Identification of hazardous sites

Fall 2021

INTRODUCTION

- There exists a lot of methods used for identifying high risk locations or sites that experience more crashes than one would expect.
- Although a lot of methods exist, there is still a significant amount of research currently done on this topic.
- The goal of the proposed methods consists of identifying sites that have abnormal number of crashes.
- In other words, given the characteristics of the site, it experiences more crashes than sites having the same characteristics.
- This assumption is a little tricky, because sites that experience abnormal safety records often have unusual characteristics.

General 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.
- 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.

General Characteristics

- Network screening can be conducted using either a reactive or a proactive approach.
  - The reactive approach relies on analyses of historic crash data
  - 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. (Focus of this lecture)

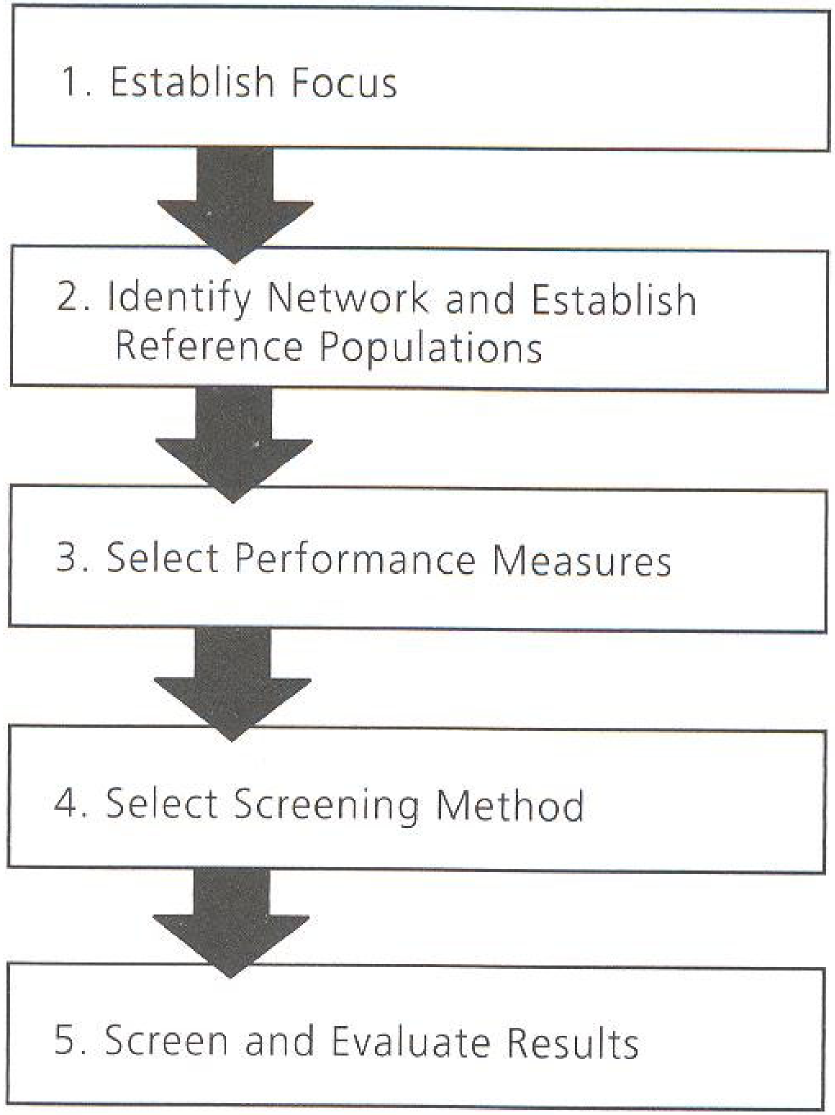
General Characteristics

- 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 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.

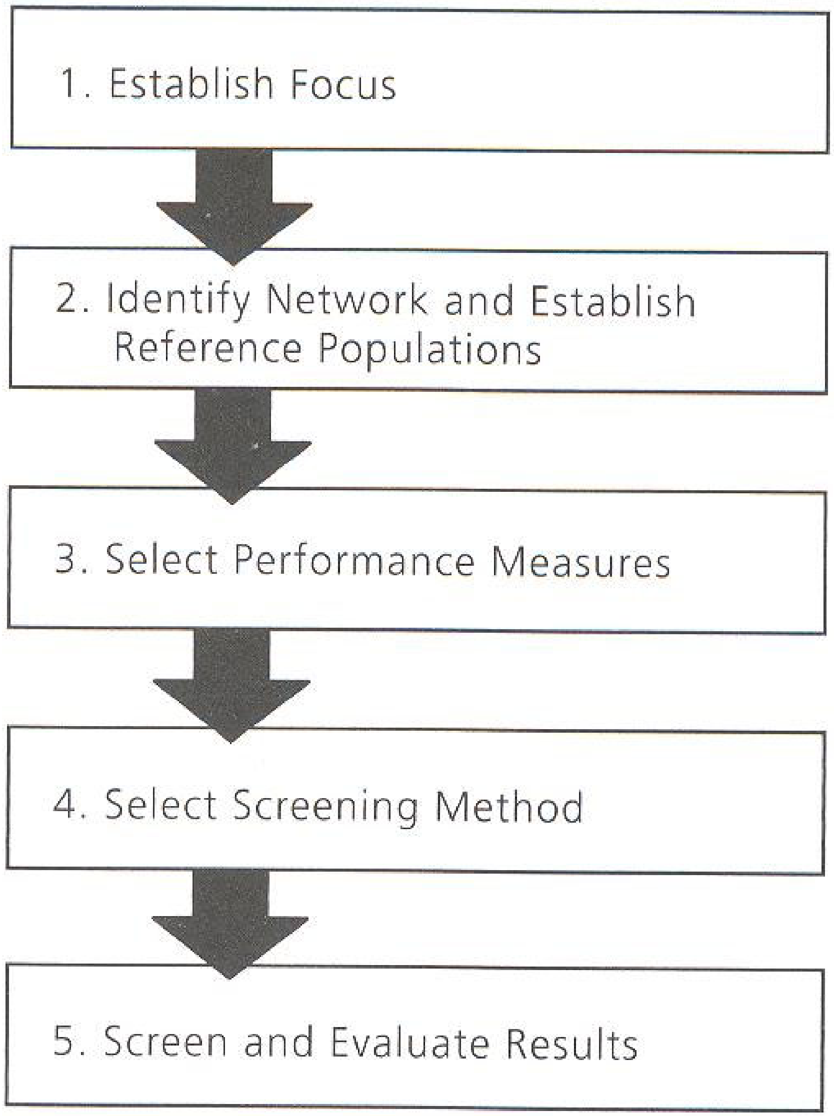
General Characteristics

- 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.

Network Screening Process - HSM

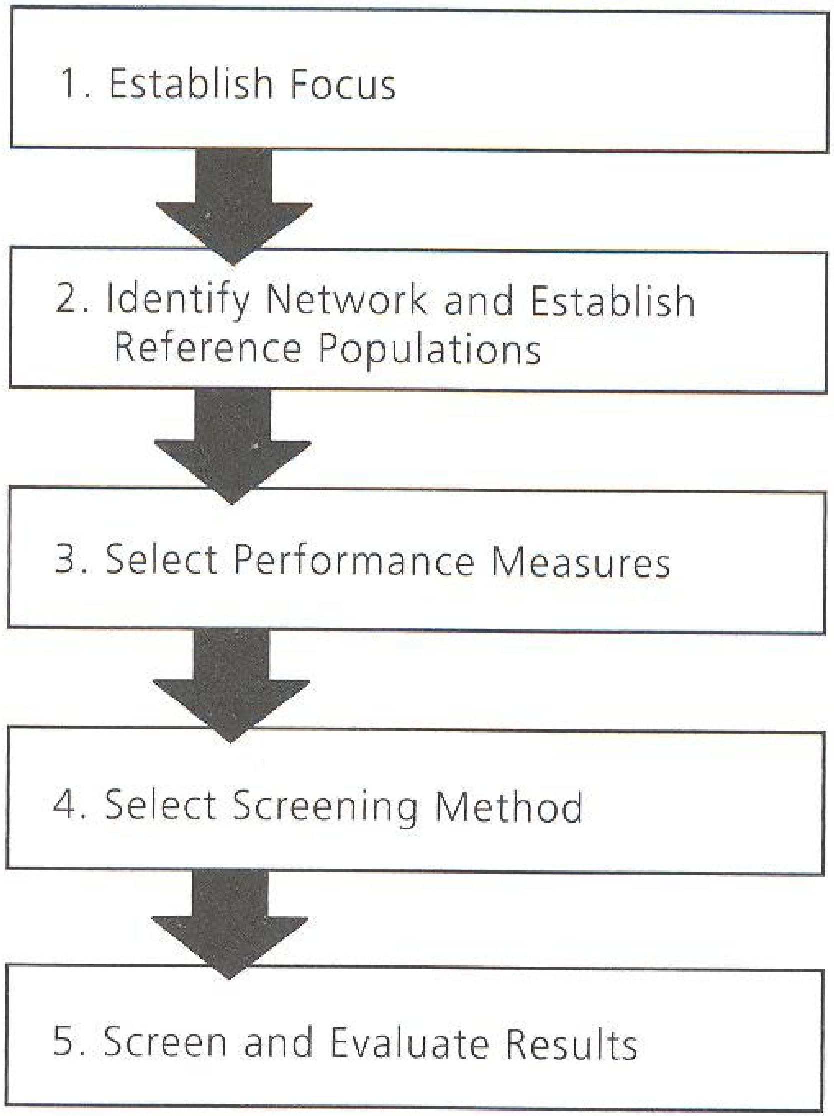


Network Screening Process - HSM



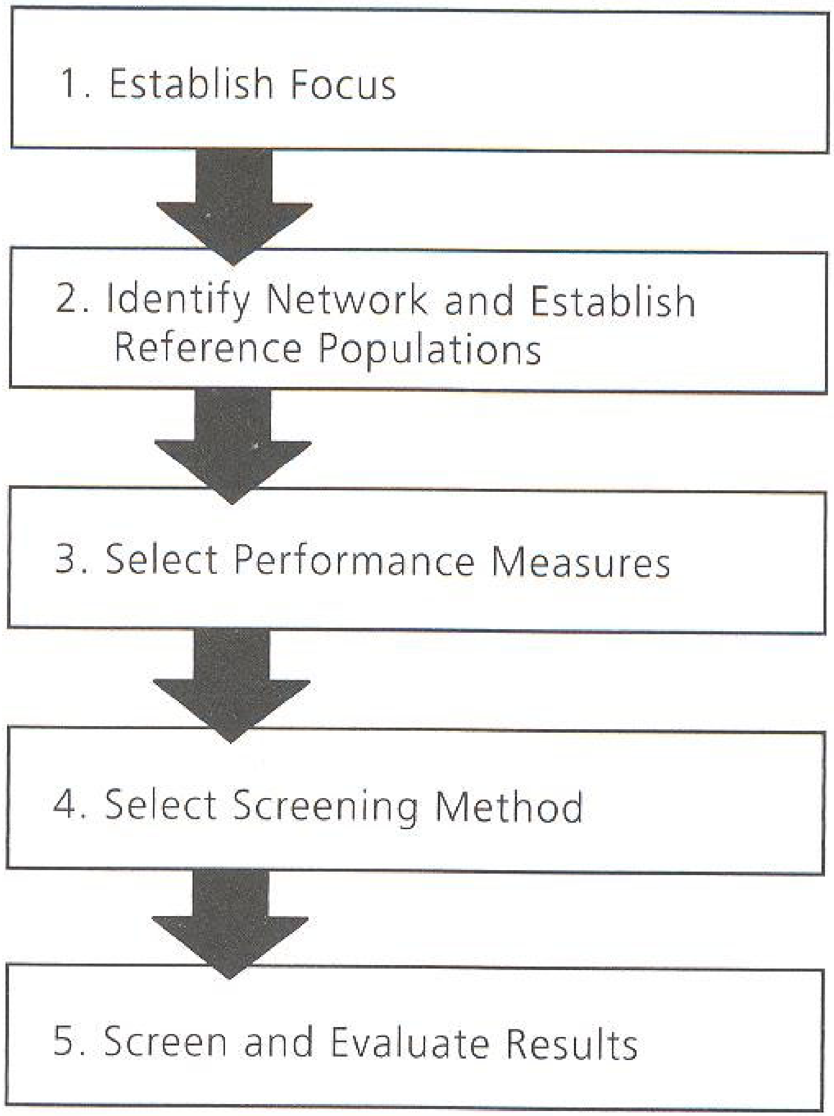
Identify purpose and target specific crashes

Network Screening Process -HSM



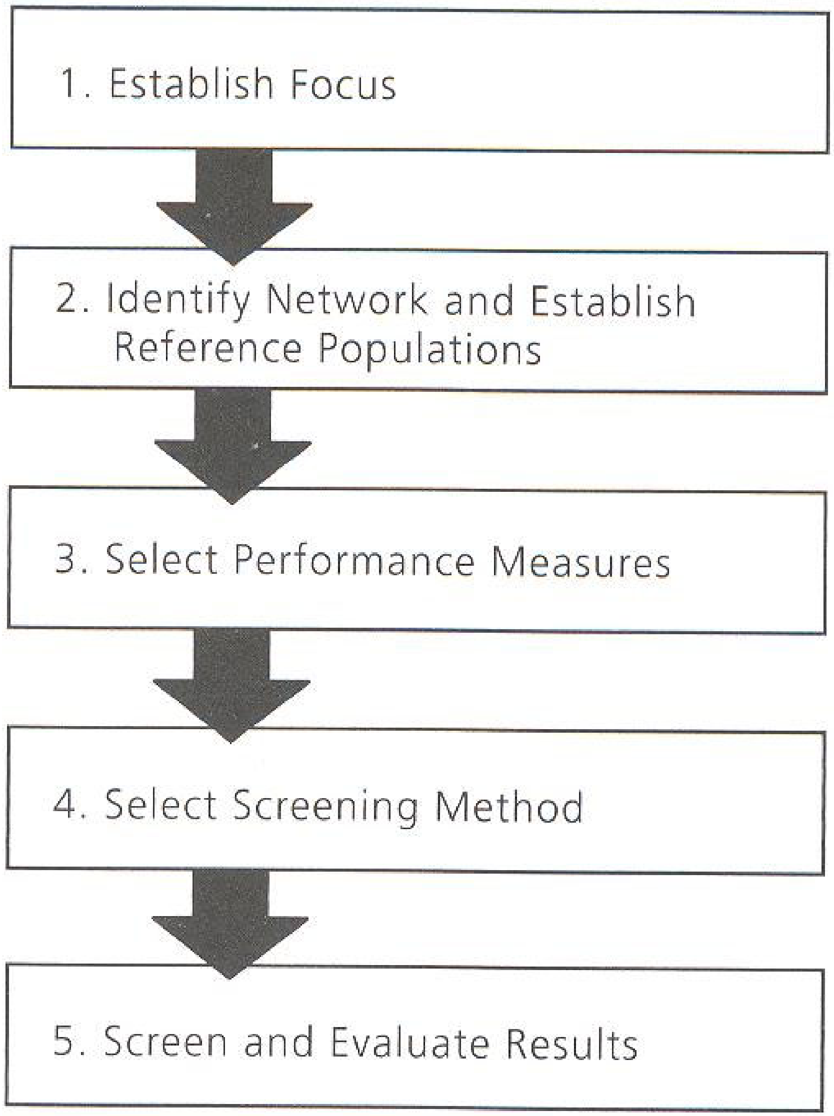
Types of sites or facilities (i.e., segments, intersections, etc.) Identify control group with sites that have similar characteristics (if needed)

Network Screening Process - HSM



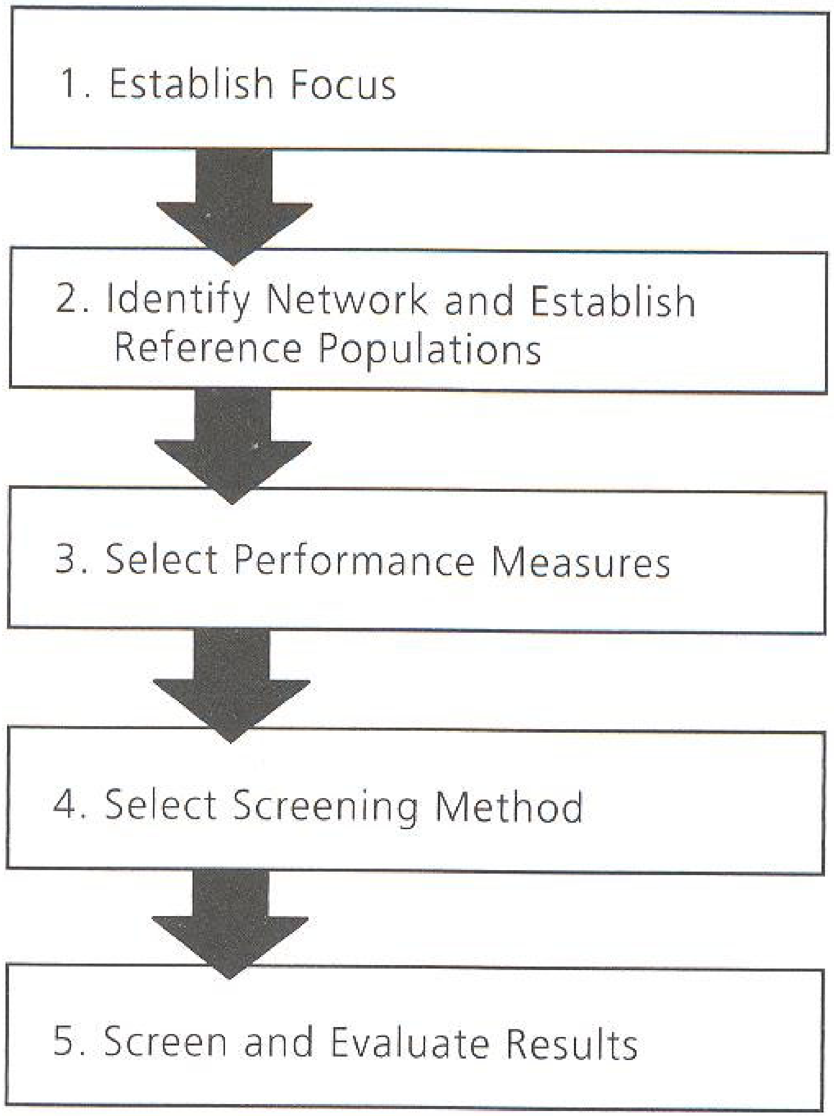
Use one or more of these methods (PM) (discussed below)

Network Screening Process - HSM



Four methods: ranking, sliding window, peak searching and continuous risk profile.

Network Screening Process - HSM



Conduct screen analysis and evaluate results

Data Used for Exercises

Intersection number	ADT on major st	ADT on minor st	Fatal crashes	Injury crashes	PDO crashes	Total crashes
1	37191	16705	0	18	26	44
2	19999	390	0	6	9	15
3	25608	2530	0	9	14	23
4	27223	8463	0	10	19	29
5	20336	7013	0	2	4	6
6	7524	1875	0	5	12	17
7	31646	17502	0	10	30	40
8	32117	9888	0	20	39	59
9	25224	17258	0	16	31	47
10	6856	5509	0	6	7	13
11	15025	4400	0	18	30	48
12	6856	390	0	0	1	1
13	14544	6720	1	15	27	43
14	9881	5806	0	9	5	14
15	39173	4324	0	8	12	20
16	14544	2741	0	14	14	28
17	9544	5089	0	9	11	20
18	8855	7704	0	4	7	11
19	8855	1714	0	2	1	3

50 sites in the dataset

PM: CRASH FREQUENCY

Characteristics:

Simplest method of identification

Sites ranked by crash frequency

Advantages:

Very simple

Sites with high frequency readily identified

Disadvantages:

Bias towards high volume sites (site selection effects)

Do no consider long-term mean

PM: CRASH FREQUENCY

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.

PM: CRASH RATE

Characteristics:

Ratio between crashes and exposure

Advantages:

Common method used by DOTs

Includes traffic exposure

Disadvantages:

Traffic volume needs to be known for every site

Does not include long-term mean

Non-linear relationship between crashes and exposure

PM: CRASH RATE

One site:

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

Average of all sites:

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

PM: CRASH RATE

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.

PM: RATE QUALITY CONTROL METHOD

Characteristics:

- Developed by industrial engineers for quality control purposes
- Sites higher than threshold identified as abnormal

Advantages:

- Consider randomness of crashes
- Includes traffic exposure

Disadvantages

- Complex methodology (for practicing engineers)
- Non-linear relationship between crashes and exposure

PM: RATE QUALITY CONTROL METHOD

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

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).

PM: RATE QUALITY CONTROL METHOD

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.

PM: EQUIVALENT PROPERTY DAMAGE ONLY

Characteristics:

Assign weights to different crash severity

PDO: 1, Minor Injury: 3.5, Serious Injury: 9.5

$$f_s = \frac{Cost_s}{Cost_{PDO}} \qquad EPDO\ score = \sum_s f_s \times C_s$$

Advantages:

Takes into consideration crash severity

Disadvantages:

Does not include exposure

Does not consider long-term mean

Bias towards high-speed sites

PM: EQUIVALENT PROPERTY DAMAGE ONLY

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.

PM: RELATIVE SEVERITY INDEX

Characteristics:

Consider severity of trauma sustained in any given crashes (to compute crash costs)

Assign weights to the average crash severity of certain types

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

C → Crashes

Advantages:

Takes into consideration crash severity

Reduces outside influences on crash severity (e.g. age of driver)

Disadvantages:

Does not include exposure

Does not consider long-term mean

Bias towards high-speed sites

PM: RELATIVE SEVERITY INDEX

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.

PM: COMBINED CRITERIA

Characteristics:

Avoid using the pitfalls of one single method

Combined Threshold:

More than one method used at the same time (e.g., 5+ frequency and 3+ for crash rate)

Individual Threshold and Minimum Criteria:

Sites are ranked by one method and sites ranked high are investigated using another method

PM: COMBINED CRITERIA

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

With

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

$$CT = \sum Cost_t \times N_v$$

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).

PM: STATISTICAL MODELS/PREDICTION

Characteristics:

Develop statistical model(s) using the reference population

Compare observed value with predicted value

Advantages:

Account for non-linear relationship between exposure and crashes

More accurate

Disadvantages:

Relatively complex

Do not account for long-term mean (for the comparison)

PM: STATISTICAL MODELS/PREDICTION

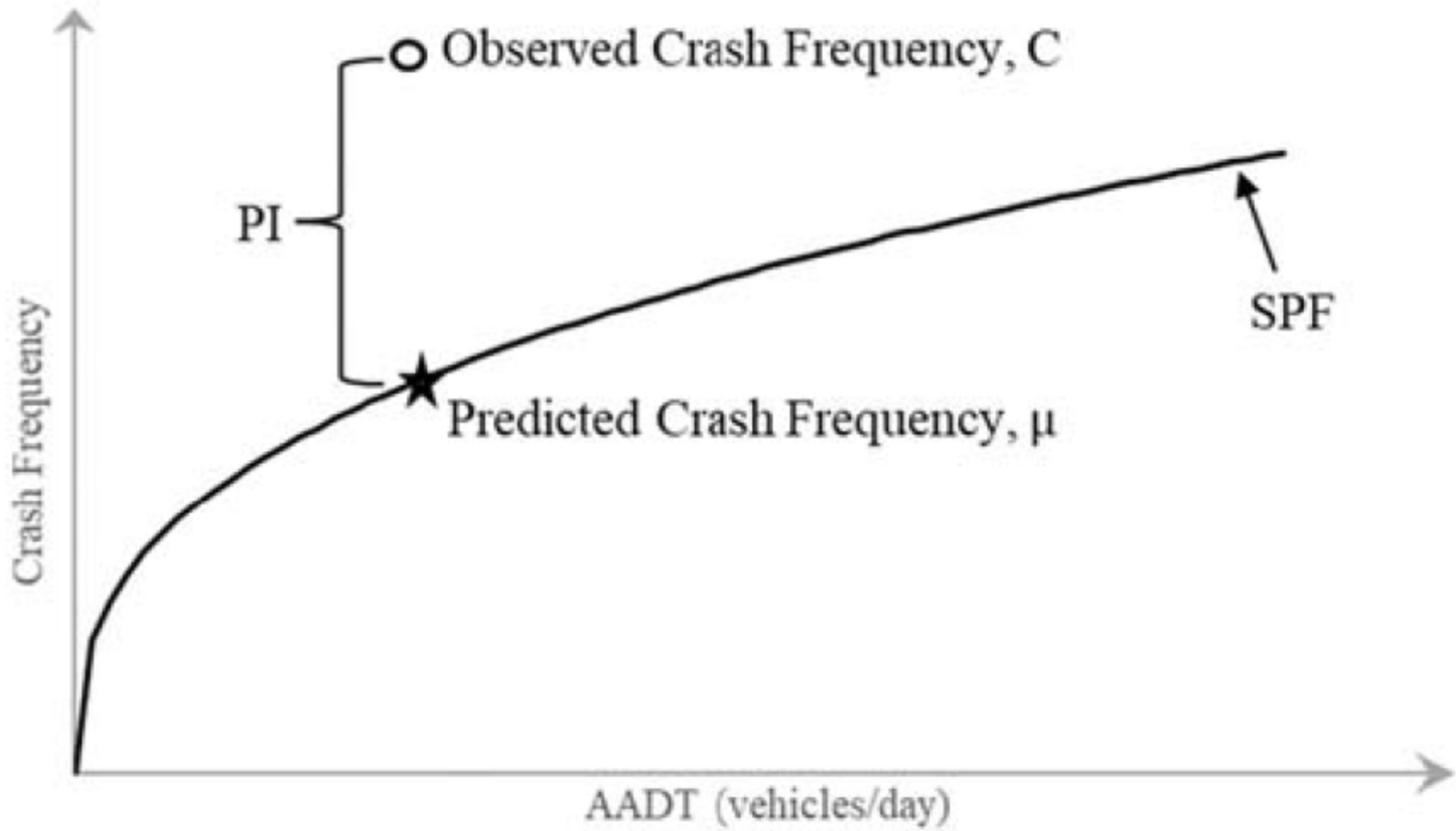


FIGURE 8.1 Graphical representation of potential for improvement using SPF.

PM: STATISTICAL MODELS/PREDICTION

$$PI = C - \mu$$

The potential for safety improvement method has also been defined as "identification of sites with promise." This method consists of comparing the observed or predicted values at given site with predicted values estimated from the reference population. The difference between the two indicates that the site could potentially reduce its number of crashes to those of the reference population.

PM: POTENTIAL FOR SAFETY IMPROVEMENT

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.

PM: LEVEL OF SERVICE OF SAFETY

Characteristics:

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.

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.

Advantages:

Account for non-linear relationship between exposure and crashes

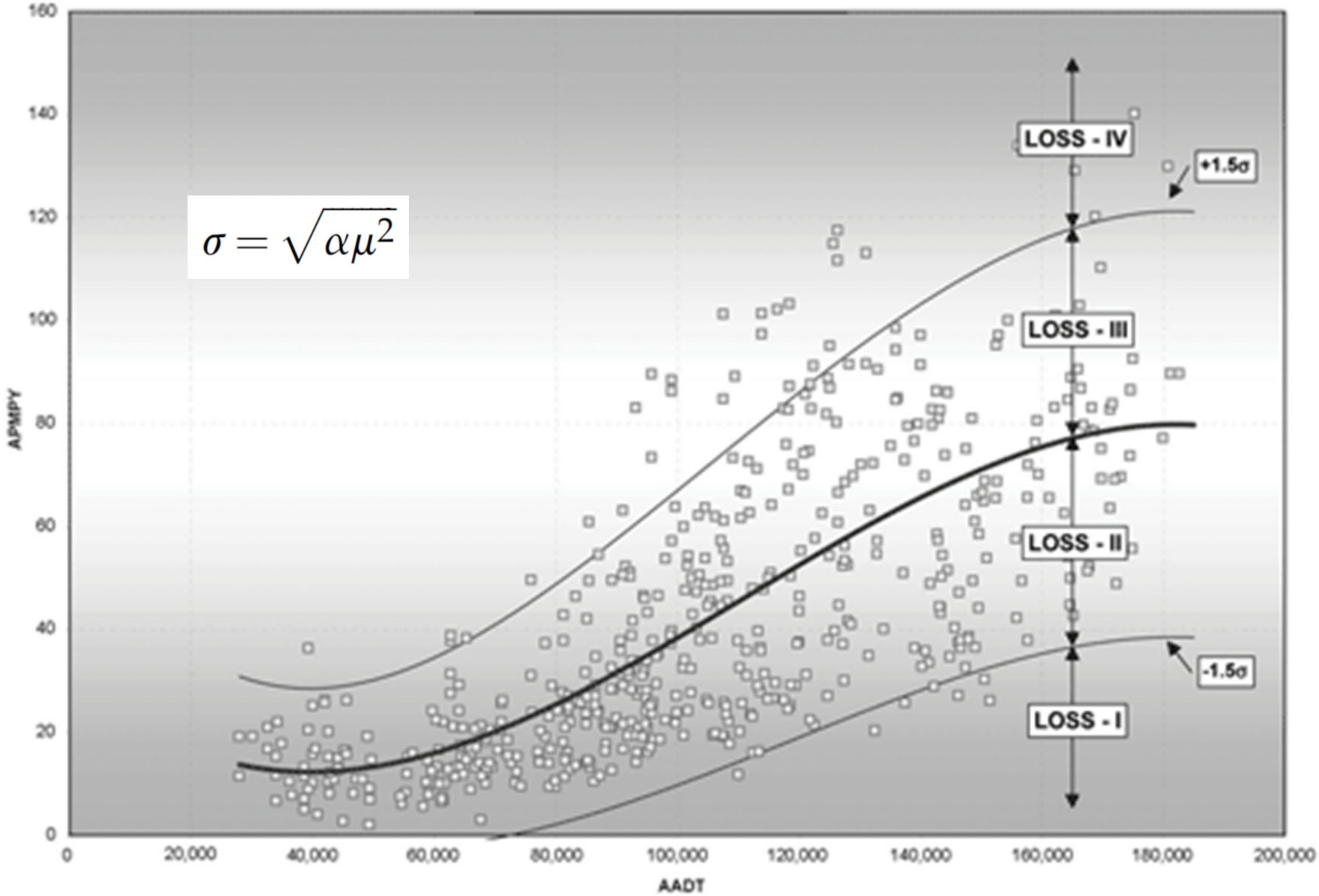
Provide more information than the previous method

Disadvantages:

Relatively complex

Do not account for long-term mean (although the method has recently been updated by including the EB method)

PM: LEVEL OF SERVICE OF SAFETY



- LOSS-I - Indicates low potential for crash reduction;
- LOSS-II - Indicates low to moderate potential for crash reduction;
- LOSS-III - Indicates moderate to high potential for crash reduction; and
- LOSS-IV - Indicates high potential for crash reduction.

PM: LEVEL OF SERVICE OF SAFETY

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	$(\mu - 1.5\sigma) \leq C < \mu$	Low to moderate potential for crash reduction
III	$\mu \leq C < (\mu + 1.5\sigma)$	Moderate to high potential for crash reduction
IV	$C \geq (\mu + 1.5\sigma)$	High potential for crash reduction

PM: LEVEL OF SERVICE OF SAFETY

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.

PM: EMPIRICAL BAYES METHOD

Characteristics:

Use information from the reference population and the observed at the site

Characteristics of the reference population can be estimated via the method of moments or statistical models

Advantages:

So far, most accurate method

Take into consideration long-term mean

Disadvantages:

Relatively complex

PM: EMPIRICAL BAYES METHOD

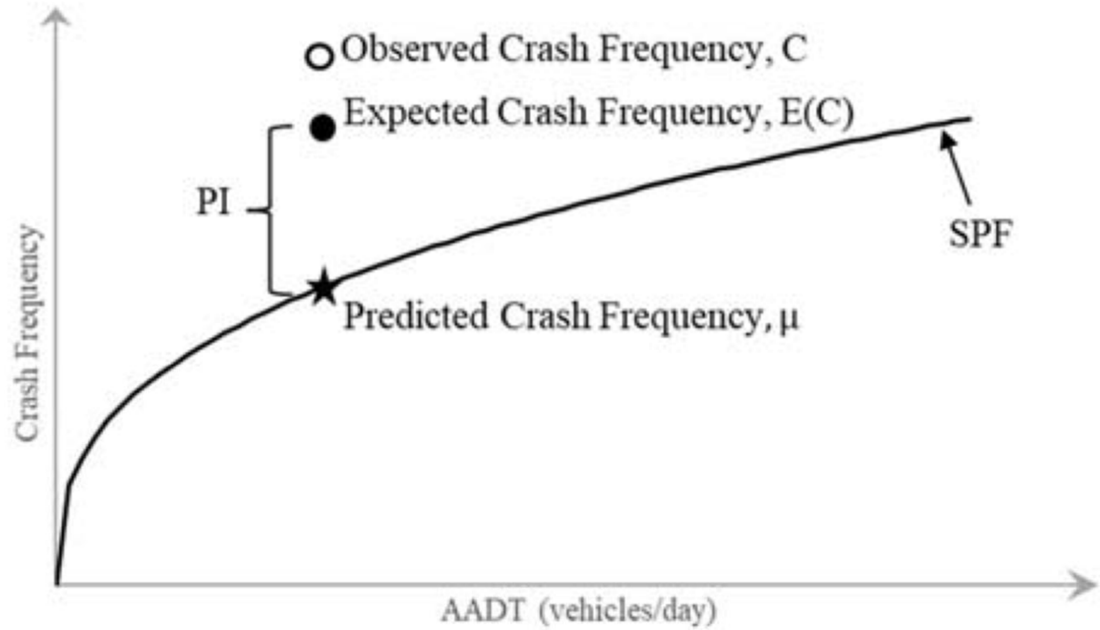


FIGURE 8.2 Graphical representation of potential for improvement using EB method.

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

$$E(C) = w \times \mu + (1 - w) \times C$$

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

PM: EMPIRICAL BAYES METHOD

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.

PM: FULL BAYES METHOD

Characteristics:

Relatively new method that ranks sites using posterior probabilities that a site experience more crashes than expected

Advantages:

Includes all covariates of the model for the ranking process

Provide probably best estimate for identification purposes

Disadvantages:

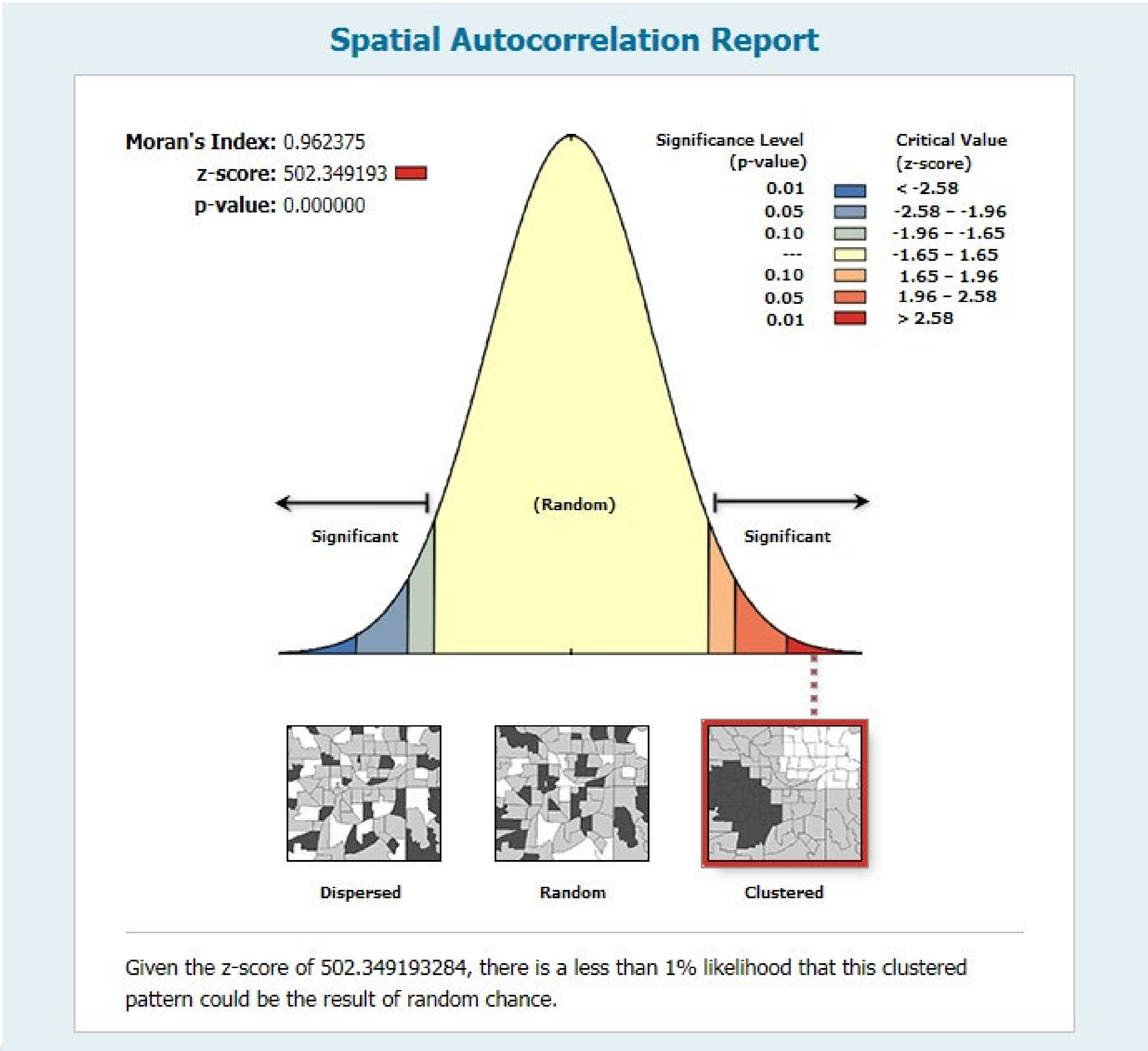
Highly complex

See Miaou and Song (Vol. 37(4), 2005, pp. 699-720) and Miranda-Moreno et al. (TRR 2102, 2009, pp. 53-60) for additional information

PM: GEOSTATISTICAL METHODS

- Clustering methods
  - K-means clustering
  - Ripley's K-function
  - Nearest neighborhood clustering
  - Moran's I index
  - Getis-Ord general G*(d)
- Kernel density estimation
  - See Chapter 9 for a detailed description

PM: GEOSTATISTICAL METHODS



https://blogs.oregonstate.edu/geo599spatialstatistics/2016/06/08/spatial-autocorrelation-morans/

Screening Methods

Ranking

Performance measures are applied to all the sites and ranked with each other.

Sliding Window

A window with a specified length (e.g., 0.3 mile) is conceptually moved along a road from beginning and end in increments of a specified size (e.g., 0.1 mile). Only valid for highway segments (unless intersections are included as part of the segment).

Peak Searching Method

Similar to the sliding window. In this case, you divide each segment into small windows of equal length (say 0.1 mile), use one of the PMs, calculate the average and variance, and estimate the coefficient of variation (COV). If the COV is smaller than a predetermine value (0.25) for one or more sites, then the sites are identified as hazardous and are ranked by order.

$$COV = \frac{\sqrt{Var(\text{Performance Measure})}}{Mean(\text{Performance Measure})}$$

Continuous Risk Profile

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.

Screening Methods

Sliding Window

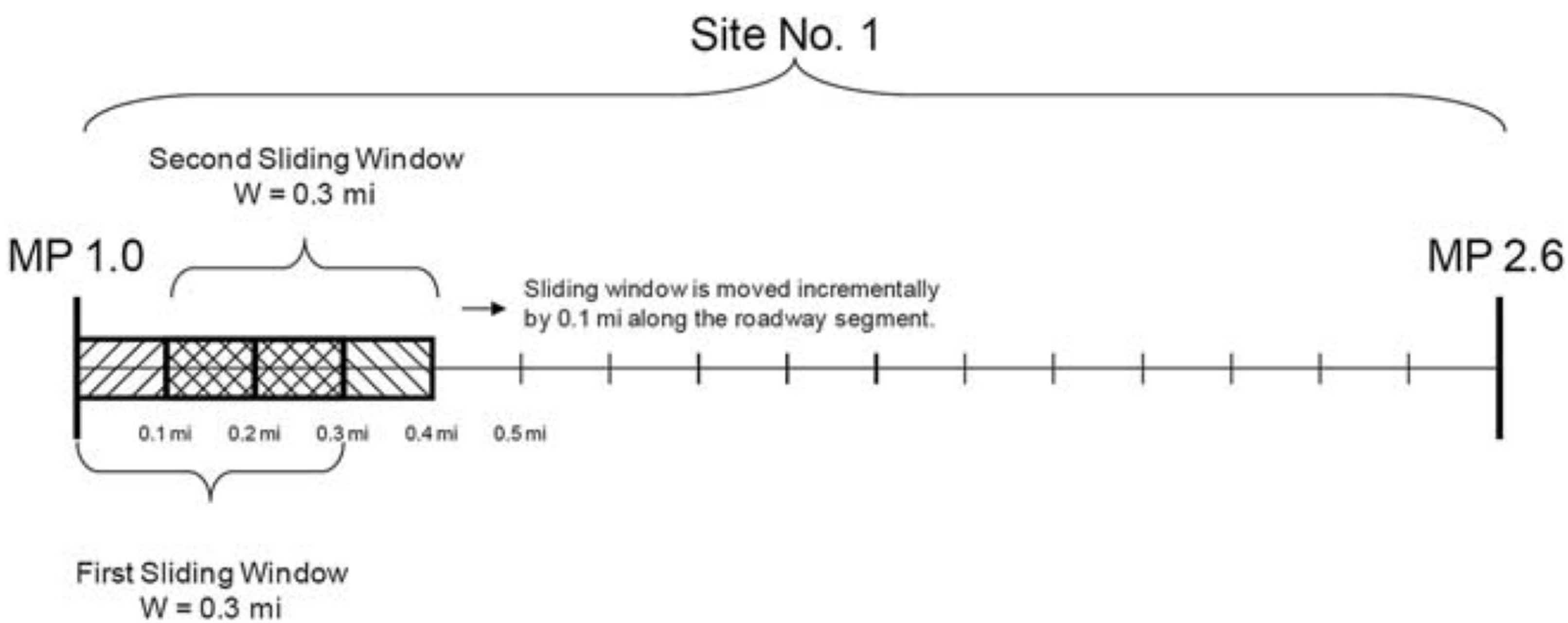
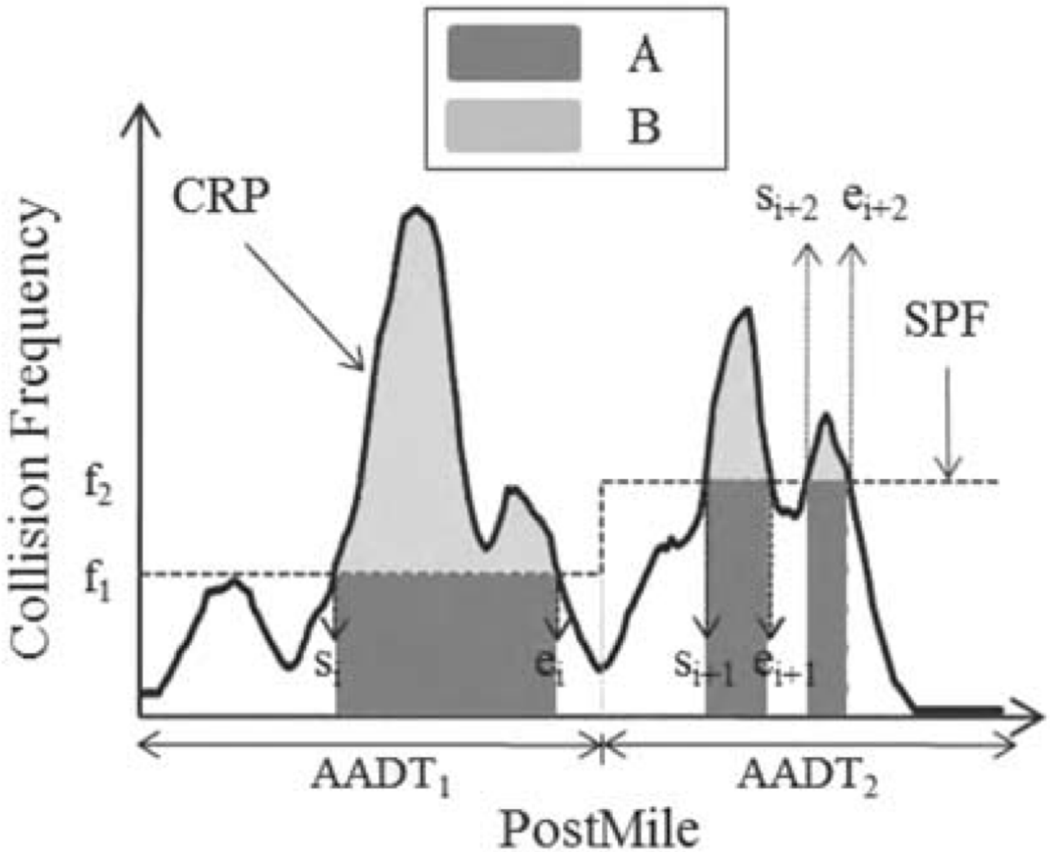


FIGURE 8.3 Illustration of sliding window method (Harwood et al., 2010).

Screening Methods

Continuous Risk Profile

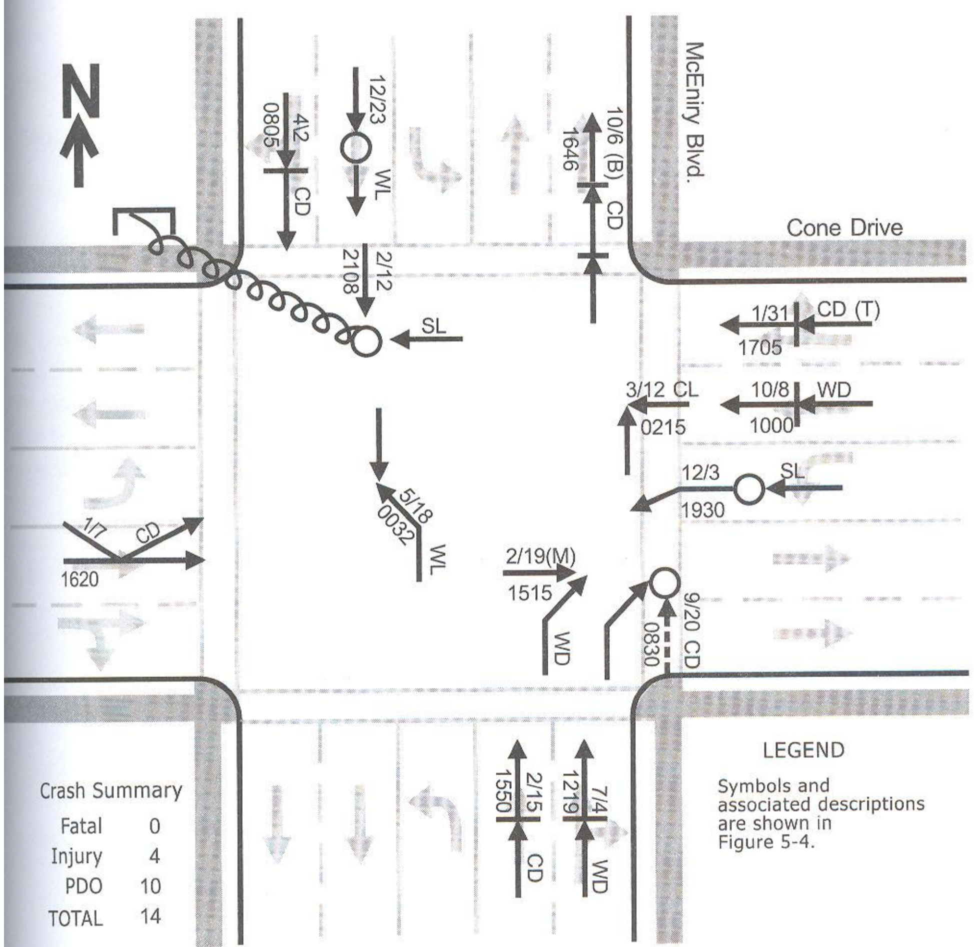


Diagnostics

- Step 1: Safety Data Review
  - Review crash types, severity and environmental conditions for the sites identified by one or a combination of Performance Measures. Conduct exploratory analyses (discussed previously)
- Step 2: Assess Supporting Documentation
  - Review past studies and plans covering the site vicinity for know issues, opportunities and constraints (if they exist or available).
- Step 3: Assess Field Conditions
  - Visit site and observe multimodal facilities and services in the area. (more below)

Diagnostics

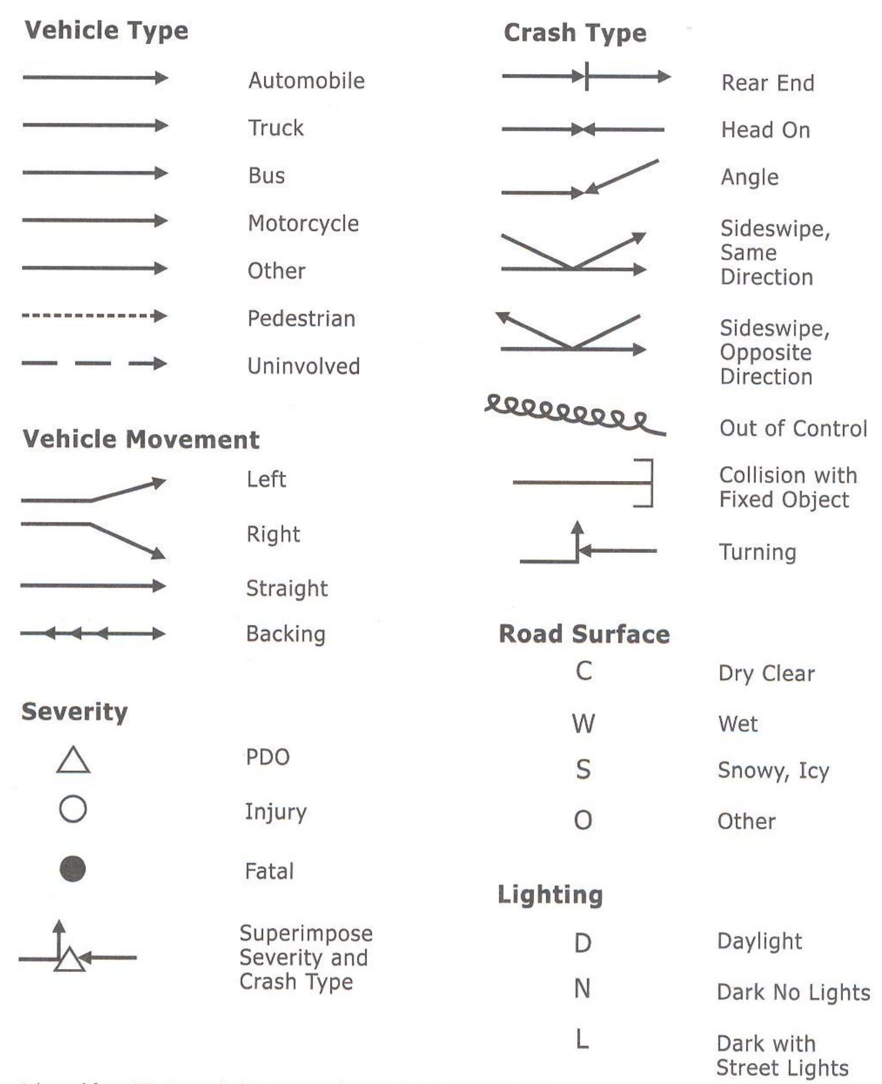
Step 1: Safety Data Review



dapted from ITE *Manual of Transportation Engineering Studies* (4)

gure 5-3. Example of an Intersection Collision Diagram

Diagnostics



Adapted from ITE *Manual of Transportation Engineering Studies* (4)

Figure 5-4. Example Collision Diagram Symbols

Diagnostics

- Step 3: Assess Field Conditions
  - Roadway and roadside characteristics
    - Signs, pavement conditions, sight distance, roadside features, etc.
  - Traffic conditions
    - Travel conditions, queue storage, excessive vehicular speeds, etc.
  - Traveler behavior
    - Drivers, pedestrians, cyclists
  - Roadway consistency
  - Land use
  - Evidence of problems
    - Skid marks, broken glass, damaged guardrail or landscape