumed value or calculated using the screening method itself. The sections below provide details about a few commonly used methods for the selection process of hazardous sites. Some of the methods and examples described below are adapted from the Road Safety Manual (RSM) (PIARC, 2003) and the Highway Safety Manual (HSM) (AASHTO, 2010). For the methods not presented in this chapter, the interested reader is referred to the RSM and HSM.

This chapter first discusses various hazardous site selection methods that rely on observed crashes, predicted crashes, or expected crashes. The discussion includes each method’s strengths and weaknesses. Then, the chapter presents geospatial hotspot methods that consider the effects of unmeasured variables by accounting for spatial autocorrelation between the crash events over a geographic space. As the hazardous site selection methods may not efficiently identify the point locations where the deficiency exists, this chapter documents a list of the high crash concentration location methods. Next, the proactive approach methods are presented as they are gaining increasing attention due to the nature of crash prevention. Lastly, the screening evaluation methods are discussed in detail.

8.2 Observed crash methods

Observed crash methods rely, as the name implies, on historic crashes that occurred at a particular site. Typically, those sites with the highest crash measures are analyzed in detail to identify potential safety countermeasures or treatments. For the methods described in this section, crash counts are used as the estimate for the long-term mean of the site.

8.2.1 Crash frequency method

The crash frequency method is the simplest of all network screening methods. A site that experiences a higher number of crashes over a period of time indicates a safety issue and warrants further investigation. Multiple crashes occurring at a location are called "clustering." Crash clusters can be identified by a corridor, or a specific segment or an intersection on the corridor.

Crash frequency is defined as the number of crashes occurring at a particular site, facility, or network per unit of time, usually in a 1-year period. Crash frequency for a roadway segment is measured in the number of crashes per year per mile. This method does not provide a specific performance threshold. Typically, twice the average crash frequency for the reference population is used as the threshold (which is equivalent to $2 \times mean$). If the crash frequency at a site is higher than the threshold, then the site is considered to be hazardous.

The crash frequency method only requires crashes and their locations, which makes it an attractive quantitative hazardous site selection method. This method does not require traffic volume or roadway feature data. Although it is easy to implement, the crash frequency method does have shortcomings. As the crash frequency method does not consider traffic volume, it has an intrinsic bias toward higher volume locations because crashes are correlated with traffic volume. Hence, it is more likely that higher volume locations experience more crashes than lower volume locations.

Exercise 8.1

Using dataset 8.1, identify hazardous intersections based on the crash frequency method.

Calculate the average crash frequency in the reference population using Eq. 5.2:

$$\mu_r = \frac{1}{N}\sum_{i=1}^{N} x_i = \frac{1024}{50} = 20.48$$

Determine the threshold:

$$\mu(threshold) = 2 \times \mu_r = 40.96$$

The intersections with a total crash frequency greater than the threshold are identified as hazardous locations. Based on the crash frequency method, intersections 1, 8, 9, 11, 13, and 46 are detected.

8.2.2 Crash rate method

The crash rate method takes exposure data of a segment or intersection into account when evaluating the safety performance compared with other similar segments, or intersections. For exposure data, this method typically uses traffic volume and mileage. The other uncommon exposure measures used are population, lane, or roadway miles, and licensed drivers within a community. The following formula is typically used to calculate the crash rate:

$$R_i = \frac{C_i \times 10^6}{N \times 365 \times AADT_i \times L_i} \tag{8.1}$$

where R_i is the crash rate for site i (in million-vehicles miles), C_i is the crash frequency at site i, N is the number of years of crash data considered, $AADT_i$ is the annual average daily traffic for site i. (note: total entering volumes is considered for the intersections), and, L_i is the length of site i (note: L is not considered for intersections).

This method does not provide a specific performance threshold. Typically, twice the average crash rate for the reference population is used as the threshold. The crash rate for the reference population is calculated using the following equation.

$$R_r = \frac{\sum_i C_i \times 10^6}{N \times 365 \times \sum_i (AADT_i \times L_i)} \tag{8.2}$$

where R_r is the average crash rate in the reference population.

The crash rate method has an ability to identify low-volume sites that may not necessarily experience a high number of crashes but are high-crash risk locations. However, the crash rate method has an intrinsic bias towards lower traffic volume sites (i.e., it inflates the ratio between crash count and volume). Hence, the crash rate of a site should be compared with other sites that have similar characteristics such as functional class, number of lanes, and traffic volume.

Exercise 8.2

Using dataset 8.1, identify hazardous intersections based on the crash rate method.

Calculate the average crash rate in the reference population using Eq. (8.2).

$$R_r = \frac{\sum_i C_i \times 10^6}{N \times 365 \times \sum_i (AADT_i)} = \frac{1024 \times 10^6}{2 \times 365 \times 1,177,116} = 1.19$$

Determine the threshold

$$R(threshold) = 2 \times R_r = 2.38$$

The intersections with a total crash rate greater than the threshold are identified as hazardous locations. Based on the crash rate method, intersections 6, 11, 13, and 31 are detected.

8.2.3 Rate quality control method

The rate quality control (RQC) method uses a critical rate, which is calculated for each site. If the crash rate at the site is higher than the critical rate, then it is considered to be hazardous. The following equation is used to calculate the critical rate of a road segment (AASHTO, 2010).

$$R_c = R_r + \left[p \times \sqrt{\frac{R_r \times 10^6}{AADT \times N \times 365 \times L}} + \frac{10^6}{2 \times AADT \times N \times 365 \times L} \right] \tag{8.3}$$

where R_c is the critical crash rate, R_r is the crash rate for the reference population, and, p is the P-value (=1.036, 1.282, 1.645, and 2.326 for a level of confidence of 85%, 90%, 95%, and 99%, respectively).

Exercise 8.3

Using dataset 8.1, identify hazardous intersections based on the rate quality control method.

Calculate the critical rate for each site based on Eq. (8.3). For intersection 1, with a level of confidence of 95%, the critical crash rate is:

$$R_c = R_r + \left[p \times \sqrt{\frac{R_r \times 10^6}{AADT \times N \times 365 \times L}} + \frac{10^6}{2 \times AADT \times N \times 365 \times L}\right]$$

$$= 1.19 + \left[1.645 \times \sqrt{\frac{1.19 \times 10^6}{53,896 \times 2 \times 365}} + \frac{10^6}{2 \times 53,896 \times 365}\right] = 1.49$$

For intersection 1, the crash rate is 1.12. As it is lower than the critical crash rate, it is detected as nonhazardous intersection.

Based on the RQC method, intersections 6, 8, 11, 13, 16, 17, 21, 26, 29, 31, 36, and 46 are detected.

8.2.4 Equivalent property damage only method

The earlier measures rely on just the crash frequency and not on the injury severity of crashes. As the injury severity is different for each crash, there is a need to account for and differentiate among crash severities. The equivalent property damage only (EPDO) score considers both frequency and severity of crashes. It is important to note that the crash severity refers to the worst level of injury sustained by one of the crash victims. It does not account for the lower level injuries sustained by other occupants in the crash.

The EPDO method calculates a single combined severity score by assigning weighting factors to crashes by severity. The societal costs of crashes by severity are used to develop an EPDO score. As the weight is dependent on the crash costs, this measure is heavily influenced by fatal crashes. Even one fatal crash at a site can make the site to be ranked high compared to other sites with no fatal crashes but many nonfatal crashes. To overcome this limitation, fatal and injury crashes are combined into one category in some instances to avoid overemphasizing fatal crashes. The weighting factors are calculated as follows:

$$f_s = \frac{Cost_s}{Cost_{PDO}} \tag{8.4}$$

where f_s is the weighting factor for crash severity s, $Cost_s$ is the crash cost for crash severity s, and $Cost_{PDO}$ is the crash cost for PDO crash severity.

The EPDO score for a particular site is developed using the following equation:

$$EPDO\ score = \sum_s f_s \times C_s \tag{8.5}$$

where C_s is the number of crashes of severity s.

This method does not provide a specific performance threshold. Typically, twice the EPDO score for the reference population is used as the performance threshold.

Exercise 8.4

Using dataset 8.1, identify hazardous intersections based on the EPDO method. Use the following weights (AASHTO, 2010).

Severity	Weight
Fatal (K)	542
Injury (A/B/C)	11
PDO (O)	1

Calculate the EPDO score for each site based on Eq. (8.5). For intersection 1, the EPDO score is:

$$EPDO\ score = \sum_s f_s \times C_s = 542 \times 0 + 11 \times 18 + 1 \times 26 = 224$$

The average EPDO score in the reference population ($EPDO\ score_r$) is calculated as 119.52.

Determine the threshold:

$$EPDO\ score(threshold) = 2 \times EPDO\ score_r = 239.04$$

The intersections with EPDO score greater than the threshold are identified as hazardous locations. Based on the EPDO method, intersections 8, 13, 33, and 46 are detected.

8.2.5 Severity index method

Severity index (SI) is another measure that considers the EPDO score but standardizes it based on the total number of crashes at the site. The severity index is calculated using the following equation:

$$SI = \frac{EPDO\ score}{C} \tag{8.6}$$

where SI is the severity index.

This method does not provide a specific performance threshold. Typically, twice the SI score for the reference population is used as the performance threshold. The EPDO index method presented in the RSM is the same as the SI method, whereas the EPDO average crash frequency method presented in the HSM is equivalent to the EPDO score method. The EPDO score and SI methods may not provide the same results so they should not be used interchangeably.

Exercise 8.5

Using dataset 8.1, identify hazardous intersections based on the SI method.

Calculate the SI for each site based on Eq. (8.6). For intersection 1, the SI is:

$$SI = \frac{EPDO\ score}{C} = \frac{224}{44} = 5.09$$

The average SI in the reference population (SI_r) is calculated as 5.62.

Determine the threshold:

$$SI(threshold) = 2 \times SI_r = 11.24$$

The intersections with SI greater than the threshold are identified as hazardous locations. Based on the SI method, intersections 13 and 33 are detected.

A few studies have compared the performance of all observed crash-based methods in the identification of hazardous sites (e.g., see Montella (2010); Lim and Kweon (2013)). Among the observed crash methods, Montella (2010) found that the crash frequency method performed better than others and also has more appealing theoretical arguments. Furthermore, when he compared successive years, sites that experienced high crash counts were consistently ranked over the time period analyzed. On the other hand, he cautioned agencies against using the crash rate method for the same reason (wide variation in the ranking was observed over successive years). In addition, he showed that the EPDO method is largely inconsistent. Lim and Kweon (2013) found that the crash frequency method performed the best in identifying the top 1% of hazardous intersections and the RQC method performed better in identifying the top 5% and 10% of hazardous intersections.

8.2.6 Composite safety score

Some studies have combined different methods such as crash frequency, severity, and collision type in the network screening process (Qin et al., 2009; MAG, 2010). For the selection of hazardous intersection locations, these studies used a composite safety score that included crash severity and collision type measures with crash frequency. The crash rate was not included due to its correlation with crash frequency. Note that, although this method is used for identifying hazardous intersection locations, it can be used for the selection of other types of sites as well. The composite score is calculated as (Qin et al., 2009):

$$FS = \frac{1}{5} \times \frac{CF}{\mathrm{Max}(CF)} + \frac{3}{5} \times \frac{CS}{\mathrm{Max}(CS)} + \frac{1}{5} \times \frac{CT}{\mathrm{Max}(CT)} \tag{8.7}$$

With

$$CS = 40C_K + 9C_A + 5C_B + 2C_C + C_{PDO} \tag{8.8}$$

$$CT = \sum Cost_t \times N_v \tag{8.9}$$

where FS is the composite safety score; CF is the total crash frequency at the site; CS is the total crash severity index for the site; CT is the total crash type score for the site; $\mathrm{Max}(CF)$, $\mathrm{Max}(CS)$, and $\mathrm{Max}(CT)$ are the maximum values recorded for any intersection in the network; C_s is the frequency of crashes with severity s; $Cost_t$ is the crash cost of collision type t; and, N_v is the number of vehicles involved in each crash. For more details about this method, the interested reader is referred to Qin et al. (2009).

The Maricopa Association of Governments' Strategic Transportation Safety Plan employed a slightly modified composite score and called it Index of Intersection Safety Score (ISS) (MAG, 2010). ISS is calculated using the following equation.

$$ISS = \frac{1}{4} \times \frac{CF}{\mathrm{Max}(CF)} + \frac{1}{2} \times \frac{CS}{\mathrm{Max}(CS)} + \frac{1}{4} \times \frac{CT}{\mathrm{Max}(CT)} \tag{8.10}$$

This method does not provide a specific performance threshold. In general, sites are ranked according to the composite safety score and selected based on budget availability. The composite safety score method is similar to the Relative Severity Index (RSI) presented in the RSM and HSM. As opposed to the composite safety score, the RSI does not consider the actual severity but relies on the average severity of several crashes that occurred under similar conditions (PIARC, 2003).

Table 8.1 describes the overall strengths and limitations of each observed crash method that needs to be considered while selecting a method.

8.3 Predicted crash methods

Crash occurrence is a random process and crashes naturally fluctuate over time at any given site. This year-to-year variability in crash frequencies adversely affects the hazardous site selection if the screening method is based on just observed crash data. To circumvent this problem, it is important to determine the expected safety of a site based on the safety performance of other similar sites. All factors such as traffic exposure, operational and geometric features, and weather factors that influence the site safety should be considered using a statistical or crash-frequency model. These types of models that are used to estimate the expected safety performance are called crash-frequency models, crash prediction models, or safety performance functions (SPFs) in the highway safety literature. For the rest of this section, we refer to crash-frequency models as SPFs to be consistent with the literature on this topic. SPFs are typically developed using cross-sectional data from a group of similar sites and consider changes in the roadway, environmental, and operational factors that can influence crash occurrence (Hauer, 1997). SPFs not only account for nonlinear relationships between crashes and exposure, but also capture and quantify the random and uncertain nature of the crash occurrence. When the site's observed safety is worse than the expected safety obtained from the SPFs then the site is considered to be hazardous.

8.3.1 Potential for improvement using predicted crashes

The site's observed average crash frequency is compared with a predicted average crash frequency from an SPF. The SPF in this case represents the average estimated or predicted number of crashes for sites having similar characteristics. The excess predicted crash frequency, or the potential for improvement (PI), is the difference between the observed and predicted crash frequencies (see Fig. 8.1). When the PI is greater than zero, a site experiences more crashes than predicted. When the PI is less

TABLE 8.1 Strengths and limitations of observed crash methods (Tsapakis et al., 2017).

Method	Strengths	Limitations
Crash frequency	• Simple • High crash frequency sites are necessarily detected	• Does not account for the severity. • Does not properly capture the long-term mean of the sites. • Does not estimate a threshold to indicate sites experiencing more crashes than predicted for sites with similar characteristics • Does not account for traffic volume • Does not identify low-volume collision sites where low-cost countermeasures could be easily applied
Crash rate	• Simple • Accounts for traffic exposure	• Does not account for the severity. • Does not properly capture the long-term mean of the sites. • Does not estimate a threshold to indicate sites experiencing more crashes than predicted for sites with similar characteristics • Assumes linear relationship between crashes and traffic volume • Bias towards low volume and low collision sites
Rate quality control	• Reduces exaggerated effect of sites with low volumes • Accounts for variance in crash data • Estimates a threshold for comparison	• Does not account for the severity. • Does not properly capture the long-term mean of the sites. • Assumes linear relationship between crashes and traffic volume
EPDO score	• Simple • Accounts for crash severity	• Does not account for traffic volume. • Does not properly capture the long-term mean of the sites. • Does not estimate a threshold to indicate sites experiencing more crashes than predicted for sites with similar characteristics • Bias toward high-speed rural sites • May overemphasize locations with a small number of severe crashes depending on weighting factors used

Continued

TABLE 8.1 Strengths and limitations of observed crash methods (Tsapakis et al., 2017).—cont'd

Method	Strengths	Limitations
Severity index	• Simple • Accounts for crash severity • Comparisons can be easily made across sites due to standardization based on crash frequency	• Does not properly capture the long-term mean of the sites. • Does not estimate a threshold to indicate sites experiencing more crashes than predicted for sites with similar characteristics • Does not account for traffic volume • May overemphasize locations with a small number of severe crashes depending on weighting factors used
Composite safety score	• Accounts for crash frequency • Accounts for crash severity • Accounts for collision type	• Does not properly capture the long-term mean of the sites. • Does not estimate a threshold to indicate sites experiencing more crashes than predicted for sites with similar characteristics • Does not account for traffic volume • Requires rigorous analysis

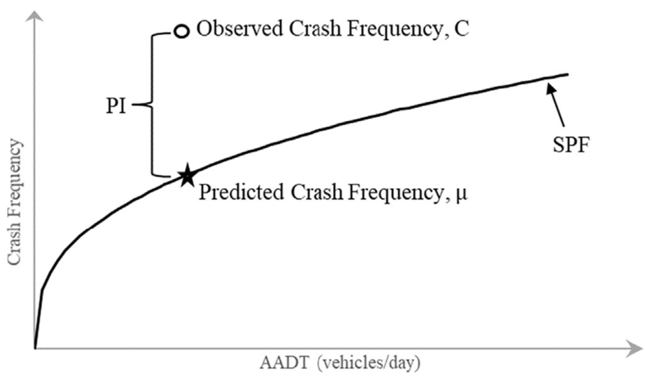

FIGURE 8.1 Graphical representation of potential for improvement using SPF.

than zero, a site experiences fewer crashes than predicted. The PI for a site is calculated using the following equation.

$$PI = C - \mu \tag{8.11}$$

where PI is the potential for improvement for the site; C is the observed crash frequency; and, μ is the predicted crash frequency from the SPF.

This method is highly dependent on the quality of the SPF and it does not take into account the random nature of crashes. This method does not provide a specific performance threshold. All the sites that have PI greater than zero can be considered hazardous but the preference will be given to the sites with the highest PI.

Exercise 8.6

Using dataset 8.1, identify hazardous intersections based on the potential for improvement using the SPF method. For simplicity, let us assume that the crashes are a function of just the major and minor street flows. Use the following functional form and parameter estimates.

$$\mu = e^{\beta_0} AADT_{maj}^{\beta_{maj}} AADT_{min}^{\beta_{min}}$$

Parameter	Estimate	Std. error	*P*-value
β_0 (Intercept)	−4.3049	1.3370	0.0013
β_{maj} (Major street AADT)	0.5969	0.1410	<0.0001
β_{min} (Minor street AADT)	0.1850	0.0614	0.0026
α (Dispersion)	0.2423	0.0610	<0.0001

Calculate the predicted crashes for each site based on the above functional form. For intersection 1, the predicted number of crashes are:

$$\mu = e^{-4.3049} 37191^{0.5969} 16705^{0.1850} = 43.6 \text{ crashes}$$

The *PI* for intersection 1 is estimated as 44 − 43.6 = 0.4 crashes.

If sites with *PI* greater than 20 crashes are considered to be hazardous locations, then intersections 8, 11, 13, and 46 are detected.

8.3.2 Level of service of safety

With the level of service of safety (LOSS) method, the sites are ranked based on their safety performance relative to the predicted average crash frequency for the reference population under consideration. The LOSS is divided into four categories based on the degree of deviation of observed safety performance from the predicted average crash frequency. Each site is assigned a LOSS category based on the difference between the average crash frequency observed at each site and the predicted average crash frequency of the reference population. Sites with high potential for crash reduction are flagged for further investigation. The first step involves calculating the standard deviation of the predicted crashes. The estimate of standard deviation is based on the assumption of the negative binomial distribution for the SPF and is calculated using the following equation (Kononov and Allery, 2003; AASHTO, 2010).

$$\sigma = \sqrt{\alpha\mu^2} \tag{8.12}$$

where σ is the standard deviation of predicted crashes, α is the over-dispersion parameter of the SPF (recall that $Var(y) = \mu + \alpha\mu^2$), and μ is the predicted crash frequency from the SPF.

The limits for the four LOSS categories are shown in Table 8.2.

Exercise 8.7

Using dataset 8.1, identify hazardous intersections based on the LOSS method. Use the same functional form and parameter estimates as in Exercise 8.6.

Calculate the standard deviation of predicted crashes using Eq. (8.12). For intersection 1, the standard deviation is estimated as follows:

$$\sigma = \sqrt{\alpha\mu^2} = \sqrt{0.2423 \times 43.6^2} = 21.5 \text{ crashes}$$

The LOSS for intersection 1 falls under category III.

The sites with LOSS IV are considered as hazardous locations, so intersections 11, 13, and 46 are detected.

Table 8.3 describes the overall strengths and limitations of each predicted crash method.

TABLE 8.2 LOSS categories (Kononov and Allery, 2003; AASHTO, 2010).

LOSS	Condition	Description
I	$\sigma < C < (\mu - 1.5\sigma)$	Low potential for crash reduction
II	$(\mu - 1.5\sigma) \le C < \mu$	Low to moderate potential for crash reduction
III	$\mu \le C < (\mu + 1.5\sigma)$	Moderate to high potential for crash reduction
IV	$C \ge (\mu + 1.5\sigma)$	High potential for crash reduction

TABLE 8.3 Strengths and limitations of predicted crash methods (Tsapakis et al., 2017).

Method	Strengths	Limitations
PI using predicted crashes	• Accounts for traffic volumes • Estimates a threshold for comparison	• Results may not properly capture the long-term mean of the sites (use average of the sites)
Level of service of safety	• Accounts for variance in crash data • Accounts for traffic volumes • Estimates a threshold for measuring potential to reduce crash frequency	• Results may not properly capture the long-term mean of the sites (use average of the sites)

8.4 Bayesian methods

Statistical modeling using Bayesian techniques have been widely used for analyzing traffic safety data and identifying hazardous sites. Bayesian techniques can be classified into two categories: Empirical Bayes (EB) and Full Bayes (FB) modeling methods. In this section, we will refer to the Bayes method as the FB method to distinguish this approach from the EB method. Both are described in Chapter 7—*Before–After Studies in Safety*.

8.4.1 Empirical Bayes method

The observed crash frequency of a site may not necessarily provide enough information about its safety because the short-term crash frequency may vary significantly from the long-term crash mean. For sites with low crash frequencies, the effect is magnified because the changes due to variability in crash frequencies represent an even larger fluctuation compared with the expected average crash frequency. As a result, it is difficult to determine whether changes in the observed crash frequency are due to the changes in the site's performance or due to random fluctuations. To overcome this issue, the EB method is applied to estimate the expected crash frequency. It provides a better estimate of the long-term mean of the site. As explained in the previous chapter, the EB method combines the site's crash history with the predicted crash frequency from the crash-frequency model or SPF to calculate its expected crash frequency.

Although a few hazardous identification methods using the EB method have been proposed in the literature (PIARC, 2003; AASHTO, 2010), the

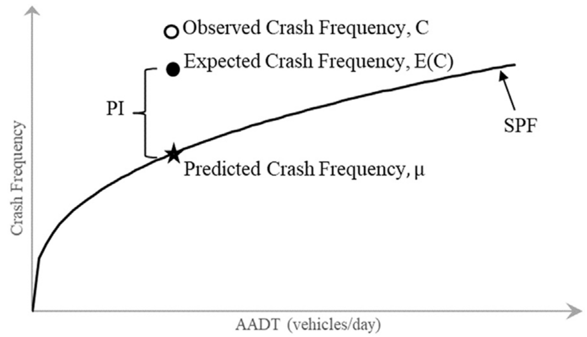

FIGURE 8.2 Graphical representation of potential for improvement using EB method.

potential for improvement using expected crashes (or PI based on EB) is a method that has been favored over the other EB-based methods. With this method, the site's expected average crash frequency is compared with the predicted average crash frequency from an SPF. The difference between the expected and predicted crash frequencies is the PI (see Fig. 8.2). Note the difference when compared with Fig. 8.1 is due to the random fluctuation with the crash occurrence. If this is not accounted for, then the crash reduction is overestimated, as shown in Fig. 8.1. When the PI is greater than zero, a site experiences more crashes than predicted. When the PI is less than zero, a site experiences fewer crashes than predicted. The PI based on the EB expected crashes for a site is calculated using the following equation.

$$PI_{EB} = E(C) - \mu \tag{8.13}$$

with,

$$E(C) = w \times \mu + (1 - w) \times C \tag{8.14}$$

$$w = \frac{1}{1 + (\alpha \times \mu)} \tag{8.15}$$

where PI_{EB} is the potential for improvement based on the expected crashes for the site, $E(C)$ is the expected crash frequency, and w is the weight of the predicted crash frequency.

This method does not provide a specific performance threshold. All the sites that have PI_{EB} greater than zero can be considered hazardous but the preference will be given to the sites with the highest PI_{EB}.

Exercise 8.8

Using dataset 8.1, identify hazardous intersections based on the potential for improvement using EB method. Use the same functional form and parameter estimates as in Exercise 8.6.

Calculate the weight of the predicted crash frequency at each site using Eq. (8.15). For intersection 1, the weight of the predicted crash frequency is

$$w = \frac{1}{1 + (\alpha \times \mu)} = \frac{1}{1 + (0.2423 \times 43.6)} = 0.086$$

The expected crash frequency $E(C)$ and PI for intersection 1 is estimated as

$$E(C) = 0.086 \times 43.6 + (1 - 0.086) \times 44 = 43.97 \ \text{crashes.}$$

$$PI_{EB} = 43.97 - 43.6 = 0.37$$

If sites with PI greater than 20 crashes are considered as hazardous locations, then intersections 11 and 13 are detected.

For roadway segments with different lengths, it is necessary to standardize the PI_{EB} because the excess crashes are a function of the expected crash frequency at the site. Recently, some researchers proposed considering the ratio instead of the difference, as shown in the following equation (Wunderlich et al., 2019; Geedipally et al., 2020).

$$PI_{EB} = \frac{E(C)}{\mu} \tag{8.16}$$

This new method also does not provide a specific performance threshold. All the sites that have PI_{EB} greater than one can be considered hazardous but the preference will be given to the sites with the highest PI_{EB}.

Srinivasan et al. (2011) determined that the EB method for hazardous site selection works well with longer road segments. The authors recommended aggregating the adjacent segments as long as the homogeneity still exists with respect to key variables such as road classification, traffic volume, terrain, number of lanes, and road width. Recently, Shirazi et al. (2020) developed guidelines on how spatial aggregation can be conducted (for more details, see Chapter 6—*Cross-Sectional and Panel Studies in Safety*). With aggregation, an entire segment would be identified as the site for investigation, as opposed to a small section where a cluster of crashes occurred. The methods that are used to identify the clusters where the crashes are concentrated are described in Section 8.6.

8.4.2 Bayes method

As discussed in the previous section, with the EB method, it is assumed that the covariate effect on crashes is known with certainty. This assumption is not always true as there could be a need to account for the uncertainty in inferences (e.g., the posterior distribution shown in Fig. 2.7). The Bayes or FB method can accommodate all uncertainties in the model (posterior estimates). The Monte Carlo Markov Chain simulation can be used to determine the posterior distribution by assuming a prior distribution initially and then iteratively computing and updating the posterior marginal. Some researchers have proposed using the FB method for identifying hazardous sites (e.g., Huang et al. (2009); Miranda-Moreno et al. (2009)). In general, the FB method is used to improve the accuracy of the crash estimates and as such it can be applied for any hotspot method that involves estimated crashes. The details for estimating crash-frequency models using the Bayes method can be found in Chapter 2—*Fundamentals and Data Collection*.

Huang et al. (2009) compared the EB method and the FB approach with different model specifications for the hazardous site identification using some of the evaluation methods presented in Section 8.9. The authors found that the FB models that accounted for cross-site random effects and/or serial correlation between crash observations in successive time periods at a specific site significantly outperformed the EB method. Guo et al. (2019) compared the performance of the EB method and the FB model that Fawcett et al. (2017) proposed for identifying hazardous sites. The authors found that the FB model outperformed the EB method in most cases, and the FB model improves the accuracy of hazardous site identification significantly. However, it is important to note that the FB method cannot be implemented as easily as the EB method. A safety analyst needs to weigh-in factors such as sample size, model complexity and the performance of the method when selecting an approach.

Compared with the previous methods, the Bayesian methods have the advantage of more accurately estimating the long-term mean of a site, although they are more complex to implement. This is true as long as the dataset used to calculate the prior estimates has the same characteristics as the dataset used for identifying sites with potential for improvements. In some cases, the models may need to be recalibrated for local conditions, as discussed in Chapter 6 – *Cross-Sectional and Panel Studies in Safety*. Between the EB and FB, the FB has been shown to be more reliable than the EB method, as described earlier. In addition, the FB method usually requires fewer observations than the EB method as the latter requires the development of a crash-frequency model using the maximum likelihood estimation method, which is more data intensive (Lord and Miranda-Moreno, 2008).

8.5 Combined criteria

It is common that each method may provide a different set of hazardous sites due to the fact that every network screening method gives a priority to a different factor. Every method has its own strengths and limitations so it is hard for a safety analyst to select one method over the other.

The RSM proposed using various combinations of criteria to reduce the limitations of individual screening methods. The following are three variants described in the RSM (PIARC, 2003):

1. *Combined thresholds*—each method's thresholds must be met for the site to be detected.
2. *Individual threshold*—if at least one method's threshold is met then the site is detected.
3. *Individual threshold and minimum criteria values*—sites are ranked in descending order using one method, and once the site reaches the minimum thresholds set for other criteria then it is detected.

Tsapakis et al. (2017) recommended a process for combining the results of different methods. First, they assigned a rank to each site provided by a specific method. For example, a particular site with the highest potential for improvement in the list is given the first rank. However, this site may be third in the list when sorted by crash rate. After the ranks were assigned, they allocated weights to each method based on their strengths and limitations and calculated the combined ranking weight. Higher weights were assigned to the more rigorous measures that account for more key factors mentioned earlier. On the contrary, methods just based on the observed crash data were assigned the lowest weights. The combined rank weight is calculated using the following equation.

$$W_r = \sum_{k=1}^{K} w_k \times r \tag{8.17}$$

where W_r is the combined rank weight for the site; w_k is the weighting factor assigned to a network screening method k; and, r is the rank of the site according to the network screening method *k*. The sum of weights for all screening methods must be equal to one. An analyst needs to assign the weights based on their judgment.

8.6 Geostatistical methods

With the evolution of geographical information systems, geospatial hotspot methods have gained increased attention due to their simplicity in application. The traditional methods, except the methods that rely only on observed crash data, typically need a large amount of data for identifying hazardous sites. However, traditional methods do not show a spatial correlation between the adjacent sites. Even if they account for the correlation, it is hard to visualize the safety of sites in a network. Geospatial methods consider the effects of unmeasured variables by accounting for spatial autocorrelation between the crash events over a geographic space and are considered the most popular tools for the visualization of crash data and hotspot analysis. A complete description of all the spatial data models is provided in Chapter 9—*Models for Spatial Data*. Clustering methods and Kernel density estimation are the common geospatial methods that have been applied in hazardous site identification.

8.6.1 Clustering methods

Clustering methods rely on grouping crash events into multiple clusters where events within a cluster share high similarities but there will be a minimal similarity between clusters. These methods detect whether a given point (i.e., crash) distribution differs from a complete random fashion (i.e., complete spatial randomness (CSR)) throughout the study area. For the identification of clusters, these methods evaluate the degree of association of a crash event with its surrounding events. The positive spatial correlation indicates that the crashes are clustered. These clustered areas are classified as hazardous areas for safety improvement.

8.6.1.1 K-means clustering

K-means clustering is a nonhierarchical clustering technique that has been used to analyze patterns in the distribution of crashes and identify hazardous sites (Levine et al., 1995; Kim and Yamashita, 2007; Anderson, 2009; Mauro et al., 2013; Selvi and Caglar, 2018). For the K-means method, a dataset is divided into K clusters with at least one object in each cluster. The clusters are continuously renewed in a cyclic pattern until the most suitable solution is achieved. As a result, the intracluster similarity of events is high and the intercluster similarity of events will be low. While using K-Means, the analyst needs to define an optimal number of K clusters before the procedure starts. The procedure attempts each combination of K clusters such that the sum of the distance from every point to each of the K clusters centers is minimized (Kim and Yamashita, 2007). A combination that provides the minimal sum of all distances (or all squared distances) is finally selected.

8.6.1.2 Ripley's K-function

The Ripley's K-function is a spatial analysis method that is used for identifying crash clusters (Yamada and Thill, 2004; Fan et al., 2018; Kuo and Lord, 2019). The K-function characterizes crash patterns and describes how these patterns occur over a given area of interest. Ripley's K-function informs whether crashes observed within a given radius are more or less than what would be expected under complete spatial randomness. By this, we can determine if the crashes appear to be dispersed, clustered, or randomly distributed throughout the study area. However, this method does not inform where crashes cluster.

The Ripley's K function is also called a multidistance spatial cluster analysis and is used to depict the spatial distribution of data points. This method defines the spatial patterns of data points (i.e., either clustered or feature-dispersed) over a range of distances. A number of variations of Ripley's original K function have been proposed. One common transformation of K function is in the following equation (Kuo and Lord, 2019).

$$K(d) = \sqrt{A\left(\sum_{i}\sum_{j,j\neq i}(K_{ij})\right)/(\pi \times n(n-1))} \tag{8.18}$$

where d is the distance between the points i and j, n is the total number of points, A is the area of the region containing all points, and K_{ij} is the weight. If the observed K value is larger than its expected K value for a particular distance (d), then the point pattern is considered as clustered instead of randomly distributed. When the observed K value is outside the 95% confidence interval, the distribution of the data points was significantly different from a random distribution at distance d.

8.6.1.3 Nearest neighborhood clustering

Nearest neighborhood cluster analysis is one of the classification techniques that has been used to identify crash clusters (Kim and Yamashita, 2007; Lv et al., 2009; Iranitalab and Khattak, 2017; Kuo and Lord, 2019). This hierarchical clustering technique is used to identify the cluster of points within each spatial distribution. The clusters of one distribution are then compared with clusters of the other distribution. The Euclidean distance is used to identify sets of points that are clustered more closely than would be expected by random chance. If the mean random distance between them is less than a minimum distance based on the standard error of a random distribution, then points are considered to be clustered. The average nearest neighbor (ANN) index is calculated using the following equation (Kuo and Lord, 2019).

$$ANN = \frac{\overline{d}}{\overline{\overline{\delta}}} = \frac{\overline{d}}{0.5 \times \sqrt{A/n}} \tag{8.19}$$

where $\overline{d}$ is the mean distance between each crash point and its nearest neighbor, δ is the mean distance of the points distributed randomly, A is the study area, and n is the number of crash points. When the ANN value is less than 1 and its Z-score significant, the dataset contains clustered points. If the ANN value is larger than 1 and its Z-score significant, the dataset points are dispersed. Otherwise, the observed points are randomly distributed.

8.6.1.4 Moran's I index

Moran's I index is one of the most commonly used indexes for identifying hazardous sites by measuring the autocorrelation between spatial units (Moons et al., 2009; Truong and Somenahalli, 2011; Yu et al., 2014). This index considers both attribute similarity and location in measuring the autocorrelation. The similarity of attribute values of two points is computed with reference to the mean value in the cluster. For the spatial correlation between units i and j, the Moran's I index is calculated using the following equation.

$$I(d) = \frac{\sum_i \sum_j w_{ij}(x_i - \overline{x})\left(x_j - \overline{x}\right)}{W \sum_i (x_i - \overline{x})^2 / n} \tag{8.20}$$

where $\overline{x}$ is the mean of x; w_{ij} is a matrix of spatial weights with zeroes on the diagonal (i.e., $w_{ii} = 0$), and W is the sum of all w_{ij}, $W = \sum\limits_i \sum\limits_j w_{ij}$. Distance d is used to determine the neighbors j.

The value for this index ranges from −1 to +1. A value closer to +1 indicates that the crashes occurred in clusters and a value closer to −1 indicates crashes are dispersed. A value closer to zero means data is spread randomly. In general, this method evaluates the extent to which the crash counts in a specific spatial unit vary with the crash counts in its neighboring spatial units. The hazardous road segments can be identified by aggregating the continuous spatial units that share similar traits. It should be noted that the global Moran's I statistic does not inform where crashes cluster. In such instances, local Moran's I should be used to identify local clusters. Chapter 9—*Models for Spatial Data* provides additional details about the spatial clustering using Moran's I Index.

8.6.1.5 Getis-Ord general $G^*(d)$

Getis-Ord Gi* (or simply $G^*(d)$) statistics is a measure that is used to investigate the variation and associations between spatial units. $G^*(d)$ statistics can be either global or local and is useful in identifying crash hot and cold spots within a geographical area (Songchitruksa and Zeng, 2010; Prasannakumar et al., 2011; Truong and Somenahalli, 2011; Soltani and Askari, 2017). Similar to the global Moran's I statistic, global $G^*(d)$ statistic does not inform where crashes cluster. In such instances, local $G^*(d)$ should be used to identify local clusters. More details about this statistic are provided in Chapter 9 – *Models for Spatial Data*.

8.6.2 Kernel density estimation

The kernel density estimation (KDE) is the most commonly used method for hazardous site selection because of its accuracy and consistency in prediction (Anderson, 2009; Kuo et al., 2013; Thakali et al., 2015). As opposed to clustering methods, KDE can define the extent of risk of a crash. Using the KDE method, the risk surrounding each crash can be calculated and the crash risk density is defined. When the distance is closer to zero, the crash risk density reaches the highest value and decreases with increased distance.

The two factors that influence the outcome of KDE are bandwidth and cell size. Bandwidth is used to define the uncertainty about the exact position of a crash. The values of bandwidth and cell size are defined using study area conditions and surrounding data characteristics. In the KDE method, a circular area (known as kernel) of defined bandwidth is created around each crash point. The distance from the point to a reference location based on a mathematical function is calculated. For each location, distances are summed for all the areas, including those at which no crashes of the indicator variable are recorded. This process is repeated for all successive crash points. As result, a kernel is placed over each crash event, and the sum of these individual kernels gives the density estimate for the distribution of road crashes. When the neighboring spatial units sharing a value of the local density estimate is higher than a given threshold, the site is identified as hazardous (Famili et al., 2018).

8.7 Crash concentration location methods

The selection of a high crash concentration location method is dependent on whether the analysis is related to nodes (e.g., intersections, or ramp terminals), segments, or facilities (i.e., combination of nodes and segments). For nodes, the influence area boundaries are predefined (e.g., 250 ft radius around the center of an intersection) and sites are simply screened by their ranks. For roadway segments, three methods are available.

8.7.1 Sliding window method

In the sliding window method, a window of a certain length is conceptually moved along a study segment from one end to another at specified increments. The selected hazardous site selection method's performance measure (e.g., crash rate or PI) is then calculated for each position of the window. From all the windows analyzed, the windows are ranked based on the values of the site selection method. At any window position, each segment is characterized by the maximum value calculated within or overlapping the beginning of the adjacent segment. With this, there is an increased chance of detecting a hazardous site at the screening stage if the safety problem in a window overlaps the adjacent site. Fig. 8.3 shows an example of conducting the sliding window method using a window length of 0.3 miles and an increment distance of 0.1 miles.

8.7.2 Peak searching method

The individual roadway segments are divided into windows of similar length and the windows do not span multiple segments. The window begins at the left limit of a roadway segment and increases in length incrementally until it reaches the end. At every increment, we have a

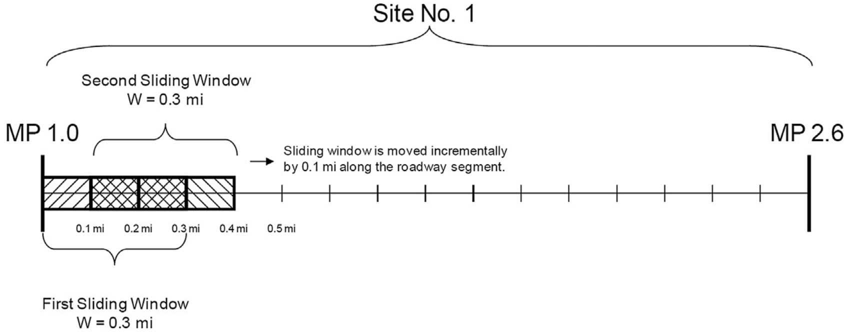

FIGURE 8.3 Illustration of sliding window method (Harwood et al., 2010).

specific window where an estimated measure (e.g., crash rate, EPDO score) can be calculated. The window that has the highest value of the estimate is identified and tested for statistical significance. The coefficient of variation (CV) is used as the test of significance. After determining the limited values of CV, the values below it pass the test. When a window passes the test, the entire road segment is ranked by the largest value of the estimate per mile. Else, the window size is increased and the process starts again for the road segment. With this method, localized safety problems are not overlooked by using too large a window and, in addition, the statistical test ensures that those are reliable estimates and not due to some randomness in the data.

Fig. 8.4 illustrates the main steps of the method. The roadway is first subdivided into 0.1-mile windows, with the exception of the last window, which may overlap with the previous window. The selected method's performance measure is then calculated for each window, and the resulting value is subject to a desired level of precision. If none of the 0.1-mile segments meet the desired level of precision, the segment window is increased to 0.2 miles, and the process is repeated until a desired precision is reached or the length of the window equals the entire segment length. For example, if the desired level of precision is 0.2, and the calculated coefficient of variation for each segment is greater than 0.2, then none of the segments meet the screening criterion and the segment length should be increased.

8.7.3 Continuous risk profile

The continuous risk profile (CRP) method has been initially proposed by Chung and Ragland (2009) that is based only upon observed crashes. The motivation for the development of the CRP method was to overcome the following: (1) risk is assumed to be a constant throughout the extension of the window; and, (2) all factors leading to high risk are assumed to reside within that window. The first assumption may not always hold true because of a particular factor at a specific point (e.g., utility pole closer to the road). For the second assumption, it is sometimes possible that the crashes within a window could result from factors that reside outside the window (e.g., secondary crash incidents due to another crash upstream).

The CRP method includes three main steps as described below:

- Filter out random noise using the weighted moving average technique and plot continuous crash risk profile along a study section of the highway using field data
- Calculate the predicted crash frequency for the study section based on the traffic flow (AADT) and corresponding SPFs, as shown in

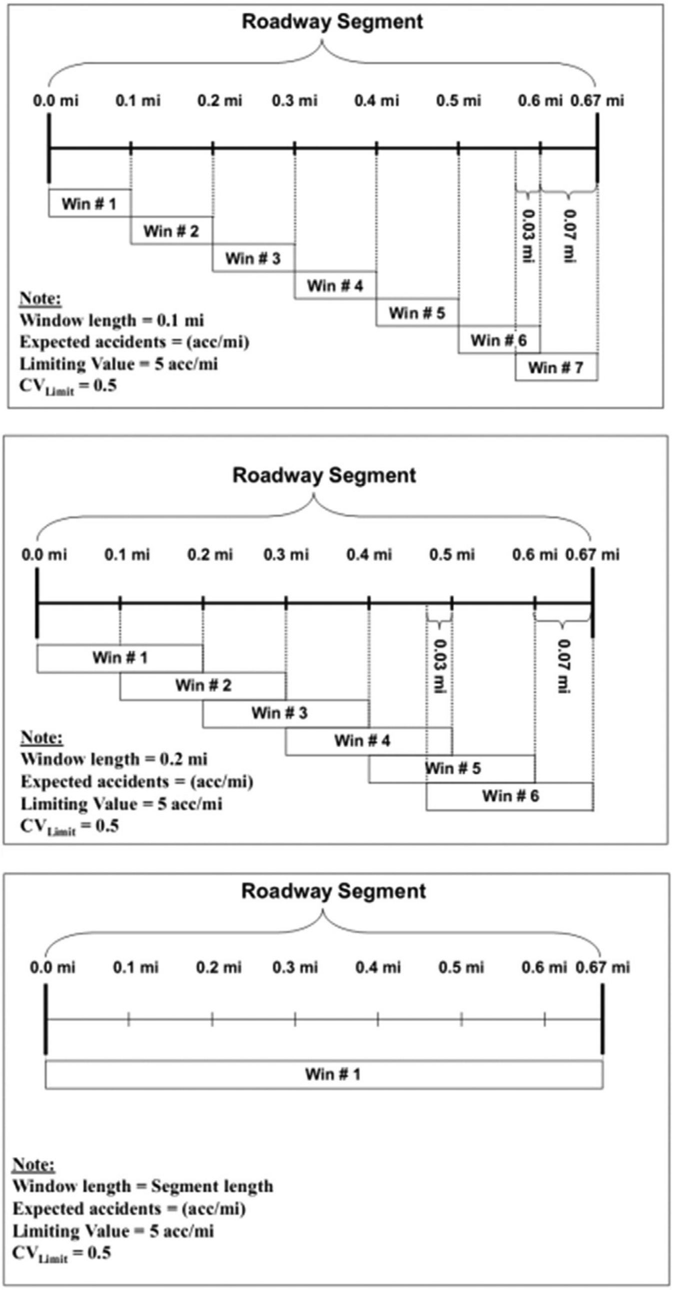

FIGURE 8.4 Illustration of peak searching method (Harwood et al., 2010).

Fig. 8.1. The predicted crash frequency should be in the unit that is used to plot the crash risk profile.
- Compare the predicted crash frequency (dark gray area labeled A in Fig. 8.5) with crash risk profile (i.e., compare the dotted line labeled SPF with the CRP). The location where the profile exceeds the predicted crash frequency is designated as the endpoints of a study site (see locations labeled *s* and *e*) (Grembek et al., 2012). The light gray area labeled B denotes the excess crash frequency.

8.8 Proactive methods

The traditional crash-based network screening methods are considered reactive in nature because sites are selected for potential safety improvement after the sites experience crashes (often over several years). As crash occurrence is a function of exposure, sites selected for potential safety improvement tend to be located in urban areas where traffic volumes are higher. For example, Fig. 8.6 shows a distribution of intersection safety performance based on the expected fatal and injury crash frequency at stop-controlled and signalized intersections in New Hampshire (Gross et al., 2016). If the crash-based methods are considered, then the selection for safety treatment will be mostly biased toward urban intersections. The crashes at rural intersections tend to be more dispersed around the network, so these intersections are seldom selected and rural areas may receive a disproportionally low safety investment. In rural areas, the types of crashes remain relatively consistent from year to year across the system while the locations of crashes tend to fluctuate.

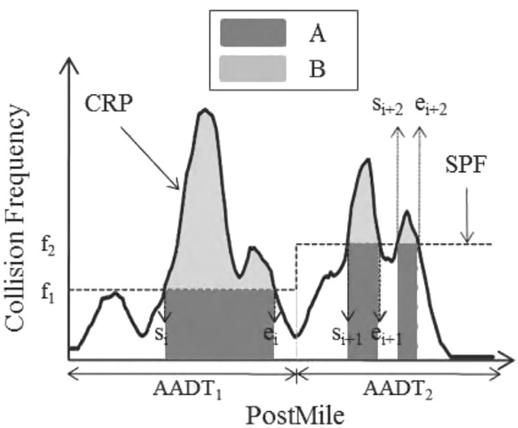

FIGURE 8.5 Illustration of the CRP method. *From Grembek, O., Kim, K., Kwon, O.H., Lee, J., Liu, H., Park, M.J., Washington, S., Ragland, D., Madanat, S.M., 2012. Experimental Evaluation of the Continuous Risk Profile (CRP) Approach to the Current Caltrans Methodology for High Collision Concentration Location Identification.*

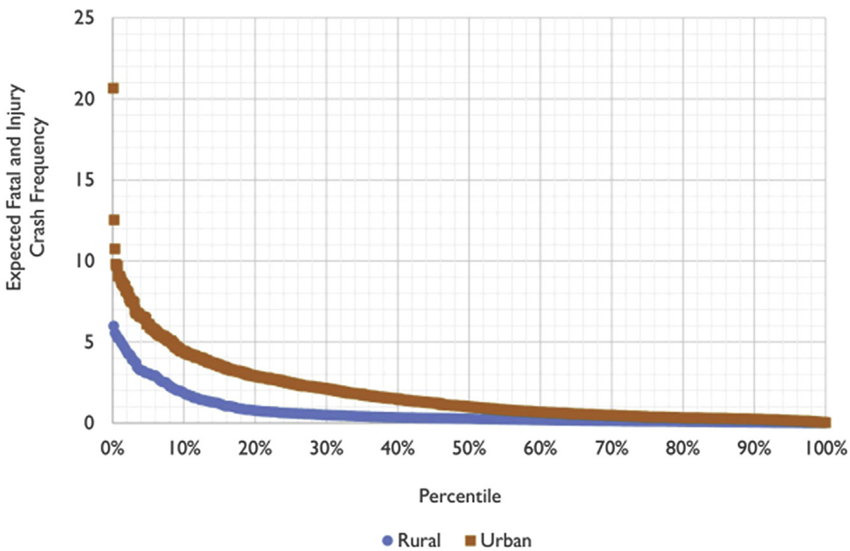

FIGURE 8.6 Distribution of intersection safety performance. *From Gross, F.B., Harmon, T., Bahar, G.B., Peach, K., 2016. Reliability of Safety Management Methods: Systemic Safety Programs. United States. Federal Highway Administration. Office of Safety.*

In recent years, agencies started supplementing traditional crash-based methods with systemic safety methods. Instead of relying on crash history, the systemic approach focuses on the selection and treatment of sites based on site-specific attributes such as roadway geometry and cross-sectional design, roadside and area features, operational characteristics, and others that are known to influence crash risk (Preston et al., 2013). The systemic approach targets crash types that occur with high frequency across the roadway network but are not concentrated at individual locations. These types of locations are often overlooked when ranking sites using traditional crash-history-based screening methods. The systemic approach is proactive in nature as sites can be prioritized for safety improvements even if they do not have a history of crashes (as its emphasis is on the characteristics of the site). Systemic safety analysis focuses on low-cost safety countermeasures that can be implemented broadly across the roadway network.

The following are the three main steps involved in the systemic approach for selecting the sites for safety treatment (Preston et al., 2013):

(1) Identify focus crash types and facility types,
(2) Develop risk factors, and
(3) Screen and prioritize candidate locations

8.8.1 Identify focus crash types and facility types

Focus crash types are the most prevalent severe crashes occurring in a given geographical region. These crash types are generally included in the highway safety plans developed by states, metropolitan areas, and other regional organizations. The roadways on which these focus crash types occur are the focus facility types. Agencies can use a decision tree diagram to determine the facility types where the focus crash types are most prevalent (for e.g., see Walden et al., 2015). Example of focus crash type is the roadway departure crashes, and the focus facility type is the rural two-lane highways.

8.8.2 Develop risk factors

Risk factors are roadway characteristics that are associated with an increased risk of the focus crash types. In this regard, crash-frequency models/SPFs can be used to prioritize the variables that are significant in influencing the focus crash types. Once the influential variables are identified, risk factors are determined by comparing the statistics of crash incidence (frequency and severity) with the exposure measures (mileage, vehicle miles traveled, or number of intersections). Different variables that are significant in influencing the focus crashes should be evaluated to identify the risk factors. As an example, for the horizontal curve density variable, Fig. 8.7 shows a comparison of roadway departure crashes with the proportion of existing highway vehicle miles traveled (VMT) (VMT is calculated as a product of segment length and the AADT) within the respective range. The figure clearly states that roadway departure crashes are overrepresented on roadways with a higher number of horizontal curves.

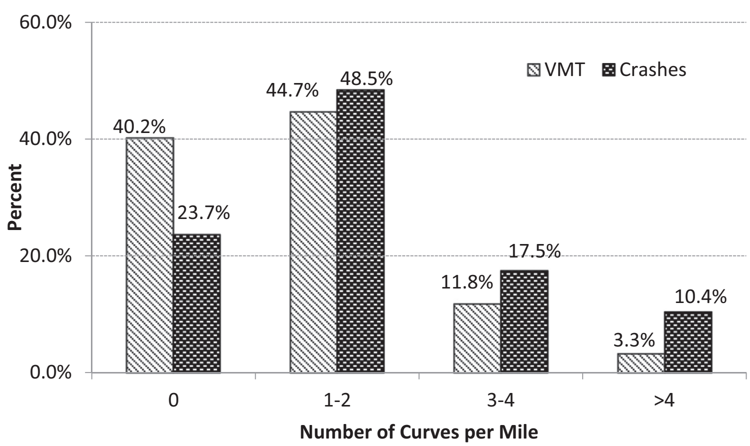

FIGURE 8.7 Proportion of roadway departure crashes and VMT as a function of curve density (Avelar et al. (2020)).

8.8.3 Screen and prioritize candidate locations

A simpler way to prioritize is to consider the locations that met the criteria for all risk factors. For example, all locations with at least four horizontal curves per mile, less than 11 ft lane, no shoulders and speed above 55 miles per hour will be prioritized. If the sample size is small, then the locations are prioritized even if they meet the criteria for fewer risk factors.

Alternatively, a scoring system can help to prioritize a variable and consequently, a safety project. For the variable of interest, a score is given for each predefined range based on the crash frequency and the over- or underrepresentation. Table 8.4 provides an example for the scoring based on the proportion of crash frequency and crash overrepresentation when compared to the roadway VMT (Walden et al., 2015).

Based on the scores provided in Table 8.4, the total score for a particular risk factor can be calculated using the following equation.

$$W_t = \begin{cases} 10 + CT + CO, & \mathit{if\ over - representation} \\ 10 + CT - CU, & \mathit{if\ under - representation} \end{cases} \tag{8.24}$$

where W_t is the computed total score, CT is the score based on crash total, CO is the score based on crash overrepresentation, and, CU is score based on crash underrepresentation.

Once the scores are computed for all sites, each location is ranked based on its computed total score compared with others. The sites that receive top scores are prioritized for systemic improvement. The selection of sites for systemic safety treatment depends on the funds availability, but top 5% of sites are typically considered a priority.

Surrogate measures of safety are another kind of proactive methods that can be used for identifying hazardous sites. These measures are based not on crashes, but on traffic conflicts, and are described in detail in Chapter 11 − *Surrogate Safety Measures.*

8.9 Evaluating site selection methods

As noted earlier, different site selection methods may provide a different set of hazardous sites because each method gives a priority to a different factor. Every method has its own strengths and limitations so it is hard for an analyst to choose one method over the other. A safety analyst may want to know which method performs better than other methods. Below are some of the methods that can be used to assess the superiority of a particular method. Cheng and Washington (2008) introduced the site consistency test, method consistency test, and total rank differences test, and Montella (2010) proposed the total score test. Note that Guo et al. (2020) has expanded on the methods proposed by

TABLE 8.4 Risk factor weight criteria (Walden et al., 2015).

	Score										
Category	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Crash frequency	≥0% and <10%	≥10 and <20%	≥20 and <30%	≥30 and <40%	≥40 and <50%	≥50 and <60%	≥60 and <70%	≥70 and <80%	≥80 and <90%	≥90 and <100%	100%
Crash overrepresentation	0%	>0% and <2%	≥2% and <3%	≥3% and <4%	≥4% and <5%	≥5% and <6%	≥6% and <7%	≥7% and <8%	≥8% and <9%	≥9% and <10%	≥10% and ≤100%
Crash underrepresentation	0%	>0% and <2%	≥2% and <3%	≥3% and <4%	≥4% and <5%	≥5% and <6%	≥6% and <7%	≥7% and <8%	≥8% and <9%	≥9% and <10%	≥10% and ≤100%

Cheng and Washington (2008) by allowing the methods to be evaluated over several years.[1] A few other evaluation methods in regards to hot spot identification such as Pearson's correlation coefficient percentage of dispersion deviations, mean absolute deviation, mean absolute error are also employed in previous studies but are not discussed here. For more details about these methods, the interested reader is referred to Huang et al. (2009) and Lim and Kweon (2013).

8.9.1 Site consistency test

The site consistency test (SCT) is used to measure the ability of a screening method to consistently identify a hazardous site in subsequent observational periods. This test is based on the logic that if there are no significant changes at the site that is identified as hazardous in time period i, then in the subsequent time period $i+1$, it should be inferior in terms of safety performance. The screening method that reports a larger number of crashes for all sites combined in the time period $i+1$ is considered as the superior method in identifying the sites with poor safety performance. The SCT statistic for roadway segments is calculated using the following equation.

$$SCT_k = \frac{\sum_{r=1}^{R} C_{r,k,i+1}}{\sum_{r=1}^{R} L_{r,k}} \tag{8.25}$$

where $C_{r,k,i+1}$ is the number of crashes at a site in the time period $i+1$ that is ranked r as identified by network screening method k, $L_{r,k}$ is the length of roadway segment for the r^{th} ranked site, and R is the rank threshold that is used as cut-off in the site identification (note that $r=1$ corresponds to the site with the poorest safety performance). The segment length in the denominator is needed because the number of hazardous sites and their total length identified by the screening methods may vary from each other. The SCT statistic for intersections is calculated using the following equation.

$$SCT_k = \frac{\sum_{r=1}^{R} C_{r,k,i+1}}{N_k} \tag{8.26}$$

where N_k is the total number of intersections identified as hazardous sites by network screening method k.

[1] These authors have created a tool that can automatically compare different methods. It can be accessed here: https://ceprofs.civil.tamu.edu/dlord/HSID_Evaluation/index.htm.

Jiang et al. (2014) pointed out that the above statistic is reasonable for crash frequency based site selection methods only. Other methods such as those based on the potential for improvement do not always identify high crash frequency sites. A site can experience a large number of crashes due to high vehicle exposure but may not necessarily be a high-risk site. This is especially true for freeways that experience significant safety issues compared to rural two-lane highways. Jiang et al. (2014) proposed a modified SCT statistic as shown in the following equation.

$$SCT_k = \frac{\sum_{r=1}^{R} PI_{r,k,i+1}}{\sum_{r=1}^{R} L_{r,k}} \tag{8.27}$$

where $PI_{r,k,i+1}$ is the potential for improvement at a site in the time period $i+1$ that is ranked r as identified by hazardous site selection method k.

8.9.2 Method consistency test

The method consistency test (MCT) measures the number of hazardous sites commonly identified in time period i, and in the subsequent time period $i+1$. This test is also based on the logic that in the absence of significant changes at the site that is identified as hazardous in the time period i, the site should remain hazardous in the subsequent time period $i+1$. The screening method that identifies higher number of common sites in two consecutive periods is considered as the superior method. The MCT statistic is calculated using the following equation.

$$MCT_k = \{x_{r=1}, x_{r=2}, \ldots, x_{r=R}\}_i \cap \{x_{r=1}, x_{r=2}, \ldots, x_{r=R}\}_{i+1} \tag{8.28}$$

where $x_{r=1}, x_{r=2}, \ldots, x_{r=R}$ are the hazardous sites that are ranked $1, 2, \ldots, R$ respectively based on their (poor) safety performance identified by the hazardous site selection method k.

8.9.3 Total rank differences test

The total rank difference test (TRDT) is the measure that compares the ranking of a site in time period i, with its ranking in the subsequent time period $i+1$. Similar to the previous two tests, this test is based on the assumption that, in the absence of significant changes, the rank of a hazardous site in the time period i, should remain the same in the subsequent time period $i+1$. The screening method that yields a smaller total rank difference is considered a superior method. The TRDT statistic is calculated using the following equation.

$$TRDT_k = \sum_{x=1}^{N} \left| R(x_{k,i}) - R(x_{k,i+1}) \right| \tag{8.29}$$

where $R(x_{k,i})$ is the rank of site x in time period i, and $R(x_{k,i+1})$ is the rank of site x in time period $i+1$ identified by the network screening method k. N is the total number of hazardous sites.

8.9.4 Total score test

The total score test (TST) combines all the above three methods to provide a synthetic index that gives an effectiveness measure relative to the methods being compared. The TST statistic is based on the assumption that all three tests above carry equal weight (Montella, 2010). The TST is calculated using the following equation.

$$TST_k = \frac{100}{3} \times \left[\left(\frac{SCT_k}{maxSCT} \right) + \left(\frac{MCT_k}{maxMCT} \right) + \left(1 - \frac{TRDT_k - minTRDT}{maxTRDT} \right) \right] \tag{8.30}$$

where $maxSCT$, $maxMCT$, and $maxTRDT$ refer to the maximum values of the respective test statistic calculated from all the performance measures being compared. The term $minTRDT$ refers to the minimum TRTD value obtained from all the performance measures being compared. The screening method that has highest TST value is considered as a superior method.

8.9.5 False identification test

The false identification test is used to compute the false positives (selecting a safe site as a hazardous site) and false negatives (selecting a hazardous risk site as a safe site) identified by a screening method. Many earlier studies have used this test to compare different site selection methods (e.g., see Cheng and Washington (2005) and Geedipally and Lord (2010)). For conducting this test, it requires that truly safe and hazardous sites are known a priori. This means real-world empirical data cannot be used because we will never know a priori what sites are truly hazardous. Simulation is usually preferred to empirical data because the hazardous sites can be known beforehand. Using simulation, we can establish sites that are a priori hazardous and can assess if the site selection methods work well for identifying these locations. This in turn helps to count the number of false positives and negatives. For details about the simulation design, the reader is referred to Geedipally and Lord (2010). According to a given site selection method, the outcome can be classified as shown in Table 8.5. These outcomes are used for evaluating the performance of different methods for identifying the true hazardous sites.

TABLE 8.5 Outcomes classified according to a given hazardous site selection method.

	Number of "true" hazardous sites	Number of "true" nonhazardous sites	
Number of sites "detected" as hazardous	T_P	F_P	D
Number of sites "detected" as nonhazardous	F_N	T_N	$n-D$
	H	$n-H$	n

where: n—total number of sites in the dataset under analysis; H—number of "true" hazardous; T_P—number of sites correctly classified as hazardous (true positives); F_P—number of false positives or Type I errors; F_N—number of false negatives or Type II errors; T_N—number of sites correctly classified as nonhazardous (true negatives).

The following measures can be used to evaluate the comparative performance of different screening methods in terms of the power to detect the "true" hazardous sites.

False positive rate. False positive rate (FPR) (also known as probability of false alarm) is the proportion of Type I errors among the "detected" hazardous sites. The method with a smaller FPR is considered to be the best among all available methods. The FPR statistic is calculated using the following equation.

$$FPR = \frac{F_P}{F_P + T_N} = \frac{F_P}{n - H} \tag{8.31}$$

False negative rate. False negative rate (FNR) (also known as Miss Rate) is the proportion of Type II errors among the "detected" nonhazardous sites. It is expected that the FNR is relatively small if the method performs well. The FNR statistic is calculated using the following equation.

$$FNR = \frac{F_N}{T_P + F_N} = \frac{F_N}{H} \tag{8.32}$$

Sensitivity. Sensitivity (SENS) (also known as true positive rate (TPR), power, or probability of detection) is the proportion of sites that have been correctly detected as hazardous sites. This criterion is interpreted as the capacity of the method to detect a "true" hazardous site in a group under analysis. This value should be closer to one if the method performs well. A method with low SENS only selects a few hazardous sites for safety treatment and thus, results in low safety improvement. The SENS statistic is calculated using the following equation.

$$SENS = \frac{T_P}{T_P + F_N} = \frac{T_P}{H} \tag{8.33}$$

Specificity. Specificity (SPEC) (also known as true negative rate or selectivity) represents the proportion of nonhazardous sites that have been correctly classified as "true" nonhazardous. This criterion gives the capacity of a method to detect "true" nonhazardous in a group under analysis. This value should be closer to one if the method performs well. A method with low SPEC selects a few nonhazardous sites for safety treatment and thus, results in wasting safety resources. SPEC is equivalent to 1-FPR. The SPEC statistic is calculated using the following equation.

$$SPEC = \frac{T_N}{T_N + F_P} = \frac{T_N}{n - H} \tag{8.34}$$

Risk. Risk (RISK) is the proportion of the total number of errors (Types I and II) and the number of sites under analysis. The RISK values close to 0 are expected when a method or a model performs well. The RISK statistic is calculated using the following equation.

$$RISK = \frac{F_P + F_N}{n} \tag{8.35}$$

Area under the ROC curve. The receiver operating characteristic (ROC) curve is obtained by plotting the FPR (i.e., 1—SPEC) on x-axis and TPR (i.e., SENS) on the y-axis (see Fig. 8.8). Each point on the ROC represents an FPR and TPR pair values corresponding to a specific decision threshold. For a screening method that has the highest accuracy, the ROC curve will be closer to the upper left corner. The two-dimensional area underneath the entire ROC curve is called area under the ROC curve.

8.9.6 Poisson mean differences

Poisson mean differences (PMD) gives different weights for each falsely identified site. More precisely, it is used to differentiate various

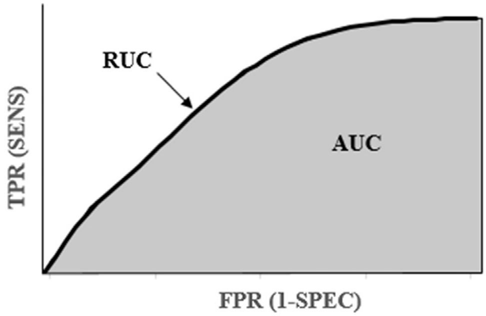

FIGURE 8.8 Area under the ROC curve.

false identifications and to quantify more clearly the adverse consequences that result from erroneous identifications of unequal importance (Cheng and Washington, 2008). It is the sum of the absolute difference of true Poisson means (TPMs) associated with the falsely identified sites and critical TPMs (for more details about this test, the reader is referred to Cheng and Washington, 2008). It should be noted that this method relies on simulation rather than observed data as the Poisson mean is not known with certainty. Although the PMD method is rarely used, it is included here for completeness. The model with a larger value of PMD is less desirable.

References

AASHTO, 2010. Highway Safety Manual, first ed. American Association of State Highway and Transportation Officials, Washington, D.C.

Anderson, T.K., 2009. Kernel density estimation and k-means clustering to profile road accident hotspots. Accid. Anal. Prev. 41 (3), 359−364.

Avelar, R., Geedipally, S., Das, S., Wu, L., Kutela, B., Lord, D., Tsapakis, I., 2020. Evaluation of Roadside Treatments to Mitigate Roadway Departure Crashes. Texas Department of Transportation.

Cheng, W., Washington, S., 2008. New criteria for evaluating methods of identifying hot spots. Transport. Res. Rec. 2083, 76−85.

Cheng, W., Washington, S.P., 2005. Experimental evaluation of hotspot identification methods. Accid. Anal. Prev. 37 (5), 870−881.

Chung, K., Ragland, D.R., 2009. The Continuous Risk Profile Approach for the Identification of High Collision Concentration Locations on Congested Highways. Transportation and Traffic Theory 2009: Golden Jubilee. Springer, pp. 463−480.

Famili A., Sarasua W.A., Ogle J. H., Shams A., and Asadi N., 2018. GIS based spatial analysis of pedestrian crashes: A case study of South Carolina, Proceedings of International Conference on Transportation and Development 2018: Connected and Autonomous Vehicles and Transportation Safety, Pittsburg, PA, pp. 368−376.

Fan, Y., Zhu, X., She, B., Guo, W., Guo, T., 2018. Network-constrained spatio-temporal clustering analysis of traffic collisions in jianghan district of wuhan, China. PLoS One 13 (4).

Fawcett, L., Thorpe, N., Matthews, J., Kremer, K., 2017. A novel bayesian hierarchical model for road safety hotspot prediction. Accid. Anal. Prev. 99 (Pt A), 262−271.

Geedipally, S.R., Le, M., Fitzpatrick, K., Avelar, R., 2020. Identifying High Risk Locations to Address Pedestrian Safety Issues. 99th Annual Meeting of the Transportation Research Board, Washington, D.C.

Geedipally, S.R., Lord, D., 2010. Identifying hot spots by modeling single-vehicle and multi-vehicle crashes separately. Transport. Res. Rec. 2147 (1), 97−104.

Grembek, O., Kim, K., Kwon, O.H., Lee, J., Liu, H., Park, M.J., Washington, S., Ragland, D., Madanat, S.M., 2012. Experimental Evaluation of the Continuous Risk Profile (CRP) Approach to the Current Caltrans Methodology for High Collision Concentration Location Identification.

Gross, F.B., Harmon, T., Bahar, G.B., Peach, K., 2016. Reliability of Safety Management Methods: Systemic Safety Programs. Federal Highway Administration. Office of Safety, United States.

Guo, X., Wu, L., Lord, D., 2020. Generalized Criteria for Evaluating Hotspot Identification Methods. Working Paper, Texas A&M University, College Station, TX, USA.

Guo, X., Wu, L., Zou, Y., Fawcett, L., 2019. Comparative analysis of empirical bayes and bayesian hierarchical models in hotspot identification. Transport. Res. Rec. 2673 (7), 111−121.

Harwood, D.W., Torbic, D.J., Richard, K.R., Meyer, M.M., 2010. Safetyanalyst: software Tools For Safety Management of Specific Highway Sites. Turner-Fairbank Highway Research Center.

Hauer, E., 1997. Observational Before-After Studies in Road Safety: Estimating the Effect of Highway and Traffic Engineering Measures on Road Safety Pergamon (Tarrytown, N.Y., U.S.A).

Hauer, E., Kononov, J., Allery, B., Griffith, M.S., 2002. Screening the road network for sites with promise. Stat. Methodol. App. Des. Data Anal. Eval. (1784) 27−32.

Huang, H., Chin, H.C., Haque, M.M., 2009. Empirical evaluation of alternative approaches in identifying crash hot spots. Transport. Res. Rec. 2103, 32−41.

Iranitalab, A., Khattak, A., 2017. Comparison of four statistical and machine learning methods for crash severity prediction. Accid. Anal. Prev. 108, 27−36.

Jiang, X., Abdel-Aty, M., Alamili, S., 2014. Application of Poisson random effect models for highway network screening. Accid. Anal. Prev. 63, 74−82.

Kim, K., Yamashita, E.Y., 2007. Using ak-means clustering algorithm to examine patterns of pedestrian involved crashes in honolulu, Hawaii. J. Adv. Transport. 41 (1), 69−89.

Kononov, J., Allery, B., 2003. Level of service of safety: conceptual blueprint and analytical framework. Transport. Res. Rec. 1840 (1), 57−66.

Kuo, P.-F., Lord, D., 2019. A promising example of smart policing: a cross-national study of the effectiveness of a data-driven approach to crime and traffic safety. Case Stud. Transp. Policy 7 (4), 761−771.

Kuo, P.-F., Lord, D., Walden, T.D., 2013. Using geographical information systems to organize police patrol routes effectively by grouping hotspots of crash and crime data. J. Transport Geogr. 30 (0), 138−148.

Levine, N., Kim, K.E., Nitz, L.H., 1995. Spatial analysis of honolulu motor vehicle crashes: I. Spatial patterns. Accid. Anal. Prev. 27 (5), 663−674.

Lim, I.-K., Kweon, Y.-J., 2013. Identifying high-crash-risk intersections: comparison of traditional methods with the empirical bayes−safety performance function method. Transport. Res. Rec. 2364 (1), 44−50.

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.

Lv, Y., Tang, S., Zhao, H., 2009. Real-time highway traffic accident prediction based on the k-nearest neighbor method. In: Proceedings of the 2009 International Conference on Measuring Technology and Mechatronics Automation, pp. 547−550.

MAG, 2010. Network Screening Methodology for Intersections. The Maricopa Association of Governments, Phoenix, AZ.

Mauro, R., De Luca, M., Dell'acqua, G., 2013. Using a k-means clustering algorithm to examine patterns of vehicle crashes in before-after analysis. Mod. Appl. Sci. 7 (10), 11.

Miranda-Moreno, L.F., Fu, L.P., Ukkusuri, S., Lord, D., 2009. How to incorporate accident severity and vehicle occupancy into the hot spot identification process? Transport. Res. Rec. 2102, 53−60.

Montella, A., 2010. A comparative analysis of hotspot identification methods. Accid. Anal. Prev. 42 (2), 571−581.

Moons, E., Brijs, T., Wets, G., 2009. Identifying hazardous road locations: hot spots versus hot zones. Trans. Comput. Sci. VI 288−300. Springer.

PIARC, 2003. Roadway Safety Manual World Road Association.

Prasannakumar, V., Vijith, H., Charutha, R., Geetha, N., 2011. Spatio-temporal clustering of road accidents: GIS based analysis and assessment. Proced. Soc. Behav. Sci. 21, 317−325.

Preston, H., Storm, R., Dowds, J.B., Wemple, B., Hill, C.M., Systematics, C., 2013. Systemic Safety Project Selection Tool. Federal Highway Administration. Office of Safety, United States.
Qin, X., Laracuante, L., Noyce, D.A., Chitturi, M.V., 2009. Systemwide Intersection Safety Prioritization Development and Assessment.
Selvi, H.Z., Caglar, B., 2018. Using cluster analysis methods for multivariate mapping of traffic accidents. Open Geosci. 10 (1), 772−781.
Shirazi, M., Geedipally, S.R., Lord, D., 2020. A Simulation Analysis to Study the Temporal and Spatial Aggregations of Safety Datasets with Excess Zero Observations.
Soltani, A., Askari, S., 2017. Exploring spatial autocorrelation of traffic crashes based on severity. Injury 48 (3), 637−647.
Songchitruksa, P., Zeng, X., 2010. Getis−ord spatial statistics to identify hot spots by using incident management data. Transport. Res. Rec. 2165 (1), 42−51.
Srinivasan, R., Ullman, G., Finley, M., Council, F., 2011. Use of empirical bayesian methods to estimate temporal-based crash modification factors for construction zones. In: Proceedings of the Transportation Research Board 90th Annual Meeting.
Thakali, L., Kwon, T.J., Fu, L., 2015. Identification of crash hotspots using kernel density estimation and kriging methods: a comparison. J. Mod. Transp. 23 (2), 93−106.
Truong, L.T., Somenahalli, S.V., 2011. Using GIS to identify pedestrian-vehicle crash hot spots and unsafe bus stops. J. Pub. Transp. 14 (1), 6.
Tsapakis, I., Dixon, K., Li, J., Dadashova, B., Holik, W., Sharma, S., Geedipally, S., Le, J., 2017. Innovative Tools and Techniques in Identifying Highway Safety Improvement Projects: Technical Report. Texas A&M Transportation Institute.
Walden, T.D., Lord, D., Ko, M., Geedipally, S., Wu, L., 2015. Developing Methodology for Identifying, Evaluating, and Prioritizing Systemic Improvements.
Wunderlich, R., Dixon, K., Wu, L., Geedipally, S., Dadashova, B., Shipp, E., 2019. Making Every Day Count: Applying Data-Driven Safety Analyses in a Txdot District. Texas Department of Transportation.
Yamada, I., Thill, J.-C., 2004. Comparison of planar and network k-functions in traffic accident analysis. J. Transport Geogr. 12 (2), 149−158.
Yu, H., Liu, P., Chen, J., Wang, H., 2014. Comparative analysis of the spatial analysis methods for hotspot identification. Accid. Anal. Prev. 66, 80−88.