METHODS FOR IDENTIFYING HIGH COLLISION CONCENTRATION LOCATIONS (HCCL) FOR POTENTIAL SAFETY IMPROVEMENTS – PHASE II: EVALUATION OF ALTERNATIVE METHODS FOR IDENTIFYING HCCL

Final Report

CFS Number 2078A DRI

Prepared by:

Raghavan Srinivasan
University of North Carolina Highway Safety Research Center, Chapel Hill, North Carolina
Phone: 919-962-7418; Email: srini@hsrc.unc.edu

Craig Lyon
Persaud and Lyon, Inc., Toronto, Canada

Bhagwant Persaud
Persaud and Lyon, Inc., Toronto, Canada

Carol Martell
University of North Carolina Highway Safety Research Center, Chapel Hill, North Carolina

Jongdae Baek
University of North Carolina Highway Safety Research Center, Chapel Hill, North Carolina

Prepared for:
California Department of Transportation
Sacramento, California

January 21, 2011

TABLE OF CONTENTS

ACKNOWLEDGMENTS

This work was funded by the California Department of Transportation. The authors would like to thank the members of the Table C task force, the Caltrans Research Panel, and other Caltrans staff for their support and guidance throughout this effort. The authors appreciate the assistance provided by HSIS staff Forrest Council, David Harkey, Yusuf Mohamedshah, and Behrang Hejazi in extracting the necessary data and in importing the data into SafetyAnalyst. The authors also thank Darren Torbic and Karen Richard from the Midwest Research Institute for their help regarding SafetyAnalyst before and during the SafetyAnalyst training course in Sacramento in Spring 2009.

DISCLAIMER

ABSTRACT

The objective of network screening should ideally be to not only identify sites for safety investigation but also to prioritize those sites efficiently. Using roadway, intersection, and collision data from California, this study compared the performance of methods based on the EB procedure, the LOSS method, and the Table C method. Two intersection types (rural four leg stop controlled and rural four leg signalized) and two roadway types (rural two lane roads and urban freeways) were included in the evaluation.

Safety Performance Functions (SPFs) were estimated using the California data for different collision types. The results of the evaluation indicated that compared to the Table C method, methods based on the EB procedures (EB expected and EB expected excess) tend to identify sites that have a higher AADTs and higher expected collisions. It is also clear that the top ranked sites that are identified based on the EB Expected and EB Expected Excess collisions methods have more collisions in the future compared to the top ranked sites from the Table C method. One of the evaluation approaches compared the ability of each method to flag and prioritize the locations previously investigated using the results of the Table C method, using data prior to the actual selection, and considering whether or not those locations were recommended for improvement. Since the sites investigated were selected based on the results of the Table C method, the Table C method did quite well compared to the other methods. At the same time, in many of the cases, the other methods did equally well and in some cases better than the Table C procedure.

The cumulative residual plots indicate that the SPFs directly calibrated from the California data are better than the default SafetyAnalyst SPFs that were recalibrated with the same California data. Hence, whenever possible, the SPFs directly calibrated from the most recent California data are recommended instead of using the default SPFs from SafetyAnalyst.

The methods based on the EB procedure work better with longer road segments. Hence, contiguous road segments could be aggregated once they remain homogenous with respect to AADT and key characteristics such as road classification, terrain, number of lanes and road width. With expanded lengths, an entire segment would be flagged and prioritized for safety investigation, not just the location (window) with the collision history that triggered the investigation. This is useful since the source of the problem may be quite removed from the triggering site and the aggregation of adjacent segments could potentially mitigate the spatial correlation that may exist due to secondary collisions in the vicinity of the segments.

The “proportions” method in SafetyAnalyst can be used as a diagnostic tool and possibly in combination with the EB methods for network screening. Network screening should be done on an annual basis and based on the most recent 5 years of data. The SPFs used for this purpose should also be re-calibrated annually to the most recent 5 years, whether they are California-specific or the default ones in SafetyAnalyst. Recalibration is automatic within SafetyAnalyst. New SPFs should be estimated every 5 years, recognizing that there are other SafetyAnalyst applications for SPFs. The SPFs estimated in this study for selected roadway and intersection types can be used for before-after evaluation of engineering treatments in addition to network screening.

EXECUTIVE SUMMARY

One of the first steps in effectively managing a road network is to identify sites that require safety investigations. It is important that the identification process is efficient, otherwise scarce resources may be wasted on sites that are incorrectly identified as collision concentration sites while roadway locations with a truly high potential for cost-effective safety improvement may not be flagged in this process. Conventional methods that make use of just collision counts or collision rates (per unit of exposure) are now known to have problems because they do not effectively account for the potential bias due to regression-to-the-mean phenomenon in which sites with a randomly high account could be incorrectly identified as having a high potential for improvement, and vice versa. Another problem with conventional methods that make use of collision rates is the implicit assumption that collision frequency and traffic volume are linearly related. Many recent studies have shown that the relationship between collisions and volume depends on the type of facility but tends to be non-linear.

Approach

Using roadway, intersection, and collision data from California, this study compared the performance of methods based on the empirical Bayes (EB) procedure (EB expected and EB expected excess), the Level of Service of Safety (LOSS) method, and the CALTRANS "Table C" method. Two intersection types (rural four leg stop controlled and rural four leg signalized) and two roadway types (rural two lane roads and urban freeways) were included in the evaluation. The following three approaches were used in the evaluation:

> **Approach 1**. Compare the ability of each method to rank those locations that are more likely to have high collision frequencies in the future.
>
> **Approach 2**. Compare, retrospectively, the performance of each method in selecting and ranking locations that were investigated and recommended for improvement (correct positives) and those that were investigated and not recommended for improvement (false positives).
>
> **Approach 3**. Compare the characteristics of top ranked locations by each method.

Initially, in approach 2, the plan was to not only use the information about whether a location was recommended for improvement, but the cost-effectiveness of these improvements. However, CALTRANS indicated that such cost-effectiveness data were not readily available for use in our evaluation.

In addition to comparing the methods using the three approaches this effort also investigated and demonstrated the significance of the regression to the mean issue by comparing the collision frequency of top ranked sites in 2000-2003 with the collision frequency for the same sites in 2004-2007.

Data

Roadway, intersection and collision data files were obtained from the Highway Safety Information System (HSIS) for 2000 to 2007. The influence area of an intersection was assumed to be within a radius of 250 feet from the intersection. To define interchange influence areas it was decided to apply a 0.3 mile radius around all ramps and use this area as the interchange influence area.

Preparation of the intersection data started with using the 2007 intersection file as the base file. The 2007 roadway file was merged to this file using the county, route, and milepost variables in order to add the roadway class variable that was used to distinguish between rural and urban environments. Next, each year‟s intersection file was merged by the county, route number, and milepost variables to add the major and minor road AADTs for each year as well as traffic control and number of lanes on each roadway. Only intersections which could be matched for each year and whose traffic control and number of lanes did not change were included for analysis. Information on the following collision types was extracted: total collisions, injury and fatal collisions, sideswipe, rear end, and broadside.

Before matching the collision data with the road segments, those collisions coded as taking place on a ramp were first removed from the data. Information on the following collision types was extracted: total collisions, injury and fatal collisions, head-on, sideswipe, rear-end, hit object, and overturn‟.

Caltrans provided an excel file that recorded the site investigations triggered by the current Table C application. A six year history from 2003 to 2008 was made available. The information provided for each site includes District, County, Route, Postmiles, Direction, Hwy/Int/Ramp, Log # (Table C All locations end in A, Wet end in W), Initiation Date, Approval Date, No Action or Improvement Recommended, Improvement Completion Date, and Investigation Date. This information was linked to the intersection and road segment databases created.

Safety Performance Functions

The EB methods and the LOSS method require the development of Safety Performance Functions (SPFs) which are mathematical equations that relate collision frequency (of different types) to site characteristics, in particular, traffic volume. To develop the safety performance functions (SPFs), generalized linear modeling was used to estimate model coefficients assuming a negative binomial error distribution, which is consistent with the state of research in developing these models. The dependent variable was either collisions per mile-year (for roadway segments) or collisions per intersection-year (for intersections). SPFs were estimated for different collision types.

Conclusions and Recommendations

- It is clear that the bias due to regression to the mean could be significant even if four years of data are used, especially for low volume sites such as rural two lane roads and rural stop controlled intersections. The state of the art EB method accounts for this possible bias. The other methods considered in this evaluation (LOSS and Table C) do not account for this possible bias.

- The results of the evaluation indicated that, compared to the Table C method, methods based on the EB procedures (EB expected and EB expected excess) tend to identify sites that have higher AADTs and higher expected collisions. In addition, the top ranked sites that are identified based on the EB procedures have more collisions in the future compared to the top ranked sites from the Table C method.

- One of the evaluation approaches compared the ability of each method to flag and prioritize the locations previously investigated using the results of the Table C method considering whether or not those locations were recommended for improvement. Since the sites investigated were selected based on the results of the Table C method, the Table C method, on the whole, did quite well compared to the other methods. At the same time, in many of the cases, the other methods did equally well and in some cases better than the Table C procedure.

- The analysis also examined the false positives for intersections, i.e., intersections that were selected and investigated using the Table C method for which no treatment was recommended. The top ranked sites from the EB expected method tend to have fewer false positives compared to the other methods.

- A key side benefit of the research is the indication that the SPFs directly calibrated from the California data are better than the default SafetyAnalyst SPFs that were recalibrated with the same California data. Hence, whenever possible, SPFs directly calibrated from the most recent California data should be used instead of the default SPFs from SafetyAnalyst.

- The methods based on the EB procedure work better with longer road segments. Hence, contiguous road segments could be aggregated once they remain homogenous with respect to AADT and key characteristics such as road classification, terrain, number of lanes and road width. With expanded lengths, an entire segment would be flagged and prioritized for safety investigation, not just the location (window) with the collision history that triggered the investigation. This is useful since the source of the problem may be quite removed from the triggering site and the aggregation of adjacent segments could potentially mitigate the spatial correlation that may exist due to secondary collisions in the vicinity of the segments.

- The “proportions” method in SafetyAnalyst was also investigated and was seen as having potential for use as a diagnostic tool and in combination with the EB methods for network screening.

- Finally, it is recommended that network screening be done on an annual basis and based on the most recent 5 years of data. The SPFs used for this purpose should also be re-calibrated annually to the most recent 5 years, whether they are California-specific or the default ones in SafetyAnalyst. Recalibration is automatic within SafetyAnalyst. New SPFs should be estimated every 5 years, recognizing that there are other SafetyAnalyst applications for SPFs. For example, the SPFs estimated in this study for selected roadway and intersection types can be used for before-after evaluation of engineering treatments in addition to network screening.

1. BACKGROUND

One of the first steps in effectively managing a road network is to identify sites that require safety investigations. It is important that the identification process is efficient, otherwise scarce resources may be wasted on sites that are incorrectly identified as collision concentration sites while roadway locations with a truly high potential for cost-effective safety improvement may not be flagged in this process. Conventional methods that make use of just collision counts or collision rates (per unit of exposure) are now known to have problems because they do not effectively account for the potential bias due to regression-to-the-mean phenomenon in which sites with a randomly high account could be incorrectly identified as having a high potential for improvement, and vice versa. Another problem with conventional methods that makes use of collision rates is the implicit assumption that collision frequency and traffic volume are linearly related. Many recent studies have shown that the relationship between collisions and volume depends on the type of facility but tends to be non-linear. For example, a 20% increase in volume will not necessarily result in a 20% increase in collisions. For most facilities, the relationship implies a smaller increase in collisions than the increase in volume with the result that the lower volume sites will have the highest collision rates and will tend to be flagged by the conventional collision rate method.

One method that has been proposed to overcome the pitfalls of conventional methods is the empirical Bayes (EB) procedure. Part of the reason that the EB method and other state of the art methods are not widely used by state agencies is the limited validation and testing of these approaches in the context of identifying a prioritized list of sites that provides the greatest opportunity for safety improvement in a cost effective manner. Another reason has been the unavailability of appropriate software tools for applying the EB methodology for network screening. With the availability of SafetyAnalyst (a software initially developed by FHWA and being supported as an AASHTOWare product), the EB procedure has become available to state and local agencies for their use for network screening and other safety management functions. The objective of this project is to evaluate various methods using data from California and identify the method(s) that are optimal for identifying locations for improvement. A related objective is the assessment of what it takes to be able to use California data with SafetyAnalyst.

2. OVERVIEW OF METHODS FOR NETWORK SCREENING

This section provides an overview of the methods discussed in Phase 1 of the project (Ragland and Chan, 2008) as well as additional methods potentially of relevance to the project. The methods reviewed include:

- Table C method currently used by CALTRANS
- The Level of Service of Safety (LOSS) method
- Empirical Bayes methods
- Continuous Risk Profile (CRP) for highway segments

The additional methods include:

- Screening based on high proportions of specific collision types
- Detection of safety deterioration over time
- Full Bayes methods

All of these additional methods (except full Bayes) are discussed in the upcoming Highway Safety Manual (HSM) or in SafetyAnalyst. Following is a discussion of each method along with its advantages and disadvantages. The last part of the section shows the methods that were identified for evaluation in the study.

2.1 OVERVIEW OF METHODS

2.1.1 Table C Method from CALTRANS

The Table C method identifies, in a given time period, sites that have experienced significantly more collisions per unit of ADT than the statewide average. Locations screened and identified in 12, 6, and 3 months period are sent out to Caltrans districts for investigation. For this method, sites are screened within rate groups of similar sites. There are currently 67 rate groups for highway segments, 30 for intersections, and 80 for ramps. For highway segments, the roadway is screened by a sliding window of size 0.2 miles and in increments of 0.02 miles. If a segment is flagged then the window position slides an increment of 0.2 miles. Thus there is no overlapping of flagged segments. When the highway rate group changes along a roadway the process stops and restarts at the beginning of the new rate group. The segments of roadway considered within an intersection influence area are not considered in the highway segment screening, nor are ramps.

For intersections, the influence area is predetermined, usually 250 ft., and all collisions within that area are considered. For ramps, only ramp collision data are included.

The criteria for flagging a site for investigation are: a) the observed collision frequency is greater than the average for the rate group with 99.5% confidence in either the 3, 6, or 12 months period, and b) there are 4 or more collisions in the time period.

The minimum number of observed collisions required for significance (N_R) is found by using the following formula:

$$N_R = N_E + 2.576(N_E)^{1/2} + 1.329$$

Where, N_E is the average number of collisions for the rate group calculated as:
$N_E = (ADT)(t)(L\ )(\ R_E)/\ 1{,}000{,}000$

ADT = Average Daily Traffic, vehicle per day
t = time, in days = number of quarters x days/quarter x days/time period
L = length, in miles (= l for Ramps and Intersections)
R_E = Average Collision Rate, in Collision/million vehicle (ACCS/MV) or Collision/million vehicle mile (ACCS/MVM)
= Base Rate + ADT factor

Each Rate Group has a Base Rate that is determined by looking at all collisions in a three year time period. Some highway segment rate groups also include an ADT factor which adjusts the base rate given a site‟s ADT. While the procedure is relatively straightforward, is easy to apply, and does consider some measure of statistical significance, several factors may reduce its efficiency as discussed below.

Accounting for Regression-To-The-Mean (RTM)

Using only the observed collision rates means that regression-to-the-mean is not being accounted for. This is particularly of concern in the present context, since time periods as short as three months and maximum of only 12 months are being used in the Table C identification process. Thus sites with randomly high collision counts (and rates) in such short time periods can be mistakenly flagged for site investigation while other, more deserving locations with randomly low counts (and rates) may escape detection and follow up investigation. The problem of RTM in Table C method may be mitigated to some extent by adapting a longer analysis time period (e.g., minimum of 36 months).

Use of Constant Collision Rates for Most Rate Groups

For all intersection and ramp and many highway segment rate groups, a constant value for average collision rate is used. This assumes that the relationship between collision frequency and traffic volumes is strictly a linear one. This relationship has in fact been shown to be non-linear, with low ADT sites usually tending to have higher collision rates than higher ADT sites. Thus, comparing collision rates to a single average base rate may lead to sites with low ADT and relatively few collisions being flagged over more deserving locations with more collisions and higher ADTs, but lower rates.

Accounting for Collision Severity

The present method only accounts for total collisions and wet weather collisions. Thus, opportunities for improving sites with more severe collisions to produce greater safety benefits may be missed.

Rate Group vs. SPF
Rate group categories do account for many variables expected to impact collision risk, including area type, number of lanes, and ADT. The implicit assumption that expected collision frequency (and rate) is constant across the range of ADT that defines the rate group may affect estimation accuracy. For some highway segment rate groups this difficulty in addressed by including an ADT factor which adjusts the base rate given a site‟s ADT. However, the use of Safety Performance Functions (SPFs) to be discussed under EB methods may be better in this respect in that it allows the direct estimation of expected collision frequency for a specific ADT. That the SPF can be updated in SafetyAnalyst with new data provides an added advantage to the use of SPFs where this software is used.

Use of Constant Window Size
Use of a fixed window size to screen segments is a significant problem since it leads to sites with randomly high collision counts being falsely selected if the window size is too small, or missing localized safety problems that could be “averaged out” if the window size is too large.

No Ranking of Flagged Sites
Once a site is flagged then it is investigated. This does not reflect the reality that some sites may be more deserving than others and should receive higher priority. One measure that could be considered for o ranking the flagged sites could be the difference between the observed collision count and N_R.

2.1.2 Level of Service of Safety

The Level of Service of Safety (LOSS) concept was introduced by Kononov and Allery (2003). As proposed, this method is similar to the Table C method in that the observed collision count is compared to an expected collision count and the level of deviation is measured. The Table C method considers whether the deviation is large enough to conclude with statistical certainty that more collisions occurred than would be expected for the average site. In the LOSS method, the deviation from the expected for an average site is described by creating 4 bins, or level of service levels.

The expected level of safety for similar sites is determined by using SPFs. Safety Performance Functions (SPFs) are mathematical equations that relate the expected collision frequency (of different types and severity) to site characteristics. At the basic level, the site characteristics in an SPF may just include traffic volume (in this report called type 1 SPF). A more complicated SPF will include other site characteristics such as number of number of lanes, lane width, presence/absence of turn lanes, in addition to traffic volume (called type 2 SPF). These SPFs would be used to predict the average collision frequencies for a combination of traffic volume and other site characteristics.

Applying the LOSS method involves the following steps:

Step 1
Apply the appropriate SPF to estimate the expected number of collisions, κ, for the site under consideration.

Step 2
Calculate the standard deviation of the estimate in Step 1.

$\sigma(\kappa) = (\phi\kappa^2)^{0.5}$; if SPF assumes a negative binomial distribution of collision counts.

Where ϕ is the overdispersion parameter of the SPF.

Step 3
Compare the observed number of collisions, K, to the limits for the 4 LOSS categories.

LOSS	**Condition**	**Description**
I	$0<K<(\kappa - 1.5\sigma(\kappa))$	indicates a low potential for collision reduction
II	$(\kappa - 1.5\sigma(\kappa)) \leq K < \kappa$	indicates better than expected safety performance
III	$\kappa \leq K < (\kappa + 1.5\sigma(\kappa))$	indicates less than expected safety performance
IV	$K \geq (\kappa + 1.5\sigma(\kappa))$	indicates a high potential for collision reduction

Sites with a LOSS of IV are ranked highest for further safety investigations, followed by LOSS III, LOSS II, and then LOSS I.

Through the use of safety performance functions, the LOSS method would improve upon the current Table C method by eliminating the use of constant collision rates across ADT and by the potential inclusion of additional variables which impact the expected collision rates of sites. There are however some potential drawbacks to the method as it is currently applied.

Accounting for Regression-To-The-Mean
Using the observed collision counts means that regression-to-the-mean is not being accounted for. As previously discussed, the result is that sites with randomly high collision counts in a short time period are likely being mistakenly flagged for site investigation and that other, more deserving locations, are not being investigated.

Accounting for Collision Severity
Improving sites with more severe collisions will lead to greater benefits. For sites with the same collision frequency, it would therefore be advantageous to assign some higher ranking to sites where the collisions tend to be more severe. If the collision data and SPFs are available by severity type, a LOSS for different severities could be determined. However, there is no logical method for creating a mixed rating of, say a LOSS of II for severe injury collisions, with a rating of III for PDO collisions.

No Ranking of Flagged Sites

There is no ranking of individual sites within a LOSS category, ignoring that within a category some locations will be more deserving of further investigation than others. However, by taking the difference between the observed collision frequency and the average collision frequency (from an SPF), one can use that value to come up with a rank.

2.1.3 Empirical Bayes Methods

The empirical Bayes (EB) methods refer to a suite of screening methods that are based on the empirical Bayes method of estimating the *long-term* expected collision frequency for a location. These methods have been adopted for the Federal Highway Administration‟s (FHWA) *SafetyAnalyst* software which, among other analyses, performs network screening. It is also documented as a preferred methodology in the recently published Highway Safety Manual.

The empirical Bayes estimate of expected collision frequency for a location is a weighted combination of the prediction from a safety performance function (SPF) and the observed collision count for the location. The weights are calculated based on the EB procedure that makes use of the overdispersion parameter that is an outcome of the SPF development using negative binomial regression. If data used to calibrate the SPFs are spatially correlated, the statistical significance of the parameters would be overestimated if the spatial correlation is not accounted for using special procedures such as Generalized Estimation Equations (GEE) (Lord and Persaud, 2000). Whether these issues have practical significance is not clear, but the assumption in developing the SafetyAnalyst SPFs appears to be that they do not.

Sites are ranked in descending order of the expected collision frequency (E) or, alternatively, the expected excess collision frequency, which is the difference between E and the SPF prediction.

> *Estimate of Expected Collisions for a site (E) = w × (SPF prediction) + (1 – w) × (Observed collision frequency)*
>
> where:
> $0 \leq w \leq 1$
>
> *Estimate of Expected Excess Collisions for a site = (Estimate of Expected Collisions for a site) – (SPF Prediction)*

Screening may be conducted for all collision types or for specific collision types and severities. Screening may also be done by weighting the expected collision frequency using relative unit cost estimates for collisions of various severity and the expected collision frequencies by severity.

The method of screening available depends on the site type. For intersections and ramps, the influence area boundaries are defined and sites are simply screened by the expected or excess collision frequency. For roadway segments[1] that can have varying lengths, two approaches are

---

[1] In this discussion, segment and site are used interchangeably. Segments refer to pieces of pavement derived from California‟s roadway inventory file.

available. One is the sliding window approach where a window of fixed length moves in defined increments and the calculations are performed at each window location. Each segment is characterized by the maximum value calculated at any window position within or overlapping the beginning of adjacent segment. In so doing, there is an increased chance of detecting a high risk site at the screening stage if the collision problem manifests itself in a window overlapping the adjacent site.

The second is the peak search approach. This approach makes use of incrementally growing window lengths that are selected so no windows span multiple roadway segments. The window starts at the left boundary of a road segment and increases in length incrementally until it reaches the end. At each increment, we have a specific window where an estimated collision count can be calculated. For example, a segment of 0.5 mile can produce windows with lengths of 0.1, 0.2, 0.3, 0.4, and 0.5 miles assuming an increment length of 0.1 mile (SafetyAnalyst version 4.0.0 uses an increment of 0.1 mile in the peak search method). The window with the largest value of the estimate of expected or expected excess collisions per mile (or some measure weighted by collision severity), such as Equivalent Property Damage Only (EPDO) is then tested for statistical significance. The test of significance is the coefficient of variation, CV, equal to the standard error of the estimate divided by the estimate. A limiting value of the CV is specified by the analyst, and values of CV below the limiting value pass the test. If the window passes the test then the entire road segment is ranked by the largest value of the estimate per mile. If the test is not passed then the window size is increased and the process starts again for the road segment. The advantage of this method is that localized safety problems are not overlooked by using too large a window yet the statistical test ensures that they are in fact reliable estimates and not due to some randomness in the data.

Figure 2.1, taken from the functional specification for module 1 (network screening) in *SafetyAnalyst*[2], illustrates the sliding window approach based on EPDO collision frequency. This diagram shows all possible windows for two adjacent segments (sites). Site No. 23 is ranked by window number 3 which has the highest value for all windows which overlap that site. Site No. 24 is ranked by window number 8.

Figure 2.2 taken from the slides presented during a SafetyAnalyst training course, illustrates the peak search method. A segment (site) that is 0.67 miles long is shown. The shortest window possible window size is 0.1 miles long and this window increases in size in 0.1 mile increments. The final window is equal to the length of the segment (i.e., 0.67 miles). So, with this segment, window lengths of 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, and 0.67 are possible. Figure 2.2 shows how the window lengths of 0.1, 0.2, and 0.67 are used. In Figure 2.2, CV_{Limit} is the limiting value of CV that is discussed above.

---

[2] Draft Functional Specification for Module 1 - Network Screening, Midwest Research Institute, et al., May 2003, Contract No. GS-23F-0379K, Task No. DTFH61-01-F-00096.

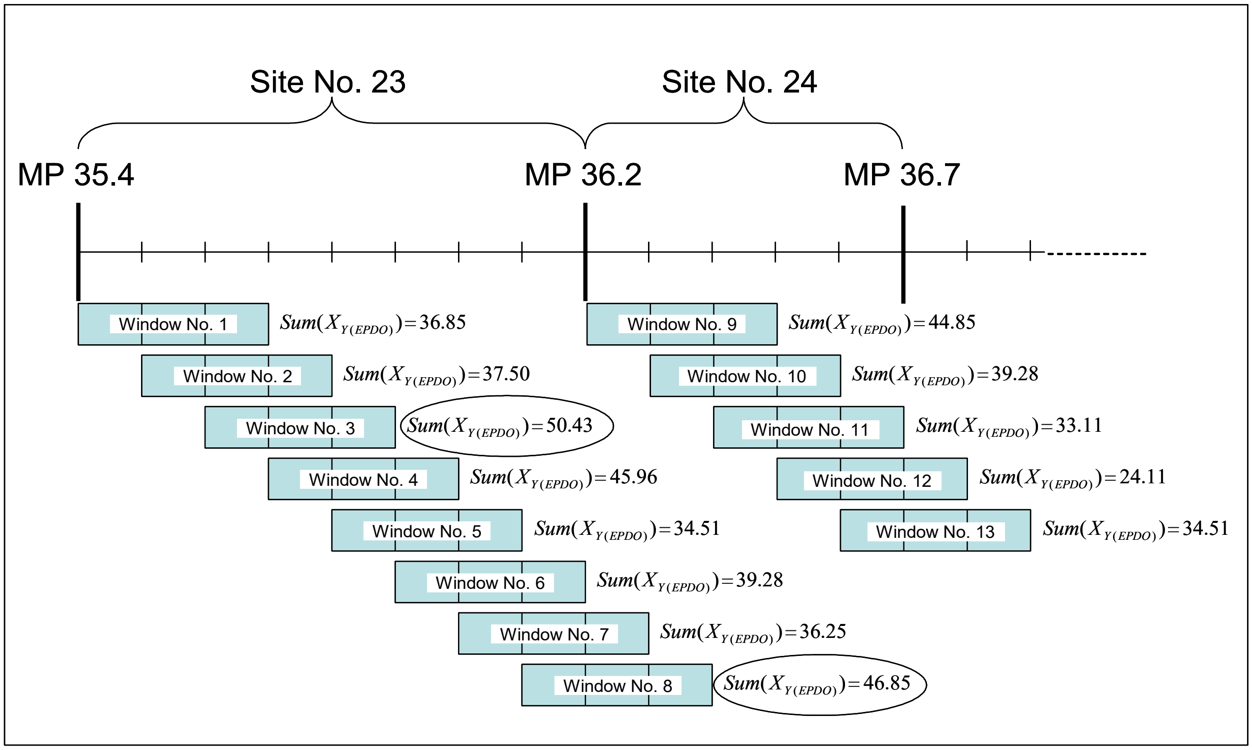
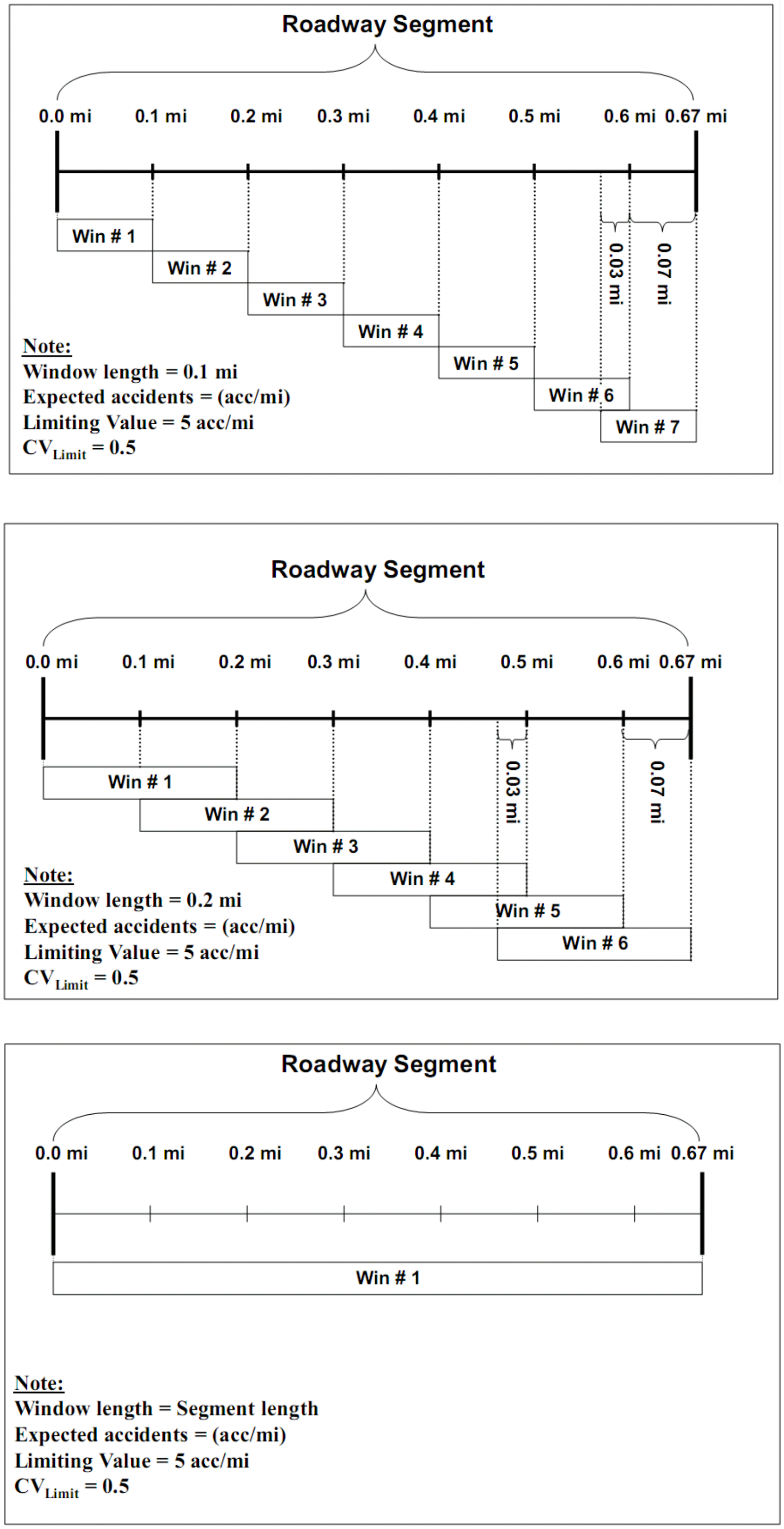


Figure 2.1: Illustrating the sliding window approach



Note:
Window length = 0.1 mi
Expected accidents = (acc/mi)
Limiting Value = 5 acc/mi
$CV_{Limit} = 0.5$

Note:
Window length = 0.2 mi
Expected accidents = (acc/mi)
Limiting Value = 5 acc/mi
$CV_{Limit} = 0.5$

Note:
Window length = Segment length
Expected accidents = (acc/mi)
Limiting Value = 5 acc/mi
$CV_{Limit} = 0.5$

Figure 2.2: Illustrating the peak search method (Source: SafetyAnalyst training materials)

The EB based methods as applied in *SafetyAnalyst* offer a number of advantages, including:

- Properly accounting for regression-to-the-mean, thus avoiding flagging locations based on short-term randomly high collision counts.
- Use of SPFs to properly consider the impact of traffic volume and other factors on expected collision frequencies.
- Consideration of collision severity by weighting severity specific estimates (optional in *SafetyAnalyst*).
- The peak search method for segments of varying length allows varying window sizes to be used between roadway segments, thus identifying localized areas of risk while still considering statistical significance of the estimates.
- All sites are ranked by their unique estimate as opposed to flagging a group of sites for investigation.

There are some remaining issues related to these methods that have yet to be resolved, including:

- Is it better to rank by the expected or excess collision frequency? There has been some debate in the safety community on this topic without any clear consensus. There are advantages and disadvantages for both these methods. The use of expected collisions is embedded in the concept of Collision Modification Factors (CMFs) since the benefit of a treatment can be expressed as the product of the expected collisions with (CMF – 1). On the other hand, there is no way to directly apply CMFs to expected excess collisions. However, using expected excess is attractive and intuitive because it "rests on the belief that if a site has more collisions than what is normal at similar sites, there must be site-specific causes that explain the excess, and that if causes are identified, they could be remedied, and the excess reduced" (Hauer et al., 2002). SafetyAnalyst allows the user to select either method for their network screening.
- How important is it for the SPFs to include variables in addition to AADT?
- How important is it for the SPF calibration process to account for spatial correlation?

2.1.4 Continuous Risk Profile (CRP) Method

Chung and Ragland (2007) have proposed a method called the continuous risk profile (CRP) that is based only upon observed collisions. The motivation for the development of the CRP method was based on two criticisms of the current Table C method, which flags specific windows of a fixed length:

1. "risk is assumed to be a constant throughout the extension of the window"
2. "all factors leading to high risk are assumed to reside within that window; it is possible that collisions within a window could result for example, from collisions in the vicinity or weaving patterns caused by factors that reside outside the window". This problem may be more common on urban freeways where secondary collisions could occur as a result of bottlenecks.

The method is suggested by the developers as being particularly advantageous over SPF-based methods where spatial correlation of data used to develop the SPFs may be an issue. However, this potential advantage would now appear to be moot, given recent research that uses tools such as General Estimating Equations (GEE) and Full Bayes methods to account for temporal and spatial correlation in data (Aguero-Valverde and Jovanis, 2008); Wang and Abdel-Aty, 2006). It should be noted in passing that the SafetyAnalyst SPFs do not account for spatial correlation. However, it is unlikely that this would adversely affect the network screening results in SafetyAnalyst since the coefficient for the key variable used (ADT) would not be materially affected if spatial correlation were considered.

The main rationale behind the CRP method appears to be that a continuous profile plot of risk along a roadway can help identify zones of high risk. Whereas plotting the observed collision count versus distance would appear very scattered, with most locations recording zero collisions in a given time period, a cumulative graph is smoother and visually identifies stretches of roadway that have experienced many collisions. The steps in applying the method to a particular road are as follows:

1) At each location, d, on the roadway, calculate the cumulative count of collisions.
2) From the cumulative count of collisions, subtract the cumulative expected number of collisions, equal to the cumulative distance multiplied by the average collision rate (collisions/mile). In effect this is a cumulative „excess‟ collision count. (See Figure 2.3 below.)
3) Calculate a moving average of the cumulative „excess‟ collision count using a sliding window. The size of the sliding window is up to the analyst. This moving average is employed to reduce the impact of random fluctuations in collision counts.
4) At each location, d, the positive values of the moving average, which pertain to locations with more observed collisions than expected, remain the same. If the value of the moving average is negative, a situation at locations with fewer observed collisions than expected, then a value of 0 is assigned.
5) The new values of the moving average calculated in step 4 are graphed versus the cumulative distance on the roadway. The graph allows the identification of where risk starts to increase and decrease as well as locations of local peak risk. (See Figure 2.4 below for an example using one year of data.)

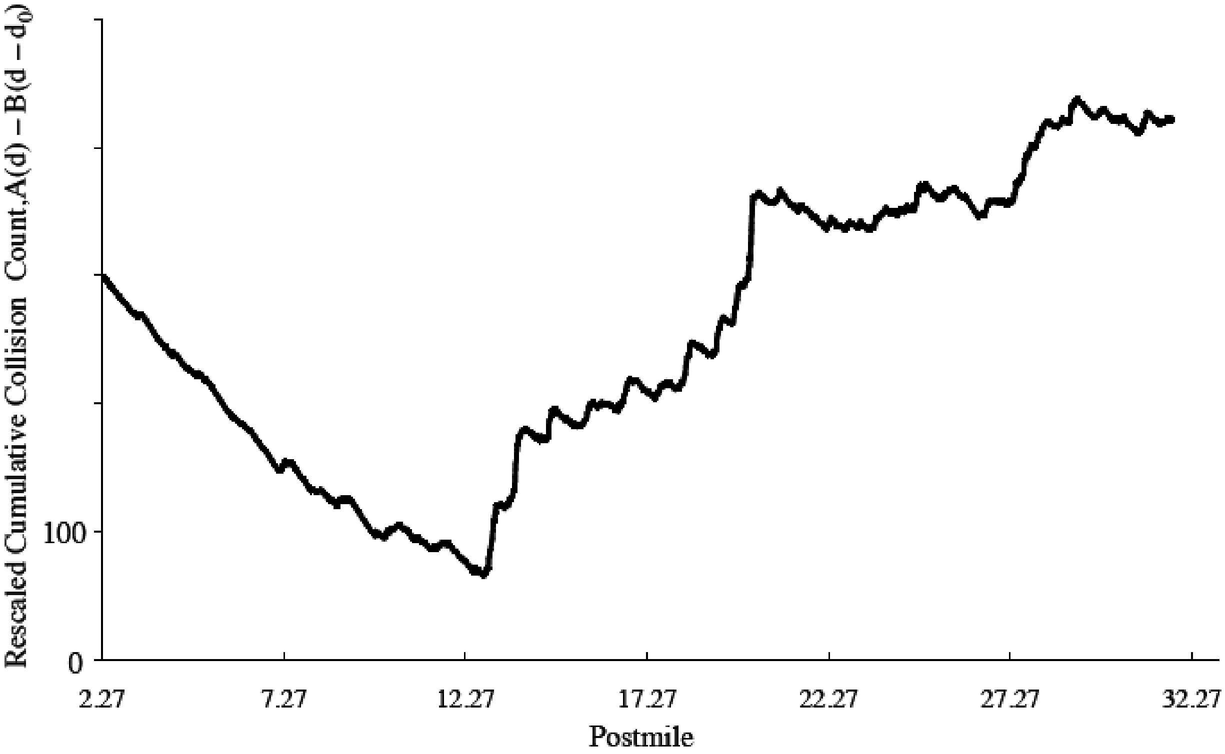


Figure 2.3: Taken from Chung and Ragland (2007)

Figure 2.4: Adapted from Chung and Ragland (2007)

The CRP method has a number of potential deficiencies.

Accounting for Regression-To-The-Mean (RTM)

Using only the observed collision rates means that regression-to-the-mean is not being accounted for. Sites with randomly high collision counts in a short time period could be mistakenly selected for site investigation and other, more deserving locations with randomly low counts, may escape detection and not be investigated. By including multiple years of data, the potential bias due to regression to the mean can be reduced to some extent, but there is evidence from other research that even five years of data would exhibit significant RTM bias.

Accounting for Traffic Volumes and Other Variables Affecting Expected Collision Frequencies

By not making use of traffic volumes the impact of exposure on expected collision frequencies is ignored. The result may be that the procedure may target locations with what appears to be abnormally high numbers of collisions but which in fact are quite normal for the level of volume. Conversely, locations with lower ADTs but unusually high collision frequencies may escape detection and not be investigated. Similar results may arise due to ignoring other variables that affect expected collision frequencies. One way of potentially accounting for traffic volume would be to base the method on expected collision rates but this would suffer from the difficulty caused by the non-linear relationship between collisions and AADT, as noted earlier.

Accounting for Collision Severity

As presented, the CRP method is applied to one collision type at a time. It is recognized however that the method could be adapted to consider multiple collision or severity types. For example, at step 3 above, the weighted averages for several severities could be combined by weighting each.

Use of Constant Window Size

Too large a window may ignore very localized areas of risk while too small a window may be biased towards locations with randomly high collision counts. However, like the SafetyAnalyst sliding window and peak search methods it does recognize that the source of a collision problem in one window may be some distance from where the problem is observed.

No Ranking of Flagged Sites

There appears to be no logic for defining which segments on the roadway should be considered as one location, neither is there a method for ranking flagged sites. One measure that could be considered for ranking the flagged sites is the area under the curve in Figure 2.4.

2.1.5 Screening Based on High Proportions

The method of screening based on high proportions identifies and ranks locations that have a proportion of a specific collision type relative to the total collisions that is higher than some average or threshold proportion value for similar road types. This method can also be applied as a diagnostic tool to identify overrepresented collision types at a site. Kononov (2002) found that looking at the percentage distribution of collisions by collision type can reveal the “existence of collision patterns susceptible to correction” that may or may not be accompanied by the overrepresentation in expected or expected excess collisions. This method was originally proposed by Heydecker and Wu (1991) and has been included as one of the methods in *SafetyAnalyst*.

The method is identical for different location types. However, only similar location types should be analyzed together because collision patterns will naturally differ. For example, the collision patterns are different for stop-controlled intersections, signalized intersections, and two-lane roads, so the method would be applied separately to the three types of facilities and separately for urban and rural environments.

The basic theory follows that the observed proportion of a collision type (p_i) at a site i with total collisions of n_i and target collision x_i is assumed to follow the binomial distribution, shown below. Although there is some true mean proportion, the observed proportion in a given time period is randomly dispersed about this value.

$$f(x_i / n_i, \mu_i) = \binom{n}{x_i} \mu_i^{x_i} (1-\mu_i)^{n_i - x_i}, 0 \leq x_i \leq n_i$$

It is further assumed that the expected proportion at a specific site, μ_i, is constant for a given site but varies between similar sites. This between site variation is assumed to follow the Beta distribution.

$$g(\mu / \alpha, \beta) = \frac{\mu^{\alpha-1}(1-\mu)^{\beta-1}}{B(\alpha, \beta)}, \quad 0 < \mu < 1$$

The α and β parameters of the Beta distribution are calculated using the same population of similar sites to be screened.

Combining these two distributions, the method calculates, for each site, the probability that it‟s true proportion is higher than the threshold value, given the observed data.

The method is applied as follows:

Step 1: Select the site type and collision or severity type to be screened.

Step 2: Select the limiting value of expected proportion of collisions, θ^*. This is a limiting value of the proportion of all collisions that may be considered.

Step 3: Find the total number of collisions of the collision type of interest during the study period at each site, x_i.

Step 4: Find the total number of all collision types at each site, n_i.

Step 5: Calculate the observed proportion, $\theta_i = x_i/n_i$, for each site for the collision type of interest for $n \geq 2$.

Step 6: Calculate the mean proportion of target collisions by type or severity for all m sites under consideration, according to:

$$\bar{\theta} = \frac{\sum_{i=1}^{m} \theta_i}{m} \qquad n \geq 2$$

Only sites with 2 or more collisions are used to estimate this value but all sites having at least one collision may be screened.

Step 7: Calculate the variance of the target collisions according to:

$$s^2 = \frac{1}{m-1}\left[\sum_{i=1}^{m}\left(\frac{x_i^2 - x_i}{n_i^2 - n_i}\right) - \frac{1}{m}\left(\sum_{i=1}^{m}\frac{x_i}{n_i}\right)^2\right], \qquad n \geq 2$$

Where:
m = number of sites
x_i = observed collisions by type or severity for the specific location
n_i = total number of collisions for the specific location

Step 8: Calculate the parameters α and β used in estimating the ranking measure in the next step:

$$\alpha = \frac{\bar{\theta}^2 - \bar{\theta}^3 - s^2\,\bar{\theta}}{s^2}$$

$$\beta = \frac{\alpha}{\bar{\theta}} - \alpha$$

Step 9: For each site with one or more collisions, solve for $P(\theta_i>\theta^*|x_i,n_i)$. This is the probability that a site‟s long-term expected proportion is higher than the limiting value proportion, θ*.

$$P(\theta_i > \theta^* \mid x,n) = 1 - \frac{1}{B(\alpha+x,\beta+n-x)}\int_{\theta=0}^{\theta^*} \theta^{\alpha+x-1}(1-\theta)^{\beta+n-x-1}\,d\theta$$

where B(.,.) is the beta function value based on the two parameters calculated in Step 8 and defined as follows:

$$B = \frac{(\alpha+x-1)!\,(\beta+n-x-1)!}{(\beta+\alpha+n-1)!}$$

These equations may be solved through any numerical integration routine. The simplest way is to use Microsoft Excel by specifying the following equation:

$P(\theta_i>\theta^*|x_i,n_i)$. = 1-betadist ($\theta^*,\alpha+x_i,\beta+n_i-x_i$).

Step 10: Rank sites in descending order by the value of $P(\theta_i>\theta^*|x_i,n_i)$.

The strengths of this method are in its use with limited data and its potential for use for screening based on a specific collision type when traffic volume data are not available. Lyon et al. (2007) compared this method to the EB methods for specific collision types and concluded that it is a workable alternative to the EB methods for network screening for specific collision types when reliable safety performance functions or exposure data are not available. The approach is also statistically based and takes into account the greater uncertainty in estimating proportions at locations with few collisions.

The main weakness of the approach is that regression-to-the-mean is not directly accounted for, although the use of statistical thresholds does account for this in some manner. Another weakness in methods in this category is that there is no logical approach to determine the threshold value. In addition, since this method focuses on proportion of collision types, it is possible that due to low proportion of other collision types, some collision types may incorrectly appear to be overrepresented.

2.1.6 Detection of Safety Deterioration over Time

SafetyAnalyst incorporates a methodology for identifying for investigation those sites that experience a gradual or sudden increase in mean collision frequency (Hauer, 1996a; and Hauer, 1996b). The description below is adapted from the *SafetyAnalys*t network screening whitepaper at www.safetyanalyst.org. To illustrate the methodology, the following example is provided. On a section of highway, the following collisions have been recorded over the past 5 years:

Year_i	Recorded collisions, x_i
1	7
2	5
3	10
4	15
5	13

Time periods are numbered *1, 2,....,T, T+1,..., L*. The number of observed collisions in each time period is denoted $x_1, x_2, \ldots, x_L$. For the end of any time period T, $(1 \le T < L)$, the difference between the "after" and "before" period collision averages is calculated as:

$$\Delta(T) = \frac{\sum_{T+1}^{L} x_i}{L-T} - \frac{\sum_{1}^{T} x_i}{T}$$

For the example road section, the results are given below.

T	Recorded collisions, x_i	Collision average before	Collision average after	*Δ(T)*
1	7	7.0	10.8	–3.8
2	5	6.0	12.7	–6.7
3	10	7.3	14.0	–6.7
4	15	9.3	13.0	–3.8

Based on these results, two tests will be conducted. The first test is to detect a potential steadily increasing trend in mean collision frequencies. The second test is to detect a potential sudden jump in the mean collision frequency. Sites meeting statistical tests for a gradual or sudden increase in mean collision rate can then be ranked as desired.

The method is statistically rigorous and is particularly applicable for situations where the safety of sites can change significantly over time. However, it does not address traffic volumes which may change significantly over time and does not directly account for regression-to-the-mean. It is relatively untested and there are no known applications. A decision has been taken to exclude this method from the Highway Safety Manual. Nevertheless it is incorporated in the *SafetyAnalyst* software so complexity of calculations should not be a deterrent to its use in applicable situations.

The method is similar in principle and purpose to one proposed more recently by Chung et al. (2011) for "proactively detecting a hot spot where its safety level slowly deteriorates over time". The basis of Chung et al.‟s method, which was illustrated using California freeway data, is the continuous risk profile (CRP) approach reviewed earlier. Normalized CRP plots from previous years are first compared to detect sites with significant changes in collision profile. Then sequential hypothesis testing is used to identify the target site whose growth factor remained above the 70th percentile for more than two years. However, as the authors acknowledge, it is unclear whether waiting two years was an adequate time period to properly address the issues that arise from the regression-to–the-mean, so further research is planned. Nevertheless, the method is promising.

2.1.7 Full Bayes Methods

Recently, some researchers have started using the full Bayes (FB) approach to modeling safety data (e.g., Pawlovich et al., 2005; Miaou and Lord, 2003; Lan et al., 2009) which allows the estimation of complex model forms that are not easily handled by conventional Generalized Linear Modeling that is typically used to estimate SPFs.

Concerning network screening, full Bayes is not a screening method on its own, but is an alternate method of predicting a desired measure of safety, for example, the expected collision frequency at a site. How full Bayes and empirical Bayes differ warrants a brief discussion. Bayesian statistics is an approach whereby previously held knowledge is included when making an estimate using current data. In the context of estimating expected collision frequencies at specific locations Bayes uses two clues: a) the observed collision frequency at the site; and b) the expected collision frequency for similar locations.

In the empirical Bayes approach, a safety performance function (SPF) is estimated using a reference group of similar sites and the prediction of this SPF is used for the estimate of the expected collision frequency for similar locations. When applying the SPF, the estimated parameters are used directly although it is acknowledged that there are standard errors associated with them. Typically, calibrated and available SPFs such as the default ones in SafetyAnalyst do not consider spatial correlation in data from different sites used to estimate the SPFs. Such correlation can be considered in special procedures such Generalized Estimating Equations (GEE), which, by and large, have not been used for network screening SPFs because the improvement in estimation is not of practical significance, especially for SFFs where AADT is the only independent variable.

In the full Bayes approach, the same reference group is used and the SPF estimated but a more complicated approach is applied whereby all possible values of the estimated parameters are considered and not just the most likely value. Additionally, if the analyst has knowledge a-priori about what the parameter estimates are expected to be then this can be considered in the modeling.

Recently, there has been some debate in the research community as to whether it is more appropriate to apply empirical Bayes or full Bayes methods. Recent research focusing on this debate includes Huang et al. (2009) and Aguero-Valvarde and Jovanis (2009). While full Bayes methods can, in some circumstances, provide advantages over empirical Bayes methods, the very significant added complexity of the approach necessitates that it provide substantial benefits to be warranted.

One of the advantages of FB is that it allows the consideration of spatial correlation, but as noted above, this is not vital for network screening SPFs used in the EB based models. Another advantage is that a distribution of likely values of expected collision frequency can be determined for each site. Thus it would be possible to screen sites not only by their mean expected value but by say the 85th percentile value if so desired.

The main disadvantage with the FB approach is that it is exceptionally complex and requires significant statistical expertise to implement. Complex software is required, in effect creating a „black box" for those without the in depth knowledge of FB techniques and the software itself. On the other hand, the EB based methods and method of high proportion may be easily conducted within spreadsheets once the required SPFs or other needed parameters are available. These calculations are straightforward enough that they may even be done on paper.

It is foreseeable that the method adopted by CALTRANS should be understandable and repeatable by engineering staff. For this reason, we conclude the FB modeling methods are not appropriate for further consideration. The same logic was used in excluding Full Bayes methods from *SafetyAnalyst* and the first edition of the Highway Safety Manual.

2.2 COMPARATIVE EVALUATIONS OF METHODS BASED ON PREVIOUS RESEARCH

Hauer et al., (2004) used five different criteria to produce five ranked lists of Sites With Promise (SWiP) for rural two-lane roads in Colorado"s mountainous terrain. The five criteria were the following:

- Criterion 1: Sites where most collisions are expected
- Criterion 2: Sites where most severity-weighted collisions are expected
- Criterion 3: Sites where most excess collisions are expected
- Criterion 4: Sites where most severity-weighted excess collisions are expected
- Criterion 5: Sites at which the product (collisions/mile-year) X (excess collisions/mile-year in standard deviations) is highest

Expected collision frequency was computed using the empirical Bayes method. At 22 of the top-ranking sites chosen by the five criteria, a detailed engineering analysis was performed to estimate the costs and safety benefits of 61 actions. When the cost-effectiveness ratios of the projects were compared, Criterion 1 and Criterion 2 performed better than the remaining 3 criteria.

Cheng and Washington (2008) proposed four quantitative evaluation tests for evaluating different screening methods, including reliability of results, ranking consistency, and false identification consistency and reliability. The tests were demonstrated using three years of Arizona road section collision data and four commonly applied screening methods (Collision Frequency Ranking, Collision Rate Ranking, Collision Reduction Potential, and Empirical Bayes (EB)). The EB method proved to be superior in most of the five evaluation tests. In contrast, identifying hot spots using Collision Rate Rankings performed the poorest.

Elvik (2008) used data for Norwegian roads to compare five techniques that embodied different degrees of control for randomness in collision counts in identifying “hazardous road locations”. As a basis for the comparison, a hazardous road location was defined as any road location that has a higher expected number of collisions than similar locations due to local risk factors present at the location. The following five techniques for identifying hazardous road locations were compared:

1. Recorded number of collisions during a specific period.
2. Observed collision rate (collisions per million vehicle kilometers) during a specific period.
3. Combination of a critical count of collisions and a collision rate above normal during a specific period.
4. Empirical Bayes estimate of the expected number of collisions at each location.
5. The size of the contribution of presumably local risk factors to the empirical Bayes estimate of the expected number of collisions at each location. Each criterion was applied to the upper 1 %, upper 2.5 % and upper 5 % of the distribution of sites according to the criterion.

The performance of the techniques was assessed in terms of sensitivity and specificity. The empirical Bayes technique was found to perform the best.

2.3 SELECTION OF METHODS FOR EVALUATION IN THIS STUDY

In selecting the methods, the intent was to eliminate methods that fall into one of two categories: (1) methods that are not conceptually sound, and (2) methods that may be conceptually sound but very difficult to implement in practice because they may require significant statistical expertise that state and local agencies may not have. The project team and CALTRANS discussed the possibility of including the CRP method as part of the evaluation, but it had to be excluded because the CRP developers did not provide access to the code that was necessary to implement the CRP method. In any case, it was felt that the SafetyAnalyst EB-based methods do address the limitations of the Table C method that the CRP method was intended to overcome. The full Bayes method was not included because it is a very complex method and very unlikely that Caltrans will implement it as there are no software tools to implement that method at this time.

The following methods were selected for evaluation:

1. Expected collisions based on the empirical Bayes (EB) approach
2. Expected Excess collisions based on the EB approach

3. Level of Service (LOSS) method
4. The Table C method currently being used by California to identify locations for further review
5. Screening based on high proportions of collision type

Since screening based on high proportions of collision type will most likely be used in combination with the other methods or when traffic volume is not available, it was mainly used as an illustration. Among the first four methods shown above, the last three methods try to identify sites based on some measure of „excess collisions‟, i.e., excess of average, expected, or critical collision frequency. The first method tries to identify sites based on the total number of expected collisions at a site. As discussed earlier, the EB methods can be applied not only to total collisions but different types of collisions and severities. However, in this evaluation, only total collisions were used, because in Table C sites are identified based on total collisions.

The first three methods make use of SPFs and hence specifically account for the fact that the relationship between collision frequency and traffic volume is not linear. As mentioned earlier, SPFs can include just traffic volume (type 1 SPFs) or other site characteristics in addition to traffic volume (type 2 SPFs). SafetyAnalyst allows only type 1 SPFs in their program. Hence, in the evaluation, some of the comparisons between the methods were done using SafetyAnalyst, but many of the comparisons were done outside of SafetyAnalyst. To allow the evaluation of SPFs that included other variables apart from traffic volume, some of the comparisons were made by developing a sliding window program to implement the Table C method and the other methods mentioned above. SafetyAnalyst was used to compare the peak search and sliding window methods for roadway segments.

3. EVALUATION APPROACH

An empirical evaluation was undertaken to measure how well each of the potential network screening methods identifies locations with promise using California State Highway data and information on locations subjected to safety investigation by CALTRANS. The approaches described below were applied to two types of intersections (rural four-leg stop-controlled and rural four-leg signalized), and two types of roadway segments (rural two-lane roads and urban freeways).

The following three approaches were considered for the evaluation:

> **Approach 1**. Compare the ability of each method to rank those locations that are more likely to have high collision frequencies in the future.
> **Approach 2**. Compare, retrospectively, the performance of each method in selecting and ranking locations that were investigated and recommended for improvement (correct positives) and those that were investigated and not recommended for improvement (false positives).
> **Approach 3**. Compare the characteristics of top ranked locations by each method.

Initially, in approach 2, we had planned to not only use the information about whether a location was recommended for improvement, but the cost-effectiveness of these improvements. However, CALTRANS indicated that such cost-effectiveness data are not readily available for use in our evaluation. This may be a topic for future research.

In performing the evaluation, we recognized that there is no perfect approach to compare the performance of different screening methods. It is possible that a method may perform better with one approach, and under certain conditions, and worse with another approach or with the same approach under other conditions. By using three approaches, we are able to obtain further insight into the performance of different methods, along with their advantages and disadvantages.

3.1 APPROACH 1

Suppose from the entire population of sites (say, intersections) of a particular class (e.g. four-legged rural signalized) in a jurisdiction there is a desire to identify, for example, the 25 intersections with the greatest need for safety investigation. Suppose analyst 1 used Method 1 to identify what they believe to be the top-ranked 25 intersections and places them in Group 1, and analyst 2 uses Method 2 to identify what they believe to be the top-ranked 25 intersections (Group 2)[3]. Several intersections will appear in both groups, but some will appear in one group and not the other. It is reasonable that the better of the two methods is the one that identifies the group that is likely to have more collisions of interest in the future. This, in effect, is the group that has the most collisions of interest in a subsequent time period. This assessment was done by

---

[3] We recognize that sites identified from Table C are not ranked. However, for the purpose of this evaluation and comparison to the other methods, the difference between observed collision frequency and N_R was used for the ranking. Similarly, for the LOSS method, the difference between the observed collision frequency and predicted collision frequency from an SPF was used for ranking.

taking one year at a time. For example, we ranked sites based on data from 2000 for each method, and then determined the number of collisions that the top ranked sites from each method experienced in the next few years.

3.2 APPROACH 2

In this approach we used available information on selected sites for safety improvement to get more insight into the ability of various methods to highly rank the sites. The list of locations identified by the Table C network screening process had resulted in a sub-set of locations recommended for safety improvement after field investigation stage. Samples of locations selected for safety improvement were used to evaluate various methods with respect to their ability to highly rank the sub-set of sites recommended for improvement. Efficient ranking methods should give a high ranking to those investigated sites that needed safety improvement treatment, and a low ranking to those sites that were investigated and did not need treatment.

In addition, for intersections we also examined false positives, i.e., locations that were investigated but not selected for improvement. Here we wanted to see if the top ranked sites from certain method(s) identified fewer false positives compared to the other methods.

3.3 APPROACH 3

The characteristics of the top ranked locations were compared to each other. Such characteristics include traffic volumes and collision frequencies, information that is readily available in the California State Highway data.

It is important to note that for the comparisons, only total collisions were considered because Table C uses only the information about total collisions. However, the other methods that were evaluated in this study can be implemented with selected collision types and severity levels. For example, the EB methods can be used to screen sites based on fatal and injury collisions or equivalent property damage only (EPDO) collisions.

In addition to comparing the methods using the three approaches we also investigated the extent of the regression to the mean issue by comparing the collision frequency of top ranked sites in 2000-2003 with the collision frequency for the same sites in 2004-2007. Further discussion of this investigation is presented in sections 6 and 7.

4. PREPARATION OF ROADWAY AND COLLISION DATA FILES FOR ANALYSIS

Roadway, intersection and collision data files were obtained from the Highway Information System (HSIS). The roadway and intersection files were obtained for 2000 to 2007.

Prior to developing the analysis databases for road segments and intersections, staff at HSIS linked the locations of intersections and ramps to the roadway segment data file in 2007. The aim of this was to identify portions of the road segments which are within the influence area of either an intersection or an interchange. It was desired to not include intersection influence areas in the analysis of roadway segments and to develop separate models for road segments within and outside of interchange influence areas. The influence area of an intersection was assumed to be within a radius of 250 feet from the intersection.

Unfortunately it is not possible to identify the center of an interchange within the roadway file. To define interchange influence areas it was decided to apply a 0.3 mile radius around all ramps and use this area as the interchange influence area.

Preparation of the intersection data started with using the 2007 intersection file as the base file. The 2007 roadway file was merged to this file using the county, route, and milepost variables in order to add the roadway class variable which was used to define rural versus urban environments. Next, each year‟s intersection file was merged by the county, route number, and milepost variables to add the major and minor road AADTs for each year as well as traffic control and number of lanes on each roadway. Only intersections which could be matched for each year and whose traffic control and number of lanes did not change were included for analysis.

Collisions within 250 feet of the intersection center were included. Counts of total collisions as well as collision type subsets were summed for each intersection from 2000 to 2007. Injury collisions were defined as those resulting in a „fatality‟, „severe injury‟ or „other visible injury‟. Other collision types queried included those defined as „sideswipe‟, „rear-end‟ and „broadside‟. The intersection types chosen for analysis were: 1) rural four-legged signalized, and 2) rural four-legged stop-controlled.

Preparation of the segment data also started with the 2007 roadway file as the base file. Previous years' road segment files were merged to this file to add the AADTs for the previous years. Due to the complexities of accounting for changing mileposts due to realignment it was assumed that mileposts did not change for the segment data. Data exploration indicated that a very low percentage of segments would be affected and it was decided that the issue was not significant for the current analysis which is focused on comparing methods for network screening.

Before matching the collision data with the road segments, those collisions coded as taking place on a ramp were first removed from the data. Total collisions in each segment as well as other collision types were summed. Injury collisions were defined as those resulting in a „fatality‟, „severe injury‟ or „other visible injury‟. Other collision types queried included those defined as

„head-on‟, „sideswipe‟, „rear-end‟, „hit object‟ and „overturn‟. The segment types chosen for analysis were: 1) rural two-lane, and 2) urban freeway. Rural two-lane segments which were within the influence area of an intersection were discarded.

4.1 INFORMATION ABOUT SITES THAT WERE SELECTED FOR INVESTIGATION AND RECOMMENDED FOR IMPROVEMENT

As discussed earlier, one of the evaluation goals was to compare the efficiency of different ranking methods to identify sites which would ultimately be selected for treatment. In other words, which methods maximize the number of correct positives and minimizes the number of false positives. This required a list of sites which were investigated, the year of the investigation and whether or not improvements were ultimately recommended.

Caltrans provided an excel file named „TableCInv6yr„ that recorded the site investigations triggered by the current Table C application. A six year history from 2003 to 2008 was made available. The information included for each site includes District, County, Route, Postmiles, Direction, Hwy/Int/Ramp, Log # (Table C All locations end in A, Wet end in W), Initiation Date, Approval Date, No Action or Improvement Recommended, Improvement Completion Date, and Investigation Date. This information was linked to the intersection and road segment databases created.

As mentioned earlier, we had initially planned on examining the cost effectiveness of treatments at these sites as part of the evaluation. However, in order to do that, we needed to know what was implemented at each site and how effective the treatment was. At a minimum, this would have required looking at individual records one at a time which was time prohibitive. Making use of information about the specific treatment that may have been implemented at a particular site may be a topic for future research.

5. SAFETY PERFORMANCE FUNCTIONS AND HIGH PROPORTION PARAMETERS

As discussed earlier, the EB methods and the LOSS method require the development of Safety Performance Functions (SPFs). Generalized linear modeling was used to estimate model coefficients assuming a negative binomial error distribution, which is consistent with the state of research in developing these models. The over-dispersion parameter (k) was also estimated in the model calibration process and used in the estimation the EB estimate of the collision frequency.

The dependent variable was either collisions per mile-year (for roadway segments) or collisions per intersection-year (for intersections). The relationship between the dependent and independent variables was assumed to be log-linear:

Dependent variable $= exp\ \beta_0 + \ {}_{i=1}^{n} \beta_i X_i$

Where, X‟s are the independent variables and β‟‟s are parameters to be estimated. For each collision type, two SPFs were developed; one type just used AADT (major and minor road AADT in the case of intersections), while the other used other available site characteristics in addition to AADT. As discussed earlier, the SPF with just AADT was called type 1 SPF, and the SPF that included other site characteristics in addition to AADT was type 2.

Details on all of the SPFs developed are provided in Appendix A. The details include the variables (X‟‟s), the parameter estimates (β‟‟s), the standard error of the estimates, the over-dispersion parameter, and the results of chi-square tests to show whether the estimates were statistically significant.

As discussed earlier, the high proportion method uses the α and β parameters of the Beta distribution for the computations. These parameters are provided in Appendix B.

6. EVALUATION RESULTS FOR INTERSECTIONS

The first part of this section investigates the regression to the mean phenomenon by comparing the collision frequencies from 2000 to 2003 with 2004 to 2007. The second part provides a discussion of the results that were obtained when the different network screening methods were compared using the three approaches that were discussed earlier. The third part is an illustration of the high proportion method using data from stop controlled intersections. The last part uses a technique called Cumulative Residual (CURE) plots to compare the performance of type 1 SPFs estimated with California data with the default SPFs in SafetyAnalyst that were recalibrated using the same California data. As mentioned earlier, the scope of this evaluation included rural four-leg minor road stop controlled intersections and rural four-leg signalized intersections.

6.1 INVESTIGATION OF THE REGRESSION TO THE MEAN PHENOMENON

To investigate the regression to the mean phenomenon, data on total collisions were compiled for each intersection for the four year period from 2000 to 2003. Based on the count of the total collisions, the intersections were divided into groups and ranked in descending order. If there is regression to the mean, then the top ranked intersections (i.e., the intersections with the most number of collisions in 2000 to 2003) will experience a decrease in collisions in 2004 to 2007, even if there were untreated (as is likely the case for most intersections), and the bottom ranked intersections (i.e., the intersections with the least number of collisions in 2000 to 2003) will experience an increase in collisions in 2004 to 2007.

Tables 6.1 and 6.2 show the results of this comparison for rural four leg stop controlled intersections and rural four leg signalized intersections. In both the tables, the first column shows the number of collisions per intersection for the different groups. The second column shows the number of intersections in that group. This is followed by columns that show the number of collisions for each intersection group and the number of collisions per intersection in each group. The last column is the % change in the number of collisions in 2004-2007 compared to the number of collisions in 2000-2003. For example, in the first row of Table 6.1, intersections in the 40+ category experienced an average of 61.75 collisions in 2000 to 2003 and 48.75 in 2004 to 2007. This represents a change of (48.75-61.75)/61.75 = -0.2105 (i.e., -21.05%).

For the stop controlled intersections, the average number of collisions per intersection in 2000 to 2003 was 3.86. It is clear that the intersection groups whose average collision frequency in 2000 to 2003 was below 3.86 experienced a significant increase in collisions in 2004 to 2007. Similarly, the intersection groups whose average collision frequency in 2000 to 2003 was above 3.86 experienced a decrease in collisions (with one exception).

For signalized intersections, the average number of collisions per intersection in 2000 to 2003 was 20.99. Due to the smaller sample of intersections, the trends are not as clear. Four of the five groups of intersections whose average collision frequency in 2000 to 2003 was below 20.99 experienced an increase in collisions in 2004 to 2007. Similarly, four out of the six groups of intersections whose average frequency in 2000 to 2003 was above 20.99 experienced a reduction in collisions.

It is possible that some of these intersections changed between 2000 and 2007 either due to the implementation of engineering treatments or decreases/increases in traffic volume. So, it is possible that some of the decrease or increase in collisions is due to these changes rather than regression to the mean. Nevertheless, it is clear that the possible bias due to regression to the mean could be substantial and needs to be accounted for even if four years of data are included.

Table 6.1 Illustration of regression to the mean in rural four-leg stop controlled intersections

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

Note: In 2000 to 2003, mean frequency was 3.86 collisions per intersection

Table 6.2 Illustration of regression to the mean in rural four leg signalized intersections

Acc/intersection in 2000 to 2003	Number of intersections	Sum of accs from 2000 to 2003	Sum of accs from 2004 to 2007	Average accs per intersection in 2000 to 2003	Average accs per intersection in 2004 to 2007	% Change
60+	4	297	369	74.25	92.25	24.24
50-59	4	218	175	54.50	43.75	-19.72
40-49	7	293	311	41.86	44.43	6.14
30-39	10	330	299	33.00	29.90	-9.39
25-29	19	520	480	27.37	25.26	-7.69
20-24	12	256	253	21.33	21.08	-1.17
15-19	16	273	289	17.06	18.06	5.86
10-14	16	195	156	12.19	9.75	-20.00

5-9	18	134	151	7.44	8.39	12.69
1-4	15	45	83	3.00	5.53	84.44
0	1	0	2	0.00	2.00	Infinite increase

Note: In 2000 to 2003, mean frequency was 20.99 collisions per intersection

6.2 RESULTS FROM THE COMPARISON OF THE METHODS USING THE THREE APPROACHES

6.2.1 Results from Approach 1

The intent of approach 1 is to compare the ability of each method to rank high those locations that are more likely to have high collision frequencies in the future. Here is an example of how this approach was implemented. Let us take year 2001 for a particular method (say the LOSS method). As mentioned earlier, in the LOSS method, the difference between the actual collisions and the predicted number of collisions (based on a SPF) is used as a way of identifying locations with promise. Using data for that year, all sites were ranked using the LOSS method. From this ranking, we chose the top X number of sites (say 10), and for these 10 sites, and computed the total number of collisions in the 'future' (i.e., after 2001: from 2002 to 2007). This process was repeated for all the methods for all the years. The first method in the list (expected collisions using the EB method) is expected to perform better in this method because the measure of interest is expected collisions unlike the other methods where the measure of interest is excess collisions.

Although the current thinking in the safety community is to use multiple years of data for network screening to account for the possible bias due to regression to the mean, the Table C procedure uses a maximum of 1 year of data, and hence 1 year of ranking data were used for the comparisons.

Tables 6.3 through 6.4 show the results from this approach. For the EB Expected, EB Expected Excess, and the LOSS methods, results are shown for type 1 SPFs with AADT as the only variable (called SPF1), type 2 SPFs with additional independent variables (called SPF2), and default SPFs from SafetyAnalyst, which have AADT as the only independent variable (designated as SPF SA). Table 6.3 shows the results when rural four leg minor road stop controlled intersections were examined using this approach. This Table shows the number of „future‟ collisions separately for each year and each method. For example, the first part of the Table shows the number of collisions from 2001 to 2007 when 2000 data was used to rank the sites in each method. The first column of the table indicates that computations were done for top 10, top 50, top 100, and top 200, in each method. If the number of collisions in a row of the table is higher for particular method, then it implies that a particular method is more effective in ranking high those locations that are more likely to have high collision frequencies in the future. For each row, the cell(s) with the highest value is highlighted in bold. Table 6.4 shows the results for rural four leg signalized intersections.

Overall, it is clear that the method that uses expected collisions (using EB) is more effective in ranking high those locations that are more likely to have high collision frequencies in the future. Based on the earlier discussion, this is not very surprising because this method identifies sites based on expected collisions unlike the other methods that identify sites based on some measure of excess collisions. Among the other methods, the Table C method is generally associated with identifying sites that have the least number of collisions in the future. Another finding is that in the majority of cases, the results are not very different regardless of the type of SPF that is used.

Table 6.3: Future Collisions When Ranked on Year Y (stop controlled)

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

Table 6.4: Future Collisions When Ranked on Year Y (signalized)

	EB Expected			EB Expected Excess			LOSS			
Top X Sites	**SPF1**	**SPF2**	**SPF SA**	**SPF1**	**SPF2**	**SPF SA**	**SPF1**	**SPF2**	**SPF SA**	**Table C**
Y = 2000										
10	**920**	**920**	**920**	769	722	777	769	722	769	598
25	1830	**1892**	1840	1664	1656	1745	1664	1671	1721	1410
50	**2918**	2906	2866	2629	2623	2644	2629	2623	2644	2296
75	3704	**3747**	3704	3268	3220	3346	3307	3220	3384	3041
Y = 2001										
10	**907**	**907**	**907**	793	793	840	811	793	793	635
25	1609	**1640**	1603	1411	1420	1522	1407	1420	1413	1294
50	**2510**	2506	2499	2374	2336	2361	2374	2336	2361	1967
75	3219	**3242**	3172	2753	2866	2763	2778	2930	2850	2574
Y = 2002										
10	**695**	**695**	**695**	570	570	570	447	447	542	447
25	**1448**	**1448**	**1448**	1235	1316	1235	1214	1170	1240	953
50	**2172**	2151	2146	1920	1956	1943	1937	1956	1943	1630
75	2658	**2661**	2641	2450	2312	2509	2481	2419	2554	2138
Y = 2003										
10	**585**	**585**	**585**	521	571	541	526	547	544	454
25	1102	**1122**	1079	992	968	994	976	914	988	835
50	**1747**	1741	1743	1607	1548	1597	1607	1548	1597	1367
75	2178	**2180**	2178	1930	1972	2005	2010	1993	2065	1866
Y = 2004										
10	**457**	**457**	**457**	328	328	350	314	314	328	322
25	832	**838**	**838**	787	797	800	774	767	779	613
50	1220	**1239**	1226	1184	1190	1208	1184	1188	1208	1043
75	1557	1563	**1571**	1468	1454	1462	1471	1461	1475	1347
Y = 2005										
10	**307**	296	**307**	295	295	295	263	259	263	221
25	536	**545**	536	502	526	502	489	522	489	420
50	820	**849**	832	785	789	785	786	789	789	677
75	1018	1018	**1020**	938	958	943	947	979	960	870
Y = 2006										
10	160	153	160	154	154	**163**	144	144	154	103
25	**262**	261	**262**	253	257	251	257	254	255	187
50	428	**430**	**430**	408	420	412	408	420	412	352
75	**527**	526	**527**	494	498	506	505	515	511	439

6.2.2 Results from Approach 2

In approach 2, the intent is to compare the ability of each method to flag and prioritize the locations previously investigated in one year, using data from the previous year as described below, and considering whether or not those locations were recommended for improvement. For implementing this approach, the first step was to compile the list of sites that were identified by the Table C procedure in each year for investigation. Let us say, we choose year 2004. We then used data from the previous year (2003) to rank these sites based on the different methods that we are evaluating. For each method, we note the number of sites that were identified as "improvement recommended". Efficient ranking methods should give a high ranking to those investigated sites that were found to be deserving of treatment, and a low ranking to those sites that were investigated and no treatment subsequently recommended. It is important to note that the starting point in this approach was in using the sites that were selected for investigation based on the results from the Table C method. Thus, the evaluation was expected to be biased in favor of Table C with respect to producing an optimal and ranked list of locations due to several factors described in section 2 of this report. However, the intent of the investigation was to see how the other methods performed when ranking the sites that were recommended for improvement.

Tables 6.5 and 6.6 show the results obtained based on this approach for stop controlled and signalized intersections. In these tables, the number of sites that were recommended for improvement is shown among the top 5 sites, top 10 sites (and so on), that were identified in each method. Results are shown for Table C, EB Expected, EB Expected Excess, and the LOSS method. For each year, in each column, cells with the highest value are shown in bold. The last column shows the total number of sites that were recommended for improvement in each year among the sites that were identified for investigation by Table C. The last column will be the same for all the methods because only sites that were identified for investigation by the Table C method were included in this evaluation. For example, in 2003, 68 stop controlled intersections were investigated and 27 were recommended for improvement (see Table 6.5). The results from Tables 6.5 and 6.6 seem to indicate that the other methods did equally well as Table C in many of the cases, and in a few cases, they did better than Table C. As before, there is no clear difference between the results obtained with the three different types of SPFs.

Table 6.5: Number of intersections selected as 'improvement recommended' (stop controlled)

	2003 Investigations				
	Top 5	Top 10	Top 20	Top 50	All (68)
Table C	3	5	**11**	22	27
EB_expected (SPF1)	3	**6**	9	**23**	27
EB_expected (SPF2)	3	**6**	9	**23**	27
EB_expected (SPF_SA)	3	**6**	9	**23**	27
EB_expected_excess (SPF1)	3	**6**	10	21	27
EB_expected_excess (SPF2)	3	**6**	10	22	27
EB_expected_excess (SPF_SA)	**4**	**6**	10	21	27
LOSS (SPF1)	**4**	4	10	22	27
LOSS (SPF2)	3	4	10	22	27
LOSS (SPF_SA)	4	**6**	9	21	27
	2004 Investigations				
	Top 5	Top 10	Top 20	Top 50	All (81)
Table C	**1**	**2**	**6**	13	18
EB_expected (SPF1)	**1**	**2**	5	10	18
EB_expected (SPF2)	**1**	**2**	5	11	18
EB_expected (SPF_SA)	**1**	**2**	5	10	18
EB_expected_excess (SPF1)	**1**	**2**	4	10	18
EB_expected_excess (SPF2)	**1**	**2**	4	10	18
EB_expected_excess (SPF_SA)	**1**	**2**	4	11	18
LOSS (SPF1)	**1**	1	4	**14**	18
LOSS (SPF2)	**1**	1	4	13	18
LOSS (SPF_SA)	0	0	3	**14**	18

	2005 Investigations				
	Top 5	Top 10	Top 20	Top 50	All (34)
Table C	**1**	1	**2**		2
EB_expected (SPF1)	**1**	1	**2**		2
EB_expected (SPF2)	**1**	**2**	**2**		2
EB_expected (SPF_SA)	**1**	**2**	**2**		2
EB_expected_excess (SPF1)	**1**	1	**2**		2
EB_expected_excess (SPF2)	**1**	1	**2**		2
EB_expected_excess (SPF_SA)	**1**	1	**2**		2
LOSS (SPF1)	**1**	1	1		2
LOSS (SPF2)	**1**	1	1		2
LOSS (SPF_SA)	**1**	1	1		2
	2006 Investigations				
	Top 5	Top 10	Top 20	Top 50	All (76)
Table C	**2**	**2**	**6**	**19**	30
EB_expected (SPF1)	0	0	4	15	30
EB_expected (SPF2)	0	0	4	14	30
EB_expected (SPF_SA)	0	0	4	15	30
EB_expected_excess (SPF1)	0	1	3	15	30
EB_expected_excess (SPF2)	0	1	3	15	30
EB_expected_excess (SPF_SA)	0	1	4	15	30
LOSS (SPF1)	1	**2**	5	16	30
LOSS (SPF2)	1	**2**	5	16	30
LOSS (SPF_SA)	**2**	**2**	5	15	30
	2007 Investigations				
	Top 5	Top 10	Top 20	Top 50	All (50)
Table C	0	**3**	**7**	17	17
EB_expected (SPF1)	0	2	6	17	17
EB_expected (SPF2)	0	2	6	17	17
EB_expected (SPF_SA)	0	2	6	17	17
EB_expected_excess (SPF1)	**1**	2	6	17	17
EB_expected_excess (SPF2)	**1**	2	6	17	17
EB_expected_excess (SPF_SA)	**1**	2	6	17	17
LOSS (SPF1)	0	**3**	6	17	17
LOSS (SPF2)	0	**3**	6	17	17
LOSS (SPF_SA)	0	**3**	6	17	17
	2008 Investigations				
	Top 5	Top 10	Top 20	Top 50	All (39)
Table C	**2**	**5**	**8**		15
EB_expected (SPF1)	**2**	3	**8**		15
EB_expected (SPF2)	**2**	3	7		15
EB_expected (SPF_SA)	**2**	3	**8**		15
EB_expected_excess (SPF1)	**2**	4	7		15
EB_expected_excess (SPF2)	**2**	4	7		15
EB_expected_excess (SPF_SA)	**2**	4	**8**		15
LOSS (SPF1)	1	3	7		15
LOSS (SPF2)	1	4	**8**		15
LOSS (SPF_SA)	1	4	**8**		15

Table 6.6: Number of intersections selected as 'improvement recommended' (signalized)

	2003 Investigations		
	Top 5	Top 10	All (16)
Table C	**2**	4	5
EB_expected (SPF1)	1	4	5
EB_expected (SPF2)	1	4	5
EB_expected (SPF_SA)	1	4	5
EB_expected_excess (SPF1)	**2**	**5**	5
EB_expected_excess (SPF2)	**2**	4	5
EB_expected_excess (SPF_SA)	**2**	4	5
LOSS (SPF1)	1	**5**	5
LOSS (SPF2)	**2**	4	5
LOSS (SPF_SA)	1	**5**	5
	2004 Investigations		
	Top 5	Top 10	All (14)
Table C	**1**	**1**	1
EB_expected (SPF1)	0	**1**	1
EB_expected (SPF2)	0	**1**	1
EB_expected (SPF_SA)	0	**1**	1
EB_expected_excess (SPF1)	0	**1**	1
EB_expected_excess (SPF2)	0	**1**	1
EB_expected_excess (SPF_SA)	0	**1**	1
LOSS (SPF1)	0	**1**	1
LOSS (SPF2)	**1**	**1**	1
LOSS (SPF_SA)	0	**1**	1
	2005 Investigations		
	Top 5	Top 10	All (5)
Table C	**1**		1
EB_expected (SPF1)	**1**		1
EB_expected (SPF2)	**1**		1
EB_expected (SPF_SA)	**1**		1
EB_expected_excess (SPF1)	**1**		1
EB_expected_excess (SPF2)	**1**		1
EB_expected_excess (SPF_SA)	**1**		1
LOSS (SPF1)	**1**		1
LOSS (SPF2)	**1**		1
LOSS (SPF_SA)	**1**		1
	2006 Investigations		
	Top 5	Top 10	All (11)
Table C	2	**5**	6
EB_expected (SPF1)	**3**	**5**	6
EB_expected (SPF2)	**3**	**5**	6
EB_expected (SPF_SA)	**3**	**5**	6

EB_expected_excess (SPF1)	**3**	**5**	6
EB_expected_excess (SPF2)	**3**	**5**	6
EB_expected_excess (SPF_SA)	**3**	**5**	6
LOSS (SPF1)	**3**	**5**	6
LOSS (SPF2)	2	**5**	6
LOSS (SPF_SA)	**3**	**5**	6
	2007 Investigations		
	Top 5	Top 10	All (9)
Table C	0		0
EB_expected (SPF1)	0		0
EB_expected (SPF2)	0		0
EB_expected (SPF_SA)	0		0
EB_expected_excess (SPF1)	0		0
EB_expected_excess (SPF2)	0		0
EB_expected_excess (SPF_SA)	0		0
LOSS (SPF1)	0		0
LOSS (SPF2)	0		0
LOSS (SPF_SA)	0		0
	2008 Investigations		
	Top 5	Top 10	All (8)
Table C	0		2
EB_expected (SPF1)	**2**		2
EB_expected (SPF2)	**2**		2
EB_expected (SPF_SA)	**2**		2
EB_expected_excess (SPF1)	1		2
EB_expected_excess (SPF2)	1		2
EB_expected_excess (SPF_SA)	1		2
LOSS (SPF1)	0		2
LOSS (SPF2)	0		2
LOSS (SPF_SA)	0		2

The analysis also examined the false positives, i.e., intersections that were selected and investigated using the Table C method for which no treatment was recommended. For this analysis, data from the previous year were used to rank all intersections (not just the intersections selected by Table C for investigation) based on the different measures. The number of false positives in the top ranked intersections from each method was thus determined. Since an intersection had to be selected by Table C to be determined as a false positive, the top ranked sites from the Table C method may, naturally, have the most false positives. However, as before, the intent of the investigation was to see how the other methods performed.

Results from this analysis are shown in Tables 6.7 and 6.8. In these tables, the number of false positives is shown among the top 5 intersections, top 10 intersections (and so on), that were identified in each method. Results are shown for Table C, EB Expected, EB Expected Excess, and the LOSS method. For each year, in each column, cells with the lowest value are shown in bold. The last column shows the total number of false positives. The last column will be the same for all the methods because the same set of intersections were used for all the methods.

For stop controlled intersections, there seems to be very little difference between the number of false positives in the top 5 and top 10 ranked sites from the different methods. However, for the top 20 and top 50 ranked sites, the top ranked sites from the EB Expected method seem to have fewer false positives compared to the other methods. One reason for this is that the EB expected method is the only method that uses the „expected‟ number of collisions whereas the other methods use some measure of „excess‟ collisions. The results are similar for signalized intersections as well where the top ranked sites from the EB Expected method seem to have fewer false positives when the top 10 ranked sites are examined.

.

Table 6.7: Number of false positives (stop controlled)

	2003 Investigations				
	Top 5	Top 10	Top 20	Top 50	All (1942)
Table C	1	3	8	21	41
EB_expected (SPF1)	2	**2**	**6**	15	41
EB_expected (SPF2)	2	**2**	7	**14**	41
EB_expected (SPF_SA)	2	**2**	7	**14**	41
EB_expected_excess (SPF1)	2	3	8	19	41
EB_expected_excess (SPF2)	2	3	7	19	41
EB_expected_excess (SPF_SA)	1	3	8	19	41
LOSS (SPF1)	**0**	4	8	22	41
LOSS (SPF2)	1	4	9	21	41
LOSS (SPF_SA)	**0**	3	8	20	41
	2004 Investigations				
	Top 5	Top 10	Top 20	Top 50	All (1942)
Table C	3	7	10	26	63
EB_expected (SPF1)	**2**	6	**8**	23	63
EB_expected (SPF2)	**2**	6	9	**21**	63
EB_expected (SPF_SA)	**2**	6	**8**	22	63
EB_expected_excess (SPF1)	3	**5**	11	26	63
EB_expected_excess (SPF2)	**2**	**5**	10	28	63
EB_expected_excess (SPF_SA)	3	**5**	10	26	63
LOSS (SPF1)	3	6	11	29	63
LOSS (SPF2)	3	6	12	28	63
LOSS (SPF_SA)	4	7	11	28	63
	2005 Investigations				
	Top 5	Top 10	Top 20	Top 50	All (1942)
Table C	**1**	4	9	15	32
EB_expected (SPF1)	2	3	**4**	**12**	32
EB_expected (SPF2)	2	3	**4**	**12**	32

EB_expected (SPF_SA)	2	3	5	13	32
EB_expected_excess (SPF1)	2	3	7	13	32
EB_expected_excess (SPF2)	2	3	7	13	32
EB_expected_excess (SPF_SA)	2	3	6	13	32
LOSS (SPF1)	**1**	3	8	13	32
LOSS (SPF2)	**1**	3	8	13	32
LOSS (SPF_SA)	2	**2**	7	**12**	32
	2006 Investigations				
	Top 5	Top 10	Top 20	Top 50	All (1942)
Table C	**3**	**8**	13	26	46
EB_expected (SPF1)	5	10	13	**19**	46
EB_expected (SPF2)	5	10	13	**19**	46
EB_expected (SPF_SA)	5	10	**12**	20	46
EB_expected_excess (SPF1)	5	9	13	25	46
EB_expected_excess (SPF2)	5	**8**	**12**	24	46
EB_expected_excess (SPF_SA)	5	**8**	13	25	46
LOSS (SPF1)	4	**8**	13	27	46
LOSS (SPF2)	4	**8**	13	27	46
LOSS (SPF_SA)	**3**	**8**	**12**	25	46
	2007 Investigations				
	Top 5	Top 10	Top 20	Top 50	All (1942)
Table C	5	6	**7**	**13**	33
EB_expected (SPF1)	5	6	9	14	33
EB_expected (SPF2)	5	6	8	14	33
EB_expected (SPF_SA)	5	6	10	14	33
EB_expected_excess (SPF1)	**4**	**5**	8	16	33
EB_expected_excess (SPF2)	**4**	6	9	15	33
EB_expected_excess (SPF_SA)	**4**	**5**	8	17	33
LOSS (SPF1)	5	7	**7**	15	33
LOSS (SPF2)	5	7	**7**	15	33
LOSS (SPF_SA)	**4**	7	**7**	**13**	33
	2008 Investigations				
	Top 5	Top 10	Top 20	Top 50	All (1942)
Table C	2	3	4	10	24
EB_expected (SPF1)	1	**1**	3	**8**	24
EB_expected (SPF2)	1	**1**	**2**	**8**	24
EB_expected (SPF_SA)	**0**	**1**	3	**8**	24
EB_expected_excess (SPF1)	1	2	3	**8**	24
EB_expected_excess (SPF2)	1	2	3	**8**	24
EB_expected_excess (SPF_SA)	1	2	3	9	24
LOSS (SPF1)	2	2	4	11	24

LOSS (SPF2)	2	2	4	11	24
LOSS (SPF_SA)	1	2	5	11	24

Table 6.8: Number of false positives (signalized)

	2003 Investigations		
	Top 5	Top 10	All (122)
Table C	3	6	11
EB_expected (SPF1)	**1**	**2**	11
EB_expected (SPF2)	**1**	**2**	11
EB_expected (SPF_SA)	**1**	**2**	11
EB_expected_excess (SPF1)	3	3	11
EB_expected_excess (SPF2)	2	3	11
EB_expected_excess (SPF_SA)	2	3	11
LOSS (SPF1)	3	5	11
LOSS (SPF2)	2	5	11
LOSS (SPF_SA)	3	5	11
	2004 Investigations		
	Top 5	Top 10	All (122)
Table C	**2**	5	13
EB_expected (SPF1)	3	**3**	13
EB_expected (SPF2)	3	**3**	13
EB_expected (SPF_SA)	3	**3**	13
EB_expected_excess (SPF1)	3	5	13
EB_expected_excess (SPF2)	**2**	6	13
EB_expected_excess (SPF_SA)	3	5	13
LOSS (SPF1)	**2**	5	13
LOSS (SPF2)	**2**	5	13
LOSS (SPF_SA)	3	6	13
	2005 Investigations		
	Top 5	Top 10	All (122)
Table C	**1**	2	4
EB_expected (SPF1)	**1**	2	4
EB_expected (SPF2)	**1**	2	4
EB_expected (SPF_SA)	**1**	2	4
EB_expected_excess (SPF1)	**1**	**1**	4
EB_expected_excess (SPF2)	**1**	**1**	4
EB_expected_excess (SPF_SA)	**1**	2	4
LOSS (SPF1)	**1**	**1**	4
LOSS (SPF2)	**1**	**1**	4
LOSS (SPF_SA)	**1**	**1**	4
	2006 Investigations		

	Top 5	Top 10	All (122)
Table C	**2**	**2**	5
EB_expected (SPF1)	**2**	**2**	5
EB_expected (SPF2)	**2**	**2**	5
EB_expected (SPF_SA)	**2**	**2**	5
EB_expected_excess (SPF1)	**2**	**2**	5
EB_expected_excess (SPF2)	**2**	**2**	5
EB_expected_excess (SPF_SA)	**2**	**2**	5
LOSS (SPF1)	**2**	**2**	5
LOSS (SPF2)	**2**	**2**	5
LOSS (SPF_SA)	**2**	**2**	5
	2007 Investigations		
	Top 5	Top 10	All (122)
Table C	3	**3**	9
EB_expected (SPF1)	**2**	**3**	9
EB_expected (SPF2)	3	**3**	9
EB_expected (SPF_SA)	**2**	**3**	9
EB_expected_excess (SPF1)	3	4	9
EB_expected_excess (SPF2)	**2**	4	9
EB_expected_excess (SPF_SA)	**2**	4	9
LOSS (SPF1)	3	4	9
LOSS (SPF2)	**2**	4	9
LOSS (SPF_SA)	3	4	9
	2008 Investigations		
	Top 5	Top 10	All (122)
Table C	3	5	6
EB_expected (SPF1)	**2**	**3**	6
EB_expected (SPF2)	**2**	**3**	6
EB_expected (SPF_SA)	**2**	**3**	6
EB_expected_excess (SPF1)	**2**	**3**	6
EB_expected_excess (SPF2)	**2**	**3**	6
EB_expected_excess (SPF_SA)	**2**	**3**	6
LOSS (SPF1)	**2**	5	6
LOSS (SPF2)	**2**	5	6
LOSS (SPF_SA)	**2**	5	6

6.2.3 Results from Approach 3

Here, the intent is to compare the characteristics of top ranked sites from each method. The two main characteristics that were selected for this comparison were the total intersection AADT and the expected number of collisions. The expected number of collisions was chosen (instead of the actual number of collisions) because it corrects for possible bias due to regression to the mean and is a better estimate of the true long-term collision propensity. To implement this approach, sites were ranked based on each method, and for the top ranked sites, the average total intersection AADT and the average expected total collisions were computed. Results are shown for the most recent 3 years of data. For EB Expected, EB Expected Excess, and the LOSS methods, computations were done using type 1 SPFs.

Table 6.9 shows the results for stop controlled intersections and Table 6.10 shows the results for signalized intersections. As expected, the top ranked sites from the EB Expected method have the highest average AADT and the highest number of expected collisions. At the same time, it is clear that the top ranked sites from the Table C method have the lowest average AADT and the lowest expected number of collisions. It is important to note that the Table C method does not account for the non-linear relationship between traffic volume and collision frequency, and that may be a reason why the top ranked sites in the Table C method have fewer expected collisions and tend to have lower average AADTs compared to the top ranked sites in the LOSS method. On average, the top ranked sites from the EB Expected Excess method have more expected collisions compared to the LOSS method, because the EB Expected Excess method explicitly accounts for regression to the mean.

Table 6.9: Average total intersection AADT and average expected collisions for top ranked sites (stop controlled intersections)

	EB Expected		**EB Expected Excess**		**LOSS**		**Table C**	
Top X Sites	**AADT**	**Expected Total**	**AADT**	**Expected Total**	**AADT**	**Expected Total**	**AADT**	**Expected Total**
Y = 2004								
10	30553	11.2	23301	10.7	20474	10.2	18031	9.8
50	23945	6.6	20301	6.2	16720	5.5	11123	4.9
100	22109	5.2	16985	4.9	15456	4.5	10308	3.8
200	18874	3.9	15016	3.6	13199	3.3	7786	2.5
Y = 2005								
10	25120	10.4	24308	10.1	21535	9.7	19626	9.2
50	23525	6.3	17711	6.0	15952	5.4	12280	5.0
100	21356	5.0	16077	4.7	13978	4.3	10532	3.8
200	19761	3.8	14246	3.4	12468	3.2	7696	2.5
Y = 2006								
10	27282	10.5	25065	9.3	17223	7.8	15879	8.2
50	22424	6.2	18347	5.8	17000	5.5	11009	4.5
100	21538	4.7	16527	4.4	14124	4.0	10422	3.5
200	19668	3.6	14795	3.3	12602	2.9	7020	2.2

Table 6.10: Average total intersection AADT and average expected collisions for top ranked sites (signalized intersections)

	EB Expected		**EB Expected Excess**		**LOSS**		**Table C**	
Top X Sites	**AADT**	**Expected Total**	**AADT**	**Expected Total**	**AADT**	**Expected Total**	**AADT**	**Expected Total**
Y = 2004								

10	45589	17.5	31879	14.2	30219	13.7	24119	12.2
25	34996	12.4	30657	11.7	28801	11.0	20699	9.3
50	27796	9.4	24173	8.7	24173	8.7	21087	8.0
75	25670	7.8	22358	7.3	22415	7.3	20640	6.9
Y = 2005								
10	44134	17.3	40464	17.0	40053	16.0	23613	11.6
25	34225	11.7	29413	10.9	28749	10.6	22276	9.3
50	28451	8.7	24084	8.1	24167	8.2	20112	7.1
75	25936	7.3	21715	6.7	21880	6.7	19606	6.1
Y = 2006								
10	44766	14.7	40166	14.2	31987	12.3	24070	10.0
25	34858	10.8	30157	10.1	29389	9.7	21398	7.7
50	29253	8.1	25268	7.6	25268	7.6	21175	6.5
75	26098	6.6	22428	6.1	22820	6.2	20004	5.5

6.3 ILLUSTRATION OF THE HIGH PROPORTION METHOD

Where traffic volumes, a necessity for the methods based on SPFs and for the Table C method, are not available, the high proportion method may be useful for screening locations for specific collision types. As mentioned earlier, Kononov (2002) argues that proportion by collision type can be useful even under situations where traffic volume data are available and SPFs can be developed. This method is available in Module 1 (network screening) of SafetyAnalyst.

To illustrate this method, the database of 1,942 rural four-legged stop-controlled intersections was used (this is the same database of stop controlled intersections used in the other comparisons). The database was split into two time periods, 2000 to 2003 and 2004 to 2007. The first time period was used to rank sites. Two collision types, rear-end and broadside, were separately screened for using the EB estimate of expected collisions and the high proportion method. The high proportion method screened by the probability that a site‟s proportion of a specific collision type is higher than the proportion for the location type. The alpha and beta parameters from Appendix B were used, and the mean proportions were 24% and 43% for rear-end and broadside collisions respectively.

Table 6.11 below shows for the top 20 sites selected by the EB expected collision method (using type 1 SPFs), where these sites were ranked by the high proportion method. For rear-end collisions, 11 of the top 20 sites by the EB method were also ranked in the top 20 by the high proportion method. Although not shown in the table, 14 of 20 were ranked in the top 26. For broadside collisions, 11 of the top 20 sites ranked by the EB methods were also in the top 20 when ranked by the high proportion method.

Table 6.11 Illustration of High Proportion Method

Ranking by EB Expected Rear-End	Ranking by Probability Expected Proportion of Rear-End Exceeds Mean Proportion	Ranking by EB Expected Broadside	Ranking by Probability Expected Proportion of Broadside Exceeds Mean Proportion
1	4	1	1
2	2	2	2
3	1	3	600
4	43	4	33
5	23	5	12
6	1902	6	6
7	5	7	3
8	32	8	28
9	11	9	118
10	19	10	17
11	148	11	115
12	12	12	96
13	130	13	16
14	25	14	7
15	17	15	30
16	6	16	8
17	10	17	9
18	26	18	4
19	83	19	116
20	14	20	68

Further, the number of collisions in the 2004 to 2007 time period were summed for the top 10, 25, 50, and 75 sites ranked by both the EB expected and high proportions methods for both rear-end and broadside collisions. It can be seen in Tables 6.12 and 6.13 below that the high proportions method is not as effective as the EB method in identifying sites that have higher collisions in the future. This is not surprising since the high proportion method is screening by proportions and not the frequency of these collision types, but it does suggest that where AADTs (and correspondingly, SPFs) are available, the EB methods can be very effective for network screening.

Table 6.12: Number of Rear-End Collisions in the Future (2004 to 2007) for top ranked sites in EB Expected and High Proportion Method

Method for Rear-End	Top 10	Top 25	Top 50	Top 75
EB_expected	131	300	437	540
High Proportion	89	235	373	422

Table 6.13: Number of Broadside Collisions in the Future (2004 to 2007) for top ranked sites in EB Expected and High Proportion Method

Method for Broadside	Top 10	Top 25	Top 50	Top 75
EB_expected	125	278	558	754
High Proportion	111	266	491	657

6.4 USE OF CURE PLOTS TO COMPARE SAFETY PERFORMANCE FUNCTIONS

Given the potential application of SafetyAnalyst in California it was sought to compare the performance of the default SPFs from SafetyAnalyst with those developed in the current project directly from data in California. SafetyAnalyst allows the user to use either the default SPFs or type 1 SPFs directly estimated with local data.

It is important to note that default SPFs from SafetyAnalyst (including the results discussed earlier), were recalibrated to the California data collected for this project using the calibration procedure documented in the Highway Safety Manual (SafetyAnalyst also uses the same procedure for calibrating default SPFs to local data). In this procedure, a calibration factor (multiplier) is calculated as the ratio of the sum of collision counts for the calibration data to the sum of the predictions from the model. The method of comparison used Cumulative Residual (CURE) Plots. In this method, documented by Hauer and Bamfo (1997), the cumulative residuals (the difference between the observed and predicted collisions for each location) are plotted in increasing order for each covariate, e.g. AADT, separately. Also plotted are graphs of the 95% confidence limits. If there is no bias in the model, the plot of cumulative residuals should stay inside of these limits. The graph shows how well the model fits the data with respect to each individual covariate. The CURE plots provide the opportunity to evaluate how a model performs over the range of covariates. CURE plots should be constructed for each continuous variable within the SPF (CURE plots are not very effective in assessing the performance associated with categorical variables). In most cases, this will be AADT for segments and major and minor road AADT for intersections. Due to the large number of SPFs calibrated for the current project and the absence of collision type SPFs in SafetyAnalyst only the SPFs using AADTs as the independent variables for total collisions were compared.

6.3.1 Rural Four-Leg Stop-Controlled Intersections

Figure 6.1 plots the cumulative residuals of both the default SafetyAnalyst (SA) model and the model estimated from California data against the major road AADT. The 95% confidence limits are not provided in order to make the figure less cluttered. The figure shows that the two models provide very close predictions across the range of major road AADT.

Figure 6.2 shows the cumulative residuals for minor road AADT. In contrast, there is a stark difference in the performance of the two models. The plot of cumulative residuals for the SafetyAnalyst model tend to show an underprediction at lower AADTs and a overprediction at higher AADTs, with the plot of cumulative residuals maxing out at about +800. The Caltrans model tends to overpredict in the mid-range of AADTs from about 1,000 to 1,500 and the plot of cumulative residuals maxes out at approximately -600.

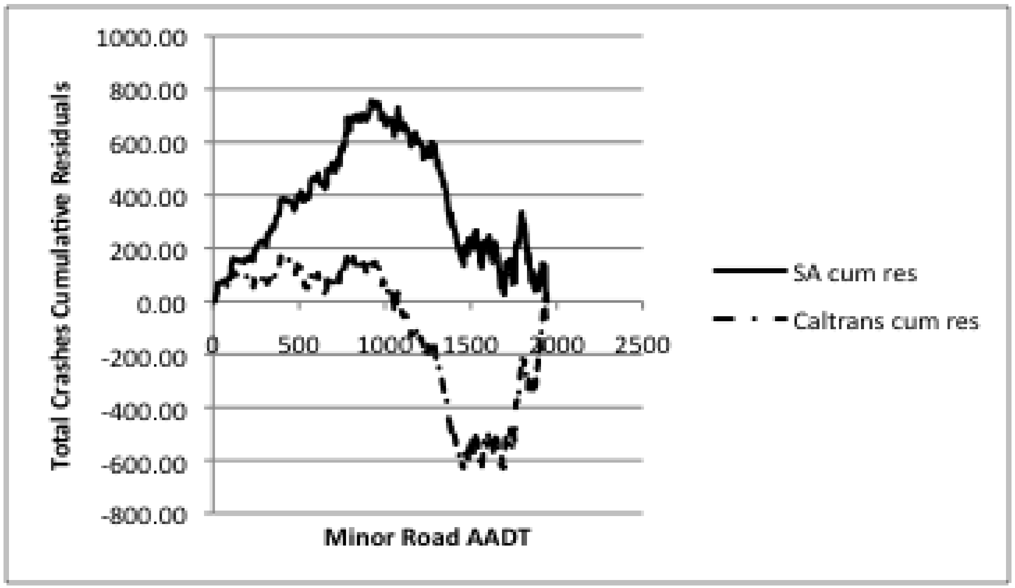


Figure 6.1 CURE Plot for Rural Four-Leg Stop-Controlled Intersections: SafetyAnalyst and Caltrans Models vs. Major Road AADT

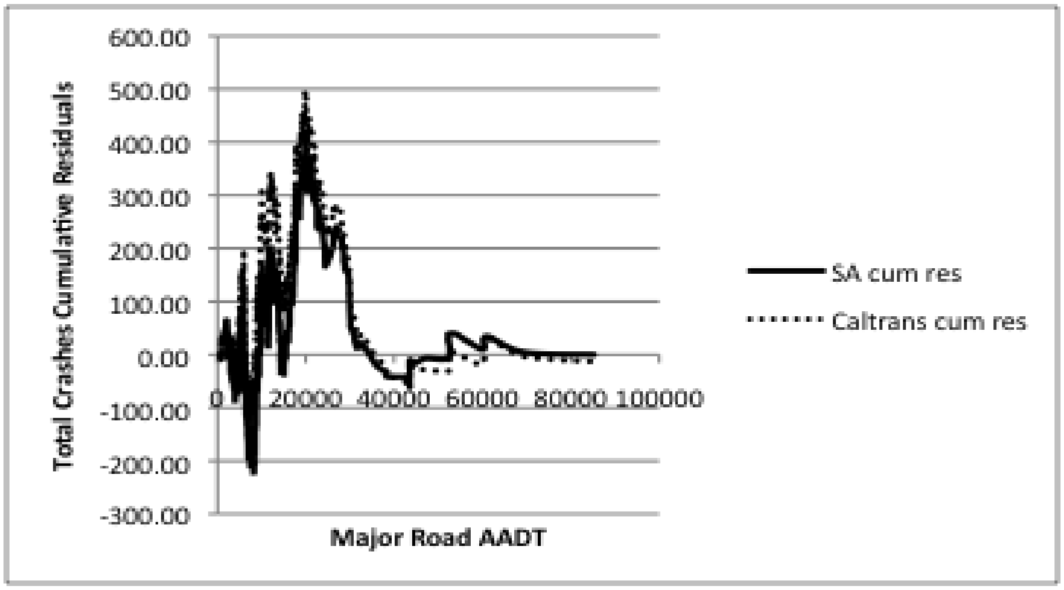


Figure 6.2 CURE Plot for Rural Four-Leg Stop-Controlled Intersections: SafetyAnalyst and Caltrans Models vs. Minor Road AADT

Because the two models do not perform similarly versus minor road AADT it was of interest to see how each performs compared to the 95% confidence limits (see Figures 6.3 and 6.4). Both models exhibit some prediction bias within certain ranges of AADT as can be seen by the plot of cumulative residuals straying outside the 95% confidence limits. The Caltrans model does perform better though, oscillating closer around 0 and with a smaller maximum deviation from 0.

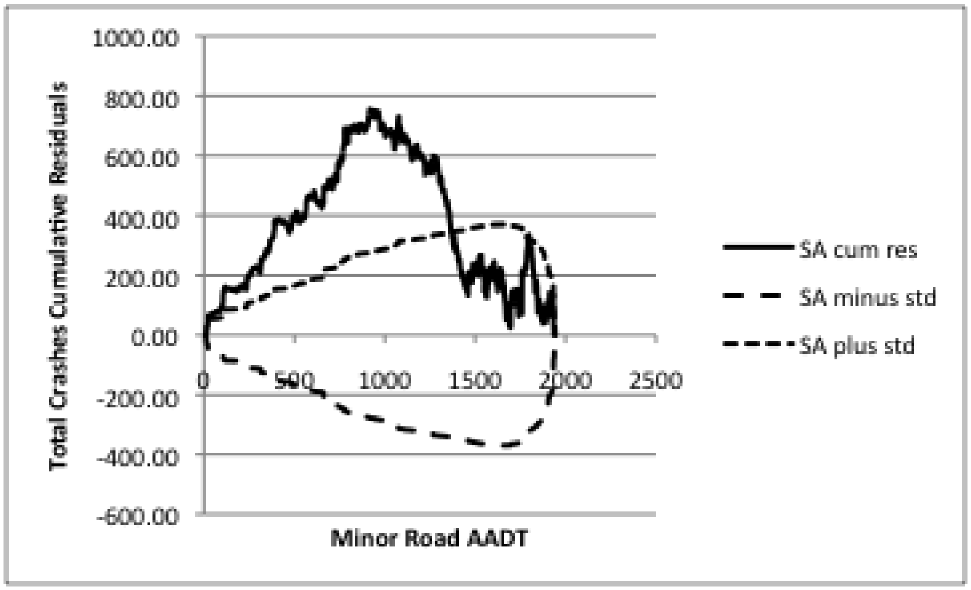


Figure 6.3 CURE Plot for Rural Four-Leg Stop-Controlled Intersections: SafetyAnalyst Model vs. Minor Road AADT

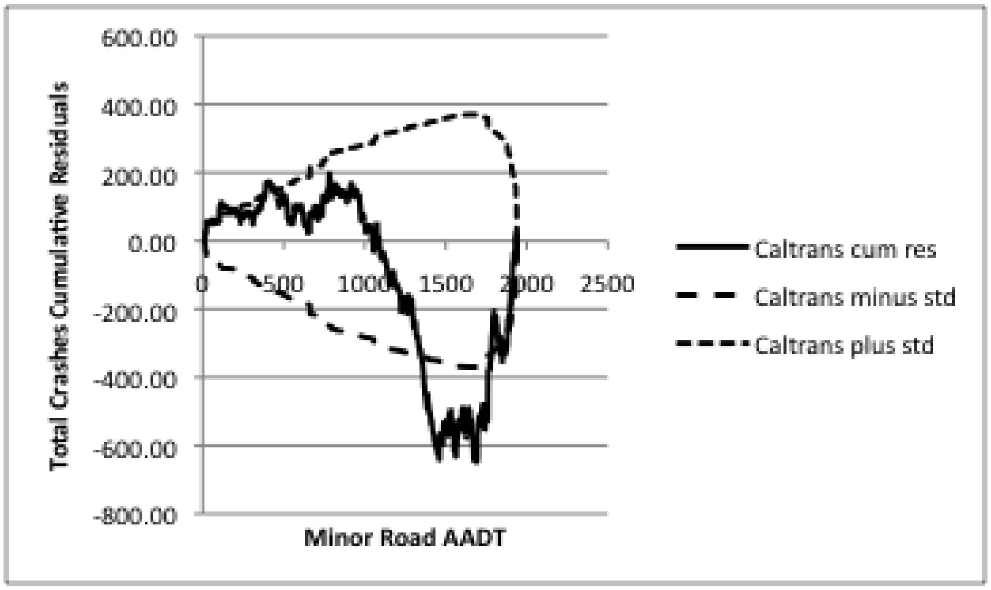


Figure 6.4 CURE Plot for Rural Four-Leg Stop-Controlled Intersections: Caltrans Model vs. Minor Road AADT

6.3.2 Rural Four-Leg Signalized Intersections

Figure 6.5 plots the cumulative residuals of both the SafetyAnalyst (SA) model and the model directly estimated with California data against the major road AADT. Again, the plots of the 95% confidence limits are not provided in order to make the figure less cluttered. The figure shows that the Caltrans model does perform better than the SafetyAnalyst model across the range of AADT.

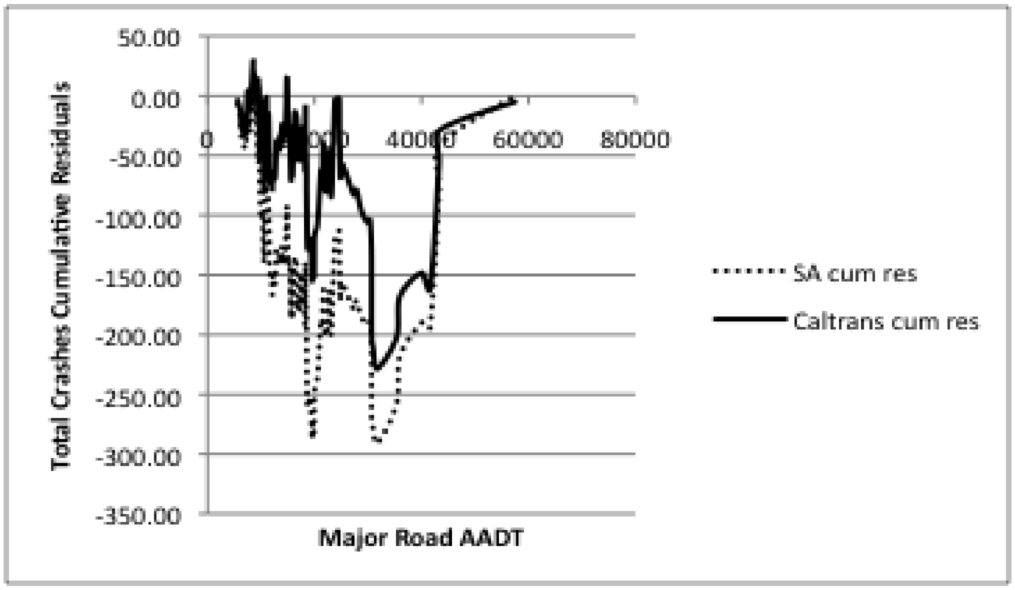


Figure 6.5 CURE Plot for Rural Four-Leg Signalized Intersections: SafetyAnalyst and Caltrans Models vs. Major Road AADT

Figures 6.6 and 6.7 plot the cumulative residuals for both models and the boundary lines. Both models perform quite well with the cumulative residuals largely staying within the 95% confidence limits.

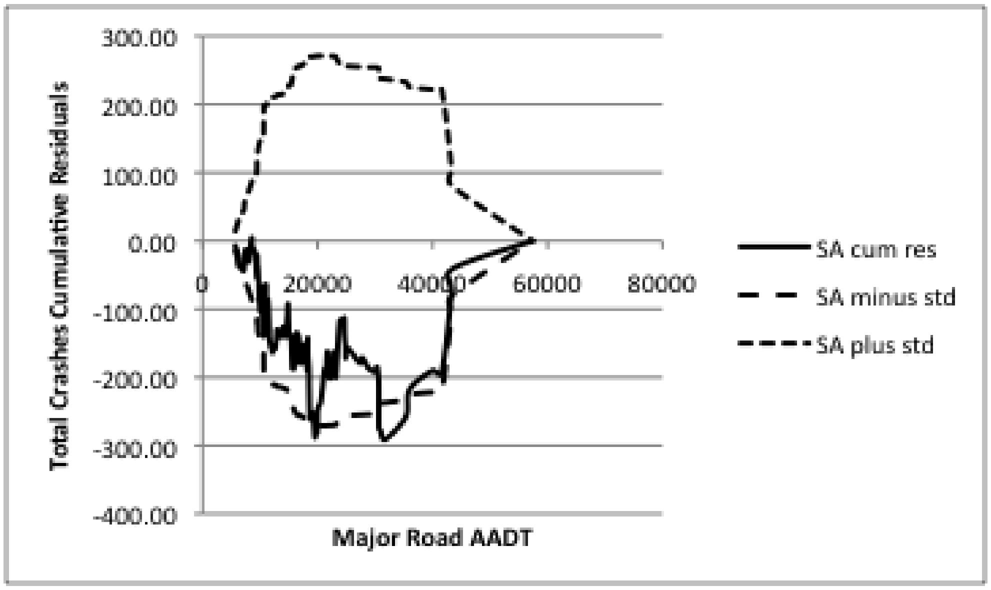


Figure 6.6 CURE Plot for Rural Four-Leg Signalized Intersections: SafetyAnalyst Model vs. Major Road AADT

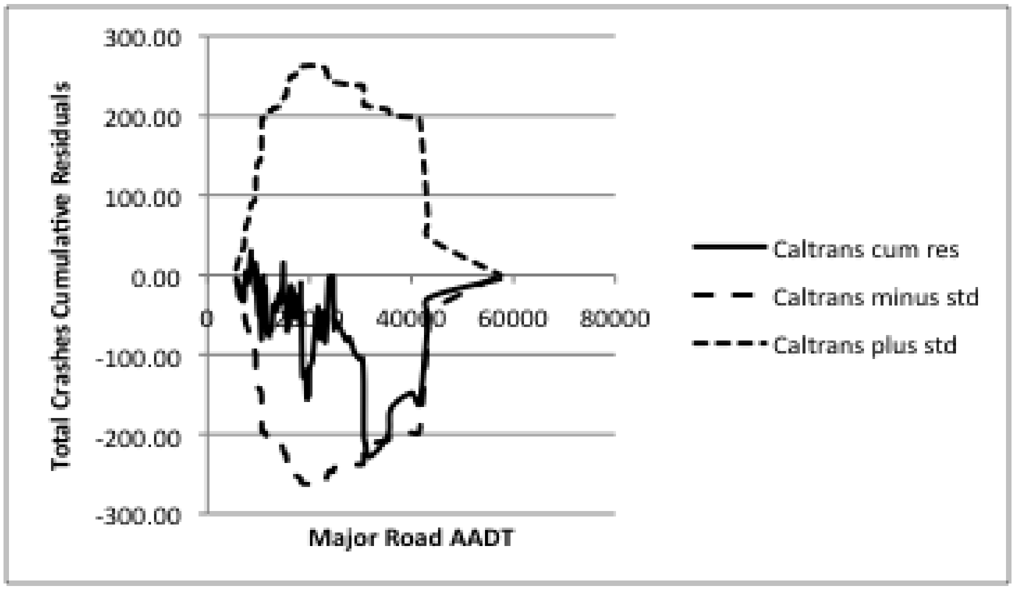


Figure 6.7 CURE Plot for Rural Four-Leg Signalized Intersections: Caltrans Model vs. Major Road AADT

Figure 6.8 plots the cumulative residuals of both the SafetyAnalyst (SA) model and the new model (Caltrans) against the minor road AADT. The figure shows that the Caltrans model performs somewhat better than the SafetyAnalyst model across the range of AADT. Figures 6.9 and 6.10 plot the cumulative residuals for both models and the boundary lines. Both models perform quite well with the cumulative residuals staying within the 95% confidence limits.

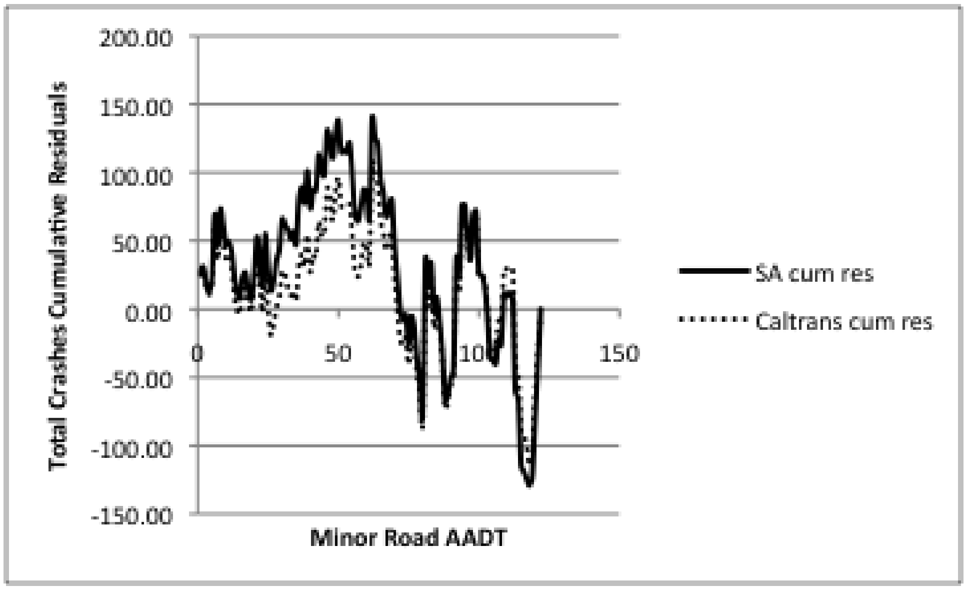


Figure 6.8 CURE Plot for Rural Four-Leg Signalized Intersections: SafetyAnalyst and Caltrans Models vs. Minor Road AADT

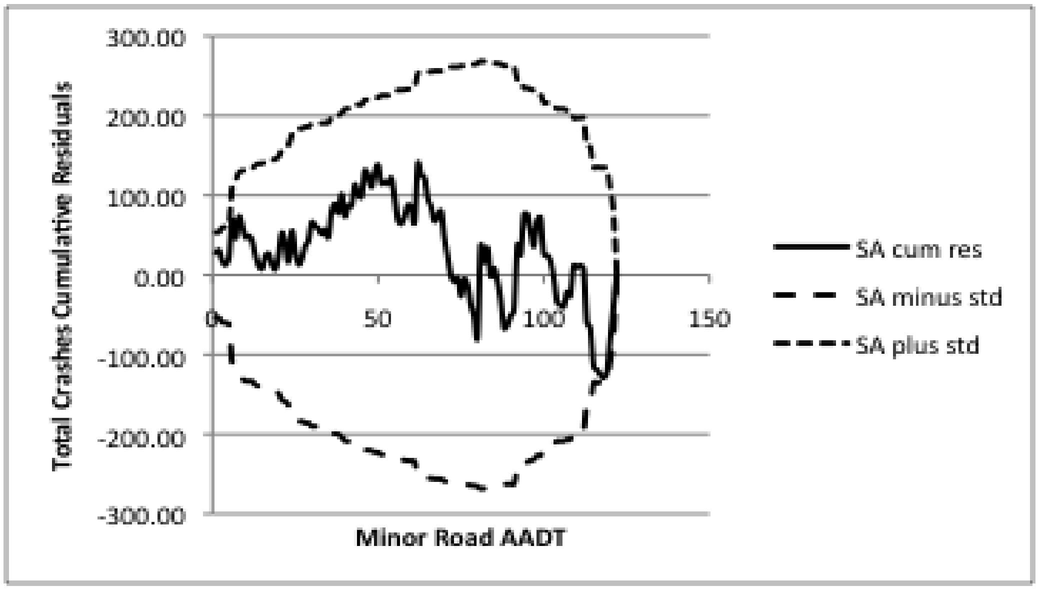


Figure 6.9 CURE Plot for Rural Four-Leg Signalized Intersections: SafetyAnalyst Model vs. Minor Road AADT

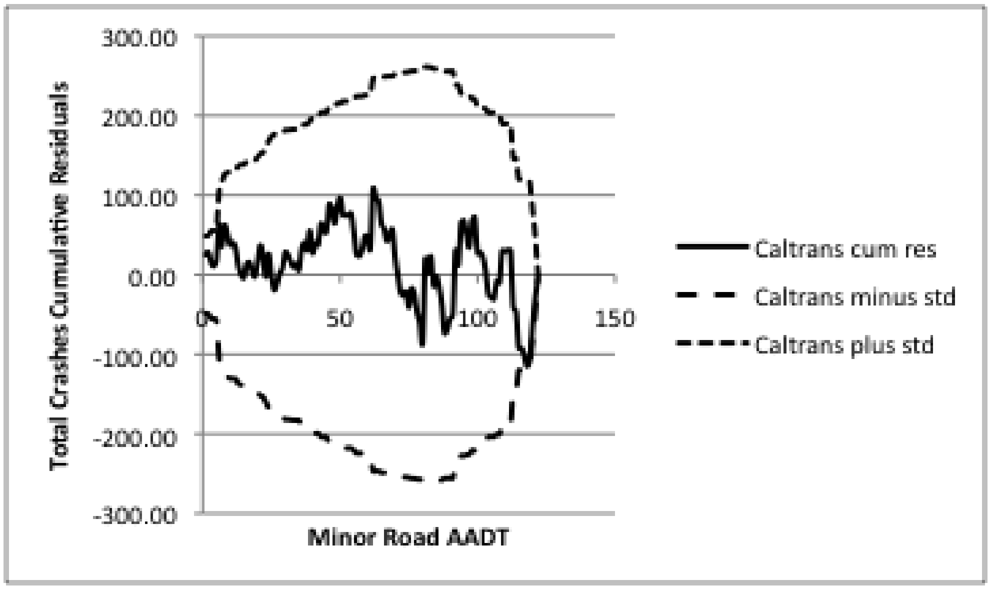


Figure 6.10 CURE Plot for Rural Four-Leg Signalized Intersections: Caltrans Model vs. Minor Road AADT

In general, the SPFs directly estimated using the Caltrans data showed a better fit to the data then the recalibrated SafetyAnalyst models. The improvement was small in some cases, however, in others it was significant. Given these results it is recommended to apply these new Caltrans-based models in lieu of the default ones in SafetyAnalyst should SafetyAnalyst be applied in California.

7. EVALUATION RESULTS FOR ROADWAY SEGMENTS

The first part of this section provides a brief overview of the sliding window program that was developed as part of this effort to allow the comparison of methods based on SPFs and the Table C method. The next part of this section investigates the regression to the mean phenomenon by comparing the collision frequencies from 2000 to 2003 with 2004 to 2007. Next is a discussion of the results that were obtained when the different network screening methods were compared using the three approaches that were discussed earlier. Following that is a discussion of the process that was used to import data into SafetyAnalyst and the results of the comparison between sliding window and peak search methods. Finally, there is a discussion of the Cumulative Residual (CURE) plots that were used to compare the performance of type 1 SPFs estimated with California data with the default SPFs in SafetyAnalyst that were recalibrated with California data. As mentioned earlier, the scope of this evaluation included rural two lane roads and urban freeways.

7.1. DEVELOPMENT OF SLIDING WINDOW PROGRAM

In order to compare the performance of the Table C methods with the EB methods that use type 2 SPFs, a sliding window program was developed. This program used a 0.2 mile window that moved 0.02 miles with every increment (similar to the approach used in the Table C procedure). A new window was started whenever the following features changed in a particular route: number of lanes, divided/undivided, area type (rural versus urban) within /outside the influence of ramps, and terrain. For each window, the following measures were computed for each year: EB Expected Collisions, EB Expected Excess Collisions, difference between Observed Frequency and Predicted Frequency (for the LOSS method), and the difference between Observed Frequency and N_R (for the Table C method). For isolated segments shorter than the window length (i.e., shorter than 0.2 miles), all the measures were calculated and converted to a per-mile basis. For the methods that used SPFs, measures were computed using type 1 SPFs, type 2 SPFs, and the default SPFs from SafetyAnalyst. Each segment was then assigned the maximum value calculated at any window position overlapping the segment. It is important to note that this is a slight departure from the Table C method as implemented in California, where the results are examined at the window level rather than at the segment level. However, the EB methods as applied in SafetyAnalyst have traditionally focused flagging an entire segment rather than the triggering window, and to be consistent, the results were examined at the segment level for the Table C method as well.

7.2 INVESTIGATION OF THE REGRESSION TO THE MEAN PHENOMENON

To investigate the regression to the mean phenomenon, data on total collisions were compiled for each 0.2 mile window for the four year period from 2000 to 2003. Based on the count of the total collisions, the windows were divided into groups and ranked in descending order. If there is regression to the mean, then the top ranked windows (i.e., the windows with the most number of collisions in 2000 to 2003) will experience a decrease in collisions in 2004 to 2007 and the bottom ranked windows (i.e., the windows with the least number of collisions in 2000 to 2003) will experience an increase in collisions in 2004 to 2007.

Tables 7.1 and 7.2 show the results of this comparison for rural two lane roads and urban freeways. In both the tables, the first column shows the number of collisions per 0.2 mile window for the different groups. The second column shows the number of windows in that group. This is followed by columns that show the number of collisions for each window group and the number of collisions per window in each group. The last column is the % change in the number of collisions in 2004-2007 compared to the number of collisions in 2000-2003.

For the rural two lane roads, the average number of collisions per 0.2 mile window in 2000 to 2003 was 1.79. It is clear that the groups whose average collision frequency in 2000 to 2003 was below 1.79 experienced a significant increase in collisions in 2004 to 2007. Similarly, the groups whose average collision frequency in 2000 to 2003 was above 1.79 experienced a decrease in collisions.

For urban freeways, the average number of collisions per intersection in 2000 to 2003 was 55.88. All the seven groups whose average collision frequency in 2000 to 2003 was below 55.88 experienced an increase in collisions in 2004 to 2007. However, four of the groups whose average frequency in 2000 to 2003 was above 55.88 also experienced an increase in collisions in 2004 to 2007, although the percentage increase for these four group were substantially lower than the percentage increase for the seven groups whose average group collision frequency in 2003 to 2003 was below 55.88. The top three ranked groups experienced a reduction in collisions in 2004-2007.

Again, as discussed in the previous section on intersections, it is possible that some of these locations changed between 2000 and 2007 either due to the implementation of engineering treatments or decreases/increases in traffic volume. So, it is possible that some of the decrease or increase in collisions is due to these changes rather than regression to the mean. Nevertheless, it is clear that by accounting for the possible bias due to regression to the mean, we can reduce the chances of selecting sites that have randomly high collision counts.

Table 7.1 Illustration of regression to the mean in rural two lane road data

Acc/window in 2000 to 2003	**Number of windows**	**Sum of accs from 2000 to 2003**	**Sum of accs from 2004 to 2007**	**Average accs per window in 2000 to 2003**	**Average accs per window in 2004 to 2007**	**% Change**
40+	43	1831	981	42.58	22.81	-46.42
30-39	173	5962	5019	34.46	29.01	-15.82
25-29	229	6048	5160	26.41	22.53	-14.68
20-24	588	12904	9661	21.95	16.43	-25.13
15-19	1511	25063	19665	16.59	13.01	-21.54
10-14	4836	55548	45541	11.49	9.42	-18.02
9	2280	20520	16498	9.00	7.24	-19.60
8	3078	24624	20740	8.00	6.74	-15.77
7	4403	30821	25527	7.00	5.80	-17.18
6	6183	37098	30895	6.00	5.00	-16.72
5	9249	46245	37419	5.00	4.05	-19.09
4	14551	58204	47196	4.00	3.24	-18.91

3	24355	73065	61702	3.00	2.53	-15.55
2	40927	81854	74857	2.00	1.83	-8.55
1	74385	74385	86164	1.00	1.16	15.84
0	122641	0	78126	0	0.63703	infinite

Note: In 2000 to 2003, mean frequency was 1.79 collisions per 0.2 mile window

Table 7.2 Illustration of regression to the mean in urban freeway data

Acc/window in 2000 to 2003	Number of windows	Sum of accs from 2000 to 2003	Sum of accs from 2004 to 2007	Average accs per window in 2000 to 2003	Average accs per window in 2004 to 2007	% Change
401+	189	89879	84955	475.55	449.50	-5.48
301-400	482	162379	149522	336.89	310.21	-7.92
201-300	2311	552739	540993	239.18	234.09	-2.13
151-200	3158	544156	551448	172.31	174.62	1.34
101-150	6280	767852	810533	122.27	129.07	5.56
76-100	5604	487189	526040	86.94	93.87	7.97
51-75	9265	574350	628569	61.99	67.84	9.44
41-50	5417	245438	273439	45.31	50.48	11.41
31-40	7097	249825	286879	35.20	40.42	14.83
21-30	9593	242516	288366	25.28	30.06	18.91
11-20	12358	189038	235056	15.30	19.02	24.34
6-10	6605	52899	71345	8.01	10.80	34.87
1-5	5855	18728	31499	3.20	5.38	68.19
0	536	0	1342	0.00	2.50	Infinite

Note: In 2000 to 2003, mean frequency was 55.88 collisions per 0.2 mile window

7.2 RESULTS FROM THE COMPARISON OF THE METHODS USING THE THREE APPROACHES

7.2.1 Results from Approach 1

The intent of this approach is to determine if any of the methods are more effective in identifying sites that have more collisions in the future. This approach was applied in the same way it was applied for intersections by taking one year at a time and computing the collisions in the future for the top ranked sites. Tables 7.3 and 7.4 show the number of collisions per mile for the top ranked sites in each method for urban freeways and rural two lane roads, respectively. All segments including those shorter than 0.1 mile were included in the analysis, but only the results from sites 0.1 mile or longer are included in these Tables to prevent the findings from being skewed by extremely short sections. In these tables, the cell with the highest value in each row is highlighted in bold.

It is clear from Tables 7.3 and 7.4 that for the large majority of cases, the top ranked sites from EB Expected method are associated with the highest future collisions per mile. For urban freeways, the top ranked sites from the other three methods have similar overall future collisions per mile, especially if a larger group of sites (e.g., sites ranked 100 or lower) are considered.

However, for rural two lane roads, there is a definite trend where the top ranked sites from the EB Expected Excess method have the second highest future collisions followed by the LOSS method and then the Table C method. As mentioned earlier, the LOSS method does not account for regression to the mean and the Table C method does not account for regression to the mean and the non-linear relationship between collision frequency and traffic volume. That may be part of the reason why the LOSS and Table C methods are not as effective in identifying sites with a large number of future collisions. Another finding is that in the majority of cases, the results are not very different regardless of the type of SPF that is used.

Table 7.3 Future collisions per mile when ranked on year Y (urban freeways)

	EB Expected			EB Expected Excess			LOSS			Table C
Top X	SPF1	SPF2	SPF SA	SPF1	SPF2	SPF SA	SPF1	SPF2	SPF SA	
Y=2000										
10	1575.3	1575.3	1575.3	**1842.5**	1840.0	**1842.5**	1734.5	1734.5	1734.5	1734.5
25	1496.5	**1500.9**	1496.5	1315.2	1315.2	1315.2	1315.2	1307.1	1295.1	1304.8
50	1314.8	1314.6	1324.1	1342.7	1361.3	1363.2	1342.7	1351.5	1359.5	**1364.1**
100	**1273.9**	1250.9	1238.6	1210.0	1188.1	1161.6	1147.8	1157.5	1140.7	1167.8
250	1039.1	**1044.5**	1037.4	1004.0	992.3	974.5	996.9	987.1	966.9	1025.5
500	**916.3**	914.7	916.1	893.5	888.4	858.8	871.2	872.5	835.1	854.6
1000	**770.9**	**770.9**	769.9	733.3	732.7	725.6	725.1	728.1	712.9	737.5
Y=2001										
10	**1800.4**	**1800.4**	1755.5	1411.7	1428.0	1315.3	1428.0	1428.0	1315.3	1315.3
25	**1574.5**	**1574.5**	**1574.5**	1552.2	1440.3	1440.3	1458.6	1440.3	1309.6	1401.5
50	1311.0	1311.0	**1314.5**	1236.8	1207.0	1219.8	1239.2	1249.3	1257.3	1255.6
100	1064.6	1090.5	1084.4	1062.8	1055.4	1016.0	1061.7	1059.3	1016.2	**1090.6**
250	904.5	908.0	**917.2**	894.9	887.3	873.8	881.5	864.7	862.5	884.3
500	795.2	796.0	**797.8**	776.9	771.3	767.6	764.3	765.1	739.7	769.9
1000	678.6	**679.1**	678.5	656.3	656.1	646.5	650.8	654.4	636.2	640.8
Y=2002										
10	**1203.5**	**1203.5**	**1203.5**	1061.4	1061.4	1108.3	1061.4	989.3	989.3	1061.4
25	1309.6	1309.6	**1321.3**	1279.8	1227.9	1171.0	1227.9	1171.0	1093.4	1159.4
50	**1154.1**	**1154.1**	1152.0	1129.6	1075.3	1075.1	1075.1	1075.1	1065.0	1097.9
100	883.3	883.3	887.6	872.9	878.0	879.4	862.2	862.6	868.3	**900.0**
250	**754.2**	747.2	753.0	741.4	734.6	731.9	734.6	736.0	725.7	744.7
500	662.5	**665.4**	662.8	644.9	643.5	628.8	632.5	632.7	624.3	635.7
1000	568.4	569.1	**570.2**	559.5	556.0	544.5	550.6	545.7	537.7	545.7
Y=2003										
10	**1548.8**	**1548.8**	**1548.8**	1407.1	1407.1	1356.3	1236.2	1236.2	1356.3	1236.2
25	**1146.9**	**1146.9**	**1146.9**	1144.6	1046.9	1081.3	1135.9	1135.9	1063.5	1135.9
50	**830.1**	**830.1**	**830.1**	815.6	815.0	789.4	801.0	815.1	764.5	812.4
100	**735.2**	730.4	722.1	709.3	712.8	703.0	690.7	686.1	692.3	725.8
250	593.9	595.9	598.7	**604.9**	602.0	597.9	601.9	599.3	589.6	591.5
500	**538.9**	536.1	537.3	519.9	520.1	514.4	516.5	515.8	512.4	518.6
1000	456.2	455.8	**456.6**	448.6	447.4	443.9	444.7	442.6	436.3	441.5
Y=2004										
10	1088.4	1135.1	1105.3	**1145.8**	**1145.8**	**1145.8**	**1145.8**	**1145.8**	**1145.8**	1090.2
25	**888.9**	**888.9**	**888.9**	869.1	838.4	813.2	858.4	813.2	804.8	863.5
50	669.1	669.1	**672.1**	640.7	628.3	641.0	628.3	646.5	617.9	652.6

100	555.8	**556.8**	545.4	543.1	545.5	520.4	530.8	531.4	514.5	536.8
250	450.7	450.0	**452.7**	445.9	442.0	445.0	444.6	446.6	445.6	444.5
500	401.8	402.1	**402.3**	394.6	396.1	385.6	387.8	389.9	380.3	388.4
1000	340.1	341.8	**341.9**	331.5	332.4	324.8	328.0	328.3	321.1	327.3
Y=2005										
10	**694.0**	**694.0**	**694.0**	**694.0**	**694.0**	**694.0**	**694.0**	**694.0**	**694.0**	566.8
25	**606.8**	**606.8**	**606.8**	564.2	584.1	579.8	584.1	579.8	518.8	544.0
50	**464.3**	**464.3**	457.0	455.6	449.5	442.3	443.9	443.9	442.3	460.7
100	378.3	372.7	376.2	376.3	390.8	387.6	381.7	393.2	385.4	**402.8**
250	**302.9**	301.9	302.6	296.1	297.1	297.8	291.3	293.2	289.3	295.5
500	**273.6**	273.5	**273.6**	264.2	262.6	259.4	262.8	260.5	255.5	262.7
1000	**224.4**	**224.4**	223.9	219.8	219.2	216.8	216.7	217.5	213.7	217.6
Y=2006										
10	**450.0**	**450.0**	**450.0**	387.4	387.4	401.4	387.4	387.4	343.6	387.4
25	**305.1**	**305.1**	**305.1**	297.9	297.9	279.2	287.6	291.4	288.9	287.7
50	247.9	247.9	247.9	250.1	243.4	253.1	245.2	**250.9**	232.6	237.6
100	**199.9**	199.0	199.4	193.3	185.7	192.2	191.6	187.7	187.3	193.3
250	154.8	155.1	**156.2**	151.7	150.4	148.5	150.3	151.3	150.1	150.2
500	135.3	**135.7**	135.4	132.4	132.3	129.4	131.3	131.5	128.1	130.7
1000	**112.2**	**112.1**	112.0	110.0	109.9	107.9	109.6	109.2	107.0	110.3

Table 7.4 Future collisions per mile when ranked on year Y (rural two lane roads)

	EB Expected			EB Expected Excess			LOSS			Table C
Top X	SPF1	SPF2	SPF SA	SPF1	SPF2	SPF SA	SPF1	SPF2	SPF SA	
Y=2000										
10	84.9	69.4	84.9	**91.9**	69.4	72.6	55.8	55.8	55.8	47.1
25	74.0	65.5	74.6	**75.1**	66.1	75.1	62.4	62.8	62.4	51.2
50	62.2	61.8	63.2	63.0	58.7	**63.3**	55.4	55.4	55.4	50.2
100	**58.4**	55.6	57.7	52.5	55.5	52.5	48.5	47.6	48.6	38.3
250	46.1	47.0	46.5	45.4	**47.6**	45.7	37.8	37.9	37.8	30.0
500	38.3	**39.8**	38.6	37.1	38.1	36.6	32.2	31.9	32.2	23.9
1000	31.1	**32.5**	31.1	29.1	29.6	28.9	24.6	24.4	24.7	17.8
Y=2001										
10	69.3	70.1	69.3	69.3	75.9	69.3	**89.1**	**89.1**	**89.1**	61.6
25	66.1	60.4	66.1	63.3	60.0	**67.5**	47.8	46.1	49.4	39.1
50	65.9	55.2	**66.5**	62.8	54.3	62.9	46.5	45.6	46.6	38.4
100	51.7	49.3	**52.4**	48.8	48.5	49.2	43.1	41.2	43.1	30.1
250	**41.0**	**41.0**	40.3	39.8	40.6	39.6	32.5	32.6	32.6	27.0
500	34.1	**35.4**	34.2	32.6	32.7	32.2	26.6	26.4	26.9	21.0
1000	26.6	**27.9**	26.3	23.7	24.8	23.6	21.0	20.8	21.0	15.2
Y=2002										
10	67.9	60.0	67.9	**82.3**	71.0	80.3	59.3	56.5	59.3	56.5
25	61.2	57.1	61.2	**65.8**	61.4	64.1	51.6	51.6	51.6	44.7
50	54.1	**56.9**	52.9	51.9	52.6	52.1	48.1	46.3	48.1	37.3
100	45.2	**45.4**	**45.4**	44.6	44.0	43.8	41.4	38.5	41.7	30.3
250	**35.6**	34.5	35.1	34.7	34.8	34.6	30.3	30.0	30.3	22.1
500	29.7	**30.3**	29.8	28.9	29.4	29.0	23.3	23.1	23.3	16.8
1000	22.3	**23.9**	22.1	20.1	20.9	20.1	17.5	17.1	17.6	12.6

Y=2003										
10	53.2	49.9	51.4	61.2	45.8	**61.8**	45.6	45.6	45.6	44.5
25	44.0	37.5	44.0	**46.1**	41.4	44.9	37.1	37.1	39.3	35.6
50	39.6	39.3	**40.1**	39.5	38.8	39.4	37.3	37.5	37.3	34.9
100	**37.3**	35.7	**37.3**	36.1	34.7	36.3	34.1	34.2	34.7	27.9
250	30.1	29.7	**30.6**	30.5	28.8	30.0	24.7	24.7	25.9	19.8
500	23.2	**24.2**	23.1	22.6	22.4	22.6	18.7	18.5	18.8	13.7
1000	17.7	**19.0**	17.8	15.9	16.8	15.9	14.0	13.5	14.0	10.3
Y=2004										
10	35.4	35.5	35.4	36.1	38.6	36.1	**45.1**	43.3	**45.1**	40.1
25	33.0	31.0	33.0	34.6	32.9	34.6	**34.8**	33.0	**34.8**	31.2
50	30.8	29.4	30.9	31.2	27.7	**31.7**	29.6	29.5	29.6	26.2
100	25.8	25.3	26.1	25.5	**26.3**	25.2	24.8	24.9	24.8	21.3
250	20.1	20.4	20.0	19.9	20.0	20.0	18.5	18.3	**18.7**	15.4
500	17.2	**17.8**	17.2	16.6	16.4	16.8	14.1	14.1	14.2	11.5
1000	13.8	**14.0**	13.9	12.6	12.8	12.6	11.1	10.9	11.2	7.7
Y=2005										
10	**23.7**	21.6	**23.7**	**23.7**	21.7	22.5	23.6	22.5	23.6	22.5
25	**25.5**	21.0	22.7	24.1	22.3	23.8	21.1	21.1	21.1	20.3
50	22.5	19.4	**22.6**	22.0	21.3	21.9	22.2	22.0	22.2	16.6
100	20.0	17.4	**20.1**	18.6	17.8	18.7	16.5	16.8	16.9	14.0
250	13.8	14.1	14.1	**14.4**	14.2	14.2	12.4	12.5	12.8	9.7
500	11.3	**11.7**	11.5	10.9	11.3	11.3	9.4	9.3	9.5	7.4
1000	8.9	**9.3**	8.9	8.1	8.4	8.0	7.2	7.0	7.2	5.2
Y=2006										
10	**12.3**	11.4	**12.3**	**12.3**	10.8	**12.3**	11.0	11.0	11.0	11.0
25	10.8	9.8	10.9	10.2	10.6	10.2	**11.4**	**11.4**	**11.4**	8.9
50	9.7	9.5	9.7	10.0	**10.2**	10.0	9.9	9.8	9.9	8.7
100	**8.6**	**8.6**	8.4	8.1	8.3	8.1	8.0	7.8	8.0	6.4
250	6.8	**6.9**	**6.9**	6.7	**6.9**	6.7	5.8	5.7	5.8	4.5
500	5.5	**5.7**	5.6	5.5	5.5	5.5	4.5	4.4	4.6	3.5
1000	4.3	**4.5**	4.3	4.0	4.1	4.0	3.4	3.4	3.4	2.5

7.2.2 Results from Approach 2

In approach 2, the intent is to compare the ability of each method to flag and prioritize the locations previously investigated, using data prior to the actual selection, and considering whether or not those locations were recommended for improvement. This method was implemented the same way as for intersections by considering 1 year at a time. As mentioned in the summary for intersections, since the starting point is the sites that were selected for investigation based on results from Table C, we expect the Table C method to do quite well in this evaluation. However, we also want to see how the other methods did in ranking the sites that were recommended for improvement.

Tables 7.5 and 7.6 show the results obtained based on this approach for urban freeways and rural two lane roads. For each year, in each column, cells with the highest value are shown in bold. The last column shows the total number of sites that were recommended for improvement in each year among the sites that were identified for investigation by Table C. The results from Tables 7.3 and 7.4 seem to indicate that the other methods did equally well as Table C in some of the cases (especially, in urban freeways), and in a few cases, they did better than Table C. There is no clear difference between the results obtained with the three types of SPFs.

Table 7.5 Number of segments selected as improvement recommended (urban freeways)

	2003 Investigations						
	Top 10	Top 50	Top 100	Top 500	Top 1000	Top 2000	All (4025)
Table C	0	1	7	**51**	**93**	194	354
EB_expected (SPF1)	0	**5**	10	48	85	191	354
EB_expected (SPF2)	0	**5**	**11**	48	87	192	354
EB_expected (SPF SA)	**1**	**5**	10	50	85	**196**	354
EB_expected_excess (SPF1)	**1**	4	9	45	88	183	354
EB_expected_excess (SPF2)	**1**	3	8	41	87	180	354
EB_expected_excess (SPF SA)	0	3	8	37	81	169	354
LOSS (SPF1)	0	3	8	46	87	181	354
LOSS (SPF2)	0	2	8	40	84	173	354
LOSS (SPF SA)	0	3	8	38	80	166	354
	2004 Investigations						
	Top 10	Top 50	Top 100	Top 500	Top 1000	Top 2000	All (5765)
Table C	2	7	13	**44**	**84**	**165**	429
EB_expected (SPF1)	**3**	8	**16**	43	73	146	429
EB_expected (SPF2)	**3**	8	15	**44**	74	146	429
EB_expected (SPF SA)	**3**	**9**	**16**	**44**	74	146	429
EB_expected_excess (SPF1)	3	**9**	14	43	78	149	429
EB_expected_excess (SPF2)	2	8	12	39	73	144	429
EB_expected_excess (SPF SA)	2	7	11	37	73	131	429
LOSS (SPF1)	2	8	10	42	81	151	429
LOSS (SPF2)	2	8	10	39	75	145	429
LOSS (SPF SA)	2	7	9	34	74	136	429
	2005 Investigations						
	Top 10	Top 50	Top 100	Top 500	Top 1000	Top 2000	All (3340)
Table C	0	2	4	**31**	**77**	**169**	250
EB_expected (SPF1)	0	1	5	26	54	134	250
EB_expected (SPF2)	0	1	**6**	26	55	137	250
EB_expected (SPF SA)	0	1	5	26	61	135	250
EB_expected_excess (SPF1)	0	**3**	**6**	27	64	145	250
EB_expected_excess (SPF2)	0	2	**6**	27	61	142	250
EB_expected_excess (SPF SA)	0	2	5	24	59	137	250
LOSS (SPF1)	0	2	**6**	26	71	148	250
LOSS (SPF2)	0	2	**6**	28	64	148	250
LOSS (SPF SA)	0	1	3	25	64	136	250
	2006 Investigations						
	Top 10	Top 50	Top 100	Top 500	Top 1000	Top 2000	All (6145)
Table C	0	1	7	41	99	**208**	579
EB_expected (SPF1)	0	**6**	12	58	109	197	579
EB_expected (SPF2)	0	**6**	**13**	**59**	110	197	579
EB_expected (SPF SA)	0	5	12	**59**	**111**	200	579

EB_expected_excess (SPF1)	0	3	6	49	95	173	579
EB_expected_excess (SPF2)	0	4	6	47	91	169	579
EB_expected_excess (SPF SA)	0	2	7	39	85	157	579
LOSS (SPF1)	0	3	4	43	91	177	579
LOSS (SPF2)	0	1	6	40	89	175	579
LOSS (SPF SA)	0	2	4	35	83	157	579
	2007 Investigations						
	Top 10	Top 50	Top 100	Top 500	Top 1000	Top 2000	All (4791)
Table C	0	3	7	28	70	156	381
EB_expected (SPF1)	**1**	4	9	33	72	154	381
EB_expected (SPF2)	**1**	**5**	**10**	36	**74**	155	381
EB_expected (SPF SA)	**1**	4	9	34	73	**158**	381
EB_expected_excess (SPF1)	0	**5**	8	**39**	68	150	381
EB_expected_excess (SPF2)	0	**5**	6	37	66	149	381
EB_expected_excess (SPF SA)	0	**5**	7	30	65	132	381
LOSS (SPF1)	0	**5**	6	**39**	69	141	381
LOSS (SPF2)	0	3	6	36	63	144	381
LOSS (SPF SA)	0	**5**	7	30	59	130	381
	2008 Investigations						
	Top 10	Top 50	Top 100	Top 500	Top 1000	Top 2000	All (4375)
Table C	0	**4**	**5**	19	38	76	178
EB_expected (SPF1)	0	3	**5**	**31**	50	76	178
EB_expected (SPF2)	0	3	**5**	**31**	**52**	77	178
EB_expected (SPF SA)	0	3	**5**	30	51	**79**	178
EB_expected_excess (SPF1)	**1**	**4**	**5**	26	42	66	178
EB_expected_excess (SPF2)	0	**4**	**5**	25	40	66	178
EB_expected_excess (SPF SA)	0	**4**	4	26	41	67	178
LOSS (SPF1)	0	**4**	4	25	40	69	178
LOSS (SPF2)	0	**4**	**5**	23	38	69	178
LOSS (SPF SA)	0	3	4	24	37	68	178

Table 7.6 Number of segments selected as improvement recommended (rural two lane roads)

	2003 Investigations						
	Top 10	Top 25	Top 50	Top 75	Top 100	Top 200	All (258)
Table C	1	**5**	**15**	**23**	**30**	**63**	75
EB_expected (SPF1)	1	**5**	9	13	22	54	75
EB_expected (SPF2)	1	4	9	12	18	54	75
EB_expected (SPF SA)	1	**5**	10	13	22	53	75
EB_expected_excess (SPF1)	1	4	8	17	22	55	75
EB_expected_excess (SPF2)	**2**	4	6	17	23	54	75
EB_expected_excess (SPF SA)	1	4	9	16	21	54	75
LOSS (SPF1)	1	4	9	19	23	57	75
LOSS (SPF2)	1	4	11	17	20	57	75
LOSS (SPF SA)	1	3	8	19	22	55	75
	2004 Investigations						

	Top 10	Top 25	Top 50	Top 75	Top 100	Top 200	All (364)
Table C	**2**	**5**	**17**	**26**	**28**	**51**	83
EB_expected (SPF1)	0	1	2	9	12	37	83
EB_expected (SPF2)	0	1	3	8	10	36	83
EB_expected (SPF SA)	0	1	2	9	12	34	83
EB_expected_excess (SPF1)	0	2	5	9	12	40	83
EB_expected_excess (SPF2)	0	2	5	9	12	40	83
EB_expected_excess (SPF SA)	0	2	4	9	12	39	83
LOSS (SPF1)	1	3	6	12	19	41	83
LOSS (SPF2)	1	3	8	15	21	44	83
LOSS (SPF SA)	0	3	6	10	18	40	83
	2005 Investigations						
	Top 10	Top 25	Top 50	Top 75	Top 100	Top 200	All (147)
Table C	1	3	**11**	16	22	32	32
EB_expected (SPF1)	1	**5**	9	13	19	32	32
EB_expected (SPF2)	1	4	9	12	19	32	32
EB_expected (SPF SA)	**2**	**5**	9	13	19	32	32
EB_expected_excess (SPF1)	0	**5**	10	13	22	32	32
EB_expected_excess (SPF2)	0	3	10	15	22	32	32
EB_expected_excess (SPF SA)	0	**5**	10	13	22	32	32
LOSS (SPF1)	1	3	8	14	**23**	32	32
LOSS (SPF2)	0	3	9	**17**	**23**	32	32
LOSS (SPF SA)	0	3	7	14	**23**	32	32
	2006 Investigations						
	Top 10	Top 25	Top 50	Top 75	Top 100	Top 200	All (299)
Table C	2	**6**	**18**	**30**	**39**	**76**	104
EB_expected (SPF1)	2	4	11	18	23	57	104
EB_expected (SPF2)	2	4	10	20	23	57	104
EB_expected (SPF SA)	2	4	10	19	23	57	104
EB_expected_excess (SPF1)	**3**	4	12	17	25	66	104
EB_expected_excess (SPF2)	**3**	4	13	19	25	67	104
EB_expected_excess (SPF SA)	2	5	12	17	26	65	104
LOSS (SPF1)	**3**	4	10	21	29	68	104
LOSS (SPF2)	2	5	10	22	30	70	104
LOSS (SPF SA)	**3**	3	10	21	29	66	104
	2007 Investigations						
	Top 10	Top 25	Top 50	Top 75	Top 100	Top 200	All (212)
Table C	1	5	8	23	**34**	60	63
EB_expected (SPF1)	**2**	6	14	23	33	58	63
EB_expected (SPF2)	**2**	6	13	**24**	33	57	63
EB_expected (SPF SA)	**2**	**7**	14	**24**	33	58	63
EB_expected_excess (SPF1)	**2**	6	14	21	29	60	63
EB_expected_excess (SPF2)	1	5	13	23	30	**61**	63
EB_expected_excess (SPF SA)	1	6	**15**	21	28	59	63
LOSS (SPF1)	**2**	5	11	20	28	60	63

LOSS (SPF2)	**2**	4	10	19	28	61	63
LOSS (SPF SA)	**2**	5	11	20	28	59	63
	2008 Investigations						
	Top 10	Top 25	Top 50	Top 75	Top 100	Top 200	All (211)
Table C	**1**	**3**	**9**	15	20	36	40
EB_expected (SPF1)	**1**	2	8	12	18	38	40
EB_expected (SPF2)	**1**	2	8	13	19	38	40
EB_expected (SPF SA)	**1**	2	8	14	17	38	40
EB_expected_excess (SPF1)	0	1	8	15	20	**40**	40
EB_expected_excess (SPF2)	0	1	5	14	20	39	40
EB_expected_excess (SPF SA)	**1**	2	8	15	21	39	40
LOSS (SPF1)	0	2	8	**18**	**22**	**40**	40
LOSS (SPF2)	0	1	6	14	19	38	40
LOSS (SPF SA)	0	2	8	**18**	21	38	40

7.2.3 Results from Approach 3

Here, the intent is to compare the characteristics of top ranked sites from each method. The two main characteristics that were selected for this comparison were AADT and the expected number of collisions per mile. As was done in the case of intersections, the expected number of collisions was chosen (instead of the actual number of collisions) because it corrects for possible bias due to regression to the mean and is a better estimate of the true long-term collision propensity. This approach was implemented the same way as for intersections for the three recent years of data.

Table 7.7 shows the results for freeways and Table 7.8 shows the results for rural two lane roads. As expected, the top ranked sites from the EB Expected method have the highest average AADT and the highest number of expected collisions. For freeways, the values for the other three methods are quite close. However, for rural two lane roads, the top ranked sites from Table C clearly have fewer expected collisions per mile and a lower average AADT. Again, it is important to remember that the Table C method does not account for the non-linear relationship between traffic volume and collision frequency, and that may be a reason why the top ranked sites in the Table C method have fewer expected collisions and tend to have lower AADTs compared to the top ranked sites in the other methods.

Table 7.7 Average AADT and average expected collisions per mile for top ranked sites (urban freeways)

	EB Expected		**EB Expected Excess**		**LOSS**		**Table C**	
Top X Sites	**AADT**	**Expected Total**	**AADT**	**Expected Total**	**AADT**	**Expected Total**	**AADT**	**Expected Total**
Y = 2004								
10	311584	979.2	304896	1011.6	304896	1011.6	304620	979.6
25	293028	902.2	283411	829.1	253379	831.9	257360	837.9
50	269251	720.2	255733	650.0	243467	648.3	250235	677.7
100	245838	564.0	236170	546.6	231733	536.7	231175	548.8
250	246887	460.2	235086	460.7	230241	453.3	227400	451.5

500	242701	397.8	233077	390.9	225569	384.9	224775	389.2
1000	227435	319.1	216146	310.4	212184	305.3	211965	306.9
Y = 2005								
10	266022	848.9	266022	848.9	266022	848.9	262796	817.0
25	291869	815.8	272898	762.8	266814	765.0	237738	715.3
50	271258	600.0	252935	592.2	242632	601.6	246402	618.1
100	263716	514.9	246509	519.1	245343	529.5	246044	543.7
250	247911	435.2	240829	427.6	235447	416.9	229797	420.4
500	244470	374.5	227653	362.2	222003	358.8	225417	357.3
1000	226260	307.7	215027	299.4	210685	295.1	213961	295.4
Y = 2006								
10	310039	847.7	284719	740.8	284719	740.8	284719	740.8
25	279717	746.7	274941	745.5	269844	680.9	248630	679.2
50	267991	630.9	265867	635.4	250764	633.1	247835	623.9
100	258112	540.0	240924	540.0	232259	529.5	233475	542.0
250	247696	439.4	233545	444.5	226066	438.4	225405	431.6
500	238481	374.4	226576	366.0	222690	363.3	223204	361.6
1000	226829	306.7	215479	298.9	212919	297.1	215752	299.5

Table 7.8 Average AADT and average expected collisions per mile for top ranked sites (rural two lane roads)

	EB Expected		**EB Expected Excess**		**LOSS**		**Table C**	
Top X Sites	**AADT**	**Expected Total**	**AADT**	**Expected Total**	**AADT**	**Expected Total**	**AADT**	**Expected Total**
Y = 2004								
10	26702	7.8	26545	8.1	12457	8.1	6323	5.4
25	26421	6.4	23729	6.7	10077	5.7	6924	4.0
50	26395	7.0	20551	6.1	12910	4.2	6004	3.3
100	21631	5.6	16939	4.8	10349	3.6	6700	2.8
250	17479	4.1	15184	3.7	9477	2.9	6522	2.1
500	15833	3.4	12711	2.9	7921	2.2	5227	1.5
1000	13473	2.7	10151	2.2	6917	1.7	4176	1.0
Y = 2005								
10	27295	7.3	27295	7.3	7322	4.4	6083	4.1
25	27195	9.7	22311	9.4	10050	4.9	5634	3.9
50	28117	7.2	19350	6.8	12756	5.9	6160	3.5
100	25782	6.5	17900	5.8	11223	4.1	6563	2.8
250	19994	4.5	15505	4.1	9268	3.0	6216	2.1
500	16628	3.4	12668	3.0	7687	2.2	5153	1.6
1000	13895	2.7	10015	2.1	6822	1.7	4314	1.1
Y = 2006								
10	22588	10.0	22588	10.0	10219	6.5	10219	6.5
25	22500	7.3	20576	7.2	13170	6.1	10336	4.6
50	24012	6.3	20432	6.6	10546	5.1	8623	3.9
100	22639	5.4	19184	5.3	9154	3.6	6217	2.6
250	18999	4.3	15632	3.8	9445	2.6	5519	1.6
500	15669	3.3	12969	3.0	7719	2.0	5037	1.4

1000	13583	2.5	10157	2.1	6847	1.6	4185	1.0

7.3 USE OF SAFETYANALYST AND COMPARISON OF SLIDING WINDOW WITH PEAK SEARCH METHOD

SafetyAnalyst links the latest roadway inventory data with multiple years of collision and traffic data. In 2003, Caltrans changed the way the county_route variable was coded. Hence, only data from 2003 to 2007 for rural two lane roads and urban freeways were imported into SafetyAnalyst. The variables from the California HSIS files were recoded to match the levels for the variables in roadway and collision files in SafetyAnalyst. Appendix C shows the details about the recoding that was done.

Version 4.0.0 of SafetyAnalyst was used for this task. We believe this was one of the later versions before SafetyAnalyst was turned over to AASHTO to be marketed as an AASHTOWare product.

SafetyAnalyst was used to compare the results from the sliding window and the peak search method for the combined urban freeway and rural two lane road dataset. Data from 2003 to 2005 were imported into SafetyAnalyst, calibrated, and used for network screening. A window size of 0.2 mile with a 0.1 mile increment was used for the sliding window method. The minimum window size for the peak search method was 0.1 mile (for the peak search method in SafetyAnalyst, this is the default and cannot be changed in Version 4.0.0). Type 1 SPFs estimated with California data were used for this evaluation. The measure for comparison was the number of collisions per mile in the future (i.e., 2006 and 2007) for the top ranked sites in each method. Table 7.9 shows the results. The top part of the table shows the results when all segments are included to identify the top ranking sites. The next part of the table shows the results when sections shorter than 0.1 miles were excluded, and the last part of the table shows the results when sections shorter than 0.2 miles were excluded. When all segments (including those that are shorter than 0.1 miles) are included, it is clear that peak search method that makes use of the EB Expected Collisions method is more effective in identifying sites that are prone to have more collisions in the future. However, when the shorter sections are excluded, the results from the peak search and sliding window method are closer if the EB Expected Collisions method is considered.

Table 7.9 Collisions per mile in 2006 and 2007 for top ranked sites (combined rural two lane and urban freeway segments)

All Sites				
Top X sites	Sliding Window		Peak Search	
	EB Expected	EB Expected Excess	EB Expected	EB Expected Excess
10	813.0	813.0	1291.6	1291.6
100	771.7	789.1	805.7	792.6
250	545.0	540.1	605.3	595.5
500	440.3	437.1	496.4	477.2

1000	376.5	370.5	401.6	394.2
Sites 0.1 miles or longer				
Top X sites	Sliding Window		Peak Search	
	EB Expected	EB Expected Excess	EB Expected	EB Expected Excess
10	699.4	699.4	725.7	689.2
100	399.8	420.4	435.5	438.8
250	340.6	334.2	352.8	353.2
500	283.8	286.7	298.5	293.4
1000	239.5	237.8	244.7	215.4
Sites 0.2 miles or longer				
Top X sites	Sliding Window		Peak Search	
	EB Expected	EB Expected Excess	EB Expected	EB Expected Excess
10	482.6	482.6	464.8	469.4
100	259.6	257.8	264.2	255.8
250	213.0	212.1	219.5	212.9
500	175.6	172.7	175.7	118.9
1000	126.9	120.5	124.1	39.6

7.4 USE OF CURE PLOTS TO COMPARE THE SAFETY PERFORMANCE FUNCTIONS

Similar to the use of CURE plots to compare the intersection SPFs, CURE plots were used to compare the SPFs for rural two lane roads and urban freeways.

7.4.1 Rural Two-Lane Road Segments

Figure 7.1 plots the cumulative residuals for the SafetyAnalyst (SA) model. The figure shows that the model overpredicts at lower AADTs and underpredicts at higher AADTs. The plot of cumulative residuals strays far outside the 95% confidence limits showing that the model exhibits bias. Figure 7.2 plots the performance of the SPFs directly estimated with Caltrans data (called as Caltrans model) which performs better. Although the cumulative residuals stray outside the 95% confidence limits, this is less so than for the SafetyAnalyst model and the maximum deviation from 0 is about half that of the SafetyAnalyst model.

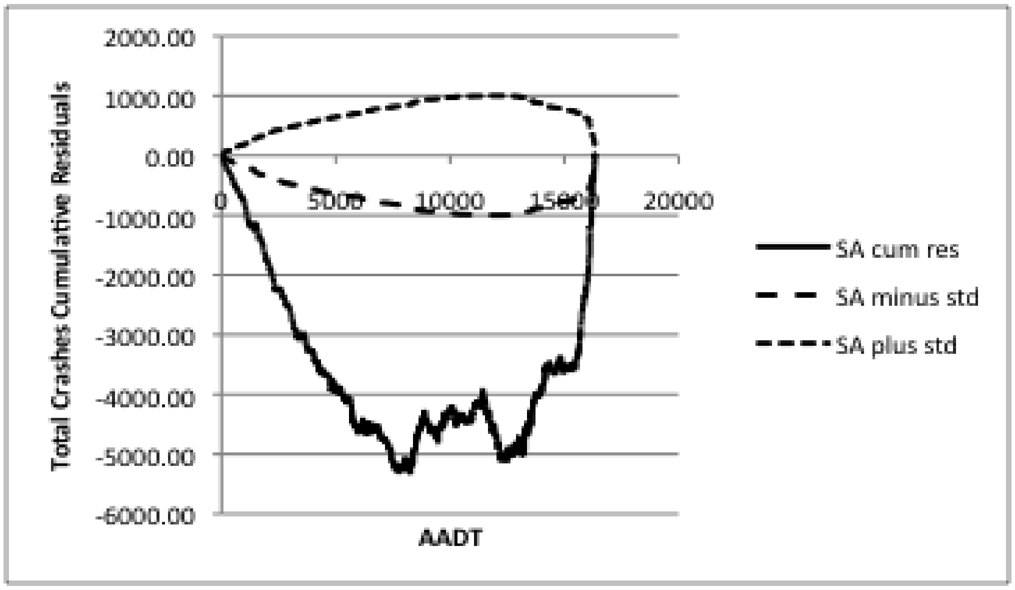


Figure 7.1 CURE Plot for Rural Two-Lane Road Segments: SafetyAnalyst Model vs. AADT

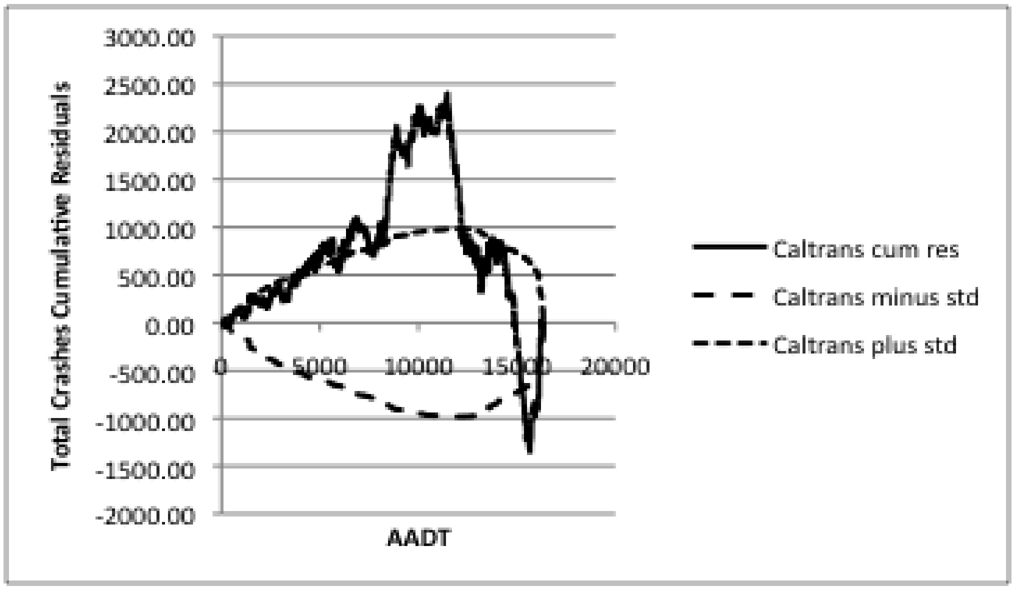


Figure 7.2 CURE Plot for Rural Two-Lane Road Segments: Caltrans Model vs. AADT

7.4.2 Urban Freeway Road Segments

Figure 7.3 plots the cumulative residuals for the SafetyAnalyst (SA) model. The figure shows that the model overpredicts at lower AADTs and under predicts at higher AADTs. The plot of cumulative residuals strays far outside the 95% confidence limits showing that the model exhibits bias. Figure 7.4 plots the performance of the Caltrans model which performs better. Although the cumulative residuals stray outside the 95% confidence limits, this is less so than for the SafetyAnalyst model and the maximum deviation from 0 is about half that of the SafetyAnalyst model.

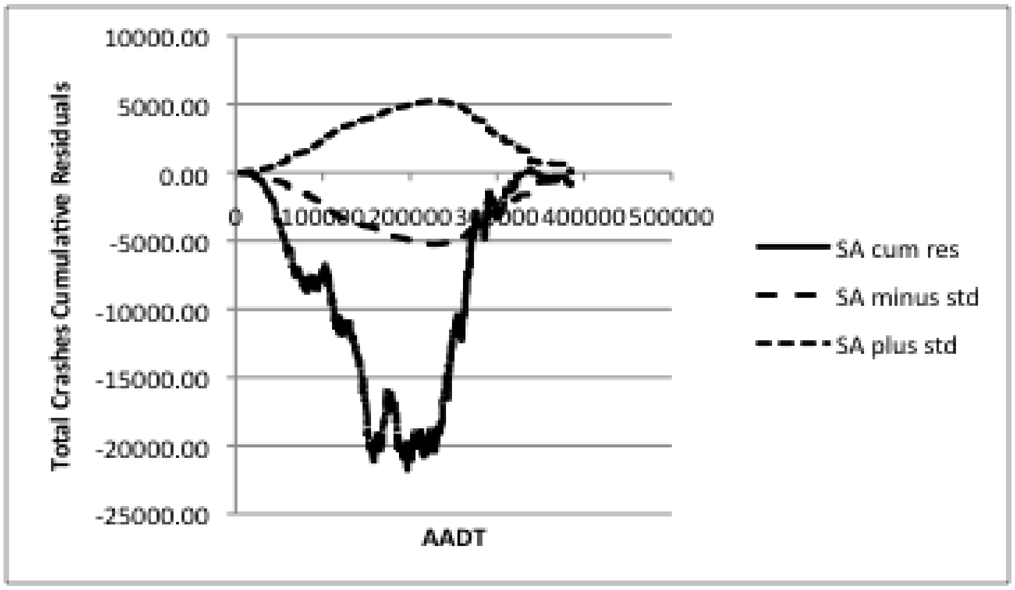


Figure 7.3 CURE Plot for Urban Freeway Segments: SafetyAnalyst Model vs. AADT

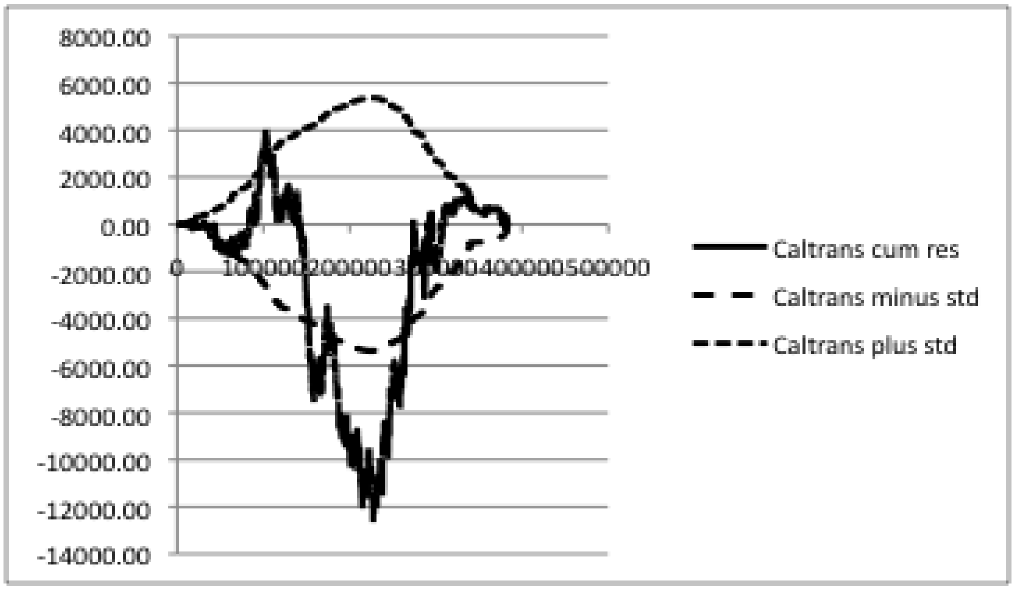


Figure 7.4 CURE Plot for Urban Freeway Segments: Caltrans Model vs. AADT

For both rural two-lane and urban freeway segments the new models developed using the Caltrans data showed a better fit to the data then the recalibrated SafetyAnalyst models. Given these results it is recommended to apply these new Caltrans-based models in lieu of the default ones in SafetyAnalyst should SafetyAnalyst be applied in California.

8. CONCLUSIONS AND RECOMMENDATIONS

The objective of network screening should ideally be to not only identify sites for safety investigation but also to prioritize those sites efficiently. An efficient process identifies sites that can be subjected to the most cost-effective treatments. In so doing, it will recognize two basic properties of collision data: random fluctuation and the non-linear relationship with traffic volume. All of these fundamental principles are embodied in the state of the art EB methodologies in SafetyAnalyst. These principles are not embodied in the Table C method, nor in other methods, such as the CRP method that have been considered for use by CALTRANS. While these methods do have certain desirable properties, by and large, it would be desirable for CALTRANS to adopt SafetyAnalyst to improve the efficiency of their network screening and their safety management process as a whole.

Here are the conclusions and recommendations based on the evaluation:

1. The bias due to regression to the mean could be significant even if four years of data are used, especially for low volume sites such as rural two lane roads and rural stop controlled intersections. The state of the art EB method accounts for this possible bias.

2. It is clear that compared to the Table C method, methods based on the EB procedures (EB expected and EB expected excess) identify sites that have higher AADTs and higher expected collisions. It is also clear that the top ranked sites based on the EB procedure have more collisions in the future compared to the top ranked sites from the Table C method.

3. One of the evaluation approaches compared the ability of each method to flag and prioritize the locations previously investigated, using data prior to the actual selection, and considering whether or not those locations were recommended for improvement. Since the sites investigated were selected based on the results of the Table C method, the Table C method did quite well compared to the other methods. At the same time, in many of the cases, the other methods did equally well and in some cases better than the Table C procedure.

4. This study also examined the false positives for intersections, i.e., intersections that were selected and investigated using the Table C method for which no treatment was recommended. The top ranked sites from the EB expected method tend to have fewer false positives compared to the other methods.

5. Future research could investigate the implications of including the cost effectiveness data in the analysis along with the possibility of investigating top ranked sites from different methods. The project team has provided Caltrans with the list of top ranked locations from the EB methods that can be investigated by the Caltrans districts.

6. In the majority of the cases, the evaluation results from the network screening methods were quite similar regardless of which type of SPF was used. However, the cumulative

residual plots indicate that the SPFs directly calibrated from the California data are better than the default SafetyAnalyst SPFs that were recalibrated with the same California data. Hence, whenever possible, if sufficient samples are available for a particular roadway or intersection type, the SPFs directly calibrated from the most recent California data are recommended instead of using the default SPFs from SafetyAnalyst. This project developed SPFs for two roadway types and two intersection types. Further research can involve the development of SPFs for other roadway and intersection types using California data.

7. The methods based on the EB procedure work best with longer road segments. Hence, contiguous road segments could be aggregated once they remain homogenous with respect to AADT and key characteristics such as road classification, terrain, number of lanes, and road width. With expanded lengths, an entire segment would be flagged and prioritized for safety investigation, not just the location (window) with the collision history that triggered the investigation. This is useful since the source of the problem may be quite removed from the triggering site. Modern safety investigation procedures such as those provided in the SafetyAnalyst diagnosis module do consider this collision "migration" phenomenon. The CRP method also aims to address this issue.

8. In prioritizing sites for safety investigation, consideration could be given to applying weights such that all types of sites can be considered for cost-effective treatments. SafetyAnalyst does this to a very limited extent (by weighting urban and rural sites) and with somewhat arbitrary weights. This enhanced weighting may have to be done outside SafetyAnalyst using the prioritized lists produced by SafetyAnalyst for various site types. This may be a topic for future research as well.

9. The "proportions" method in SafetyAnalyst can be used as a diagnostic tool and possibly in combination with the EB methods for network screening. Future research can try to assess whether using combinations of methods in network screening would lead to the identification of sites that can be subjected to the most cost effective treatments.

10. Network Screening should be done on an annual basis and based on the most recent 5 years of data. The Safety Performance Functions used for this purpose should also be re-calibrated annually to the most recent 5 years, whether they are California-specific or the default ones in SafetyAnalyst. Calibration of SPFs is automatic within SafetyAnalyst. New SPFs should be estimated every 5 years, recognizing that there are other SafetyAnalyst applications for SPFs.

11. CALTRANS should develop a database of interchange locations along with a link to the ramps from each interchange. This will allow them to determine the influence area of interchanges more accurately.

12. CALTRANS should compile data on the lengths of individual ramps so that they can be analyzed using SafetyAnalyst.

13. In talking to CALTRANS staff, it is clear that the minor road AADT at some intersections may not be very accurate and may be based on data that has not been updated for many years. Updating this information will provide more accurate results.

REFERENCES

Aguero-Valverde J. and P. Jovanis (2008), Analysis of Road Crash Frequency with Spatial Models, *Journal of the Transportation Research Board* 2061, pp. 55-63, 2008.

Aguero-Valverde J. and P. Jovanis (2009), Bayesian Multivariate Poisson Log-Normal Models for Crash Severity Modeling and Site Ranking, *Presented at the 2009 Annual Meeting of the Transportation Research Board*.

Cheng, W. and S. Washington (2008), New Criteria for Evaluating Methods of Identifying Hot Spots, *Journal of the Transportation Research Board*, 2083, pp. 76-85.

Chung K. and D. Ragland (2007), *A Method for Generating a Continuous Risk Profile for Highway Collisions*. UC Berkeley Traffic Safety Center.Paper UCB-TSC-TR-2007-6, http://repositories.cdlib.org/its/tsc/UCB-TSC-TR-2007-6, 2007.

Chung K., Jang K., Madanat, S., and S. Washington (2011). Proactive detection of high collision concentration locations on Highways, *19th International Symposium on Transportation and Traffic Theory (ISTTT)*, Berkeley, California, July 2011.

Elvik, R. (2008), Comparative Analysis of Techniques for Identifying Locations of Hazardous Roads, *Journal of the Transportation Research Board* 2083, pp 72-75.

Hauer, E. (1996a), Detection of Safety Deterioration in a Series of Accident Counts. *Transportation Research Record* 1542, 38-43.

Hauer, E. (1996b), Statistical Test of the Difference between Expected Accident Frequencies, *Transportation Research Record* 1542, 24-29.

Hauer, E. and J. Bamfo (1997), Two Tools for Finding What Function Links the Dependent Variable to the Explanatory Variables, *Published in Proceedings of ICTCT 97 Conference*, November 5-7, 1997, Lund, Sweden.

Hauer, E., Kononov, J., Allery, B., and M.S. Griffith (2002), Screening the Road Network for Sites with Promise, *Transportation Research Record* 1784, pp. 27-32.

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

Heydecker, B. J., and J. Wu (1991), Using the Information in Road Accident Records Proc., *19th PTRC Summer Annual Meeting*, London.

Huang H., Chin H., and M. Haque (2009), Empirical Evaluation of Alternative Approaches in Identifying Crash Hotspots: Naive Ranking, Empirical Bayes and Full Bayes. *Presented at the 2009 Annual Meeting of the Transportation Research Board.*

Kononov, J. (2002), Identifying Locations with Potential for Collision Reductions: Use of Direct Diagnostics and Pattern Recognition Methodologies, *Transportation Research Record* 1784, pp. 153-158.

Kononov, J. and B. Allery (2003), Level of Service of Safety: Conceptual Blueprint and Analytical Framework. *Transportation Research Record* 1840, pp 57-66.

Lan, B., Persaud B., Lyon C. and R. Bhim (2009), Validation of a Full Bayes Methodology for Observational Before-After Road Safety Studies and Application to Evaluation of Rural Signal Conversions, *Accident Analysis and Prevention*, Vol. 41(3), pp. 574-580.

Lord, D., and B.N. Persaud (2000), Accident Prediction Models With and Without Trend: Application of the Generalized Estimating Equations (GEE) Procedure, *Transportation Research Record 1717*, pp. 102-108.

Lyon, C., Gotts, B., Wong, W.K.F., and Persaud, B. (2007), Comparison of alternative methods for identifying sites with high proportion of specific accident types, *Transportation Research Record* 2019, pp. 212-218.

Miaou, S. and Lord, D. (2003), Modeling Traffic Crash-Flow Relationships for Intersections: Dispersion Parameter, Functional Form, and Bayes versus Empirical Bayes Methods. *Transportation Research Record* 1840, 31-40.

Pawlovich, M.D., Li, W., Carriquiry, A., and Welch, T. (2005), Iowa"s Experience with "Road Diet" Measures: Impacts on Crash Frequencies and Crash Rates Assessed Following a Bayesian Approach, *Presented at the 2005 Annual Meeting of the Transportation Research Board*.

Ragland, D.R. and Chan, C.-Y. (2008), *High Collision Concentration Location: Table C Evaluation and Recommendations*, Prepared for CALTRANS, Task Order 5215-6215, February 2008.

Wang X and M.Abdel-Aty (2006), Temporal and spatial analyses of rear-end crashes at signalized intersections. *Accident Analysis & Prevention*, Volume 38, Issue 6, pp. 1137-1150.

APPENDIX A: SAFETY PERFORMANCE FUNCTIONS

SPFs for Rural Two-Lane Roadways

Total collisions (type 1 SPF)

Variable	Estimate	Standard Error	Chi-square	P-value
β_0	-5.5580	0.0690	6495.77	<0.0001
[4]ln(AADT)	0.7266	0.0083	7709.59	<0.0001
Dispersion (k)	0.6730	0.0121		

Total collisions (type 2 SPF)

Variable	Estimate	Standard Error	Chi-square	P-value
β_0	-5.8982	0.0930	4026.49	<0.0001
ln(AADT)	0.8566	0.0087	9805.91	<0.0001
Flat terrain	-0.2201	0.0214	105.64	<0.0001
Mountainous terrain	0.3011	0.0197	233.99	<0.0001
Rolling terrain	0.0000	0.0000		<0.0001
Lane width (ft.)	-0.0431	0.0064	44.89	<0.0001
Shoulder width (ft.)	-0.0655	0.0030	464.45	<0.0001
Dispersion (k)	0.5517	0.0107		

Injury (type 1 SPF)

Variable	Estimate	Standard Error	Chi-square	P-value
β_0	-5.3046	0.0882	3614.81	<0.0001
ln(AADT)	0.5499	0.0106	2692.71	<0.0001
Dispersion (k)	0.7264	0.0193		

Injury (type 2 SPF)

Variable	Estimate	Standard Error	Chi-square	P-value
β_0	-5.5687	0.1249	1987.94	<0.0001
ln(AADT)	0.6972	0.0111	3973.92	<0.0001
Flat terrain	-0.2757	0.0277	98.94	<0.0001
Mountainous terrain	0.3362	0.0249	182.85	<0.0001
Rolling terrain	0.0000	0.0000		<0.0001
Lane width (ft.)	-0.0606	0.0089	46.46	<0.0001
Shoulder width (ft.)	-0.0687	0.0039	304.92	<0.0001
Dispersion (k)	0.5552	0.0167		

---

[4] Ln (or LN) is the natural logarithm, i.e., $\log_e$

Head-On (type 1 SPF)

Variable	Estimate	Standard Error	Chi-square	P-value
β_0	-9.8333	0.1804	2971.71	<0.0001
ln(AADT)	0.8891	0.0210	1789.09	<0.0001
Dispersion (k)	1.3763	0.0678		

Head-On (type 2 SPF)

Variable	Estimate	Standard Error	Chi-square	P-value
β_0	-11.2566	0.1929	3406.43	<0.0001
ln(AADT)	1.0917	0.0225	2352.80	<0.0001
Flat terrain	-0.1751	0.0517	11.49	0.0007
Mountainous terrain	0.5617	0.0473	141.26	<0.0001
Rolling terrain	0.0000	0.0000		<0.0001
Median width (ft.)	-0.0252	0.0071	12.57	0.0004
Shoulder width (ft.)	-0.0998	0.0074	182.22	<0.0001
Dispersion (k)	1.0095	0.0561	1.1194	

Sideswipe (type 1 SPF)

Variable	Estimate	Standard Error	Chi-square	P-value
β_0	-9.7941	0.1473	4418.34	<0.0001
ln(AADT)	0.9346	0.0171	2976.23	<0.0001
Dispersion (k)	0.9037	0.0407		

Sideswipe (type 2 SPF)

Variable	Estimate	Standard Error	Chi-square	P-value
β_0	-10.6973	0.1584	4559.52	<0.0001
ln(AADT)	1.0227	0.0180	3216.53	<0.0001
Flat terrain	-0.1021	0.0407	6.30	0.0121
Mountainous terrain	0.5046	0.0396	162.47	<0.0001
Rolling terrain	0.0000	0.0000		<0.0001
Median width (ft.)	-0.0061	0.0032	3.67	0.0553
Dispersion (k)	0.8146	0.0383		

Rear-End (type 1 SPF)

Variable	Estimate	Standard Error	Chi-square	P-value
β_0	-15.6210	0.1733	8125.32	<0.0001
ln(AADT)	1.6479	0.0196	7101.94	<0.0001
Dispersion (k)	1.1344	0.0382		

Rear-End (type 2 SPF)

Variable	Estimate	Standard Error	Chi-square	P-value
β_0	-16.0031	0.1851	7475.79	<0.0001
ln(AADT)	1.6861	0.0207	6657.87	<0.0001
Flat terrain	-0.0421	0.0394	1.14	0.2854
Mountainous terrain	0.2082	0.0423	24.24	<0.0001
Rolling terrain	0.0000	0.0000		<0.0001
Dispersion (k)	1.1258	0.0380		

Hit-Object (type 1 SPF)

Variable	Estimate	Standard Error	Chi-square	P-value
β_0	-4.6456	0.0930	2494.87	<0.0001
ln(AADT)	0.5078	0.0112	2041.66	<0.0001
Dispersion (k)	0.9871	0.0212		

Hit-Object (type 2 SPF)

Variable	Estimate	Standard Error	Chi-square	P-value
β_0	-5.1117	0.1241	1697.73	<0.0001
ln(AADT)	0.6963	0.0113	3776.11	<0.0001
Flat terrain	-0.4147	0.0283	214.08	<0.0001
Mountainous terrain	0.3808	0.0246	240.15	<0.0001
Rolling terrain	0.0000	0.0000		<0.0001
Shoulder width (ft.)	-0.0962	0.0040	580.88	<0.0001
Lane width (ft.)	-0.0629	0.0087	52.80	<0.0001
Dispersion (k)	0.6995	0.0174		

Overturn (type 1 SPF)

Variable	Estimate	Standard Error	Chi-square	P-value
β_0	-3.7662	0.1099	1174.03	<0.0001
ln(AADT)	0.2841	0.0134	449.21	<0.0001
Dispersion (k)	1.0629	0.0337		

Overturn (type 2 SPF)

Variable	Estimate	Standard Error	Chi-square	P-value
β_0	-3.5493	0.1608	487.43	<0.0001
ln(AADT)	0.3854	0.0141	751.83	<0.0001
Flat terrain	-0.2489	0.0362	47.15	<0.0001
Mountainous terrain	0.1952	0.0325	36.10	<0.0001
Rolling terrain	0.0000	0.0000		<0.0001
Shoulder width (ft.)	-0.0513	0.0051	100.74	<0.0001

Lane width (ft.)	-0.0703	0.0116	36.71	<0.0001
Dispersion (k)	0.9407	0.0314		

SPFs for Urban Freeway Outside the Influence of Ramps

Total collisions (type 1 SPF)

Variable	Estimate	Standard Error	Chi-square	P-value
β_0	-10.8267	0.2695	1614.37	<0.0001
ln(AADT)	1.2075	0.0234	2672.06	<0.0001
Dispersion (k)	0.6151	0.0194		

Total collisions (type 2 SPF)

Variable	Estimate	Standard Error	Chi-square	P-value
β_0	-9.5717	0.3466	762.74	<0.0001
ln(AADT)	1.2136	0.0239	2579.89	<0.0001
Divided median	0.0000	0.0000		
Undivided median	-0.5925	0.2699	4.82	0.0281
Shoulder width (ft.)	-0.0691	0.0061	130.40	<0.0001
Lane width (ft.)	-0.0646	0.0148	18.99	<0.0001
Dispersion (k)	0.5826	0.0187		

Injury (type 1 SPF)

Variable	Estimate	Standard Error	Chi-square	P-value
β_0	-9.1559	0.2780	1084.93	<0.0001
ln(AADT)	0.8889	0.0239	1381.41	<0.0001
Dispersion (k)	0.2807	0.0177		

Injury (type 2 SPF)

Variable	Estimate	Standard Error	Chi-square	P-value
β_0	-8.0895	0.4081	392.93	<0.0001
ln(AADT)	0.8746	0.0247	1257.51	<0.0001
Divided median	0.0000	0.0000		
Undivided median	-0.7662	0.3228	5.63	0.0176
Shoulder width (ft.)	-0.0289	0.0063	21.40	<0.0001
Lane width (ft.)	-0.0549	0.0210	6.83	0.0090
Dispersion (k)	0.2751	0.0175		

Head-On (type 1 SPF)

Variable	Estimate	Standard Error	Chi-square	P-value
β_0	-12.2411	1.0980	124.30	<0.0001

ln(AADT)	0.8449	0.0937	81.23	<0.0001
Dispersion (k)	0.4182	0.1863		

Sideswipe (type 1 SPF)

Variable	Estimate	Standard Error	Chi-square	P-value
β_0	-14.6935	0.3812	1486.12	<0.0001
ln(AADT)	1.3823	0.0327	1791.20	<0.0001
Dispersion (k)	0.6313	0.0289		

Sideswipe (type 2 SPF)

Variable	Estimate	Standard Error	Chi-square	P-value
β_0	-14.5365	0.3738	1512.72	<0.0001
ln(AADT)	1.4157	0.0322	1928.59	<0.0001
Shoulder width (ft.)	-0.0701	0.0070	101.64	<0.0001
Dispersion (k)	0.5892	0.0276		

Rear-End (type 1 SPF)

Variable	Estimate	Standard Error	Chi-square	P-value
β_0	-16.6328	0.3840	1875.85	<0.0001
ln(AADT)	1.6380	0.0331	2450.94	<0.0001
Dispersion (k)	0.9424	0.0320		

Rear-end (type 2 SPF)

Variable	Estimate	Standard Error	Chi-square	P-value
β_0	-16.5714	0.3768	1934.57	<0.0001
ln(AADT)	1.6826	0.0328	2626.62	<0.0001
Shoulder width (ft.)	-0.0752	0.0073	106.16	<0.0001
Dispersion (k)	0.8947	0.0307		

Hit-Object (type 1 SPF)

Variable	Estimate	Standard Error	Chi-square	P-value
β_0	-6.9181	0.2854	587.62	<0.0001
ln(AADT)	0.7504	0.0247	921.16	<0.0001
Dispersion (k)	0.4726	0.0207		

Hit-Object (type 2 SPF)

Variable	Estimate	Standard Error	Chi-square	P-value
β_0	-6.6932	0.2833	558.34	<0.0001
ln(AADT)	0.7595	0.0248	936.57	<0.0001
Divided median	0.0000	0.0000		
Undivided median	-1.1420	0.3467	16.59	<0.0001
Shoulder width	-0.0416	0.0060	47.75	<0.0001

(ft.)				
Dispersion (k)	0.4540	0.0201		

Overturn (type 1 SPF)

Variable	Estimate	Standard Error	Chi-square	P-value
β_0	-3.9085	0.4242	84.91	<0.0001
ln(AADT)	0.3080	0.0370	69.42	<0.0001
Dispersion (k)	0.5498	0.0486		

Overturn (type 2 SPF)

Variable	Estimate	Standard Error	Chi-square	P-value
β_0	-1.5275	0.7377	4.29	0.0384
ln(AADT)	0.2351	0.0384	37.49	<0.0001
Divided median	0.0000	0.0000		
Undivided median	-3.1610	1.0331	9.36	<0.0001
Shoulder width (ft.)	0.0170	0.0112	2.28	0.1311
Lane width (ft.)	-0.1382	0.0444	9.67	0.0019
Dispersion (k)	0.5176	0.0470		

Urban Freeway Within the Influence of Ramps

Total collisions (type 1 SPF)

Variable	Estimate	Standard Error	Chi-square	P-value
β_0	-11.2921	0.1099	10565.7	<0.0001
ln(AADT)	1.2788	0.0093	19008.9	<0.0001
Dispersion (k)	0.6050	0.0066		

Total collisions (type 2 SPF)

Variable	Estimate	Standard Error	Chi-square	P-value
β_0	-10.6726	0.1230	7525.59	<0.0001
ln(AADT)	1.2622	0.0102	15352.1	<0.0001
Divided median	0.0000	0.0000		
Undivided median	-1.5427	0.6306	5.98	0.0144
Shoulder width (ft.)	-0.0474	0.0021	494.49	<0.0001
Median width (ft.)	-0.0019	0.0003	54.13	<0.0001
Dispersion (k)	0.5855	0.0064		

Injury (type 1 SPF)

Variable	Estimate	Standard Error	Chi-square	P-value

β_0	-9.3011	0.1251	5530.3	<0.0001
ln(AADT)	0.9264	0.0105	7796.76	<0.0001
Dispersion (k)	0.3306	0.0071		

Injury (type 2 SPF)

Variable	Estimate	Standard Error	Chi-square	P-value
β_0	-8.7533	0.1415	3829.17	<0.0001
ln(AADT)	0.9013	0.0116	6042.61	<0.0001
Shoulder width (ft.)	-0.0255	0.0024	116.23	<0.0001
Median width (ft.)	-0.0015	0.0003	29.39	<0.0001
Dispersion (k)	0.3233	0.0070		

Head-On (type 1 SPF)

Variable	Estimate	Standard Error	Chi-square	P-value
β_0	-13.8233	0.4869	806.09	<0.0001
ln(AADT)	1.0256	0.0405	641.28	<0.0001
Dispersion (k)	0.7249	0.0820		

Head-On (type 2 SPF)

Variable	Estimate	Standard Error	Chi-square	P-value
β_0	-12.0283	0.5463	484.74	<0.0001
ln(AADT)	0.9178	0.0444	426.67	<0.0001
Shoulder width (ft.)	-0.0440	0.0084	27.52	<0.0001
Median width (ft.)	-0.0052	0.0011	23.40	<0.0001
Dispersion (k)	0.6841	0.0800		

Sideswipe (type 1 SPF)

Variable	Estimate	Standard Error	Chi-square	P-value
β_0	-14.8550	0.1462	10323.0	<0.0001
ln(AADT)	1.4407	0.0122	13841.6	<0.0001
Dispersion (k)	0.6145	0.0090		

Sideswipe (type 2 SPF)

Variable	Estimate	Standard Error	Chi-square	P-value
β_0	-14.3762	0.1623	7849.13	<0.0001
ln(AADT)	1.4322	0.0133	1538.3	<0.0001
Shoulder width (ft.)	-0.0460	0.0025	337.29	<0.0001
Median width (ft.)	-0.0008	0.0003	6.81	0.0091
Dispersion (k)	0.5973	0.0088		

Rear-End (type 1 SPF)

Variable	Estimate	Standard Error	Chi-square	P-value
β_0	-16.8036	0.1491	12707.3	<0.0001
ln(AADT)	1.6848	0.0125	18053.6	<0.0001
Dispersion (k)	0.8564	0.0098		

Rear-End (type 2 SPF)

Variable	Estimate	Standard Error	Chi-square	P-value
β_0	-16.0968	0.1674	9244.37	<0.0001
ln(AADT)	1.6640	0.0138	14460.1	<0.0001
Shoulder width (ft.)	-0.0523	0.0026	394.84	<0.0001
Median width (ft.)	-0.0019	0.0003	35.23	<0.0001
Dispersion (k)	0.8329	0.0096		

Hit-Object (type 1 SPF)

Variable	Estimate	Standard Error	Chi-square	P-value
β_0	-13.8233	0.4869	806.09	<0.0001
ln(AADT)	1.0256	0.0405	641.28	<0.0001
Dispersion (k)	0.7249	0.0820		

Hit-Object (type 2 SPF)

Variable	Estimate	Standard Error	Chi-square	P-value
β_0	-5.0660	0.1306	1504.38	<0.0001
ln(AADT)	0.6490	0.0108	3615.18	<0.0001
Divided median	0.0000	0.0000		
Undivided median	-1.5140	0.6730	5.06	0.0245
Shoulder width (ft.)	-0.0392	0.0023	301.77	<0.0001
Median width (ft.)	-0.0021	0.0003	55.81	<0.0001
Dispersion (k)	0.4581	0.0072		

Overturn (type 1 SPF)

Variable	Estimate	Standard Error	Chi-square	P-value
β_0	-2.9865	0.1848	261.11	<0.0001
ln(AADT)	0.2532	0.0157	261.01	<0.0001
Dispersion (k)	0.4585	0.0200		

Overturn (type 2 SPF)

Variable	Estimate	Standard Error	Chi-square	P-value
β_0	-3.9530	0.2143	340.22	<0.0001

ln(AADT)	0.3221	0.0175	338.85	<0.0001
Median width (ft.)	0.0042	0.0005	84.09	<0.0001
Dispersion (k)	0.4471	0.0198		

SPFs for Rural Four-Leg Stop-Controlled Intersections

Total collisions (type 1 SPF)

Variable	Estimate	Standard Error	Chi-square	P-value
β_0	-8.3834	0.2394	1226.41	<0.0001
ln(Major Road AADT)	0.6992	0.0265	697.53	<0.0001
Ln(Minor Road AADT)	0.3708	0.0149	620.48	
Dispersion (k)	0.7000	0.0301		

Total collisions (type 2 SPF)

Variable	Estimate	Standard Error	Chi-square	P-value
β_0	-8.9223	0.2890	952.99	<0.0001
ln(Major Road AADT)	0.7306	0.0291	630.42	<0.0001
Ln(Minor Road AADT)	0.3792	0.0151	632.20	
Left-turn lane on major[1]	0.1134	0.0536	4.48	0.0342
Right-turn lane on major[1]	0.1607	0.0689	5.44	0.0197
Dispersion (k)	0.6941	0.0299		

[1] takes a value of 0 if present, 1 otherwise

Injury (type 1 SPF)

Variable	Estimate	Standard Error	Chi-square	P-value
β_0	-8.9994	0.3240	771.56	<0.0001
ln(Major Road AADT)	0.6046	0.0349	300.24	<0.0001
Ln(Minor Road AADT)	0.3750	0.0204	337.53	<0.0001
Dispersion (k)	0.7678	0.0515		

Injury (type 2 SPF)

Variable	Estimate	Standard Error	Chi-square	P-value
β_0	-8.8571	0.4181	448.80	<0.0001
ln(Major Road AADT)	0.6351	0.0351	326.82	<0.0001

Ln(Minor Road AADT)	0.4031	0.0207	377.55	<0.0001
Left-turn lane on minor[1]	-0.8765	0.2365	13.73	0.0002
Presence of illumination[1]	0.4249	0.0625	46.23	<0.0001
Dispersion (k)	0.7134	0.0492		

[1] takes a value of 0 if present, 1 otherwise

Sideswipe (type 1 SPF)

Variable	Estimate	Standard Error	Chi-square	P-value
β_0	-10.4100	0.4346	573.62	<0.0001
ln(Major Road AADT)	0.7009	0.0469	223.13	<0.0001
Ln(Minor Road AADT)	0.3217	0.0256	158.20	<0.0001
Dispersion (k)	0.8751	0.0838		

Sideswipe (type 2 SPF)

Variable	Estimate	Standard Error	Chi-square	P-value
β_0	-11.6759	0.5311	483.37	<0.0001
ln(Major Road AADT)	0.7833	0.0523	224.35	<0.0001
Ln(Minor Road AADT)	0.3430	0.0261	172.65	<0.0001
Left-turn lane on major[1]	0.2653	0.0859	9.54	0.0020
Right-turn lane on major[1]	0.2721	0.1155	5.55	0.0185
Dispersion (k)	0.8528	0.0818		

[1] takes a value of 0 if present, 1 otherwise

Rear-End (type 1 SPF)

Variable	Estimate	Standard Error	Chi-square	P-value
β_0	-12.9503	0.4193	953.97	<0.0001
ln(Major Road AADT)	1.0750	0.0450	569.68	<0.0001
Ln(Minor Road AADT)	0.3128	0.0228	188.52	<0.0001
Dispersion (k)	1.1350	0.0686		

Rear-End (type 2 SPF)

Variable	Estimate	Standard Error	Chi-square	P-value
β_0	-15.5546	0.5117	924.11	<0.0001
ln(Major Road	1.2541	0.0502	624.96	<0.0001

AADT)				
Ln(Minor Road AADT)	0.3636	0.0231	24743	<0.0001
Left-turn lane on major[1]	0.6493	0.0795	66.67	0.0002
Right-turn lane on major[1]	0.3183	0.1049	9.21	<0.0001
Dispersion (k)	1.0337	0.0638		

[1] takes a value of 0 if present, 1 otherwise

Broadside (type 1 SPF)

Variable	Estimate	Standard Error	Chi-square	P-value
β_0	-9.6754	0.3544	745.14	<0.0001
ln(Major Road AADT)	0.6154	0.0383	258.46	<0.0001
Ln(Minor Road AADT)	0.5697	0.0227	630.88	<0.0001
Dispersion (k)	1.2515	0.0637		

Broadside (type 2 SPF)

Variable	Estimate	Standard Error	Chi-square	P-value
β_0	-10.0863	0.3786	709.84	<0.0001
ln(Major Road AADT)	0.6357	0.0389	267.33	<0.0001
Ln(Minor Road AADT)	0.5844	0.0231	642.71	<0.0001
Presence of illumination[1]	0.2203	0.0684	10.38	0.0013
Dispersion (k)	1.2372	0.0633		

[1] takes a value of 0 if present, 1 otherwise

SPFs for Rural Four-Leg Signalized Intersections

Total collisions (type 1 SPF)

Variable	Estimate	Standard Error	Chi-square	P-value
β_0	-7.5926	1.2191	38.79	<0.0001
ln(Major Road AADT)	0.8054	0.1192	45.69	<0.0001
Ln(Minor Road AADT)	0.1781	0.0424	17.63	<0.0001
Dispersion (k)	0.3736	0.0500		

Total collisions (type 1 SPF)

Variable	Estimate	Standard Error	Chi-square	P-value

β_0	-6.7021	1.2780	27.50	<0.0001
ln(Major Road AADT)	0.7481	0.1216	37.84	<0.0001
Ln(Minor Road AADT)	0.1534	0.0439	12.23	0.0005
Left-turn lane on minor road[1]	-0.2587	0.1210	4.57	0.0325
Dispersion (k)	0.3595	0.0485		

[1] takes a value of 0 if present, 1 otherwise

Injury (type 1 SPF)

Variable	Estimate	Standard Error	Chi-square	P-value
β_0	-4.6931	1.6243	8.35	0.0039
ln(Major Road AADT+Minor Road AADT)	0.4410	0.1631	7.31	0.0068
Dispersion (k)	0.6475	0.1074		

Injury (type 2 SPF)

Variable	Estimate	Standard Error	Chi-square	P-value
β_0	-3.1960	1.7365	3.39	0.0657
ln(Major Road AADT+Minor Road AADT)	0.3096	0.1703	3.31	0.0690
Left-turn lane on minor road[1]	-0.3486	0.1761	3.92	0.0478
Dispersion (k)	0.6183	0.1045		

[1] takes a value of 0 if present, 1 otherwise

Sideswipe (type 1 SPF)

Variable	Estimate	Standard Error	Chi-square	P-value
β_0	-9.4974	1.6997	31.22	<0.0001
ln(Major Road AADT)	0.7130	0.1697	17.66	<0.0001
Ln(Minor Road AADT)	0.2458	0.0601	16.74	<0.0001
Dispersion (k)	0.5159	0.1041		

Rear-End (type 1 SPF)

Variable	Estimate	Standard Error	Chi-square	P-value
β_0	-12.2616	1.4733	69.26	<0.0001
ln(Major Road AADT)	1.1549	0.1407	67.35	<0.0001
Ln(Minor Road AADT)	0.2350	0.0513	21.02	<0.0001

Dispersion (k)	0.5227	0.0754		

Rear-End (type 2 SPF)

Variable	Estimate	Standard Error	Chi-square	P-value
β_0	-11.8607	1.5556	58.13	<0.0001
ln(Major Road AADT)	1.1602	0.1445	64.46	<0.0001
Ln(Minor Road AADT)	0.1991	0.0501	15.78	<0.0001
Left-turn lane on minor[1]	-0.5272	0.1398	14.23	0.0002
Right-turn lane on major[1]	-0.2753	0.1467	3.52	0.0605
Right-turn lane on minor[1]	0.4127	0.1579	6.83	0.0090
Dispersion (k)	0.4415	0.0660		

[1] takes a value of 0 if present, 1 otherwise

Broadside (type 1 SPF)

Variable	Estimate	Standard Error	Chi-square	P-value
β_0	-4.3655	1.6733	6.81	0.0091
ln(Major Road AADT)	0.4138	0.1673	6.12	0.0134
Ln(Minor Road AADT)	0.0911	0.0546	2.79	0.0950
Dispersion (k)	0.5912	0.0874		

Broadside (type 2 SPF)

Variable	Estimate	Standard Error	Chi-square	P-value
β_0	-11.8607	1.5556	58.13	<0.0001
ln(Major Road AADT)	1.1602	0.1445	64.46	<0.0001
Ln(Minor Road AADT)	0.1991	0.0501	15.78	<0.0001
Left-turn lane on minor[1]	-0.5272	0.1398	14.23	0.0002
Right-turn lane on major[1]	-0.2753	0.1467	3.52	0.0605
Right-turn lane on minor[1]	0.4127	0.1579	6.83	0.0090
Dispersion (k)	0.4415	0.0660		

[1] takes a value of 0 if present, 1 otherwise

APPENDIX B: HIGH PROPORTION METHOD BETA DISTRIBUTION PARAMETERS

Segments

Facility Type	**Parameter**	**Injury**	**Head-on**	**Sideswipe**	**Rear-end**	**Broadside**	**Hit object**	**Overturn**	**Pedestrian**	**Other**
Rural Two-Lane	Alpha	7.89	1.60	3.27	0.66	0.74	3.23	1.39	0.49	0.80
	Beta	17.68	28.17	3.39	3.91	11.08	4.72	6.79	95.07	11.65
Urban Freeway In Interchange	Alpha	5.02	1.69	9.06	3.70	2.25	2.40	0.70	0.35	0.73
	Beta	31.95	347.45	36.29	4.17	94.14	7.30	21.49	134.55	41.19
Urban Freeway Outside Interchange	Alpha	9.81	2.40	6.29	2.72	4.59	2.95	0.70	1.20	0.62
	Beta	48.44	473.99	31.13	3.87	197.08	6.40	13.80	485.88	26.07

Intersections

Facility Type	**Parameter**	**Injury**	**Head-on**	**Sideswipe**	**Rear-end**	**Broadside**	**Hit object**	**Overturn**	**Pedestrian**	**Other**
Rural Four-Leg Signalized	Alpha	2.99	6.96	9.82	4.02	2.66	1.66	1.68	1.26	1.74
	Beta	16.95	137.10	84.13	5.36	6.43	23.19	115.23	59.26	88.13
Rural Four-Leg Stop-Controlled	Alpha	18.75	4.39	2.69	1.69	2.24	1.55	0.54	0.75	1.73
	Beta	58.27	114.89	23.36	5.69	3.58	8.10	12.53	78.38	51.73

APPENDIX C: RECODING OF VARIABLES USED IN SAFETYANALYST

This Appendix shows the list of HSIS variables that were recoded to match with the variables and levels in SafetyAnalyst. The main focus of this effort was on the variables that were required for SafetyAnalyst. In the tables that follow this page, the first column shows the levels in SafetyAnalyst, the second column shows the levels for the corresponding CA HSIS variables, and the third column shows the recoded value for each level of the CA HSIS variable to match the SafetyAnalyst levels. The tables are provided for the roadway file, collision file, intersection, and ramp files. California"s ramp database does not include length of ramps and hence could not be included in the SafetyAnalyst network screening process.

AltRoadwaySegments

SafetyAnalyst Variable	HSIS Variable	SafetyAnalyst Code
routeName	**road.CNTYRTE**	
agencyID	**to be derived (road.CNTYRTE + road.BEGMP)**	
		cntyrte + begmp
locSystem	**no corresponding variable**	
A - Route/Milepost		"A" was coded for this variable
B - Route/County/Milepost		
C - Route/Section/Distance		
D - Section/Distance		
routeType	**veh.RTE_TYPE**	
I - Interstate - Route category interstate	'1' = 'ON STATE ROUTE'	"X" was coded for this variable
US - US route - Route category US route	'2' = 'NOT ON STATE ROUTE'	
SR - State route - Route category state route	'3' = 'INTERSECTING STATE ROUTE'	
BR - Business route - Route category business route	'-' = 'DOES NOT APPLY'	
BL - Business loop - Route category business loop	'<' = 'NOT STATED'	
SP - Spur route - Route category spur route	OTHER = 'ERROR/OTHER CODES'	
CR - County road - Route category county road		
TR - Township road - Route category township road	**road.FED_PREF**	
L - Local road - Route category local road	'$' = 'NO FEDERAL AID'	
O - Other - Route category other	'I' = 'INTERSTATE'	
X - Unknown - Route category unknown	'P' = 'PRIMARY'	
	'5' = 'FAS OR FAU – FINAL ALIGNMENT'	
	'6' = 'FAS/FAU - NOT FINAL ALIGNMENT'	
	'-' = 'INVALID DATA'	
	'+' = 'NO DATA'	
	OTHER = 'ERROR/OTHER CODES'	
	HSIS-CA data set has no variable to indicate route category, i.e., Interstate, US route, State route, etc.	
startOffset	**road.BEGMP**	
endOffset	**road.ENDMP**	
segmentLength	**road.SEG_LNG**	

SafetyAnalyst Variable	**HSIS Variable**	**SafetyAnalyst Code**
areaType	**road.RURURB**	
R - Rural - Rural area type	'R' = 'RURAL'	R
U - Urban - Urban area type	'U' = 'URBAN'	U
X - Unknown - Unknown area type	else	X
terrain	**road.TERRAIN**	
L - Level	'F' = 'Flat'	L
R - Rolling	'R' - 'Rolling'	R
M - Mountainous	'M' - 'Mountainous'	M
X - Unknown	'-' - 'Invalid Data'	X
	'+' - 'No Data'	X
d1numThruLane	**road.NO_LANE1**	
d2numThruLane	**road.NO_LANE2**	
medianWidth	**road.MEDWID**	
roadwayClass1	**road.FUNC_CLS**	
1 - Principal arterial-interstate	'1' = 'RURAL PRINCIPAL ARTERIAL WITH EXTENSION INTO URBAN PA'	1
2 - Principal arterial-other freeway or expressway	'2' = 'RURAL PRINCIPAL ARTERIAL WITH EXTENSION INTO URBAN MA'	2
3 - Principal arterial-other	'3' = 'PRINCIPAL ARTERIAL WITHOUT EXTENSION'	3
4 - Minor arterial	'4' = 'MINOR ARTERIAL'	4
5 - Major Collector	'5' = 'MAJOR COLLECTOR'	5
6 - Minor Collector	'6' = 'MINOR COLLECTOR'	6
7 - Local	'7' = 'LOCAL'	7
0 - Other	'0' = 'NONE'	0
99 - Unknown	else	99
operationWay	**no corresponding variable**	
1 - One-way road or street		"2" was coded for this variable
2 - Two way road or street		
3 - One direction of travel for a divided highway		
99 - Unknown		
medianType1	**road.MED_TYPE**	
1 - Rigid barrier system - e.g., concrete	A' = 'UNDIVIDED, NOT SEPARATED OR STRIPED'	0
2 - Semi-rigid barrier system - e.g., box beam, W-beam strong post	B' = 'UNDIVIDED, STRIPED'	0
3 - Flexible barrier system - e.g., cable, W-beam weak post	C' = 'UNDIVIDED, REVERSIBLE PEAK HOUR LANE(S)'	0
4 - Raised median with curb	E' = 'DIVIDED, REVERSIBLE PEAK HOUR LANE(S)'	9
5 - Depressed median	F' = 'DIVIDED, TWO-WAY LEFT TURN LANE'	9
6 - Flush paved median	G' = 'DIVIDED, CONTINUOUS LEFT-TURN LANE'	9

SafetyAnalyst Variable	HSIS Variable	SafetyAnalyst Code
7 - HOV lane(s)	H' = 'DIVIDED, PAVED MEDIAN'	6
8 - Railroad or rapid transit	J' = 'DIVIDED, UNPAVED MEDIAN'	5
9 - Other divided	K' = 'DIVIDED, SEPARATE GRADES'	4
0 - Undivided	L' = 'DIVIDED, SEPARATE GRADES WITH RETAINING WALL'	1
99 - Unknown	M' = 'DIVIDED, SAWTOOTH (UNPAVED)'	9
	N' = 'DIVIDED, SAWTOOTH (PAVED)'	9
	P' = 'DIVIDED, DITCH'	5
	Q' = 'DIVIDED, SEPARATE STRUCTURE'	9
	R' = 'DIVIDED, RAILROAD OR RAPID TRANSIT'	8
	S' = 'DIVIDED, BUS LANES'	7
	T' = 'DIVIDED, PAVED AREA, OCCASIONAL TRAFFIC LANE'	9
	U' = 'DIVIDED, RAILROAD AND BUS LANE'	9
	V' = 'DIVIDED, CONTAINS REVERSIBLE PEAK-HOUR LANE(S)'	9
	Z' = 'DIVIDED, OTHER'	9
	else	99
accessControl	**road.ACCESS**	
1 - Full Access Control - Public authority has full access control	'C' = 'CONVENTIONAL - NO ACCESS CONTROL'	3
2 - Partial access Control - Public authority has partial access control	'E' = 'EXPRESSWAY - PARTIAL ACCESS CONTROL'	2
3 - No Access Control - Public authority has no access control	'F' = 'FREEWAY - FULL ACCESS CONTROL'	1
99 - Unknown - Public authority has unknown access control	'S' = 'ONE-WAY CITY STREET - NO ACCESS CONTROL'	3
	'-' = 'INVALID DATA'	99
	'+' = 'NO DATA'	99
	OTHER = 'ERROR/OTHER CODES'	99
interchangeInfluence	**no corresponding variable**	
Y - Yes - Roadway is within interchange influence area		"Y" or "N" was coded based on the location of ramps - 0.3 miles on either side of ramp was considered within influence of interchange
N - No - Roadway is not within interchange influence area		
X - Unknown - Unknown whether roadway is within interchange influence area		

AltCollision

SafetyAnalyst Variable	HSIS Variable	SafetyAnalyst Code
agencyID	**acc.CASENO**	
locSystem	**no corresponding variable**	
A - Route/Milepost		"A" was coded for this variable
B - Route/County/Milepost		
C - Route/Section/Distance		
D - Section/Distance		
routeType	**veh.RTE_TYPE**	
I - Interstate - Route category interstate	'1' = 'ON STATE ROUTE'	
US - US route - Route category US route	'2' = 'NOT ON STATE ROUTE'	
SR - State route - Route category state route	'3' = 'INTERSECTING STATE ROUTE'	
BR - Business route - Route category business route	'-' = 'DOES NOT APPLY'	
BL - Business loop - Route category business loop	'<' = 'NOT STATED'	
SP - Spur route - Route category spur route	OTHER = 'ERROR/OTHER CODES'	
CR - County road - Route category county road		
TR - Township road - Route category township road	**road.FED_PREF**	
L - Local road - Route category local road	'$' = 'NO FEDERAL AID'	
O - Other - Route category other	'I' = 'INTERSTATE'	
X - Unknown - Route category unknown	'P' = 'PRIMARY'	
	'5' = 'FAS OR FAU – FINAL ALIGNMENT'	
	'6' = 'FAS/FAU - NOT FINAL ALIGNMENT'	
	'-' = 'INVALID DATA'	
	'+' = 'NO DATA'	
	OTHER = 'ERROR/OTHER CODES'	
	HSIS-CA data set has no variable to indicate route category, i.e., Interstate, US route, State route, etc.	
routeName	**acc.CNTY_RTE**	
locOffset	**acc.MILEPOST**	
collisionDate	**acc.ACC_DATE**	

SafetyAnalyst Variable	**HSIS Variable**	**SafetyAnalyst Code**
collisionSeverity1	**acc.SEVERITY**	
K - Fatal Injury	'1' = 'FATAL'	K
A - Incapacitating Injury	'2' = 'SEVERE INJURY'	A
B - Non-Incapacitating Injury	'3' = 'OTHER VISIBLE INJURY'	B
C - Possible Injury	'4' = 'COMPLAINT OF PAIN'	C
P - Property-Damage-Only	'0' = 'PDO - PROPERTY DAMAGE ONLY'	P
X - Unknown	else	X
numberOfFatalities	**acc.TOT_KILL**	
numberOfInjuries	**acc.TOT_INJ**	
junctionRelationship	**acc.INT_RMP**	
1 - Non-junction - Non-junction	'1' = 'Ramp intersection (exit), last 50 ft'	5
2 - At intersection - At intersection	'2' = 'Mid-Ramp'	5
3 - Intersection-related - Intersection-related	'3' = 'Ramp entry, first 50 ft'	5
4 - At driveway or driveway-related - At driveway or driveway-related	'4' = 'Ramp area/intersection street, within 100 ft'	3
5 - Entrance/exit ramp - Entrance/exit ramp	'5' = 'In intersection'	2
6 - Other part of interchange - Other part of interchange	'6' = 'Outside intersection (non-State route), within 250 ft'	3
7 - Railroad/highway grade crossing - Railroad/highway grade crossing	'<' = 'Not stated or undetermined'	99
8 - Crossover related - Crossover related	'-' = 'Does not apply'	1
9 - Other - Other	OTHER = 'ERROR/OTHER CODES'	9
99 - Unknown - Unknown		
	HSIS NOTE: Code "4" refers to locations on non-State intersecting routes (which are not mileposted) between 150 ft "outside" the ramp-related intersection to 150 ft outside the opposing ramp-related intersection. Similarly, code "6" is for locations on non-State routes within + 150 ft of the intersection. Thus, in both cases, the collisions are mileposted to the ramp or State route. In contrast, similar locations on intersecting State routes are mileposted to the intersecting State route. This variable will allow for identification of such "area 4" collisions if desirable in the analysis. See AREA4 in the Ramp File.	
collisionType	**acc.ACCTYPE**	
1 - Collision with parked motor vehicle	A' = 'HEAD-ON'	22
2 - Collision with railroad train	B' = 'SIDESWIPE'	25
3 - Collision with bicyclist	C' = 'REAR END'	21
4 - Collision with pedestrian	D' = 'BROADSIDE'	24

SafetyAnalyst Variable	HSIS Variable	SafetyAnalyst Code
5 - Collision with animal	E' = 'HIT OBJECT'	1 if VEH_INVL is 'E' 2 if VEH_INVL is 'F' 3 if VEH_INVL is 'G' and BIKE_FLAG is 'Yes' 4 if VEH_INVL is 'B' and PED_FLAG is 'Yes' 5 if VEH_INVL is 'H' 6 if VEH_INVL is 'I' 7 if VEH_INVL is 'J'
6 - Collision with fixed object	F' = 'OVERTURNED'	9
7 - Collision with other object	G' = 'AUTO-PEDESTRIAN'	4
8 - Other single-vehicle collision	H' = 'OTHER'	99
9 - Overturn	else	99
10 - Fire or explosion		
11 - Other single-vehicle non-collision	SafetyAnalyst separates sideswipe same direction and opposite direction. However, HSIS CA data has only level called sideswipe	
21 - Rear-end		
22 - Head-on		
23 - Rear-to-rear		
24 - Angle		
25 - Sideswipe, same direction		
26 - Sideswipe, opposite direction		
27 - Other multiple-vehicle collision		
99 - Unknown		
numVehicles	**acc.NUMVEHS**	
v1initialTravelDirection	**veh.DIR_TRVL**	
NB - Northbound	N' = 'N, NE OR NW BND'	NB
SB - Soutbound	S' = 'S, SE OR SW BND'	SB
EB - Eastbound	E' = 'EASTBOUND'	EB
WB - Westbound	W' = 'WESTBOUND'	WB
NO - Not on roadway	else	XX
XX - Unknown		
v2initialTravelDirection	**veh.DIR_TRVL**	
NB - Northbound	N' = 'N, NE OR NW BND'	NB
SB - Soutbound	S' = 'S, SE OR SW BND'	SB
EB - Eastbound	E' = 'EASTBOUND'	EB
WB - Westbound	W' = 'WESTBOUND'	WB
NO - Not on roadway	else	XX
XX - Unknown		

SafetyAnalyst Variable	HSIS Variable	SafetyAnalyst Code
v1vehicleManeuver	**veh.MISCACT1**	
1 - Movements essentially straight ahead	'A' = 'STOPPED'	12
2 - Backing	'B' = 'PROCEEDING STRAIGHT'	1
3 - Changing lanes	'C' = 'RAN OFF ROAD'	9
4 - Overtaking/passing	'D' = 'MAKING RIGHT TURN'	5
5 - Turning right	'E' = 'MAKING LEFT TURN'	6
6 - Turning left	'F' = 'MAKING U TURN'	7
7 - Making U-turn	'G' = 'BACKING'	2
8 - Entering traffic lane	'H' = 'SLOWING, STOPPING'	11
9 - Leaving traffic lane	'I' = 'PASSING OTHER VEHICLE'	4
10 - Parked	'J' = 'CHANGING LANES'	3
11 - Slowing	'K' = 'PARKING MANEUVER'	14
12 - Stopped in traffic	'L' = 'ENTERING TRAFFIC FROM SHOULDER, MEDIAN...'	14
13 - Negotiating a curve	'M' = 'OTHER POTENTIAL TURNING'	14
14 - Other	'N' = 'CROSSED INTO OPPOSING LANE'	14
99 - Unknown	'O' = 'PARKED'	10
	'P' = 'MERGING'	8
	'Q' = 'TRAVELING WRONG WAY'	14
	'R' = 'OTHER'	14
	else	99
v2vehicleManeuver	**veh.MISCACT1**	
1 - Movements essentially straight ahead	'A' = 'STOPPED'	12
2 - Backing	'B' = 'PROCEEDING STRAIGHT'	1
3 - Changing lanes	'C' = 'RAN OFF ROAD'	9
4 - Overtaking/passing	'D' = 'MAKING RIGHT TURN'	5
5 - Turning right	'E' = 'MAKING LEFT TURN'	6
6 - Turning left	'F' = 'MAKING U TURN'	7
7 - Making U-turn	'G' = 'BACKING'	2
8 - Entering traffic lane	'H' = 'SLOWING, STOPPING'	11
9 - Leaving traffic lane	'I' = 'PASSING OTHER VEHICLE'	4
10 - Parked	'J' = 'CHANGING LANES'	3
11 - Slowing	'K' = 'PARKING MANEUVER'	14
12 - Stopped in traffic	'L' = 'ENTERING TRAFFIC FROM SHOULDER, MEDIAN...'	14
13 - Negotiating a curve	'M' = 'OTHER POTENTIAL TURNING'	14
14 - Other	'N' = 'CROSSED INTO OPPOSING LANE'	14
99 - Unknown	'O' = 'PARKED'	10
	'P' = 'MERGING'	8
	'Q' = 'TRAVELING WRONG WAY'	14
	'R' = 'OTHER'	14
	else	99

AltSegmentTraffic

SafetyAnalyst Variable	HSIS Variable	SafetyAnalyst Code
agencyID	**to be derived (road.CNTYRTE + road.BEGMP)**	
calendarYear	**no corresponding variable - to be derived**	
aadtVPD	**road.AADT**	

AltIntersection

SafetyAnalyst Variable	HSIS Variable	SafetyAnalyst Code
agencyID	**to be derived (road.CNTYRTE + int.MILEPOST)**	
areaType	**road.RURURB**	
R - Rural - Rural area type	'R' = 'RURAL'	R
U - Urban - Urban area type	'U' = 'URBAN'	U
X - Unknown - Unknown area type	else	X
intersectionType1	**int.TYPEDESC**	
1 - Tee intersection	'F' = '4 - LEGGED'	3
2 - Y intersection	'M' = '> 4 LEGS'	5
3 - Four-leg intersection	'S' = 'OFFSET'	99
4 - Traffic circle/roundabout	'T' = 'TEE'	1
5 - Multileg intersection, five or more legs	'Y' = 'WYE'	2
0 - Other	'Z' = 'OTHER'	0
99 - Unknown	else	99
majorRoadLocSystem	**no corresponding variable**	
A - Route/Milepost		"A" was coded for this variable
B - Route/County/Milepost		
C - Route/Section/Distance		
D - Section/Distance		
majorRoadOffset	**int.MILEPOST**	
routeName	**int.CNTYRTE**	
routeType	**veh.RTE_TYPE**	
I - Interstate - Route category interstate	'1' = 'ON STATE ROUTE'	"X" was coded for this variable
US - US route - Route category US route	'2' = 'NOT ON STATE ROUTE'	
SR - State route - Route category state route	'3' = 'INTERSECTING STATE ROUTE'	
BR - Business route - Route category business route	'-' = 'DOES NOT APPLY'	
BL - Business loop - Route category business loop	'<' = 'NOT STATED'	
SP - Spur route - Route category spur route	OTHER = 'ERROR/OTHER CODES'	

SafetyAnalyst Variable	**HSIS Variable**	**SafetyAnalyst Code**
CR - County road - Route category county road		
TR - Township road - Route category township road	**road.FED_PREF**	
L - Local road - Route category local road	'$' = 'NO FEDERAL AID'	
O - Other - Route category other	'I' = 'INTERSTATE'	
X - Unknown - Route category unknown	'P' = 'PRIMARY'	
	'5' = 'FAS OR FAU – FINAL ALIGNMENT'	
	'6' = 'FAS/FAU - NOT FINAL ALIGNMENT'	
	'-' = 'INVALID DATA'	
	'+' = 'NO DATA'	
	OTHER = 'ERROR/OTHER CODES'	
	HSIS-CA data set has no variable to indicate route category, i.e., Interstate, US route, State route, etc.	
trafficControl1	**int.TRF_CNTL**	
1 - No control	'A' = 'NO CONTROL'	1
2 - Stop signs on cross street only	'B' = 'STOP SIGNS ON CROSS STREET ONLY'	2
3 - Stop signs on mainline only	'C' = 'STOP SIGNS ON MAINLINE ONLY'	3
4 - All-way stop signs	'D' = 'FOUR-WAY STOP SIGNS'	4
5 - Two-way flasher (red on cross street)	'E' = 'FOUR-WAY FLASHER (RED ON CROSS STREET)'	5
6 - Two-way flasher (red on mainline)	'F' = 'FOUR-WAY FLASHER (RED ON MAINLINE)'	6
7 - All-way flasher (red on all)	'G' = 'FOUR-WAY FLASHER (RED ON ALL)'	7
8 - Yield signs on cross street only	'H' = 'YIELD SIGNS ON CROSS STREET ONLY'	8
9 - Yield signs on mainline only	'I' = 'YIELD SIGNS ON MAINLINE ONLY'	9
10 - Other non-signalized	'J' = 'SIGNALS PRETIMED (2 PHASE)'	11
11 - Signals pre timed (2 phase)	'K' = 'SIGNALS PRETIMED (MULTI-PHASE)'	12
12 - Signals pre timed (multi-phase)	'L' = 'SIGNALS SEMI-TRAFFIC ACTUATED, 2 PHASE'	13
13 - Signals semi-actuated (2 phase)	'M' = 'SIGNALS SEMI-TRAFFIC ACTUATED, MULTI-PHASE'	14
14 - Signals semi-actuated (multi-phase)	'N' = 'SIGNALS FULL TRAFFIC ACTUATED, 2 PHASE'	15
15 - Signals fully actuated (2 phase)	'P' = 'SIGNALS FULL TRAFFIC ACTUATED, MULTI-PHASE'	16
16 - Signals fully actuated (multi-phase)	'Z' = 'OTHER'	99
17 - Other signalized	else	99

SafetyAnalyst Variable	HSIS Variable	SafetyAnalyst Code
18 - Roundabout		
99 - Unknown		

AltMajorRoadTraffic

SafetyAnalyst Variable	HSIS Variable	SafetyAnalyst Code
agencyID	**to be derived (road.CNTYRTE + int.MILEPOST)**	
calendarYear	**no corresponding variable - to be derived**	
aadtVPD	**int.ML_AADT**	

AltMinorRoadTraffic

SafetyAnalyst Variable	HSIS Variable	SafetyAnalyst Code
agencyID	**to be derived (road.CNTYRTE + int.MILEPOST)**	
calendarYear	**no corresponding variable - to be derived**	
aadtVPD	**int.XSTAADT**	

AltLeg

SafetyAnalyst Variable	HSIS Variable	SafetyAnalyst Code
agencyID	**no corresponding variable**	
legType	**no corresponding variable**	
1 - Major road, increasing milepost direction - Major road approach in the primary increasing milepost direction		
2 - Major road, decreasing milepost direction - Major road approach in the secondary or decreasing milepost direction		
3 - Minor road, increasing milepost direction - Minor road approach to right of the primary or increasing milepost direction		
4 - Minor road, decreasing milepost direction - Minor road approach to left of the primary or increasing milepost direction		

SafetyAnalyst Variable	HSIS Variable	SafetyAnalyst Code
98 - Not Valid - Not valid, e.g., 4th (unused) leg of a three-legged intersection		
99 - Unknown - Unknown		
legDirection	**no corresponding variable**	
NB - NB approach - Directional approach of the intersecting leg is northbound		
SB - SB approach - Directional approach of the intersecting leg is southbound		
WB - WB approach - Directional approach of the intersecting leg is westbound		
EB - EB approach - Directional approach of the intersecting leg is eastbound		
X - Unknown - Directional approach of the intersecting leg is unknown		

AltLegTraffic

SafetyAnalyst Variable	HSIS Variable	SafetyAnalyst Code
agencyID	**no corresponding variable**	
calendarYear	**no corresponding variable**	
aadtVPD	**no corresponding variable**	

AltRamp

SafetyAnalyst Variable	HSIS Variable	SafetyAnalyst Code
agencyID	**to be derived (ramp.CNTYRTE+ramp.MILEPOST)**	
locSystem	**no corresponding variable**	
A - Route/Milepost		"A" was coded for this variable
B - Route/County/Milepost		
C - Route/Section/Distance		
D - Section/Distance		
routeType	**veh.RTE_TYPE**	
I - Interstate - Route category interstate	'1' = 'ON STATE ROUTE'	"X" was coded for this variable
US - US route - Route category US route	'2' = 'NOT ON STATE ROUTE'	
SR - State route - Route category state route	'3' = 'INTERSECTING STATE ROUTE'	
BR - Business route - Route category business route	'-' = 'DOES NOT APPLY'	

SafetyAnalyst Variable	**HSIS Variable**	**SafetyAnalyst Code**
BL - Business loop - Route category business loop	'<' = 'NOT STATED'	
SP - Spur route - Route category spur route	OTHER = 'ERROR/OTHER CODES'	
CR - County road - Route category county road		
TR - Township road - Route category township road	**road.FED_PREF**	
L - Local road - Route category local road	'$' = 'NO FEDERAL AID'	
O - Other - Route category other	'I' = 'INTERSTATE'	
X - Unknown - Route category unknown	'P' = 'PRIMARY'	
	'5' = 'FAS OR FAU – FINAL ALIGNMENT'	
	'6' = 'FAS/FAU - NOT FINAL ALIGNMENT'	
	'-' = 'INVALID DATA'	
	'+' = 'NO DATA'	
	OTHER = 'ERROR/OTHER CODES'	
	HSIS-CA data set has no variable to indicate route category, i.e., Interstate, US route, State route, etc.	
routeName	**ramp.CNTYRTE**	
startOffset	**ramp.MILEPOST**	
areaType	**road.RURURB**	
R - Rural - Rural area type	'R' = 'RURAL'	R
U - Urban - Urban area type	'U' = 'URBAN'	U
X - Unknown - Unknown area type	else	X
rampType	**ramp.ON_OFFRP**	
1 - Off ramp - Exit freeway	'O' = 'ON RAMP'	2
2 - On ramp - Enter freeway	'F' = 'OFF RAMP'	1
3 - Freeway-to-freeway ramp - Connect two freeways	'Z' = 'OTHER'	0
0 - Other - Other type of ramp	'-' = 'INVALID DATA'	99
99 - Unknown - Unknown type of ramp	'+' = 'NO DATA	99
	OTHER = 'ERROR/OTHER CODES'	99
rampConfiguration	**ramp.RAMP_TYP**	
1 - Diamond - Ramp	'A' = 'FRONTAGE ROAD'	0
2 - Parclo loop - Partial clover leaf loop	'B' = 'COLLECTOR ROAD'	6
3 - Free-flow loop	'C' = 'DIRECT OR SEMI-DIRECT CONNECTOR (LEFT)'	5
4 - Free-flow outer connection	'D' = 'DIAMOND TYPE RAMP'	1
5 - Direct or semi-direct connection	'E' = 'SLIP RAMP'	0
6 - C-D road or other connector	'F' = 'DIRECT OR SEMI-DIRECT CONNECTOR (RIGHT)'	5
0 - Other	'G' = 'LOOP WITH LEFT TURN' (CONT'D)	2

SafetyAnalyst Variable	HSIS Variable	SafetyAnalyst Code
99 - Unknown	'H' = 'BOTTONHOOK RAMP'	0
	'J' = 'SCISSORS RAMP'	0
	'K' = 'SPLIT RAMP'	0
	'L' = 'LOOP WITHOUT LEFT TURN'	4
	'M' = 'TWO-WAY RAMP SEGMENT'	0
	'P' = 'DUMMY-PAIRED'	0
	'R' = 'REST AREA, VISTA POINT, TRUCK SCALE'	0
	'V' = 'DUMMY-VOLUMES ONLY'	0
	'Z' = 'OTHER'	0
	'-' = 'INVALID DATA'	99
	'+' = 'NO DATA'	99
	OTHER = 'ERROR/OTHER CODES'	0

AltRampTraffic

SafetyAnalyst Variable	HSIS Variable	SafetyAnalyst Code
agencyID	**to be derived (ramp.CNTYRTE+ramp.MILEPOST)**	
calendarYear	**no corresponding variable**	
aadtVPD	**ramp.RMP_ADT**	