HIGH RISK CRASH ANALYSIS

Final Report 558

Prepared by:
Simon Washington and Wen Cheng
Department of Civil Engineering & Engineering Mechanics
University of Arizona
Tucson, AZ 85721

December 2005

Prepared for:
Arizona Department of Transportation
206 South 17th Avenue
Phoenix, Arizona 85007
in cooperation with
U.S. Department of Transportation
Federal Highway Administration

DISCLAIMER

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
FHWA-AZ-05-558		

4. Title and Subtitle
High Risk Crash Analysis

5. Report Date
December 2005

6. Performing Organization Code

7. Author
Dr. Simon Washington and Wen Cheng

8. Performing Organization Report No.

9. Performing Organization Name and Address
University of Arizona
Tucson, AZ 85721

10. Work Unit No.

11. Contract or Grant No.
SPR-PL-1-(63) 558

12. Sponsoring Agency Name and Address
Arizona Department of Transportation
206 S. 17th Avenue
Phoenix, Arizona 85007

13.Type of Report & Period Covered
Final Report

14. Sponsoring Agency Code

15. Supplementary Notes
Prepared in cooperation with the U.S. Department of Transportation, Federal Highway Administration

16. Abstract
In agencies with jurisdiction over extensive road infrastructure it is common practice to select and rectify hazardous locations. Improving hazardous locations may arise during safety management activities, during maintenance activities, or as a result of political pressures and/or public attention. Commonly a two-stage process is used. In the first stage the past accident history of all sites is reviewed to screen a limited number of high-risk locations for further examination. In the second stage the selected sites are studied in greater detail to devise cost-effective remedial actions or countermeasures for a subset of correctable sites. Due often to limited time and resources constraints and the extensive number of candidate sites typically considered in such endeavors, it is impractical for agencies to examine all sites in detail. The current Arizona Local Government Safety Project Analysis Model (ALGSP) is intended to facilitate conducting these procedures by providing an automated method for analysis and evaluation of motor vehicle crashes and subsequent remediation of 'hot spot' or 'high risk' locations. The software is user friendly and can save lots of time for local jurisdictions and governments such as Metropolitan Planning Organizations (MPOs), counties, cities, and towns. Some analytical improvements are possible, however.

The objective of this study was to provide recommendations that will lead to improvement in the accuracy and reliability of the ALGSP software for identifying true 'hot spots' within the Arizona transportation system or network, be they road segments, ramps, or intersections.

The research resulted in 1) a survey of past and current hot spot identification (HSID) approaches, 2) evaluation of HSID methods and exploration of optimum duration of before-period crash data under simulated scenarios, 3) development of safety performance functions (SPFs) for various functional road sections within Arizona, 4) extended comparisons of alternative HSID methods based on SPFs by using real crash data, and 5) recommendations for improving the identification ability of current ALGSP model.

17. Key Words
Hot Spot Identification, High Risk Sites, Sites with Promise, Safety, Motor Vehicle Crashes

18. Distribution Statement
Document is available to the U.S. Public through the National Technical Information Service, Springfield, Virginia, 22161

23. Registrant's Seal

19. Security Classification	20. Security Classification	21. No. of Pages	22. Price
Unclassified	Unclassified	154	

SI (MODERN METRIC) CONVERSION FACTORS*

APPROXIMATE CONVERSIONS TO SI UNITS

Symbol	When You Know	Multiply By	To Find	Symbol
		LENGTH		
in	Inches	25.4	millimeters	mm
ft	Feet	0.305	meters	m
yd	Yards	0.914	meters	m
mi	Miles	1.61	kilometers	km
		AREA		
in^2	square inches	645.2	square millimeters	mm^2
ft^2	square feet	0.093	square meters	m^2
yd^2	square yards	0.836	square meters	m^2
ac	Acres	0.405	hectares	ha
mi^2	square miles	2.59	square kilometers	km^2
		VOLUME		
fl oz	fluid ounces	29.57	milliliters	mL
gal	Gallons	3.785	liters	L
ft^3	cubic feet	0.028	cubic meters	m^3
yd^3	cubic yards	0.765	cubic meters	m^3
	NOTE: Volumes greater than 1000L shall be shown in m^3.			
		MASS		
oz	Ounces	28.35	grams	g
lb	Pounds	0.454	kilograms	kg
T	short tons (2000lb)	0.907	megagrams (or "metric ton")	mg (or "t")
		TEMPERATURE (exact)		
°F	Fahrenheit temperature	5(F-32)/9 or (F-32)/1.8	Celsius temperature	°C
		ILLUMINATION		
fc	foot candles	10.76	lux	lx
fl	foot-Lamberts	3.426	candela/m^2	cd/m^2
		FORCE AND PRESSURE OR STRESS		
lbf	Poundforce	4.45	newtons	N
lbf/in^2	poundforce per square inch	6.89	kilopascals	kPa

APPROXIMATE CONVERSIONS FROM SI UNITS

Symbol	When You Know	Multiply By	To Find	Symbol
		LENGTH		
mm	millimeters	0.039	inches	in
m	meters	3.28	feet	ft
m	meters	1.09	yards	yd
km	kilometers	0.621	miles	mi
		AREA		
mm^2	Square millimeters	0.0016	square inches	in^2
m^2	Square meters	10.764	square feet	ft^2
m^2	Square meters	1.195	square yards	yd^2
ha	hectares	2.47	acres	ac
km^2	Square kilometers	0.386	square miles	mi^2
		VOLUME		
mL	milliliters	0.034	fluid ounces	fl oz
L	liters	0.264	gallons	gal
m^3	Cubic meters	35.315	cubic feet	ft^3
m^3	Cubic meters	1.308	cubic yards	yd^3
		MASS		
g	grams	0.035	ounces	oz
kg	kilograms	2.205	pounds	lb
Mg	megagrams (or "metric ton")	1.102	short tons (2000lb)	T
		TEMPERATURE (exact)		
°C	Celsius temperature	1.8C + 32	Fahrenheit temperature	°F
		ILLUMINATION		
lx	lux	0.0929	foot-candles	fc
cd/m^2	candela/m^2	0.2919	foot-Lamberts	fl
		FORCE AND PRESSURE OR STRESS		
N	newtons	0.225	poundforce	lbf
kPa	kilopascals	0.145	poundforce per square inch	lbf/in^2

SI is the symbol for the International System of Units. Appropriate rounding should be made to comply with Section 4 of ASTM E380

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

EXECUTIVE SUMMARY

In many agencies with jurisdiction over extensive road infrastructure, it is common practice to select and rectify hazardous locations. Improving hazardous locations may arise during safety management activities, during maintenance activities, or as a result of political pressures and/or public attention. Commonly a two-stage process is used. In the first stage, the past accident history of all sites is reviewed to screen a limited number of high risk locations for further examination. In the second stage, the selected sites are studied in greater detail to devise cost-effective remedial actions or countermeasures for a subset of correctable sites.

Due to limited time and resources, constraints and the extensive number of candidate sites typically considered in such endeavors, it is impractical for agencies to examine all sites in detail. The current Arizona Local Government Safety Project (ALGSP)Analysis Model, which was developed by Carey (2001) with funding from the Arizona Department of Transportation (ADOT), is intended to facilitate conducting these procedures by providing an automated method for analysis and evaluation of motor vehicle crashes and subsequent remediation of 'hot spot' or 'high risk' locations. The software is user friendly and can save large amounts of time for local jurisdictions and governments such as Metropolitan Planning Organizations (MPOs), counties, cities, and towns. However, its analytical core is based on the simple ranking of crash statistics, where the user is offered choices of crash frequency, crash rate, crash severity, or crash cost (severities associated with average costs per crash severity type). Although this method has the benefit of straightforwardness, the efficiency of identifying truly high-risk sites leaves some room for improvement. This research, funded by ADOT, aims to justify and recommend improvements to the analytical algorithms within the ALGSP model, thus enhancing its ability to accurately identify high risk sites.

Included in the results of this research are a survey of past and current hot spot identification (HSID) approaches; evaluation of HSID methods, and exploration of optimum duration of before-period crash data under simulated scenarios; development of safety performance functions (SPFs) for various functional road sections within Arizona; extended comparisons of alternative HSID methods based on SPFs by using real crash data; and recommendations for improving the identification ability of the current ALGSP model. These results are divided into the following sections:

- Literature review of HSID methods (chapter II): Through tracing the historical and conceptual development of various HSID techniques, the strengths and weaknesses associated with alternative approaches are assessed and appropriate directions of future research on HSID methods are explored and proposed. A detailed description of Bayesian approaches is also provided.

- Experimental design for evaluation of HSID methods and exploration of accident history (chapter III): In this experiment, "sites with promise" are known *a priori*. Real intersection crash data from six counties within Arizona are used to simulate crash frequency distributions at hypothetical sites. A range of real conditions is manipulated to quantify their effects. Various levels of confidences are explored.

False positives (labeling a safe site as high risk) and false negatives (labeling a high risk site as safe) are compared across the following three methods, say, simple ranking method, confidence interval method, and Empirical Bayesian (EB) method. Finally, the effect of crash history duration in these approaches is quantified.

- Safety performance functions for Arizona road segments (chapter IV): The SPFs for nine functional classifications of road sections in Arizona are created based on the crash data of Year 2000 provided by ADOT. Due to the existence of overdispersion of accidents, Negative Binomial models are utilized to develop these SPFs.

- Comparison of HSID methods based on real crash data of Arizona road segments (chapter V): On the basis of SPFs for Arizona road sections, five tests are implemented to evaluate the performances of the EB, accident reduction potential, accident frequency, and the accident rate methods. Two levels of confidences are explored under each test. In addition, the similarity of identification results of the alternative HSID methods is explored as well.

- HSID in current ALGSP model and recommended software changes (chapter VI): The algorithms for conducting HSID in the current ALGSP model are first reviewed and the software changes are then recommended. These recommendations include incorporating functional classification as an additional selection parameter, data interface improvements, accident history requirements, embed-ding the relationships between exposure and safety for various roadway functional classes, incorporation of the EB techniques to compute the expected crash count, incorporation of accident reduction potential as an additional weighting method, and incorporation of EB techniques to calculate the expected crash costs.

Based on both real and simulated data, the results in this report show significant advantages of the EB methods over other HSID methods across various confidence levels and different statistical tests. Specifically, the research found that:

- A higher percentage of truly high risk sites are identified as 'high risk.'
- A higher percentage of truly safe sites are identified as 'safe.'
- Overall misclassifications are reduced using a Bayesian approach compared to alternative methodologies.
- The Bayesian approach shows the best site consistency and method consistency among the alternative methodologies.

Although it is shown that incorporation of Bayesian techniques into the ALGSP will provide model users with more accurate prediction of hot spots, improvements are contingent upon accurate safety performance functions, which are currently unavailable in the ALGSP. Safety performance functions—the relationship between traffic volumes, road section lengths, and crashes—are provided in Appendix C for various roadway functional classifications in the state of Arizona. These safety performance functions enable the software enhancements needed to improve the ALGSP and accommodate Empirical Bayes' procedures.

CHAPTER I - INTRODUCTION

Hot spot identification is a critical contemporary transportation issue. The Intermodal Surface Transportation Efficiency Act (ISTEA) of 1991, along with the subsequent Transportation Efficiency Act for the 21st Century (TEA-21), brought HSID squarely into transportation planning activities. In particular, ISTEA requires each state to develop a work plan outlining strategies to implement Safety Management Systems (NCHRP, 2003). The objectives outlined in this management system require that several activities be undertaken by MPOs and/or DOTs:

1) The development and maintenance of a regional safety database so that safety investments can be evaluated regionally and forward in time.
2) The adoption of a defensible (i.e. state of practice) methodology for identifying safety deficiencies within a region.
3) A maintained and updated record of 'sites with promise,' including intersections, segments, interchanges, ramps, curves, etc.
4) A defensible methodology for evaluating the effectiveness of safety countermeasures.

Besides this mandate to spend safety funds wisely, there is professional pressure to conduct rigorous analyses and be held accountable for 'good number crunching.' Due to both public and professional pressures and the import associated with motor vehicle injuries and fatalities, transportation safety professionals desire analytical tools to cope with HSID.

As a powerful tool for local governments and jurisdictions, the current ALGSP model can be used to facilitate the selection of hazardous roadway locations in local jurisdictions and to aid in the evaluation of potential spot treatments of safety hazards. Its identification method is to simply rank the crash statistics in descending order and then the top ones are selected in terms of the allowed money budget. Due to a random "up" fluctuation in crash counts during the observation period, this simple ranking method is always subject to regression-to-the-mean bias, which decreases the identification accuracy. By contrast, Bayesian methods have been proposed for obviating this bias and have revealed themselves as superior for accurately identifying 'sites with promise' in considerable literature. However, much of the research was conducted on real crash data (where hazardous sites are not truly known) and comparisons across various scenarios have not been conducted. In addition, real crash data specific to Arizona regions have not been used to examine the performance of Bayesian analyses. By designing a special experiment which simulates various scenarios and using the real crash data from Arizona, this research effort evaluates and compares alternative HSID methods. All the results show the consistent superiority of Bayesian techniques for accurately identifying 'sites with promise.' This lays the solid foundation for the future incorporation of Bayesian approaches into the current ALGSP model. Moreover, safety performance functions of various classifications of road sections within Arizona are also provided in this report to facilitate the integration procedure.

This report is divided into five primary sections. In the second section of this report, Literature Review of HSID Methods, the historical and conceptual development of HSID procedures is reviewed chronologically, and for the convenience of understanding the more complicated computation procedures, the detailed description about two types of Bayesian techniques is provided.

In the third section, an experimental approach is taken to evaluate the performance of simple ranking, classical confidence intervals, and the EB techniques in terms of percent of false negatives and positives. Several practical empirical crash distributions from the state of Arizona are selected to represent a realistic range of ‘base’ crash data and several degrees of crash heterogeneity are examined in the simulation. The results demonstrate that the EB methods in general outperform the other two relatively conventional methods, especially in the low heterogeneity situations. In addition, the effect of crash history duration employed in the three HSID methods is also explored in this experiment. The moving average method is used to smooth the trend of the various duration data and to find the “knee” of the curve. Using 3 years of crash history data results in significant improvements in error rates for all three methods, and 3 through 6 years make up almost 90% of all the optimum duration.

The major focus of the fourth section is on developing the safety performances of road sections. Since design criteria and level of service vary according to the function of the highway facility**,** the safety performance function is created for each of nine types of road sections within Arizona. The data for modeling includes accident number, Annual Average Daily Traffic (AADT), and road section length. The graph showing the relationship among variables, the model form, and measures of goodness-of-fit are provided as well. It is expected that the input of the alternate SPFs would facilitate the procedure of incorporation of Bayesian techniques into the future ALGSP.

The fifth section contains a comprehensive comparison of identification performances of the EB, accident reduction potential, accident frequency, and the accident rate methods using crash data from Arizona and the SPFs developed in the previous section. Five evaluation tests including site consistency test, the method consistency test, total ranking differences test, false identification test, and false/true Poisson mean differences test are conducted. Both top 10% and top 5% locations (in terms of accident frequency) are considered as hot spots. The results across the nine types of road sections show the consistent advantage associated with the EB method, and disadvantage of the accident rate method while conducting HSID.

The final section provides recommended software changes to improve its ability to select truly hazardous locations from road network. The information of traffic volume is proposed to be incorporated in the software. As one of the factors significantly affecting road safety, it should be included in the safety performance function, which is the basis for conducting the EB analysis. Both the experimental design results based on the simulated data and the results of the evaluation tests based on Arizona crash data support the incorporation of Bayesian technique in the software. The accident reduction potential method is also recommended to be included as an additional weighting method. Finally, the recommendation of length of crash analysis period is provided.

CHAPTER II - LITERATURE REVIEW OF HSID METHODS

Identifying 'sites with promise,' also known as black spots, hot spots, or high-risk locations, has received considerable attention in the literature. This is not surprising, since there is public and professional pressure to allocate safety investment resources efficiently across the transportation system and to invest in sites that will yield safety benefits for relatively modest cost. In addition, US federal legislation requires the practice of remediating high risk locations.

It is intended that this identification stage act as an effective sieve that allows sites that do not require remedial action to pass through, while retaining sites that require remediation. This is difficult to accomplish, however, because an individual sites' safety performance (i.e. number of crashes) varies from year to year as a result of natural variation—causing two potential errors—false positives and false negatives. False positives are sites identified as needing remediation when in fact they are safe, while false negatives are sites identified as being safe when in fact they require remediation.

The following literature review comprehensively examines hot spot identification methods. It is intended to support ongoing work for the Arizona Department of Transportation aimed at improving the current ALGSP Model. It is the first of several steps toward ultimately improving the software that enables jurisdictions in the state of Arizona to identify sites for potential improvement, such as road segments, intersections, ramps, etc. This literature review is divided into two sections: the historical and conceptual development of hot-spot identification methods, and a detailed description of Bayesian techniques, the current state of the art.

HOT-SPOT IDENTIFICATION PROBLEM BACKGROUND

Due to the significant importance of identifying sites with promise, a large number of techniques have been employed to improve the detection accuracy. The historical and conceptual development of such procedures is reviewed chronologically in this section to help familiarize you with the hot-spot identification problem background.

The following notation will be useful in the discussions that follow:

X = observed accident count for a road section/site and period;
λ = expected accident count ($E\{X\}$) for the road section/site and period;
$E\{\lambda\}$ = mean of λ's for similar road sections/sites;
D = length of the road section;
Q = number of vehicles passing road section/site during period to which X pertains;
R = observed accident rate (e.g., crashes/vehicle-kilometer or crashes/million entering vehicles);
R_{EB} = accident rate estimated by the EB method;
$\bar{R}$ = average value of R for similar road sections and sites;
UCL_X = upper control limit for observed accident counts (X);
UCL_R = upper control limit for observed accident rate (R);
t = number of years of accident data to be analyzed;
α, β = parameters.

Perhaps the simplest way to identify sites with promise is by simply ranking them in descending order of their accident frequencies and/or accident rates. Although this method has the benefit of straightforwardness, the efficiency of identifying truly high-risk sites leaves considerable room for improvement. To overcome this deficiency, a substantial body of research has been devoted to providing more efficient and justifiable site identification techniques. For example, Norden et al. (1956) proposed a method to analyze accident data for highway sections based on statistical quality control techniques. Using an approximation of the Poisson distribution for crash counts, and 0.5 percent probability, they developed the equations for UCL_X and UCL_R used to identify critical thresholds. When X exceeds UCL_X (or R exceeds UCL_R), a site was identified as deviant with regard to safety. This approach drew much attention at that time, and some similar methods (with relatively minor differences) based on this procedure were proposed in subsequent years.

Researchers then began to ponder the issues of how many years (t) of accident data are necessary to conduct a defensible analysis. By finding that a 13-year average could be adequately estimated from 3 years of accident counts, May (1964) first provided the conclusion, "There is little to be gained by using a longer study period than three years." It is reasonable to use the current data instead of using old data that no longer reflect a current situation. However, considering that a sensible choice of t must depend on the magnitude of the average that is being estimated and on some knowledge of what makes past accident counts obsolete, this influential practice seems somewhat arbitrary.

Crash severity became the next issue of importance regarding HSID methods. Common sense suggested that a site with more severe crashes (all else being equal) should receive higher priority in remediation efforts. The safety index was first introduced by Tamburri and Smith (1970) and later incorporated into the practice of HSID. In essence, they said each road type (as examples, rural two-lane roads, urban freeways, etc.) had a characteristic mix (distribution) of accident severities among fatal, injury, and property damage only (PDO) crashes. On the basis of the accident severity and road type, accident costs were used to weight crashes. They also suggested that all crashes be expressed in terms of PDO equivalent accidents (for example a certain injury crash may be equivalent to 5 PDO crashes).

Deacon et al. (1975) considered the difference between identifying hot spots and sections and explored how long analysis sections should be conducted. They also presented an analysis of a sensible t, in comparison to that provided earlier by May (1964). Their conclusions suggested that a balance is sought between reliability of the crash data (longer being more reliable) and the need to detect adverse change quickly (shorter being more able to reveal adverse safety changes), and that a single t should be determined on this basis. They also recommend 9.5 as the weight for fatal and A-injury crashes, and 3.5 for B and C crashes when using a safety index.

Laughland et al. (1975) first described the ranking procedure using both the number and rate methods. The method proposed identifies hazardous locations when X exceeds some predetermined value UCL_X and R exceeds UCL_R. The claimed advantage of this procedure is that it excludes so-called hazardous locations identified as a result of *R* being as large as a result of low exposure.

Renshaw et al. (1980) argued that questions about the length of sections, duration of accident history, amount of traffic, and detection accuracy must all be considered jointly and that reliable detection is often not practical.

Hakkert and Mahalel (1978) first proposed that blackspots should be defined as those sites whose accident frequency was significantly higher than the expected at some prescribed level of significance. This point was then favored by McGuigan (1981; 1982), who put forward the concept of potential accident reduction (PAR), such as the difference between the observed accident counts and the expected number of similar sites. He stated, with some justification, that PAR should be a better basis on which to rank sites than annual accident totals (AAT), which tends to identify high flow sites which do not necessarily have the potential for accident reduction. This method is similar to the quality control method to some extent. The former represents the magnitude of the problem, that is, how many accidents can be avoided given the normal situation, and the latter represents how large the probability that the site is abnormal by using the given level of confidence.

Estimating $E\{\lambda\}$ using a multivariate model was suggested by Mahal et al. (1982). By using $E\{\lambda\}$ as the mean, they deemed a location as deviant if the probability of observing X or more accidents was smaller than some predetermined value.

Flak et al. (1982) recommended that crashes be categorized according to specific road conditions (weather, pavement material, etc.) and by accident type (turning, side-swipe, rear-end, etc.), and so forth. This concept differed from previous ones in that it seeks to identify deviant locations with regards to very specific conditions. Although appealing from an experimental design point of view, this concept is likely to produce sample sizes too small to detect significant differences for all but the largest of databases.

Hauer and Persaud (1984) proposed a concept of sieve efficiency in which the number of sites to be inspected and the expected numbers of correct positives, false positives, and false negatives serve as measures of performance. They examined the performance of various HSID techniques on the basis of performance measures that are easy to understand. They argued that the quality-control approach to HSID does not give the analyst clues about how well or how poorly the sieve is working. They also suggested that numerical methods are needed to free the procedure from reliance on the assumption that λ obeys the gamma distribution.

Regression-to-the-mean (RTM) bias associated with typical methods of site selection has been identified in the literature and some research dealing with RTM has been developed. Persaud and Hauer (1984) compared and evaluated the performance of an EB and a nonparametric method for debiasing before-and-after analyses. The results of several data sets show that the Bayesian methods in most cases yield better estimates than the other one. Wright et al. (1988) made a survey on the previous research dealing with the RTM effect. He examined the validity of assumptions associated with those methods, evaluated the robustness of the results based on the assumptions, and provided some suggestions for improving the quality of the results.

Mak et al. (1985) developed a procedure to conduct an automated analysis of hazardous locations. The procedure consists of (a) a mainframe computer program to identify and rank black-spots, (b) a microcomputer program to identify factors overrepresented in accident occurrence at these locations relative to the average for similar highways in the area, (c) a multidisciplinary approach to identify accident causative factors and to devise appropriate remedial measures, and (d) evaluation of remedial measures actually implemented. The procedure is based on accident rate (number of injury and fatal accidents per 100 million vehicle miles of travel).

Higle and Witkowski (1988) developed a Bayesian model for HSID using accident rate data rather than accident counts, which are shown to have identification criteria analogous to those used in the classical identification scheme. The comparisons between the Bayesian analysis and classical statistical analyses suggest that there is an appreciable difference among the various identification techniques in terms of HSID performance, and that some classically based statistical techniques are prone to err in the direction of excess false negatives.

Based on data from 145 intersections in Metropolitan Toronto, Hauer et al. (1988) provided Bayesian models to estimate the safety of signalized intersections on the basis of information about its traffic flow and accident history. For each of the 15 accident patterns (categorized by the movement of the vehicles), an equation is given to estimate the expected number of accidents and the variance using the relevant traffic flows. When data about past accidents are available, estimates based on traffic flow are revised with a simple equation. By applying these Bayesian models, one can estimate safety when both flows and accident history are given and, on this basis, judge whether an intersection is unusually hazardous. This method of estimation is also recommended for accident warrants in the Manual on Uniform Traffic Control Devices.

Through a simulation experiment, Higle and Hecht (1989) evaluated and compared various techniques for the identification of hazardous locations, based on classically and Bayesian statistical analyses respectively, in terms of their ability to identify hazardous locations correctly. The results reveal that the two classically based techniques suffer from some shortcomings, and the Bayesian method based on accident rate exhibits a tendency to perform well, producing lower numbers of both false negative and false positive errors.

By 1990 it was generally becoming accepted among academic circles that the Empirical Bayes approach to unsafety estimation was superior to previous HSID methods. The Bayesian approach generally makes use of two kinds of clues of an entity: its traits (such as traffic, geometry, age, or gender) and its historical crash record. It requires information about the mean and the variance of the unsafety in a “reference population” of similar entities. Obviously, this method suffers from several shortcomings: First, a very large reference population is required; second, the choice of reference population is to some extent arbitrary; and third, entities in the reference population usually cannot match the traits of the entity for which the unsafety is estimated. Hauer (1992) alleviated these shortcomings by offering the multivariate regression method for estimating the mean and the variance of unsafety in reference population. By describing its logical foundations and illustrating some numerical examples, Hauer shows how the multivariate method makes the Empirical Bayes method to unsafety estimation applicable to a wider range of circumstances and yields better estimates of unsafety than previous methods.

Persaud (1991) presented a method for estimating the underlying accident potential of Ontario road sections using accident and road related data. The comparative results indicate that the EB estimates are superior to those based on the accident count or the regression predictions by themselves, particularly for sections that might be of interest in a program to identify and treat unsafe road locations.

Brown et al. (1992) presented the convergence of HSID by police-reported data, by highway inventory, and by community reporting. Weighted injury frequencies per unit distance and weighted injury rates per 100 million vehicle-km are presented for all sites and for all numbered highway segments. Priority sites are then ranked considering injury frequencies and injury rates.

Hauer et al. (1993) explored the probabilistic properties of the process of identifying entities, such as drivers or intersections, for some form of remedial action when they experience *N* crashes within *D* units of time, the *N-D* "trigger." On the basis of the probability distribution of the "time-to-trigger," it is concluded that in road safety the problem of false positives is severe, and therefore entities identified on the basis of accident or conviction counts should be subjected to further safety diagnosis. Moreover, they found that the longer the *N-D* trigger is applied to a population, the less useful it becomes.

Tarko et al. (1996) presented a methodology of area-wide safety analyses to detect those areas (states, counties, townships, etc.) that should be considered for safety treatment. The method is implemented for Indiana at the county level and uses regression models to estimate the normal number of crashes in individual counties. The counties are priority ranked using the combined criterion including both the above-norm number of cashes and the confidence level. This combined criterion helps select counties where the excessive number of crashes is not caused solely by the randomness of the process. This application differs from previous applications in that the HSID was conducted at the planning or county level, instead of at the intersection or road segment level.

Stokes and Mutabazi (1996) traced the evolution of the formulas used in the rate-quality control method from their origin in the late 1950s to their present form, and they also presented and discussed the derivation of the basic formulas used in the method. It is suggested that, contrary to assertions in the literature, the accuracy of the equations used in the rate-quality method is not proved by eliminating the normal approximation correction factor from the original equations and the need for a correction factor is particularly apparent at higher probability levels.

On the basis of the review of previous procedures for black-spots identification, Hauer (1996) made an attempt to create some order in the thinking and made some suggestions to improve identification. In comparison with the stage of identification, he pointed out that the stage of site safety diagnosis and remediation is somewhat underdeveloped.

Persaud et al. (1999) put forward a similar concept to potential accident reduction, such as potential-for-safety-improvement (PSI). For the sake of correcting for the RTM bias, he replaced the observed accident number with the long-term mean of accident counts in the PAR previously stated.

Davis and Yang (2001) made use of Hierarchical Bayes methods combined with an induced exposure model to identify intersections where the crash risk for a given driver subgroup is relatively higher than that for some other groups. They carried out the necessary computations using Gibbs sampling, producing point and interval estimates of relative crash risk for the specified driver group at each site in a sample. The methods can also be extended to identify hazardous locations for a specified accident type. This method of HSID requires sophisticated modeling skill and software, and is currently beyond the level of most DOT staff expertise.

Kononov et al. (2002) presented the direct diagnostics method to conduct HSID and develop appropriate countermeasures. The underlying principle is that a site should be identified for further examination if there is overrepresentation of specific accidents relative to the similar sites.

With empirical Bayes gradually becoming the standard and staple of professional practice, Hauer et al. (2002) presented a tutorial on safety estimation using the EB method. This tutorial contains comprehensive illustration of using the EB procedures and can be viewed as the bridge between theory and practice for the EB application.

The above mentioned research represents only a small portion of the extensive past and current HSID research. In summary, the large body of techniques for HSID generally includes simple ranking of accident frequencies and/or accident rates, rate-quality control methods, site identification using the notion of a safety index, number-and-rate methods, accident pattern recognition method, and various applications of Bayesian approaches on both crash frequencies and crash rates. In comparison with other techniques, Bayesian techniques have been shown to offer improved ability to identify black-spots by accounting for both history and expected crashes for similar sites, which can obviate the "regression-to-the-mean" problem that simpler methods fail to correct.

This literature review summary clearly indicates that opportunities exist for possible enhancements leading to improved HSID within the recently released ALGSP model, which currently performs a simple ranking based on accident frequencies. However, as one might expect, the incorporation of Bayesian methods will increase the data collection burden: additional information about site crash histories and reference populations will need to be collected. The following section is devoted to describing the Bayesian techniques in greater detail.

BAYESIAN TECHNIQUES TO IDENTIFY HAZARDOUS LOCATIONS

An underlying characteristic of crash occurrence is the random fluctuation from year to year of crash counts under constant and unchanging traffic, weather, and roadside conditions (which of course in reality does not occur). This characteristic significantly reduces the ability to detect truly hazardous locations in the sense that a crash site may appear to represent a relatively high risk in a given year when in fact the site's underlying, inherent risk level is average or low (Hauer, 1997). A site that reveals a high observed risk in one year is on average followed by a crash count in the subsequent year that is closer to the mean—a phenomenon known as regression to the mean. However, it was shown in the previous section that Bayesian approaches, by utilizing two kinds of clues of an entity (its traits and its historical accident record), involve corrections for *RTM* and can improve significantly the efficiency of site identification. Incorporation of such techniques into the ALGSP model will offer improvements in the performance of HSID. Unfortunately, in contrast to other approaches, which are relatively straightforward, the Bayesian techniques require a greater quantity of information associated with locations inspected and also involve relatively more complicated computations – albeit trivial for a computer.

Noting that the large portion of this research is to test the performances of various HSID methods (including the somewhat typical methods and Bayesian techniques), this section describes in detail the analytical aspects of various Bayesian techniques generally accepted as 'state of the art.' The research reviews are divided into two groups: Bayesian techniques based on accident frequencies and Bayesian techniques based on accident rates.

Bayesian Techniques Based on Accident Frequencies

To alleviate the RTM bias associated with other site identification techniques, Hauer et al. (1984; 1988; 1992) discussed numerous aspects of HSID to derive what is known as the EB method. EB methods differ technically from Bayes' methods in that the former relies on empirical data as "subjective" information while the latter relies on truly subjective information (e.g. expert opinions, judgment, etc.).

The EB method rests on the following logic. Two assumptions are first needed, which can be traced back to those of Morin (1967) and Norden et al. (1956):

> Assumption 1: At one given location, accident occurrence obeys the Poisson probability law. That is, $P\langle x|\lambda\rangle$ denotes the probability of recording *x* accidents on a site where their expected number is λ, where $P\langle x|\lambda\rangle = \lambda^x e^{-\lambda} / x!$. (1)
>
> Assumption 2: The probability distribution of the λ of the population of sites is gamma distributed, where g (λ) is denoted as the gamma probability density function.

Estimation of the long term safety of an entity is obtained through using both kinds of clues, that is, the traits such as gender, age, traffic, or geometry of an entity and the historical accident record of the entity. If the count of crashes (x) obeys the Poisson probability law and the distribution of the λ's in the reference population is approximated by a Gamma probability density function, a good estimator of the λ for a specific entity is:

$$\alpha E\{\lambda\} + (1-\alpha)x, \text{ with } \alpha = E\{\lambda\}/[E\{\lambda\} + VAR\{\lambda\}] \quad (2)$$

From the above equation, we know estimates of E $\{\lambda\}$ and VAR $\{\lambda\}$ which pertain to the λ's of the reference population are needed. There are two methods to estimate the E $\{\lambda\}$ and VAR $\{\lambda\}$. One of them is the method of sample moments, the other is the multivariate regression method.

To describe the method of sample moments, let us first consider a reference population of n entities of which $n(x)$ entities have recorded X=0, 1, 2,… accidents during a specified period. With this notation, the sample mean and the sample variance are, respectively:

$$\mu = \sum xn(x) / \sum n(x) \quad (3)$$

$$s^2 = [\sum (x-\mu)^2 n(x)] / \sum n(x) \quad (4)$$

In the method of sample moments, the estimators of E $\{\lambda\}$ and VAR $\{\lambda\}$ are equal to μ and s^2- μ respectively. The larger is the reference population. These estimates are more accurate.

The primary attraction of the method is that its validity rests on a single assumption: that if λ_i remained constant, the occurrence of accidents would be well described by the Poisson probability law. However, there remain two practical difficulties: (1) It is rare that a sufficiently large data set can be found to allow for adequately accurate estimation of E $\{\lambda\}$ and VAR $\{\lambda\}$; (2) Even with very large data sets, one cannot find adequate reference populations when entities are described by several traits (e.g. geometric conditions, etc.). In order to obviate these difficulties, Hauer (1992) provided the multivariate regression method. With this correction, a multivariate model is fitted to accident data to estimate the E $\{\lambda\}$ as a function of independent variables, and the residuals (i.e., the difference between an accident count on some specific entity that served as "datum" for model fitting and the estimate E $\{\lambda\}$ calculated from the fitted model equation) are viewed as coming from a family of compound Poisson distributions:

$$VAR\{x\} = VAR\{\lambda\} + E\{\lambda\} \quad (5)$$

The E $\{\lambda\}$ of the reference population is estimated using the model equation; VAR$\{x\}$ is estimated using the squared residuals. Therefore, based on equation (5), the difference [squared residual – estimate of E $\{\lambda\}$] can be used to estimate VAR $\{\lambda\}$ for the imaginary reference population to which this datum point belongs.

As mentioned previously, it is easy to note that the primary difference between the method of sample moments and multivariate regression method is that the estimates of E $\{\lambda\}$ and VAR $\{\lambda\}$ are obtained using different analytical procedures. The method of sample moments is straightforward, while the latter one yields more precise results. Once the estimates of E $\{\lambda\}$ and VAR $\{\lambda\}$ are obtained, the expected safety of an entity is obtained using Equation 2. However, the truly hazardous locations cannot be screened based solely on the long term safety associated with each entity, a model of the entire distribution function of $\lambda|X$ is required.

On the basis of the assumptions stated previously, the probability that a site selected randomly has *x* accidents is approximated by the negative binomial (NB) probability distribution. Thus, the parameters of g (λ) are estimated using EB logic according to the following sequence of steps:

Step 1: The sample mean and variance is computed across sites. The notation *n(x)* is used to denote the number of sites that had x crashes. The estimated mean and variance are computed using:

$$\mu = \sum xn(x) / \sum n(x) \tag{6}$$

$$s^2 = [\sum (x-\mu)^2 n(x)] / \sum n(x) \tag{7}$$

Step 2: The EB weighting parameters α and β are then obtained using:

$$\alpha = \mu / (s^2 - \mu) \tag{8}$$

$$\beta = \mu * \alpha \tag{9}$$

Step 3: With the two weighting parameters obtained, the parameters of the gamma distribution are obtained such that:

$$g(\lambda) = \alpha^{\beta} \lambda^{\beta-1} e^{-\alpha\lambda} / \Gamma(\beta) \ . \tag{10}$$

The subpopulation of sites that had *x* accidents also follows a gamma probability distribution and its gamma probability density function is given by:

$$g(\lambda|x) = (1+\alpha)^{\beta+x} \lambda^{\beta+x-1} e^{-(1+\alpha)\lambda} / \Gamma(\beta + x) \, . \tag{11}$$

With the probability density functions defined, the selection of hazardous locations is now straightforward. Suppose that λ^* is the "acceptable" upper limit of accident counts, then a site *i* is identified as hazardous if the probability that λ exceeds λ^* is relatively small. Specifically, if:

$$P\left(\lambda > \lambda^* |x\right) > \delta \tag{12}$$

Where δ is the tolerance level that is contingent upon the choice of safety specialists (i.e. level of acceptable risk) and takes into account conditions in the local jurisdiction, then site *i* is identified as a truly hazardous location.

Bayesian Techniques Based on Accident Rates

In contrast to earlier papers regarding EB techniques, which were concerned with predicting the number of crashes that will occur at a particular location, Higle and Witkowski (1988) investigated using Bayesian analysis of crashes for the identification of hazardous locations based on accident rates and not frequencies. It should be noted that use of rates has been strongly discouraged by some researchers, and a growing body of literature discourages the use of rates (Hauer, 1997). Due to the similar assumptions and procedures, the research can be viewed as a complement to the previous research relying on EB approaches. Using empirical comparisons of performance between Bayesian and classical statistical analyses, Higle et al. found that there is an appreciable difference among the various identification techniques, and that some classically based statistical techniques may be prone to err in the direction of excessive false negatives.

Higle and Witkowski divided the Bayesian analysis into two steps. In the first step, crash histories are aggregated across a number of sites to get a gross estimation of the probability distribution of the accident rates across the region. In the second step, the regional distribution and the accident history at a particular site are used to obtain a refined estimation of the probability distribution associated with the accident rate at that particular site.

In performing the analysis, Higle and Witkowski made two assumptions that are similar to those made by previous researchers:

Assumption 1: At any given location, when the accident rate is known (i.e., if $\widetilde{R}_i = R$, note that $\widetilde{R}_i$ is treated as a random variable), the actual number of accidents follows a Poisson distribution with expected value $R(DQ)_i$. That is:

$$P\{X_i = X \big| \widetilde{R}_i = R(DQ)_i\} = \frac{\left(R(DQ)_i\right)^X}{X!} e^{-R(DQ)_i} \quad (13)$$

Assumption 2: The probability distribution of the regional accident rate, $f_R(R)$, is the gamma distribution. That is:

$$f_R(R) = \frac{\beta^\alpha}{\Gamma(\alpha)} R^{\alpha-1} e^{-\beta R} \quad (14)$$

Higle and Witkowski recommended that for each computation, it may be preferable to use the MME (method of moments estimates) values rather than the MLE (maximum likelihood estimates) values of α and β. Within the framework of Bayesian analysis, the site-specific parameters are: $\alpha_i = \alpha + X_i$, $\beta_i = \beta + (DQ)_i$. Based on α_i and β_i, the site-specific probability density functions were then obtained. The steps to identify the truly hazardous locations are shown as follows:

Step 1: Estimate the sample mean and variance of the observed accident rates of the population of locations:

$$\mu = \frac{1}{m} \sum_{i=1}^{m} \frac{X_i}{(DQ)_i} \quad (15)$$

$$s^2 = \frac{1}{m-1} \sum_{i-1}^{m} \left(\frac{X_i}{(DQ)_i} - \mu \right)^2 \quad (16)$$

Step 2: Estimate parameters α and β, where:

$$\beta = \mu / s^2 \quad (17)$$

$$\alpha = \mu * \beta \quad (18)$$

With the two parameters,

$$f_R(R) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} R^{\alpha-1} e^{-\beta R} \tag{19}$$

Step 3: Obtain $f_i\langle R \mid X_i,(DQ)_i \rangle$.

The subpopulation of sites that had X accidents also follows gamma distribution and its gamma probability density function is as follows:

$$f_i\langle R \mid X_i,(DQ)_i \rangle = \frac{\beta_i^{\alpha_i}}{\Gamma(\alpha_i)} R^{\alpha_i-1} e^{-\beta_i R}. \tag{20}$$

Where: $\alpha_i = \alpha + X_i$ (21)

$\beta_i = \beta + (DQ)_i$ (22)

With these probability density functions, the selection of hazardous locations is now straightforward. Suppose that λ^* is the "acceptable" upper limit accident counts, then a site i can be deemed as hazardous if the probability that λ exceeds λ^* is relatively significant. Say, if:

$$P(\lambda > \lambda^* | x) > \delta, \tag{23}$$

Where δ is the tolerance level which is contingent upon the choice of safety specialists and the actual situation of local jurisdiction. Sites above the critical threshold are then identified as truly hazardous locations.

To summarize, Bayesian techniques, by accounting for both crash history and expected crashes for similar sites, have been shown to offer improved ability to identify truly hazardous locations. The next section quantifies the differences between Bayesian techniques and other typical approaches.

CHAPTER III - EXPERIMENT DESIGN FOR EVALUATION OF HSID METHODS AND EXPLORATION OF ACCIDENT HISTORY

On the basis of the previous literature review for HSID methods, Bayesian methods revealed themselves as superior for accurately identifying sites with promise. However, much of the research was conducted on real crash data (where hazardous sites are not truly known) and comparisons across various Bayesian methods have not been conducted. This chapter is focused on examining the performances of the EB and alternative typical methods within various environments and exploring the best duration of accident history, which causes minimum false identifications.

The chapter is divided into sections as follows. Section 1, "Experiment for Evaluating HSID Method Performance," discusses the steps of an experiment designed to evaluate the performance of HSID methods. Section 2, "Experiment for Optimizing Duration of Crash History" presents the steps with regard to the optimum duration of before-period crash data. Both real data and simulated crash data are utilized in the experiments. The real data were obtained from current ALGSP users in Arizona. Simulated data correspond with a designed experiment that varies such as degree (or percentage) of difference between "correctable" and "average" sites, variability in the data, and different crash distributions. The final section provides the conclusions and recommendations that arise from the two experiments performed to evaluated HSID methods for use in the ALGSP, and translate the analytical results into practical recommendations.

EXPERIMENT FOR EVALUATING HSID METHOD PERFORMANCE

The main objective of this first experiment is to quantify and assess the predictive performance of various HSID methods, such as the simple ranking method, the method based on classical statistical confidence intervals, and the EB method, in order to identify the best one for inclusion in the ALGSP model. Of course there are many aspects of the simulation experiment that desire careful attention, such as sample sizes, nature of crash data, reliability of tests, etc. Prior to describing the detailed aspects of the experiment, HSID methods are first reviewed.

Hot Spot Identification Methods

A site (series of sites, etc.) may experience relatively high numbers of crashes due to: 1) an underlying safety problem; or 2) a random "up" fluctuation in crash counts during the observation period. Simply observing unusually high crash counts does not indicate which of the two conditions prevails at the site. It is possible to articulate the objective of HSID as follows:

> The objective of hot spot identification is to identify transportation system locations (road segments, intersections, interchanges, ramps, etc.) that possess underlying correctable safety problems, and whose effect will be revealed through elevated crash frequencies relative to similar locations.

Two aspects of the previous statement are noteworthy. First, it is possible to have truly unsafe sites that do not reveal elevated crash frequencies—these are termed ‘false negatives.’ It is also possible to have elevated crash frequencies, which do not result from underlying safety problems—these are termed ‘false positives.’ False positives, if acted upon, lead to investment of public funds with few safety benefits. False negatives lead to missed opportunities for effective safety investments. As one might expect, correct determinations include identifying a safe site as “safe” and an unsafe site as “high risk.” When considering the seriousness of errors (false positives and false negatives) with respect to safety management, one generally concludes that false negatives are the least desirable result, since a jurisdiction will fail to make wise investments and reduce fatalities, injuries (serious and minor), and property damage crashes.

For evaluation purposes, an HSID method is sought that produces the smallest proportion of false negatives and false positives. Hence, the percentages of false negatives, false positives, and overall misidentifications (false positives plus false negatives) are used to compare the performance of three commonly implemented HSID methods: 1) simple ranking of sites; 2) classically based confidence intervals; and 3) the EB methods. These three methods are now described.

The simple ranking method (denoted SR in experiments) is the most straightforward HSID method. Applying this method, a set of locations (e.g. all 4-lane signalized intersections in a jurisdiction) is ranked in descending order of crash frequencies (or counts, X), and then the top sites are identified as high-risk locations for further inspections. Typically, resources are invested to improve correctable sites from the top down until allocated funds are expended. This method, for example, is one analysis option available in the current version of the ALGSP model.

A second method for HSID is based on classical statistical confidence intervals (denoted CI in experiments) (1975). Location i is identified as unsafe if the observed accident count X_i exceeds the observed average of counts of comparison (similar) locations, μ, with level of confidence equal to δ, that is, $X_i > \mu + K_\delta S$, where S is denoted as the standard deviation of the comparison locations, and K_δ is the corresponding critical values. In practice δ is typically 0.90, 0.95, or 0.99, and depends upon the actual situation and considerations such as the number of sites, amount of safety investment resources, etc. These values serve as approximations, since they are borrowed from the normal distribution function and thus have no special meaning in terms of the distribution of true accident counts, which typically follow Poisson or negative binomial distributions. This method is commonly used in the sense that it is inferred from the classical statistics and can be performed conveniently.

Critical in the SR and CI methods is the notion of ‘comparison sites.’ Comparison sites are used to obtain an estimate of ‘expected crashes’ for similar sites. When sites are ranked using simple ranking, it is assumed that sites that are being ranked with similar geometric and traffic conditions. Geometrics and traffic play a significant role in crash potential and thus must be treated carefully. Often jurisdictions will group to the extent possible ‘similar’ sites together in the ranking; however, it is often the case that sites with different geometric and traffic conditions (i.e. exposure) are compared in the ranking method. In the confidence interval method, it is assumed that the group or set of comparison sites are similar to the site being compared. Critical to the outcome of any HSID method is the level of sophistication employed to identify comparison sites.

For the EB technique, the former section has given a detailed description. It is noteworthy that only the EB based on accident counts would be used herein. Equation 24 is followed to compute the long-term accidents of each site:

$$\lambda_i = \alpha E\{\lambda\} + (1-\alpha)x_i, \text{ with } \alpha = E\{\lambda\}/[E\{\lambda\} + VAR\{\lambda\}] \quad (24)$$

The weight parameter α is obtained by using the method of sample moments in which the estimators of $E\{\lambda\}$ and $VAR\{\lambda\}$ are equal to μ and s^2 respectively (μ denotes the sample mean and s^2 denotes the sample variance). From the above expressions, it is known that the second of the two clues, crash history, significantly affects the estimate of λ, since longer crash histories tend to be more stable (in crashes per year) than shorter crash histories. Thus, different historical accident records yield different estimators of $E\{\lambda\}$ and $VAR\{\lambda\}$, and subsequently different identification error rates (false positives and false negatives). Similarly, these different identification error rates are also supposed to be obtained under simple ranking and confidence analysis methods when utilizing various historical accident records. Because of its importance, the optimum crash history is examined in an experiment described in chapter 2 of this report.

Ground Rules for Simulation Experiment

To accomplish the evaluation of HSID methods, a simulation experiment was designed to test a variety of conditions. The simulation experiment consists of the following specific steps:

1) Generate mean crash frequencies from real data. Crash datasets from Arizona (and users of the ALGSP) which represent a range of in-situ crash conditions (i.e., intersections, road segments, etc.) are first obtained. These data are used to determine various shapes of distributions of crash means (λ’s). Gamma distributions are fit to the observed data to reflect heterogeneity in site crash means. These gamma distributed means are meant to reflect TRUTH, that is, the true state of underlying safety at various locations on a transportation network (note that in practice we do not know TRUTH—and herein lies the power of simulation). The gamma distributed means are denoted true Poisson means (TPMs), and represent the means of crashes across sites.
2) From TPMs, generate random Poisson samples. Thirty independent random numbers for each simulated site are generated. For each of the 1000 sites, the TPM is used to generate 30 crash counts that represent OBSERVED data for 30 different observation periods, which are assumed to represent years in the analysis.

3) Evaluate HSID performance. By knowing the true state of safety for sites (the TPMs), and having observed data (the randomly generated Poisson numbers), the performance of HSID methods can be tested. The following steps are used to set up the evaluation:
   a) SR, CI, and EB are applied in separate simulation runs to rank sites for improvement. These are applied by columns (a single observation period, which represents what an analyst might see in reality).
   b) For the Bayesian runs, it is assumed that rows (data across observation periods for the same site) can also be used to represent the comparison group in order to calculate E(x) and VAR(x). This implies that the analyst has accommodated for covariates and is able to estimate an expected value for a site that accounts for things such as exposure, geometrics, etc.
   c) For the various hot spot thresholds, false positives, false negatives, and total misidentifications in percent are computed. The percent of false positives will always be larger than the percent of false negatives since the latter represent hazardous sites that get identified as non-hazardous, which is much larger candidate pool of sites than hazardous sites. Recall that false positives are safe sites that are identified as hazardous, a relatively small pool of sites.
4) Evaluate effect of length of history. In the SR, CI, and EB methods the analyst must decide how long a history to use for calculations. In this experiment the effect of various accident histories (1 year until 10 years of data) on performance are evaluated based on the corresponding identification rate.
5) Make practical recommendations. The results of the previous steps are discussed and translated into practical recommendations for improving the ALGSP software.

Various aspects of the simulation experiment previously listed need to be discussed, as the quality and design of the simulated data directly impacts the quality and generalizability of the analysis results.

Generating Mean Crash Frequencies from Real Data

To support the development of simulated crash data, 6 years (January 1995 through December 2000) of crash counts from intersections in Apache, Gila, Graham, La Paz, Pima, and Santa Cruz counties in the state of Arizona are used. These data and their corresponding cumulative distributions are shown in Appendix A. Three types of characteristically different underlying cumulative distributions of TPMs were observed in the Arizona crash data: an exponential shape (denoted E), a linear shape (denoted L), and an s-shape (denoted S). In addition, two levels of heterogeneity in crash counts were observed: low heterogeneity (denoted 1) where the range in observed crash counts is less than 20 crashes, and high heterogeneity (denoted 2) where the range is in excess of 50 crashes.

Recall that the empirical distributions will be used to generate TRUTH, or the means of Poisson counts of sites with varying underlying means. In this simulation study these are denoted as TPMs. Since the data represent the true underlying safety of a site, crash counts are Poisson distributed at an individual site, and the statistic is the mean.

The cumulative distributions used to represent the TPMs are labeled as E1, E2, L1, L2, S1, and S2, respectively. For example, E2 represents an exponential shaped distribution with high heterogeneity in TPMs. These six data sets were selected from various jurisdictions within Arizona to try to represent the range of underlying characteristics related to true accident count distributions, with the intent of making the results gained from this experiment applicable across a variety of typical situations.

As stated previously, the observed data are used to inform the simulation of the TPMs. In this experiment three reasonable assumptions are required to establish the foundation for a successful simulation of crash data:

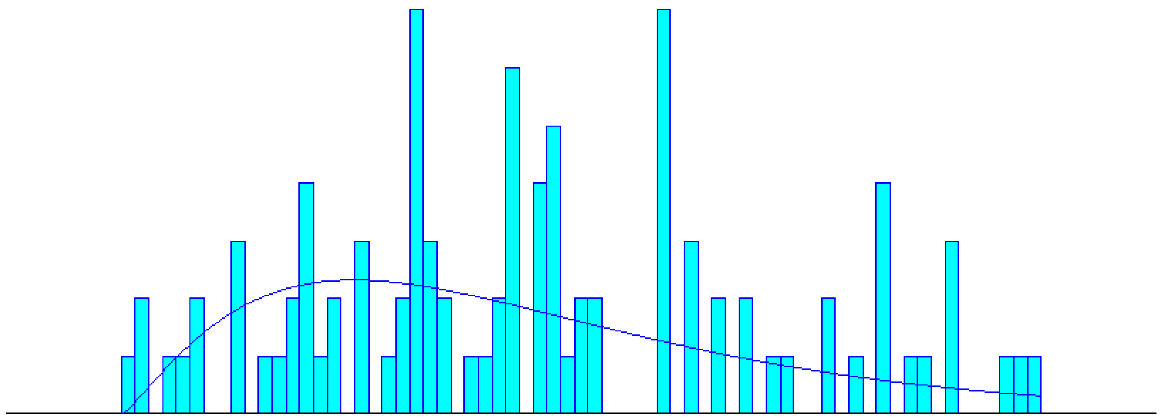
> *Assumption 1*: The empirical cumulative distributions shown in Figures 12 through 17 (see Appendix A) represent the TPMs of the underlying crash process—thus the true safety of all sites in the collection of sites is known. These data in reality are unknowable, since it is not known *a priori* which sites are "hazardous."
> *Assumption 2*: Theoretical distribution of these TPMs of the population of sites follows gamma distribution, and the probability that a site selected randomly has a given number of accidents is approximated by the negative binomial distribution.
> *Assumption 3*: The TPMs provide the basis for generating observed crash count data. Thus, for example, the median ranked site in Figure 12 (E1) that has an underlying Poisson mean of around six crashes (per observation period) is used to randomly generate a crash outcome, which could be 0, 1, 2, 3, ….etc. in any given observation period.

The result of assumptions 1 and 2 is that for each simulated site the underlying TPM (expected crash count) is known, which is then used to randomly generate the observed crash count.

Generation of Random Poisson Samples from TPMs

The empirical cumulative TPMs shown in Figures 12~17 (see Appendix A) represent the data required to meet Assumption 1 discussed previously. Using these data, observed crash counts are generated to represent observed data for a given observation period. However, due to the relatively small observed sample sizes (less than 200 sites in all six datasets) and the corresponding dispersion of crash counts, no sites would be identified as hazardous in some cases when using the three HSID methods stated previously. For example, if the top 1% of sites are identified as high risk ($\delta = 0.99$), all the sites in the datasets labeled as L1, S1, and L2 would be identified as safe when utilizing the classical confidence interval method and Bayesian method, thus leading to zero false negatives in these scenarios and damaging the regulars of results to some degree.

To solve this problem and provide sufficient sample sizes for statistical comparisons, theoretical distributions of TPMs are fitted to the six datasets. Then the sample sizes are enlarged by randomly generating the required number of sites under these gamma distributions (site specific crash means are gamma distributed whereas within-site crashes are Poisson distributed). In this experiment, 1,000 sites are simulated. Fitting specific gamma distributions to a given sequence of data can be implemented through various software packages, such as MINITAB, SAS 8.1 (1998), and Arena 7.0 (Kelton, 2003). Herein the Arena 7.0 is employed. Within the context of Arena, the curve fitting is based on the use of maximum likelihood estimators, and the quality of a curve fit is based primarily on the square error criterion. The fitting of probability density function (PDF) of a gamma distribution to the observed data is based on the histogram plot of these data. The distribution summary report also presents the expression of given probability density function, the corresponding p-value of Chi Square test and square error, etc. Figure 1 shows one example of fitting gamma distribution to the dataset. To show the fitting effect, the corresponding theoretical cumulative distribution function (CDF) is also plotted in the same graph of empirical *CDF* (Figure 2 shows the distribution of dataset E1). The figures show that the gamma distribution fits well to the observed data. The summary of all six fittings is shown in Table 1.



Distribution Summary	
Distribution:	Gamma
Expression:	3.5 + GAMM (13.4, 2.27)
Square Error:	0.020052
Chi Square Test Results	
Number of intervals	= 8
Degrees of freedom	= 5
Test Statistic	= 8.2
Corresponding p-value	= 0.16
Data Summary	
Number of Data Points	= 94
Min Data Value	= 4
Max Data Value	= 70
Sample Mean	= 33.8
Sample Std Dev	= 16.7
Histogram Summary	
Histogram Range	= 3.5 to 70.5
Number of Intervals	= 67

Figure 1: Observed and Fitted *PDF* of E1 Crash Data and Fit Summary Statistics

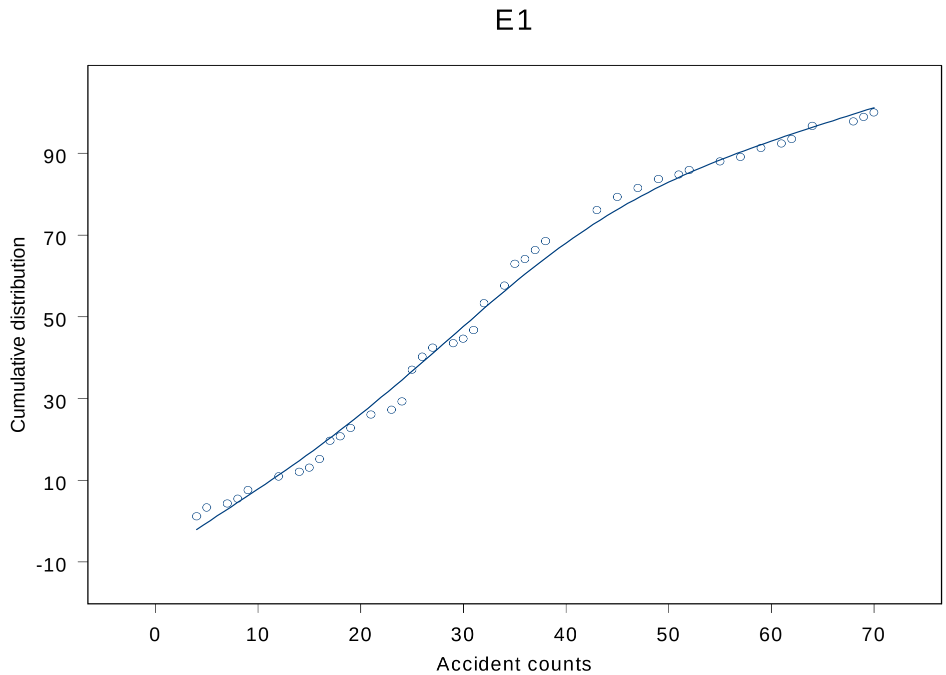


Figure 2: Fitted and Empirical CDF of E1

Table 1: Summary of Gamma Fittings of Six Datasets

Data set	Fitting Expression	Square Error	Test Statistic	*p*-value
E_1	0.5+Gamm(3.79,1.75)	0.022344	26	<0.005
E_2	1.5+Gamm(15.9,1.7)	0.011836	13.4	0.0385
L_1	0.5+Gamm(4.31,1.71)	0.038173	11.1	0.0119
L_2	3.5+Gamm(13.4,2.27)	0.020052	8.2	0.16
S_1	0.5+Gamm(2,4.3)	0.014903	33.5	<0.005
S_2	0.5+Gamm(9.06,2.57)	0.013211	23	<0.005

Note: E—Exponential shape; L—Linear shape; S—Sigmoidal shape; 1— Low heterogeneity of crash counts; 2— High heterogeneity of crash counts.

After TPMs have been simulated (the crash means across sites which reflect the true and typically unknown state of nature), the next step is to generate observed crash counts for the sites. These counts will represent the observed crash counts across observation periods for a particular site (where its true safety is known). It is well-established that crash counts fluctuate across observation periods as a result of the randomness inherent in the underlying crash process and is well approximated by a Poisson process. In other words, the count of crashes changes from one period to another even if driver demography, traffic flow, road, weather, and the like remained unchanged. To represent this natural fluctuation, a random sample of 30 observation periods (which could be months, years, etc.) associated with each location is randomly generated using a random number generator and underlying TPMs defined by the fitted distributions in Figure 12~17 (see Appendix A). A small snapshot of the data obtained from this simulation is shown in Table 2.

Table 2: Simulated Data for 30 Sites and 16 Observation Periods

SITE	TPM	PERIOD															
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	4	5	1	4	1	2	7	4	3	4	4	2	1	1	5	5	6
2	8	5	9	8	6	8	4	9	9	5	4	8	8	9	9	13	8
3	8	12	7	10	5	5	7	11	8	8	8	11	6	6	7	8	7
4	9	12	9	10	16	8	12	7	9	11	8	10	8	16	11	6	8
5	9	10	13	12	8	9	6	12	10	9	9	4	5	12	11	11	4
6	10	15	4	6	10	4	17	6	11	12	7	10	10	15	6	17	10
7	10	8	5	10	8	13	10	11	7	12	10	8	9	9	6	9	10
8	10	7	8	11	14	10	12	7	11	12	11	12	13	7	7	7	11
9	12	13	17	8	14	12	10	16	10	7	15	17	9	11	15	14	15
10	12	10	9	13	13	6	12	18	11	15	12	12	12	13	12	13	9
11	12	9	10	10	14	15	12	7	14	6	12	11	19	9	17	10	18
12	12	11	14	14	9	16	7	15	3	10	13	9	11	7	2	12	14
13	12	15	15	16	13	8	12	13	16	16	12	15	11	15	12	14	9
14	12	14	10	10	11	15	15	12	13	14	15	13	14	11	13	17	19
15	12	11	12	12	8	12	13	12	7	9	11	9	9	9	12	4	9
16	13	8	17	13	8	12	11	17	15	16	13	12	15	16	12	14	19
17	13	9	13	16	16	11	8	6	18	12	8	7	11	12	12	17	15
18	13	10	18	15	16	10	15	10	16	17	10	6	8	8	10	13	6
19	13	14	13	17	11	6	11	18	15	11	17	16	19	13	11	15	14
20	13	7	4	13	11	12	10	17	19	6	7	12	15	7	15	14	12
21	14	16	17	12	18	13	17	12	11	7	13	15	10	18	14	17	19
22	15	15	18	21	15	15	14	13	21	14	13	20	13	12	19	16	16
23	15	11	13	16	12	12	16	10	16	19	20	21	16	13	19	11	16
24	15	9	16	16	11	14	12	15	18	11	16	14	29	11	12	19	14
25	16	18	12	15	9	19	18	14	11	19	15	18	14	18	18	14	20
26	17	22	10	19	12	15	19	18	10	11	17	20	16	15	11	10	15
27	18	14	21	9	19	16	17	19	18	18	14	16	28	19	18	19	10
28	18	8	20	19	5	16	18	20	28	16	17	19	14	15	14	18	15
29	19	26	19	18	21	17	29	12	22	25	15	23	11	19	20	15	24
30	20	22	18	23	21	23	19	26	22	16	20	19	15	14	19	13	15

Note: SITE=number of site, e.g. intersection, road segment, etc.; TPM=true underlying safety of site or Poisson mean; SIMULATED DATA=observed crash count in observation period; Shaded cells represent 'truly hazardous' locations (sites 19 and 20).

Table 2 shows 16 simulated observations periods for 30 sites with TPMs given in the second column from the left. For example, the two sites with 19 or more crashes per observation period may be identified *a priori* as hazardous since the TPMs reflect the true underlying state of nature. The two sites in the shaded cells are hot spots whereas the 18 sites above the shaded area are 'safe.' In any given observation period such as observation period 5, the observed number of truly hazardous sites that recorded 19 or more crashes was two sites out of 20, where one was a truly hazardous site (site 20) and one was not (site 16, a false positive). In observation period 5 there was also a false negative, since truly hazardous site 19 revealed only 17 crashes.

So, by simulating large numbers of observation periods (30) characterized by different TPM cumulative distribution shapes, a large number of sites (1000) for each of the six observed crash distributions, the number of false negatives and positives (the sum total of the two is called false identifications) can be counted as a consequence of the three different HSID methods described previously.

Performance Evaluation Results for HSID Methods

Given knowledge of three HSID methods, the ground rules for the simulation experiment, and an explanation of how data were simulated, the three HSID methods were applied to the simulated data to evaluate their relative effectiveness at identifying hot spots.

Establishing fair comparisons among the different HSID methods is paramount. In order to objectively compare the performances of the HSID methods described previously, equivalent evaluation criteria must be used. One consideration in this regard is the use of δ, or cutoff level used to establish hazardous locations. Three values of δ are employed in the evaluations, 0.90, 0.95, and 0.99 corresponding to the top 10%, 5%, and 1% of all sites respectively. In practice, this corresponds with the amount of resources available for remediation and the number of similar sites being compared. For example, a local government wanting to remediate hot spot signalized intersections (where 75 such intersections exist) might fix 7 intersections, or 10% ($\delta = 0.90$).

All parameters of the simulation experiment have now been described. They include shapes of the TPMs (E, S, and L), levels of heterogeneity in the TPMs (1 and 2), and levels of δ (0.90, 0.95, and 0.99). Three HSID methods are assessed, SR, CI, and EB. Evaluation criteria include percent of false positives (FP), percent of false negatives (FN), and sum total percent of FP and FN, called false identifications (FI). For all of the simulations, samples sizes were 1,000 for TPMs and 30 for observation periods.

To conduct the simulation experiment with these parameters, the following steps were undertaken:

1. All the TPM cumulative distributions are divided into truly hazardous locations and non-hazardous locations, using thresholds of 0.90, 0.95, and 0.99 to represent different data separation thresholds. This step results in three “critical” crash count threshold values, $CC_{0.90}$, $CC_{0.95}$, and $CC_{0.99}$ for each combination of cumulative TPM shape and heterogeneity level. These values represent differentiation values to distinguish between known truly hazardous locations and safe locations.

2. The three different HSID methods are used to identify hot spots using the simulated data. Specifically, the SR method simply ranks observed frequencies as shown in Table 2, the CI method uses the entire sample mean and standard deviation to determine confidence intervals for ranking, and the EB method uses a weighted

average of crash history and observed frequency using Gamma distribution parameters to rank sites.

3. Simulated crash data are then compared to the values $CC_{0.90}$, $CC_{0.95}$, and $CC_{0.99}$. For the truly hazardous sites, if the randomly generated crash counts are lower than the values $CC_{0.90}$, $CC_{0.95}$, and $CC_{0.99}$, then FNs are produced. Truly hazardous sites generated observed crash counts lower than the critical crash count values. Similarly, for the collection of non-hazardous locations, when the simulated data are larger than the values $CC_{0.90}$, $CC_{0.95}$, and $CC_{0.99}$, FPs are generated. FPs and FNs are simply counted for each simulation run. Similarly, the number of FIs is the sum of the number of false negatives and positives.

4. To make the three performance metrics comparable across simulations, the percentage of FNs, FPs, and FIs are calculated. Because the FNs are the truly hazardous locations that are mistook as "safe" sites, the percentage is simply the number of simulated FNs divided by the simulated truly safe sites; similarly, the percentage of the FPs is the number of FPs divided by the truly hazardous locations. Finally, the percentage of FIs is obtained by dividing the sum of FNs and FPs by the total number of randomly generated data locations. For example, suppose there are 20 sites under inspection with the top five of them are identified as hot spots according to the corresponding information of TPM. Again, the number of simulated data for each site is assumed as 30, thus, the total truly hazardous locations would be 150, and the number of truly safe ones is 4,500. If 45 sites among the 150 truly hazardous locations are wrongly viewed as safe ones, the percent of FN would be 45/4,500*100%=1%.

5. Finally, the percentage of FPs, FNs, and FIs across simulation conditions are tallied and reported.

Tables 3 and 4 summarize the results of the errors (FNs, FPs, and FIs) produced under the variety of simulation conditions. Table 3 presents the results when heterogeneity of crash counts is relatively low, while Table 4 presents the results when heterogeneity is relatively high. Critical crash count threshold values increase from left to right in both tables. The runs labeled CI, SR, and EB refer to classical confidence interval, simple ranking, and Bayesian methods of HSID respectively. Finally, L, S, and E refer to the underlying characteristics shapes of the cumulative distributions of TPMs: linear, s-shaped, and exponential respectively.

For low heterogeneity and high heterogeneity simulations, the trends of percent errors with the increasing of value of δ are in conformance with each other, however, the values of percent errors for low heterogeneity are much higher than those for high heterogeneity. The major reasons are likely because the low heterogeneity dataset has relatively small standard deviations when compared with the other datasets. The small range of crash counts in a dataset makes it more difficult to identify hazardous locations. On the contrary, it is easy to identify hot spots when the corresponding crash counts are greatly dispersed, particularly when dispersion is large on the upper most crash count deciles.

Another prominent characteristic associated with both tables is that the percentage of false negatives decreases in the same direction as δ for the three kinds of HSID methods. In most cases the percentage of false negatives is substantially reduced using the EB method. The fairly complicated explanation for this is as follows. The threshold value divides the top 'outlying' crash counts from the remainder of the data, either the top 10%, 5%, or 1% of observed counts. By definition these counts are more likely to suffer from regression to the mean in a subsequent observation period than from counts around the TPM. Thus the crash history of the top x% of crash counts act to reduce the effect of the current crash count x when ranking these sites. As a result, sites that suffer less from regression to the mean get ranked higher in the list—sites that ordinarily would have been ranked as false negatives.

Conversely, the decrease of the percentage of the false negatives is accompanied by an increase in the percentage of the false positives (except for δ of 0.95 for L1 and L2, in these two cases, the percent error of FP under the confidence analysis method is the smallest among the three threshold values). It shows that the stricter identification criteria would select less non-hazardous sites for remedy, although it may leave the larger number of truly hazardous locations undetected. Surprisingly, the false identifications also go the same direction to the false negatives with the increase of the value of δ. Probably the best explanation for this phenomenon is that the relatively small number of false negatives can lead to more false positives, and then reduce the efficiency of the investment of local governments. In conclusion, the percent of false positives increases with the rising of thresholds, whereas the percent false negatives and false identifications decrease with the rising of thresholds. Results in almost all simulation scenarios share the same trends.

Table 3: Percent Errors for Low Heterogeneity in Crash Counts

Percent Errors: Low Heterogeneity										
δ		0.9			0.95			0.99		
Method		CI	SR	EB	CI	SR	EB	CI	SR	EB
E1	FN	2.49	3.55	2.40	1.54	2.09	1.41	0.63	0.55	0.38
	FP	62.76	31.97	21.63	82.47	39.73	26.87	114.32	54.00	37.67
	FI	7.17	6.39	4.33	5.31	3.97	2.69	2.46	1.08	0.75
L1	FN	2.21	4.44	2.91	1.39	2.40	1.73	0.15	0.62	0.45
	FP	106.14	39.97	26.20	65.24	45.67	32.80	431.62	61.00	45.00
	FI	8.75	7.99	5.24	3.62	4.57	3.28	2.10	1.22	0.90
S1	FN	0.54	6.53	5.28	0.21	3.48	2.90	0.00	0.81	0.73
	FP	753.44	58.73	47.50	1251.33	66.20	55.13	NA	80.33	72.33
	FI	10.03	11.75	9.50	6.46	6.62	5.51	1.91	1.61	1.45

Note: 1.FN—False negatives; FP—False Positives; FI—False Identifications.

2. In the table, the reason that some FPs can exceed 100% is due to non-normality of the distribution and setting of threshold, and in these cases, the CI method identifies more hazardous locations than truly exist. For the same reason, the existence of "NA" in the table is due to zero truly hazardous locations identified by confidence analysis.

3. The shaded cells show the lowest identification error rate.

Table 4: Percent Errors for High Heterogeneity in Crash Counts

Percent Errors: High Heterogeneity										
δ		0.9			0.95			0.99		
Method		CI	SR	EB	CI	SR	EB	CI	SR	EB
E1	FN	1.78	2.09	1.13	1.33	1.33	0.86	0.39	0.26	0.17
	FP	24.37	18.77	10.13	32.56	25.33	16.40	57.07	26.00	16.67
	FI	4.13	3.75	2.03	3.34	2.53	1.64	1.54	0.52	0.33
L1	FN	1.89	2.55	1.57	1.50	1.43	0.91	0.44	0.37	0.23
	FP	36.33	22.93	14.13	32.20	27.20	17.33	45.22	36.67	22.67
	FI	5.14	4.59	2.83	3.40	2.72	1.73	1.29	0.73	0.45
S1	FN	2.16	2.73	1.74	1.17	1.31	0.71	0.47	0.26	0.12
	FP	34.80	24.53	15.67	41.08	24.87	13.47	38.37	25.33	12.33
	FI	5.16	4.91	3.13	3.31	2.49	1.35	1.32	0.51	0.25

Note: 1. FN—False negatives; FP—False Positives; FI—False Identifications.

2. The shaded cells show the lowest identification error rate.

There is also some difference among the percent errors resulting from the three identification methods. Comparing to the other two traditionally methods, the Bayesian technique yields fewer false negatives in most cases in both the tables. That is, the Bayesian technique is more efficient in flagging the sites that require further analysis. Unfortunately, this higher efficiency is at the cost of the substantial number of false positives generated, which reduce the efficiency of the investment of local governments. Only in the case of budgetary constraints may the false positives not result in the unneeded repairs of the locations that are not truly hazardous. As for the confidence interval method and the simple ranking method, there is no big difference between them. Both methods generally generate higher identification error rate than does Bayesian, indicating the relatively worse performance in identifying hazardous locations.

EXPERIMENT FOR OPTIMIZING DURATION OF CRASH HISTORY

May (1964) first discussed the issue that how many years of accident data should be analyzed when determining the accident-prone locations. He explored the difference between sorts of average accident counts with "t" increasing until 13 years. The result has shown that the difference diminishes as "t" increases as well as the marginal benefit of increasing "t" declines. The "knee" of the curve is said to occur at t=3 years. Based on that information, May then came to the conclusion that "there is little to be gained by using a longer study period than three years."

In this experiment, a different logic is employed to explore the best study duration for accident data analysis. Instead of using the simple accident counts in the method presented by May, this experiment will utilize the identification error rate as an indicator, or the identification error rates associated with various "t" years compared to obtain the optimum study period. When conducting history analysis, the three identification methods are also employer, and the corresponding processes remain the same. The only difference lies in how to use the different periods of data. To show the logic clearly, another small snapshot is used again (Table 5). First, the *i*th column of data is assumed to represent the *i*th-current-year accident data. For example, for site 9, the first four data represent the accident counts during the four current years, and the rest data in the first four columns can be viewed as the accident counts associated with other similar sites during the same period. Let's consider conducting Bayesian analysis. It is known that for a given t- year period, Equation 24 is used for each site to compute the corresponding expected accident counts. However, since the TPM represent the long-term number of accidents per year, thus for the t-year period, average accident counts per year should be used in this equation. In the end of forth year, the "x" for site 10 should be 14 accidents (average of the first 4 data), and E $\{\lambda\}$ =12.88accidents (row average accident), VAR $\{\lambda\}$ =5.18 accidents^2 (row variance), α=0.713 thus the expected accident counts associated with site 10 by using the first 4-year data is 13.2 accidents. Obviously, for the 16 different observation periods, we can generate 13 Bayesian expected data associated with site 10 by using the 4-year history record. Based on these Bayesian expected accident counts of various sites, the previously stated process of the Bayesian method can then be employed to compute the percent of false negatives, false positives, and false identification for different "t" years. The similar history analysis logic can also apply to the other two identification methods. Due to a large amount of iterative computations in this experiment, a special computer code is written to calculate the various identification error rates associated with different period of accident data.

Table 5: Snapshot of the Simulated Data

Site	TPM	Simulated data															
1	3	7	3	4	3	2	1	2	3	3	4	2	3	3	4	3	2
2	3	3	5	5	2	3	1	1	4	2	2	1	2	2	7	4	5
3	5	5	7	6	5	5	6	4	4	3	4	7	4	2	4	7	2
4	7	4	6	5	9	4	6	7	4	8	10	13	6	9	7	7	3
5	8	8	6	8	6	9	9	12	7	2	3	8	11	7	5	7	7
6	9	15	10	16	12	12	8	8	6	9	12	18	15	9	7	12	8
7	9	9	10	12	8	11	5	8	9	13	9	10	12	7	7	8	5
8	12	12	5	11	18	12	12	16	12	7	10	13	10	9	11	9	13
9	13	13	13	12	10	12	12	13	14	11	7	14	13	7	16	18	7
10	14	16	14	15	11	10	12	15	9	15	15	13	11	11	16	12	11
11	15	17	15	13	15	13	13	16	16	13	11	18	14	9	12	22	18
12	16	18	19	20	11	7	14	12	10	16	18	14	17	9	15	19	18

In theory, as the "t" increases, the expected accident counts of each site, which is computed based on the simulated data, would converge to its TPM (the reason is that in the experiment each row of simulated data strictly follow the Poisson distribution) and the corresponding identification error rate would converge to zero. However, in a real situation with "t" increasing, each site would suffer from more influential factors, and thus the long period of data generally cannot represent the current situation. On the other hand, if the short period of data is used, lots of information would be missing and it is difficult to obtain the true long-term accident counts. Consequently, a trade-off should be made to find the study period that is short enough to represent the current condition and long enough to obtain the true expected accident counts. In this experiment, various identification rates are plotted versus the different "t" years. The "knee" of such a curve is expected as the optimum study period.

Considering the data is older than 10 years, it no longer reflects a current situation. In the experiment, the 30 simulated data are averagely divided into 3 groups, that is, the first 10 columns of data belong to group 1, the eleventh to twentieth column of data follows into group 2, the last 10 columns of data belong to group 3. The common characteristic shared by the three groups is assumed to reflect the true relation between identification error rate and "t" years. For each group, the three common confidence levels, 90%, 95%, and 99% are used for the three analyses.

In the diagram of *identification error rate vs. "t" year*, there still exists some fluctuations along the curve, although generally the identification error rate decreases while "t" increases. To quickly determine and eliminate the initial "warm-up" period (i.e., the period before the knee of the curve), Welch's moving average method (Kelton, 2003) is utilized. Through the moving average, this method can further out the statistical fluctuations in observations (y_i) and illustrate clearly the "warm-up" period. As shown in Figure 3, series 1 represents the original Fn rates associated with different "t." Due to the existence of two outliers (the plot of t=4 and t=6), it is difficult to obtain the "knee" of the curve. However, it is easy to know from the series 2 (the curve of moving averages) that the 5-year range is the best study period.

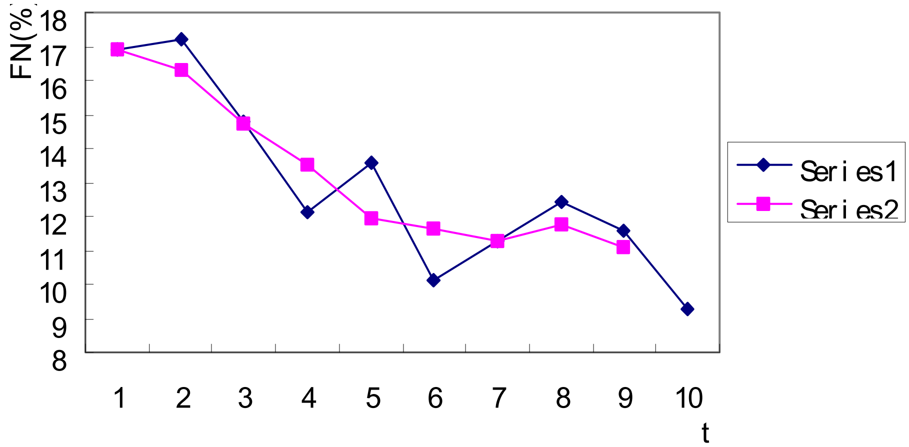


Figure 3: Moving Averages *vs.* Original Statistic

The moving average $\overline{Y}_i(w)$ (where w, the window size) of random observations is defined as follows:

$$\overline{Y}_i(w) = \begin{cases} \dfrac{y_{i-w} + \cdots + y_i + \cdots + y_{i-w}}{2w+1}, i = w+1, \cdots, m-w \\ \dfrac{y_1 + \cdots + y_i + \cdots + y_{2i-1}}{2i-1}, i = 1, \cdots, w \end{cases} \quad (25)$$

In this experiment, the window size is selected as 1.

RESULTS

Similar to the previous experiment, the three HSID methods are also performed in this experiment to explore the optimal duration of accident history. The number of various optimal "t" across the three confidence levels and three groups is shown in the Tables 6~8. For the convenience of viewing, the plots of the frequency of various t-periods for the different confidence levels and groups are illustrated in the Figures 4~6, and the plots of the cumulative results of all the confidence levels and groups are demonstrated in the Figures 7~8. Readers interested in the details of identification error rates associated with various HSID methods, confidence levels, and groups are referred to Appendix B.

Table 6: The Number of t-year Which is the "Knee" of the Curve for Group 1

Year	1	2	3	4	5	6	7	8	9	10
90%		1	22	13	6	8	2	2		
95%	1	1	23	10	8	7	2	2		
99%		2	20	8	10	6	4	3	1	
SUM	1	4	65	31	24	21	8	7	1	

Note: In this group there are 162 scenarios (3 identification methods, 3 kinds of shapes, low and high heterogeneity for crash counts, 3 threshold values for truly hazardous locations, and 3 kinds of false identifications, or FN, FP, FI).

Table 7: The Number of t-year Which is the "Knee" of the Curve for Group 2

Year	1	2	3	4	5	6	7	8	9	10
90%	2	0	28	10	4	5	3	1	1	
95%	0	3	21	11	7	6	4	2	0	
99%	0	1	27	9	5	7	2	3	0	
SUM	2	4	76	30	16	18	9	6	1	

Note: In this group there are 162 scenarios (3 identification methods, 3 kinds of shapes, low and high heterogeneity for crash counts, 3 threshold values for truly hazardous locations, and 3 kinds of false identifications, or FN, FP, FI).

Table 8: The Number of t-year Which is the "Knee" of the Curve for Group 3

Year	1	2	3	4	5	6	7	8	9	10
90%		1	22	14	6	5	2	1	1	
95%	2	2	20	7	7	8	3	4	1	
99%		3	27	11	5	5	4	1		
SUM	2	6	69	32	18	18	9	6	2	

Note: In this group there are 162 scenarios (3 identification methods, 3 kinds of shapes, low and high heterogeneity for crash counts, 3 threshold values for truly hazardous locations, and 3 kinds of false identifications, or FN, FP, FI).

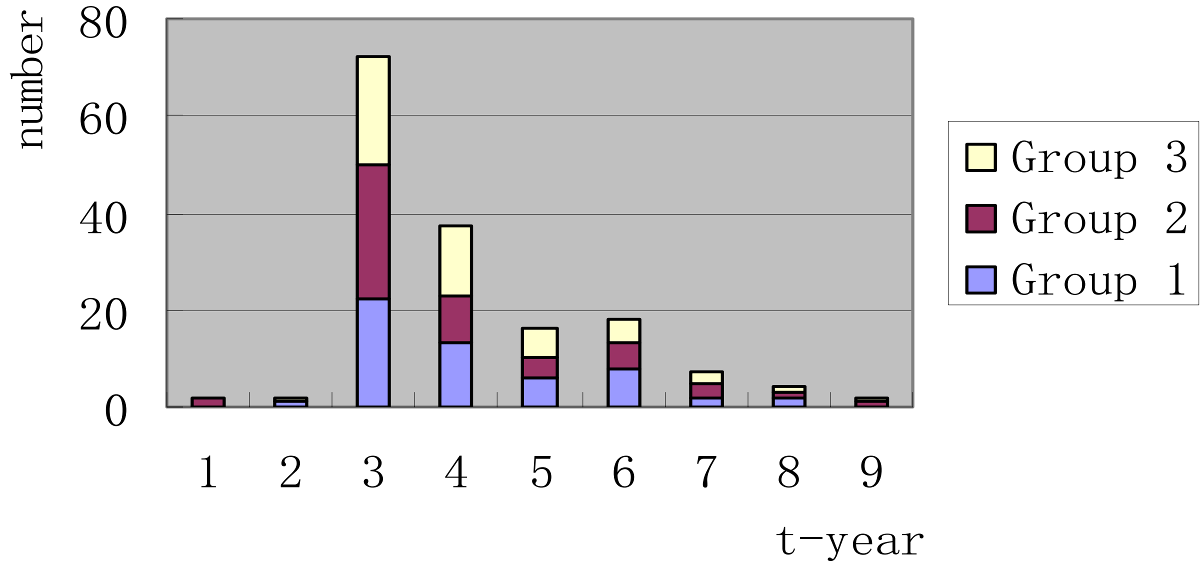


Figure 4: The Number of t-year Which is the "Knee" of the Curve for 90% Confidence Level

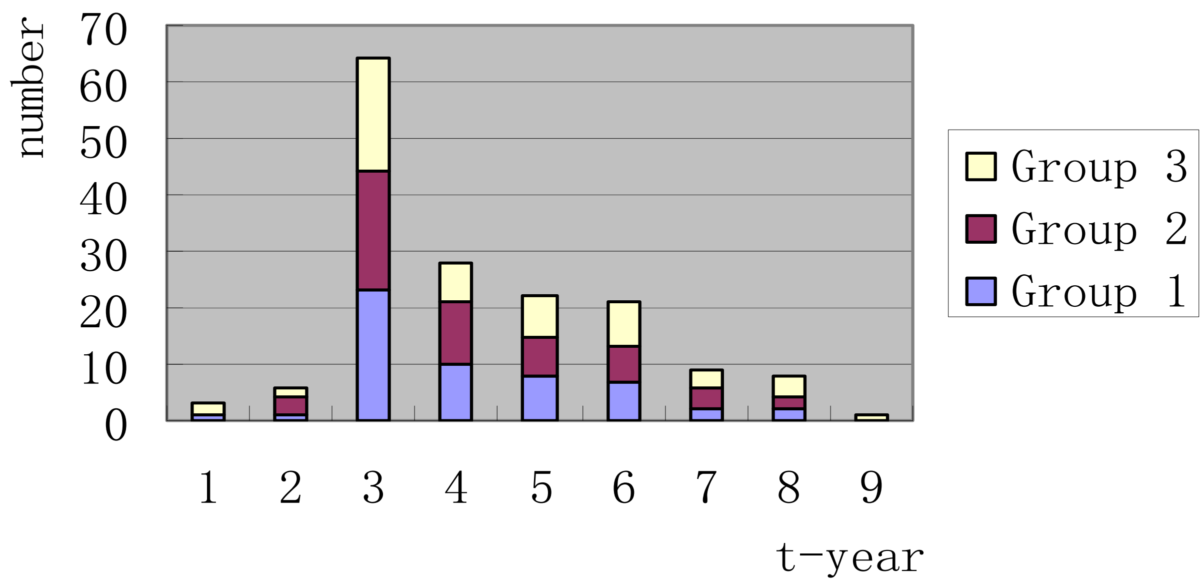


Figure 5: The Number of t-year Which is the "Knee" of the Curve for 95% Confidence Level

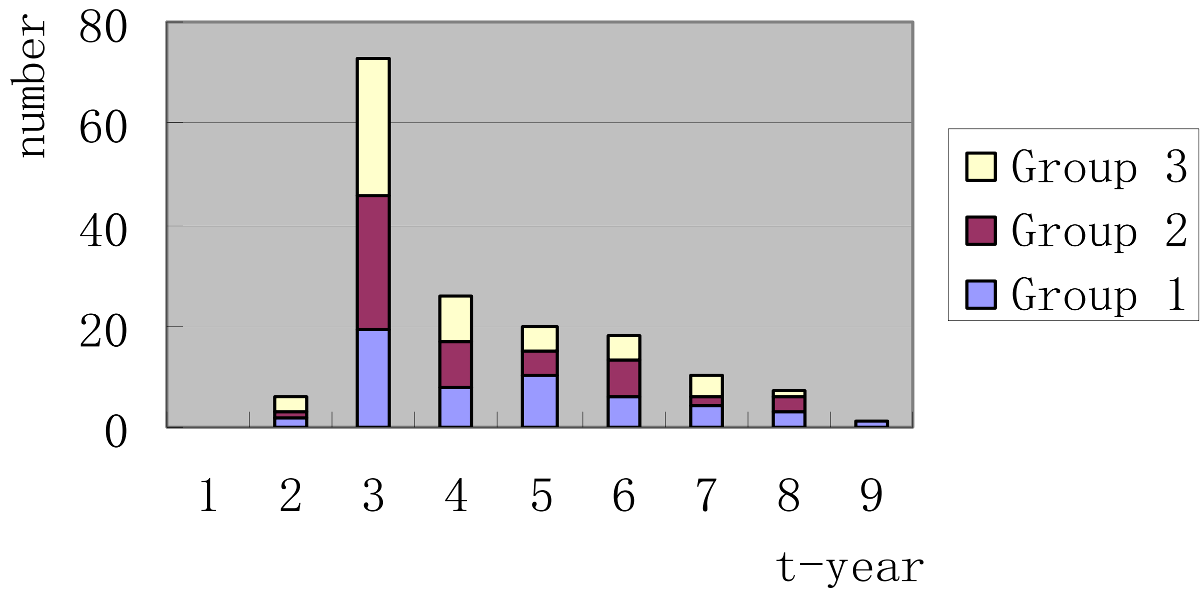


Figure 6: The Number of t-year Which is the "Knee" of the Curve for 99% Confidence Level

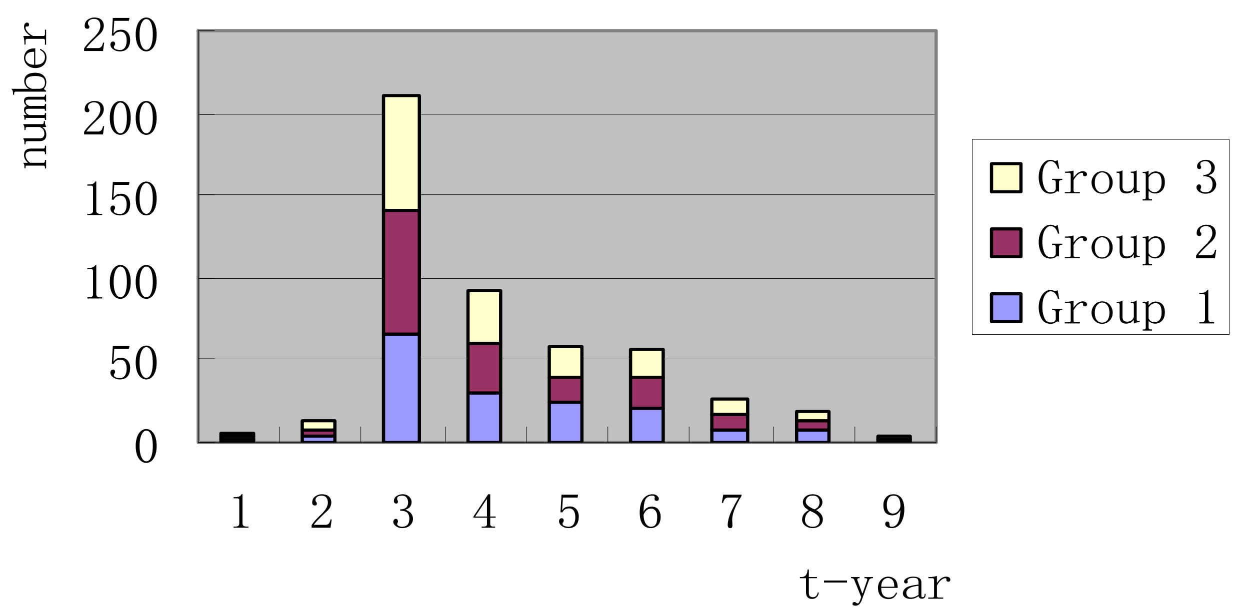


Figure 7: The Number of t-year Which is the "Knee" of the Curve for All Confidence Levels

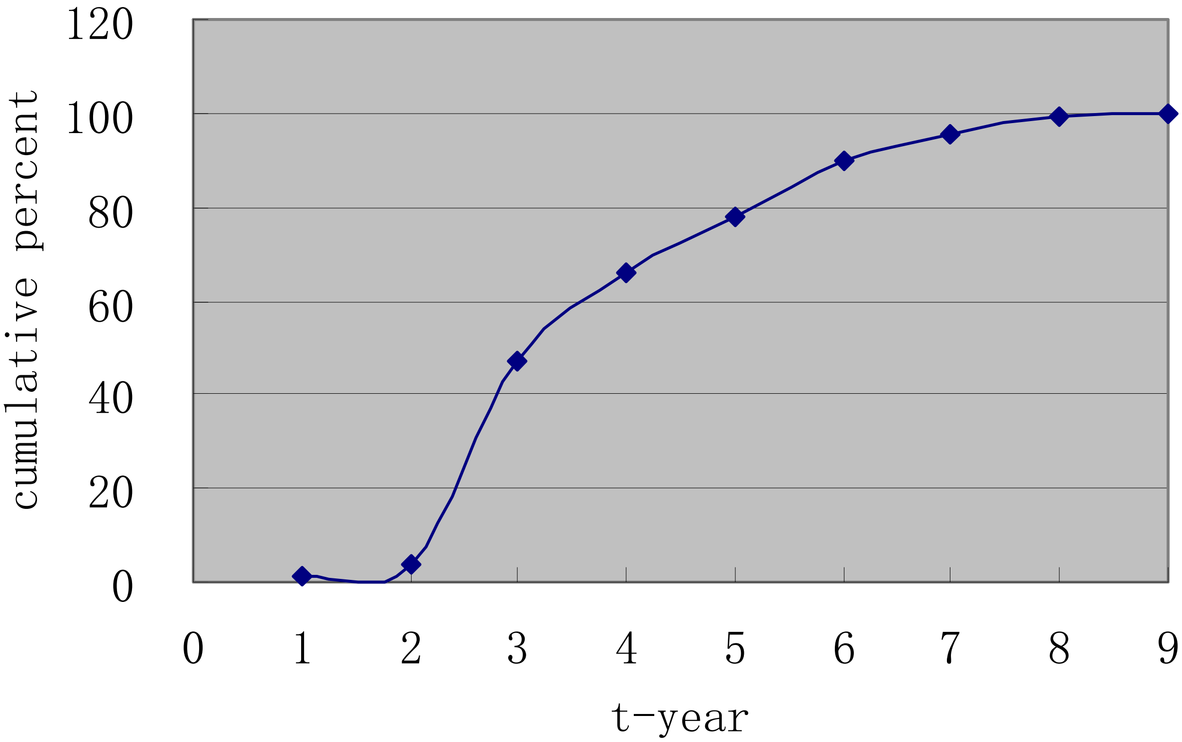


Figure 8: The Cumulative Percent Distribution of Various t-years

In terms of Figures 7 and 8, it is known that across all the simulation scenarios, a 3-year crash history represented the largest portion of "best" study period of crash history, and 3 through 6 years make up almost 90% of all the optimum t-years. Hence, as the trade-off between the long and short history record, if there is no significant physical change in the location under securitization and the long history record can be obtained, it is suggested that the most recent 6-years of crash record is sufficient to capture the majority of the beneficial effect of crash history. In contrast, 3-years of crash history data represents the 'shortest' period of time that should be used and which achieves a significant benefit of crash history (under most general conditions). Crash histories of 1 and 2 years provide relatively little benefit in the methods and under the range of conditions assessed.

To illustrate the improvement in identification performance results from using 3-year history data, Tables 9 and 10 are provided (in contrast to Tables 3 and 4). The differences lie in that Tables 3 and 4 use 1 year of crash data and the percent of identification rates are computed based on the last 30 years of data, whereas Tables 9 and 10 use 3-year data and the corresponding percent of identification rates are calculated on the basis of the current 10 years of data.

Table 9: Percent Errors for Low Heterogeneity in Crash Counts (3 Years Data)

Percent Errors: Low Heterogeneity										
δ		0.9			0.95			0.99		
Method		CI	SR	EB	CI	SR	EB	CI	SR	EB
E	FN	2.02	2.32	1.53	1.36	1.34	0.82	0.89	0.40	0.25
	FP	28.06	20.88	13.75	38.60	25.50	15.50	48.56	40.00	25.00
	FI	4.68	4.18	2.75	3.69	2.55	1.55	2.13	0.80	0.50
L	FN	2.56	2.75	2.13	1.69	1.72	1.25	0.47	0.51	0.40
	FP	33.16	24.75	19.13	50.00	32.75	23.75	91.07	50.00	40.00
	FI	5.56	4.95	3.83	4.33	3.28	2.54	0.14	0.67	0.53
S	FN	1.10	4.88	4.33	0.68	2.88	2.54	0.14	0.67	0.53
	FP	228.21	43.88	39.00	239.38	54.75	48.25	362.16	66.25	52.50
	FI	9.05	8.78	7.80	5.45	5.48	4.83	1.81	1.33	1.05

Note: 1. FN—False Negatives; FP—False Positives; FI—False Identifications; CI—Confidence Interval; SR —Simple Ranking; EB—Empirical Bayesian; E—Exponential Shape; L—Linear Shape; S— Sigmoidal Shape.

2. In the table, the reason that some FPs can exceed 100% is due to non-normality of the distribution and setting of threshold, and in these cases, the CI method identifies more hazardous locations than truly exist. For the same reason, the existing of "NA" in the table is due to zero truly hazardous locations identified by confidence analysis.

3. The shaded cells show the lowest identification error rate.

Table 10: Percent Errors for High Heterogeneity in Crash Counts (3 Years Data)

Percent Errors: High Heterogeneity										
δ		0.9			0.95			0.99		
Method		CI	SR	EB	CI	SR	EB	CI	SR	EB
E	FN	1.08	1.28	0.67	0.96	0.95	0.71	0.24	0.14	0.10
	FP	13.96	11.50	6.00	15.32	18.00	13.50	34.66	13.75	10.00
	FI	2.51	2.30	1.20	1.98	1.80	1.35	1.00	0.28	0.20
L	FN	1.72	1.63	1.36	1.19	0.96	0.87	0.41	0.21	0.20
	FP	14.37	14.63	12.25	15.07	18.25	16.50	20.11	21.25	18.25
	FI	3.08	2.93	2.45	2.14	1.83	1.65	0.86	0.43	0.38
S	FN	2.10	2.04	1.65	0.70	0.66	0.55	0.40	0.15	0.10
	FP	18.01	18.38	14.88	20.83	12.50	10.50	21.03	15.00	10.00
	FI	3.73	3.68	2.98	1.85	1.25	1.05	0.90	0.30	0.20

Note: 1.FN—False Negatives; FP—False Positives; FI—False Identifications; CI—Confidence Interval; SR —Simple Ranking; EB—Empirical Bayesian; E—Exponential Shape; L—Linear Shape; S— Sigmoidal Shape.

2. The shaded cells show the lowest identification error rate.

By comparing these tables, it is known that using 3 years of crash history data results in significant improvements in error rates for all three methods, CI, SR, and EB. Moreover, improvements are seen across most scenarios (except for two false negatives for Method CI)—with the EB method showing 10% to 20% reductions in errors on average. While the EB method still shows itself as the superior method, the SR and CI methods benefit disproportionately, more, on average, from using longer crash histories—with reductions in errors ranging from 15% to 50%.

CONCLUSIONS AND RECOMMENDATIONS

Upon reviewing the previous results of the extensive simulation experiment that is described in this report, the following conclusions and recommendations are made:

1. The EB methods, in general, outperform the other two relatively conventional methods. Under a range of practical conditions, the EB method offers 50% reductions in the percentages of false positives and false negatives compared to CI and SR methods. The EB analysis benefit, however, is contingent upon reliable and accurate safety performance functions for predicting 'expected' safety of comparison sites, which increases the demand for good geometric, traffic, and crash data, and is inherent in the analysis. It is strongly recommended that EB methods be incorporated into mainstream practice by managers of road safety who currently may use CI or SR methods.
2. In low heterogeneity situations, the benefits of the EB methods are much less pronounced. That is, when the observed differences in crashes between 'high-risk' and 'safe' sites is relatively small, the EB method offers only minor improvement compared to the SR and CI methods. This might suggest that municipalities that manage safety on systems with relatively few crashes might not experience significant improvements in performance by changing analysis platforms from SR and CI methods to the EB method.
3. The analysis of crash history suggests that a 3-year crash history constitutes the largest portion of the 486 crash history periods examined, and 3 through 6 years constitute almost 90% of all the 'optimum' crash histories. It is recommended that if possible at least 3 years of crash history duration be used in a HSID analysis, and the most recent 6-year crash history record be used if few substantive changes at the site during this period occurred.
4. Finally, a drastic improvement between the SR method (which is still used in practice) with 1 year of crash data and the EB method with 3 years of crash data is possible. For example, the percent of false negatives and false positives associated with the latter ranges between 25% and 50% less than those associated with the former. A significant improvement in the results from SR and CI methods is also possible by including 3 to 6 years of crash data, but these methods are still outperformed by the EB method.

IN SUMMARY, THE EB METHOD SHOWS ITS CONSISTENT ADVANTAGES OVER THE OTHER TWO APPROACHES IN MOST SIMULATED SCENARIOS. THE NEXT CHAPTER IS DEDICATED TO MODELING THE SAFETY PERFORMANCES OF VARIOUS CLASSIFICATIONS OF ROAD SECTIONS WITHIN ARIZONA, WHICH IS THE IMPORTANT INPUT FOR CONDUCTING EB ANALYSIS IN THE FUTURE ALGSP.

CHAPTER IV - SAFETY PERFORMANCE FUNCTIONS FOR ARIZONA ROAD SEGMENTS

The experiment described in chapter III of this report has illustrated that EB is a superior method compared to the other two conventional methods. Fewer false identifications are produced using the EB analysis from one-year through ten-year accident histories. However, the demonstrated benefits associated with Bayesian technologies are based on simulated crash data with some assumptions. The evaluation of the identification performances of EB techniques based on real crash data is lacking. As distinct from the sample moments method used in the experiment (where there is lack of detailed crash information such as traffic counts, geometric designs, etc.), the multivariate regression method, which utilizes an appropriate function of independent variables (which represent the traits of sites), is generally employed when conducting EB analysis on the real crash data. The multivariate regression models yield SPFs of various roadway classifications (e.g., two lane highways). Considering the importance of SPFs for the improvements of the current ALGSP model, the chapter is dedicated to developing SPFs of various roadway functional classifications in Arizona. This chapter is divided into four sections: data description, how to create SPFs, results of the SPFs, and conclusions.

DATA DESCRIPTION

Since design criteria and level of service vary according to the function of the highway facility, the safety performance function is developed for each functional classification of road segments. The functional classification of roadways is shown in Table 11.

The safety performance functions for various functional road sections in Arizona are created based on the crash data for the year 2000 provided by ADOT, including accident number, AADT, road section length, etc. However, in the subsequent accident data analyses using various hotspots identification methods, the three-year data, 2000-2002, are used as accident history, assuming that there is no significant change in AADT during these years.

Table 11: Functional Classification Codes

Code	Description	Code	Description
RURAL		URBAN	
1	Principal Arterial - Interstate	11	Principal Arterial - Interstate
2	Principal Arterial - Other	12	Principal Arterial-Other Freeways & Expressways
6	Minor Arterial	14	Principal Arterial - Other
7	Major Collector	16	Minor Arterial
8	Minor Collector	17	Collector
9	Local	19	Local

The basic statistics for roads of various functional classifications are shown in Table 12, in which it is illustrated that the sample sizes of rural local (9), urban collector (17), and urban local (19) are 12, 10, and 3 respectively. For the reason of small sample size, the SPFs of these three kinds of road sections are not created in this report. It is expected that they will be provided when more data are collected.

Table 12: Statistics for Roads of Various Functional Classifications

Functional Classification[1]	Number of Sections	Length (km)	Accidents (2000-2002)	Average AADT
1	403	996.113	8122	23810
2	441	1115.128	7012	7603
6	436	1132.092	5261	5483
7	628	1856.064	5285	2637
8	100	365.419	416	684
9	12	22.124	29	3194
11	207	171.647	11999	106338
12	165	199.472	9557	95931
14	429	270.617	9685	17407
16	164	139.709	2282	11499
17	10	7.524	136	4144
19	3	4.428	13	433

Note 1: The functional classification is shown in Table 11.

HOW TO CREATE SPFS?

Due to the existence of overdispersion of crashes in various classifications of road sections in Arizona and characteristics of accident occurrences, Negative Binomial models provided in S-PLUS software package are used to create these SPFs.

The advantage of using a negative binomial model is that Poisson distribution restricts the mean and the variance to be equal ($E[y_i] = VAR[y_i]$). If this equality does not hold, the data are said to be under dispersed ($E[y_i] > VAR[y_i]$) or overdispersed ($E[y_i] < VAR[y_i]$). The negative binomial model has the following expression:

$$\lambda_i = EXP(\beta x_i + \varepsilon_i) \qquad (26)$$

Where $EXP(\varepsilon_i)$ is a gamma-distributed error term with mean 1 and variance α^2. The addition of this term allows the variance to differ from the mean as below:

$$VAR[y_i] = E[y_i][1 + \alpha \cdot E[y_i]] = E[y_i] + \alpha \cdot E[y_i]^2 \qquad (27)$$

Due to data limit and noticing only the purpose of the crash prediction of various highways, only two independent variables, namely AADT and road segment length (SL) are involved in the models. Two different model forms are used. The first model form is:

$$\lambda_i = a * SL * (AADT)^b \qquad (28)$$

The second model form is:

$$\lambda_i = a * EXP(SL) * (AADT)^b \quad (29)$$

Where *a* and *b* are model parameters estimated by using NB regression model.

By using the SL as a constant factor (i.e., the number of crashes on a segment is proportional to its length), the first model form is reasonable since it ensures the predicted accidents is zero when SL is zero. However, the evaluation results showed that the second model form yielded smaller summation prediction errors (PRESS) across the nine classifications of road segments than those of the first model form. In addition, the authors found that there is a nonlinear relationship between the crash number and the length of segments in the data. Finally, SL is never zero and so the model is free to predict non-zero values at this value of SL. Hence, the second model form is selected to develop the SPFs of the road segments.

Corresponding diagnostic problems (whether or not take transformation on variables, identifying and dealing with the issue of multicollinearity, etc.) are addressed to ensure the accuracy of the SPFs. It is important to note that the data outliers and the data with high leverage values (i.e. extreme influence on model parameters) are not excluded for modeling since no errors associated with these data were observed.

RESULTS OF SPFS

The complete results of the nine NB regression models are presented in Appendix C. Criteria used to maintain variables in the models satisfy two conditions: 1) the p-value of the variable's estimated coefficient is smaller than or equal to 0.05, corresponding to a confidence level of 95% or better; and 2) the sign and magnitude of the modeled effect agrees with theoretical expectations of the crash process.

In addition to the model coefficients and associated statistics, the measurement of goodness-of-fit is also an important indicator to show the performance of the models. An equivalent measure to R2 in ordinary least squares linear regression is not available for a negative binomial regression model due to the nonlinearity of the conditional mean (E[y/X]) and heteroscedasticity in the regression (Washington, et al., 2003). As an alternative, Rp2 statistic, which is based on standardized residuals, is computed as:

$$R_p^2 = 1 - \frac{\sum_{i=1}^{n} \left[\frac{y_i - \lambda_i}{\sqrt{\lambda_i}} \right]^2}{\sum_{i=1}^{n} \left[\frac{y_i - \bar{y}}{\sqrt{\bar{y}}} \right]^2} . \quad (30)$$

where the numerator is similar to a sum of square errors and the denominator is similar to a total sum of squares.

A second method for assessing model fit is the G^2 statistic. The better model will yield with the smallest G^2 value. In order to compute G^2 values, it is necessary to estimate the individual deviances using Equation 31.

$$d_i = 2 * [y_i \ln(\frac{y_i}{\hat{\lambda}_i})] \tag{31}$$

The sum of the deviances is equal to the G^2 value, as defined in Equation 32.

$$G^2 = \sum_{i=1}^{n} d_i \tag{32}$$

It is noteworthy (from Equation 31) that d_i would take the value of negative infinite with y_i equaling to 0. Thus, all the locations with y_i =0 are excluded in order to calculate the statistic G^2.

The results of R_p^2 statistic and G^2 are also shown in the Tables 54~62 (see Appendix C).

CONCLUSIONS

All the magnitudes and signs associated with coefficients of the nine SPFs agree with engineering intuition. Generally, the overdispersion parameters of the urban road segment SPFs are significantly larger than those of rural segments; the probable reason is that there is much greater variability of AADT on urban highways and because relatively more unmeasured factors influence urban road safety. The SPF of Urban Interstate Principal Arterials has the greatest value of R_p^2, indicating that this model explains the highest proportion of variation. However, the relatively low value of R_p^2 of other SPFs does not suggest low quality of these models, and is instead explained by the somewhat lower number of variables (only two independent variables in the models) and data outliers with high leverage and influence values. In general the more data and longer accident history that are used, the more accurate the models tend to be.

It is expected that the input of these SPFs can facilitate the incorporation of EB into the ALGSP software. The next chapter contains a comprehensive comparison of alternate identification methods based on the SPFs.

CHAPTER V - COMPARISON OF HSID METHODS BASED ON REAL CRASH DATA OF ARIZONA ROAD SEGMENTS

With the safety performance functions available, the EB technique is then readily implemented to analyze the real crash data of Arizona road segments. The objective of this chapter is to examine whether or not the consistency of the advantages associated with the EB method exist in real situations. Rather than use the same three HSID methods which are explored in the previous experiment design, different HSID methods are implemented and compared in this chapter. They include the empirical Bayes' method, the accident reduction potential (the ARP) method, the accident frequency method, and the accident rate method. Five tests are proposed to compare the identification performances of the alternative HSID methods. These tests are the site consistency test, the method consistency test, the total ranking differences test, the false identification test, and the false true Poisson mean differences test. In addition, the similarity of identification results of the various HSID methods are explored as well. The remainder of this chapter first describes the procedures of implementing the EB method and the ARP method which are based on SPFs. The data used for assessment of performances are then described. Detailed machinery of the five statistical tests and relative test results are followed by the conclusions and recommendations.

HSID METHODS BASED ON SPFS

The literature review provided the underlying assumptions and general logic of Bayesian techniques, and the experiment design section presented the mechanics of the EB analysis based on the method of sample moments. However, the description of how to conduct EB analysis by using SPFs is lacking. Plus, the accident reduction potential method is based on the EB method and also requires the use of SPFs. The detailed mechanics of the SPF-based EB and the ARP methods are described in the following sections.

The EB Approach Based on SPFs

It is known from chapter II that the underlying principle of EB approach indicates that safety of a site is affected not only by some common measurable factors shared by a group of similar sites, but also by some unique characteristics associated with the specific site. Thus, the expected safety of the site (λ_i) can be expressed as follows:

$$\lambda_i = \alpha E(\lambda) + (1 - \alpha)K \qquad (33)$$

Where K is the actual crash count of the site under inspection, $E(\lambda)$ is the expected number of crashes occurring at a specific site group (similar sites), and α is the weight factor, which almost falls between 0 and 1.

The crashes expected on the similar sites (*E(λ)*) are calculated through the SPFs in which the dependent variable is μ, or the average crashes per km-year for road segments or crash counts per year for intersections, and the independent variables are various road and crash characteristics (in this report only AADT and road sections are included). As for the weight, some past studies (Hauer, et al., 2002; Persaud, et al., 1999) have provided a detailed derivation and justification. Generally, the weight is calculated as follows:

$$\alpha = [1 + (\mu * Y) / \varphi]^{-1} \quad (34)$$

Where Y is the number of years of crashes used, and φ is the overdispersion parameter which is a constant for a given model and is derived from the regression calibration process. For road sections, it is estimated per unit length.

Accident Reduction Potential Method Based on SPFs

The accident reduction potential is defined as the deference between the expected crash counts of specific site under inspection and the expected crash number of its similar sites. Thus the ARP method is to rank the value of the λ_i *-E (λ)* to identify the hazardous locations. If the λ_i is substituted by the right-hand side of Equation 33, the ARP can then be expressed as follows:

$$ARP = (1 - \alpha)(K - E(\lambda)) \quad (35)$$

where α, K, and $E(\lambda)$ remain the same as defined in Equation 33. By comparing Equation 33 and Equation 35, it is known that the higher crash counts in history (K) will increase the priority for further investigation of specific locations in both EB and the ARP methods, whereas the value *of E(λ)* poses converse impacts on the selection of hot spots. In the EB method, the greater the value of *E(λ)* , the higher the probability that the specific location is put into the list of hot spots. But in the usage of the ARP, the increase of *E(λ)* will decrease the value of the ARP, and thus yields a lower probability of being selected as a dangerous site.

Numerical Examples to Show the HSID Methods Based on SPFs

To illustrate how to use the SPF-based EB and accident reduction potential methods, two numerical examples are now provided.

Numerical Example 1: Road Segment with 1 Year of Crash Counts

A rural minor arterial segment is 2 km long, has an AADT of 5,000, and recorded 10 crashes in the year of 2000. From Table 56, it is known that the SPF for such a classification of road segment within Arizona is $0.0024 \times ADT^{0.799}$ crashes/km-year, with an over-dispersion parameter φ=3.22 /km. To estimate the safety of the road segment, one would proceed as follows:

Step 1: Average for estimates of this kind. Roads such as this have $0.0024 \times 5000^{0.799}$=2.16 crashes/km-year, on average. Therefore, segments of 2 km long are expected to have 2×2.16=4.32 crashes in one year.

Step 2: Weight. A weight is needed for joining the 10 crashes recorded on this road and the 4.32 crashes for an average of this kind. To obtain the weight, use equation 34. Here, μ=2.16 crashes / km-year, Y=1, and the estimate of φ=3.22 /km. Therefore, weight = 1/[1+(2.16×1)/3.22] = 0.599. Note that both μ and φ are in units of per unit length.

Step 3: EB estimation. Using Equation 34, the estimate of the expected crash frequency for the specific road segment at hand is: 0.599×4.32+0.401×10 = 6.60 crashes in one year. Note that 6.60 is between the average for similar sites (4.32) and the crash counts for this site (10). The EB estimator 'pulls' the crash count toward the mean and thereby accounts for the regression to the mean bias.

Step 4: Accident reduction potential (the ARP) estimation. Using Equation 35, the estimate of the accident reduction potential for this road segment is: (1-0.599)×(10-4.32) = 2.28 crashes in one year.

Numerical Example 2: Road Segment with 3 Years of Crash Counts

Suppose road segment one has 3 years of crash counts (10, 8, 11) and that the ADT in each of those three years was 5,000 vpd. To estimate the safety of the road segment, the following steps should be taken:

Step 1: Average for entities of this kind. As before, segments of this kind are expected to experience 2.16 crashes/km-year. On 2 km in 3 years, one can expect 2×2.16×3=12.96 crashes.

Step 2: Weight. The weight is 1/[1+(2.16×3)/3.22] = 0.332. Note that with one year of crash data, the weight was 0.599. As more years of crash data are used, the weight diminishes.

Step 3: EB estimation. Using Equation 34, the estimate of the expected crash frequency for the specific road segment at hand is: 0.332×12.96+0.668×(10+8+11) = 23.67 crashes in 3 years.

Step 4: Accident reduction potential (the ARP) estimation. Using Equation 35, the estimate of the accident reduction potential for this road segment is:
(1-0.332)×(29-12.96) = 10.71 crashes over 3 years.

DATA DESCRIPTION

As mentioned in chapter IV, 3-year crash data (Year 2000-2002) are utilized to evaluate the performances of various HSID methods. For the convenience of conducting corresponding evaluation tests, the 3-year crash history are separated into two groups, Period 1 (Year 2000) crashes and Period 2 (Year 2001-2002) crashes. Within the state of Arizona, the highway system is broken down into reference sections generally delimited by main intersections, thus the road section lengths vary across various road sections. To make each road section comparable, the crash density (i.e., crashes/km) is used for each HSID method. That is, the accident frequency method uses crashes/km-year to identify hazardous road sections, accident rate employs crashes/km-million vehicles to flag dangerous sections, and EB and the ARP methods also use the indicator based on crashes/km. It is important to note that more accurate results would be obtained if these sections were further divided into shorter sections, for instance, 0.5-km non-overlapping segments.

TESTS FOR COMPARISON OF HSID METHODS

In the former experiments, the false identification test was conducted to compare the performance of three HSID methods in which the percent of false negatives, false positives, and total false identifications are computed and ranked. This section presents other useful tests to examine the relative performance of alternative HSID methods, including the site consistency test, the method consistency test, total ranking differences test, and false true Poisson mean differences test. The definitions of these tests and the corresponding procedures are described in detail in the following sections. For the convenience of understanding the mechanics of each test, a sample of 20 road sections is randomly selected from the principal arterial crash data and the associated information is presented in Table 13. The sites are sorted with respect to their crash counts in the year 2000.

Table 13: Crash Information of a Sample of 20 Principle Arterial Road Sections

Site number	Period 1 (2000)				Period 2 (2001-2002)				3-year crash mean
	Crashes	Accident rate[1]	Bayesian crashes	the ARP[2]	Crashes	Accident rate[1]	Bayesian crashes	the ARP[2]	
1	1	0.157	1.21	-0.03	4	0.628	2.34	1.54	1.67
2	1	0.154	0.82	0.01	2	0.308	1.84	0.12	1.00
3	1	0.16	0.76	0.09	1	0.16	1.14	-0.08	0.67
4	2	0.308	0.83	0.44	2	0.308	1.52	0.06	1.33
5	2	0.32	0.79	0.96	1	0.16	0.84	0.02	1.00
6	4	0.52	1.71	1.42	5	0.65	2.68	1.16	3.00
7	4	0.632	1.81	1.62	10	1.58	2.94	4.3	4.67
8	5	0.771	1.86	1.50	13	2.004	3.94	2.98	6.00
9	5	0.789	1.79	3.20	4	0.632	2.98	0.56	3.00
10	6	0.741	1.92	0.20	17	2.1	4.06	6.78	7.67
11	6	0.93	1.82	1.49	11	1.706	3.9	0.22	5.67
12	6	0.94	1.89	3.75	16	2.506	3.62	7.98	7.33
13	7	1.115	1.95	0.80	15	2.39	4.24	1.12	7.33
14	7	1.103	2.01	1.82	17	2.678	4.86	8.36	8.00
15	8	1.281	1.96	1.09	20	2.402	3.78	7.56	9.33
16	8	1.263	1.99	4.37	15	2.368	4.54	5.24	7.67
17	11	1.383	3.56	1.08	16	2.012	6.36	0.16	9.00
18	11	1.733	4.33	4.06	31	3.308	6.82	13.98	14.00*
19	12	1.843	3.58	4.15	28	3.378	6.86	11.78	13.33
20	14	2.197	4.02	5.07	32	3.138	8.9	18.52	15.33*

Note: 1.The unit of accident rate is crashes/million kilometers.

2. The ARP: Accident reduction potential. The negative values result from sections whose Bayesian estimators are less than regression values of similar sections.

3. The two asterisks in last column show that sites 18 and 20 are considered as truly top 10% hazardous locations based on the 3-year crash mean, which is assumed as the true Poisson mean of each road section.

4. The shaded cells contain the information of the top 10% sections of the 20 samples in terms of the crash count in Period 1.

Site Consistency Test

This test is to compare the performances of various HSID methods in terms of the future safety of the hot spots identified by these methods. It follows that the sites identified as hazardous during an initial period would also reveal inferior safety performance in a subsequent period assuming there are no significant changes occurring at these sites. In this report, the crash data in Period 1 (assumed as "crash history") are first used with various methods to identify the hazardous road sections. The crash counts of selected sections in Period 2 (assumed as "future" period) are then measured. The higher the number of crash counts of the selected sections in Period 2, the better the performance of the HSID method. It is reasonable to employ the Period 2 crash data to validate the alternate HSID methods in the sense that only Period 1 data are used to identify the hot spots and Period 2 crash counts can be viewed as an unbiased estimate of true safety of the sites.

Based on the information of twenty road sections presented in Table 13, it is noticed that accident frequency and rate methods identify the same top 10% sections during the year 2000, or sites (19,20). Their crash counts in Year 2001-2002 are 60 (28+32). The EB method yields sites (18, 20) and the resultant total crashes in Period 2 are 63 (31+32). Whereas the ARP method selects sites (16, 20) whose total crashes in the next two years are 47. In terms of this small sized sample, the hot spots identified by the EB method show the worst safety performances in the "future" period, and remedial treatments implemented to those sections would yield the highest benefits.

Method Consistency Test

Crash counts in Period 2 are used as a benchmark to compare different HSID methods. The underlying assumption is that the identification performances of the HSID methods can be revealed through the safety performances of the corresponding identified hot spots. This test, however, is to evaluate the method performance by measuring the number of same sites identified in both periods. Since the two periods are very close, it is safely concluded that most road sections are in the same or similar operational state (similar traffic counts, same geometric designs, etc.) and their expected safety performances remain the same over the two periods. Therefore, a good HSID method is expected to identify a large number of same hot spots by using the crash data in the two different periods. The more the same hot spot is identified in the two time periods, the more reliable is the HSID method.

In a review of Table 13, it is known that for the accident frequency method, sites (19,20) and sites (18,20) are selected as top 10% hot spots in the two periods respectively. The number of same sites is thus one, or Site 20. For the accident rate method, the hazardous sections are sites (19, 20) and sites (18, 19) respectively. The same hot spot is also one, such as Site 19. It is easy to know the number of same hot spots yielded by EB and the ARP methods are both one as well. So, a large number suggests that many sites are being ranked consistently in both periods, whereas a small number suggests that only few are being ranked consistently.

Total Ranking Differences Test

In the method consistency test, the number of hot spots identified in both periods is used to measure a method's reliability. One disadvantage associated with this test is that the relative priorities (i.e., rankings of safety performances) of road sections in the two periods are not accounted for. This shortcoming can be illustrated by a simple example. Consider a group of 100 road sections, among them there is one site whose accident rate ranks 1st in Period 1 and ranks 10th in Period 2. If top 10% sites are identified as hot spots, then for the accident rate method, this site is screened out in both periods and the number of same hot spots increases by one even though the rankings of accident rate have significantly changed. To correct for this drawback and to obtain a more precise measurement of method reliability, a total ranking differences test is proposed. This test is performed by calculating the total rank differences of the hazardous road sections identified in the two periods. The less the total ranking differences, the more dependable is the HSID method. Again, this test is based on the assumption that no significant treatments are implemented on the road sections and the rankings of safety performance (it is important to note that each HSID method has the different estimator of safety performance) of each road section during the two periods remain the same. Hence, it is of great importance to make sure all the data outliers (i.e., road sections that are largely treated during Period 2) have been identified and removed, which otherwise will significantly influence the ranking differences and affect the corresponding results.

In this report, rank gives duplicate numbers the same rank. However, the presence of duplicate numbers affects the ranks of subsequent numbers. For example, in a list of integers sorted in ascending order, if the number 7 appears twice and has a rank of 4, then 8 would have a rank of 6 (no number would have a rank of 5). For some purposes one might want to use a definition of rank that takes ties into account. In the previous example, one would want a revised rank of 4.5 for the number 7. This can be easily done by adding a correction factor to the tied values and the results are expected to be similar to the results shown in this report.

Checking again the samples in Table 13, the top 10% sections (19, 20) identified by frequency method (based on data in Period 1) possess the rankings of accident frequency (19^{th}, 20^{th}) in Period 1 and rankings (18^{th}, 20^{th}) in Period 2, the total ranking differences is then 1. Whereas for the ARP method, the hazardous sections (16, 20) have the rankings of the ARP (19^{th}, 20^{th}) and rankings (13^{th}, 20^{th}) in the two periods respectively, and the resultant total ranking difference is six.

False Identification Test

This test has been conducted in the previous experiment design, that is, first the TPM of each site is specified, and then the number of FN, FP, and FI are counted. This test is also performed in this section due to two considerations. First, the results in the experiment are obtained from the simulated data, which are based on a series of strong assumptions, whereas the number of false identifications in the real world is not explored. Second, this test is the basis of the subsequent test to be described, and it would be easier for the readers to understand the next test with being familiar with the mechanics of this test. As shown in the previous experiment design, a very important issue of this test is how to identify the truly hazardous and safe locations. Considering crash counts in most road sections can be fit by Poisson distribution and the sample mean is the unbiased estimator of the Poisson parameter λ, this section of the report assumes the mean of the 3-year crash data as the TPM and the road sections with higher TPM are considered as hazardous locations. Obviously, great cautions should be taken to the results since the sample size (3) is too small, and the more accurate results are expected with longer crash history being used (it is important to note that too long of a crash history would also reduce the accuracy of the estimator since more influential factors are expected, such as change of driving population, police enforcement effects, etc.).

The last column of Table 13, 3-year crash mean, represents the TPM of each road section. Based on the ranking of these TPMs, it is known that Sites 18 and 20 are truly top 10% hazardous locations, and the rest are truly safe sites. For the accident frequency method, Site 18 in Period 1 (with a recorded crash count of 11) is viewed as safe, and the two sites in Period 2 are correctly identified, thus the number of false negatives for the accident frequency method in both periods is 1. Similarly, the truly safe Site 19 in Period 1 is viewed as hazardous, and the number of false positives for the accident frequency method in both periods is also 1. The corresponding number of false identifications is then 2. Whereas for the accident rate method, Site 18 in Period 1 (the associated rate is 1.733) and Site 20 in Period 2 (the associated rate is 3.138) are considered as safe, the total number of false negatives in the two periods is then 2. It is easy to know that the number of false positives and identifications are 2 and 4 respectively. The number of false negative, positives, and identifications for the EB and the ARP methods are both (1, 1, 2).

False True Poisson Mean Differences Test

The previously described false identification test uses the number of false negatives and false positives to assess the performances of various HSID methods. One disadvantage of this test is that each false identification is counted equally and does not reflect the different consequences resulting from various false identifications. This disadvantage can be seen in a simple example. If a site with a TPM value of 15.6 is wrongly selected for treatment instead of one with a TPM of 15.7, the "error" is really rather small; whereas if the site with a TPM of 15.6 is mistakenly selected instead of one with a TPM of 25.6, the "error" is much larger. The TPM differences associated with the two false identifications are 0.1 and 10 respectively, showing the big difference. To obviate this drawback, a false true Poisson mean differences test is designed to differentiate various false identifications and show more clearly the serious consequences brought by the erroneous identifications. The corresponding judgment indicator is the sum of the absolute difference of TPM associated with the false identified sites and "critical" TPM which separates the truly safe and hazardous locations.

It is easy to describe the mechanics of this test after describing the procedures of the false identification test. First, the thresholds of 0.90 and 0.95 used to divide the TPM cumulative distributions are specified as the "critical" TPMs. Second, when a false negative or false positive is created, the TPM of the mistakenly identified site is compared to the "critical" TPM and the absolute difference is calculated. Finally, all the TPM differences of different false identifications across alternate HSID methods are summed. The method which yields the smallest TPM differences is preferred.

It is known from the Table 13 that the "critical" TPM of the top 10% truly hazardous road section is 14. The false identification test results show that the accident frequency method results in one false negative (Site 18) and one false positive (Site 19) in the two periods. Thus, the TPM differences of false negative and positive of the accident frequency method are 0 and 0.67 (absolute value of 13.33-14) respectively. The value of 0 is obtained due to the fact that the TPM of the false negative (Site 18) is used as the "critical" TPM. The effect of this flaw diminishes quickly with the number of false negatives increasing. The total TPM differences of false identifications are then 0.67 (0+0.67). Whereas for the accident rate method, the Site 18 in Period 1 and Site 20 in Period 2 are false negatives and Site 19 in both periods are false positives. The relative TPM differences of false negatives and false positives are 1.33 (15.33-14+0) and 1.34 (0.67*2), and the total TPM differences of false identifications are 2.67 (1.33+1.34). Following the same way, it is easy to know the TPM differences of false negatives, positives, and identifications for the EB and the ARP methods are both (0, 0.67, 0.67).

COMPARISON RESULTS

To compare the relative performance of the alternative HSID methods, the five tests mentioned above are conducted. Test results of the nine classifications of road sections are first documented, and then the aggregate results are computed. In addition to the five tests, the similarity of the identification results of these HSID methods is also explored. The similarity is obtained by counting the common locations identified by various methods. In contrast to the previous experiment design, only top 10% and 5% locations are considered hot spots in this section. The case of top 1% hazardous locations is not explored due to the relative small sample size of each classification of road sections. The accumulated evaluation results of all the road sections are illustrated in Tables 14 through 20, and the assessment results of each of the nine classifications of road sections are shown in Appendix D.

Site Consistency Test Result

Table 14: Results of Site Consistency Test of Various Methods for All Classifications of Highways: Accumulated Crashes for Hot Spot Sites for Various Methods

Method	δ = 0.90			δ = 0.95		
	Crashes (2000)	Average Crashes (2001-2002)	Crash difference	Crashes (2000)	Average Crashes (2001-2002)	Crash difference
Frequency	8276	9611	1335	5639	6303	664
Rate	6899	7556	657	4257	4313	56
Bayesian	8123	9603	1480	5399	6377	978
the ARP	7314	8450	1136	5611	6260	649

Note: 1. the ARP—Method of Accident Reduction Potential.
2. δ = 0.90 and δ = 0.95 represents the cases of top 10% and 5% hazardous sections.

As discussed previously, the crashes in the Period 2 are one measure of how unsafe the selected segments really are. With this said, it is known from Table 14 that the accident frequency method outperforms other HSID methods when identifying top 10% hot spots, whereas the EB method performs best in the case of δ = 0.95. The accident rate method performs worst in both cases, whose corresponding hot spots experience the lowest number of crashes, such as 7556 and 4313 respectively.

Another observation revealed in Table 14 is that the EB method obtains the largest crash count differences (between the two periods) for both top 10% and top 5% hot spots, or, 1480 and 978 crashes respectively. In the case of $\delta = 0.90$, the road sections chosen under the EB method, during Period 1, experience 153 (8276-8123) crashes less than those chosen by the frequency method, whereas in the subsequent period these two groups of sections have almost the same crash counts. Similarly in the case of $\delta = 0.95$, the crash counts of the hot spots identified by EB rise from 5399 in Period 1 (rank 3rd among the four methods) to 6377 in Period 2 (rank 1st). The largest crash count difference indicates that the EB method has the highest ability to fight against the random effects associated with the observed crash counts and to select the true dangerous sites which tend to show the worst safety performances in the future.

Therefore, considering the both observations from Table 14, it can be concluded that the EB method shows the best performance in this test, whose corresponding hot spots show the consistently worst safety performances in the future. On the contrary, the accident rate method performs worst among the four methods. Referring to the research results shown in Appendix D, it is also found that the same observations follow from most of the disaggregate results (i.e., results for each functional classification).

Method Consistency Test Result

Table 15: Results of Method Consistency Test of Various Methods for All Classifications of Highways: Number of Sites Commonly Identified across Periods

Method	$\delta = 0.90$	$\delta = 0.95$
Bayesian	167 (56.0%)	71 (47.3%)
Frequency	148 (50.0%)	63 (42.0%)
Rate	131 (44.0%)	49 (32.7%)
the ARP	139 (46.6%)	59 (39.3%)

Note: 1. the ARP—Method of Accident Reduction Potential.
2. $\delta = 0.90$ and $\delta = 0.95$ represents the cases of top 10% and 5% hazardous sections.
3. The number means the number of locations identified by methods in both periods; the percent in the parenthesis stands for the percentage of same road sections among all the hazardous sections.

The results in Table 15 agree with the results in Table 14. The EB method shows again its superiority in terms of method consistency, and the accident rate method performs worst among the four methods once more. For both cases of top 10% and top 5%, the EB method yields the largest number of same hot spots in the two periods, or 167 and 71 respectively. The greatest figures show the highest reliability of this method considering most road sections remain the same in safety performance. The accident frequency and the ARP methods stand between the EB and rate methods.

An interesting observation from Table 15 is that the percentages of the same top 5% hot spots (shown in the parentheses) are consistently lower than those of top 10% hot spots across the four methods. This phenomenon is opposite to the original thought since the percentages of the two periods are expected to be similar. The reason is most likely that more attention has been given to top 5% sections during Period 1 by traffic engineers, and the resultant treatments performed on some of the sections improve their safety and put them down from the top 5% lists during Period 2.

The results from each individual classification of road sections are in consistent with the aggregate result.

Total Ranking Differences Test Result

Table 16: Results of Total Ranking Differences Test of Various Methods for All Classifications of Highways: Cumulative Ranking Differences of Hot Spot Sites

Method	$\delta = 0.90$	$\delta = 0.95$
Bayesian	17851	10349
Frequency	29602	15357
Rate	34869	21212
the ARP	32601	18787

Note: 1. the ARP—Method of Accident Reduction Potential.
2. $\delta = 0.90$ and $\delta = 0.95$ represents the cases of top 10% and 5% hazardous sections.

Although Table 15 shows that the difference of the number of same hot spots identified by various HSID methods is within 20% (the number ranges from 131 to 167 for top 10% hot spots, and from 49 to 71 for top 5% hot spots), the results in Table 16 exhibit the huge differences among the HSID methods in terms of total ranking differences of the hazardous locations in the two periods. In both cases of $\delta = 0.90$ and $\delta = 0.95$, the EB method reduces the total ranking differences by about 50% compared with the accident frequency method, 80% compared with the ARP method, and more than 100% compared with the accident rate method. This phenomenon means that the rankings of the safety performance of the hazardous sections identified by the EB method change slightly in between the two periods, whereas the priorities of the dangerous sections identified by the accident rate method vary significantly during the two periods. The results can be translated into the conclusion: great cautions should be taken to the accident rate method even though it might show good performance in other aspects, since the method itself is not reliable. On the contrary, the EB method possesses the highest degree of dependability, followed by the accident frequency method and the ARP method. Again, the disaggregate results are in conformance with the aggregate result.

False Identification Test Result

Table 17: Results of False Identification Test of Various Methods for All Classifications of Highways: Frequency of Errors

Method		$\delta = 0.90$	$\delta = 0.95$
Bayesian	FN	153 (2.9%)	94 (1.7%)
	FP	153 (25.6%)	94 (31.3%)
	FI	314 (5.1%)	188 (3.2%)
Frequency	FN	152 (2.8%)	91 (1.6%)
	FP	152 (25.5%)	91 (30.3%)
	FI	304 (5.0%)	182 (3.1%)
Rate	FN	333 (6.2%)	199 (3.5%)
	FP	333 (55.9%)	199 (66.3%)
	FI	666 (11.2%)	398 (6.7%)
the ARP	FN	226 (3.4%)	174(2.7%)
	FP	226 (30.9%)	174 (50.7%)
	FI	452 (6.2%)	348 (5.1%)

Note: 1. the ARP—Method of Accident Reduction Potential.
2. $\delta = 0.90$ and $\delta = 0.95$ represents the cases of top 10% and 5% hazardous sections.
3. FN—False Negatives; FP—False Positives; FI—False Identifications.
4. The number means the number of false identifications; the percent in the parenthesis stands for the corresponding percentage. FN% is defined as the number of FNs divided by the number of sites viewed as safe in various periods; FP% is defined as the number of FPs divided by the number of sites viewed as safe in various periods; FI% is defined as the number of FIs divided by all the number of sites.

One observation of Table 17 is that the results show the similar characteristics to those of a previous experiment design. That is, the percentage of false negatives and the percentage of false positives have the converse trend with increasing δ for the four HSID methods, and the false identifications go the same direction to the false negatives with the increase of the value of δ.

There is very slight performance difference between the accident frequency method and the EB method. The accident frequency method reduces the three percentages in the two levels of δ by less than 3% compared with the EB method. Compared with these two methods, the accident rate method yields many more false negatives and false positives, increasing the three percentages by more than 100%. The ARP method stands in between the three methods.

It is noteworthy that these results are obtained based on only 3-year crash data, in which the mean of the 3-year crashes is assumed as the true Poisson mean of the crash count of each road section. Combined with the previous experiment results, which use the mean of 30 simulated data following Poisson distribution as TPM, it is expected that the EB method would outperform the accident frequency method with longer crash history used.

False True Poisson Means Differences Test Result

Table 18: Results of False True Poisson Means Differences Test of Various Methods for All Classifications of Highways: Cumulative Difference in TPMs

Method		*δ* = 0.90	*δ* = 0.95
Bayesian	FN	1141.5 (7.46)	1097.3 (11.67)
	FP	1271.3 (8.31)	1169.7 (12.44)
	FI	2412.8 (7.68)	2267.0 (12.06)
Frequency	FN	1041.4 (6.85)	858.9 (9.44)
	FP	1279.0 (8.41)	1168.8 (12.84)
	FI	2320.4 (7.63)	2027.7 (11.14)
Rate	FN	4164.7 (12.54)	3175.5 (15.96)
	FP	2549.4 (7.66)	2378.2 (11.95)
	FI	6714.1 (10.08)	5553.7 (13.95)
the ARP	FN	1047.5 (4.63)	1017.0 (5.84)
	FP	1465.6 (6.48)	1358.5 (7.81)
	FI	2513.1 (5.56)	2375.5 (6.83)

Note: 1. the ARP—Method of Accident Reduction Potential.
2. δ = 0.90 and δ = 0.95 represents the cases of top 10% and 5% hazardous sections.
3. FN—False Negatives; FP—False Positives; FI—False Identifications.
4. The values in the parentheses represent average true Poisson mean difference per erroneous identification.

From the results in Table 18, it is shown that the accident rate method performs significantly worse than the other three methods. There are relatively small differences among the false TPMs associated with the accident frequency method, the EB method, and the ARP method. In the two levels of δ, the accident frequency method yields the smallest TPM differences associated with false negatives and false identifications. The EB method obtains the smallest TPM differences associated with false positives when identifying top 10% hazardous sections.

A very interesting observation is that although the TPM differences resulting from the ARP method are slightly larger than those from the accident frequency method, its number of false identifications (shown in Table 17) is much larger than those of the accident frequency method. This indicates that the ARP method yields the lowest average TPM difference per erroneous identification (shown in the parentheses), that is, each false identification resulting from the ARP method has only slight consequences.

Result of Similarity of Alternative HSID Identification Methods

Table 19: Accumulated Similarity of Various Methods for All Classifications of Highways (δ = 0.90)

	Bayesian	Frequency	Rate	the ARP[1]
Bayesian		246 (82.6%)	130 (43.6%)	224 (75.2%)
Frequency	246 (82.6%)		177 (59.4%)	265 (88.9%)
Rate	130 (43.6%)	177 (59.4%)		189 (63.4%)
the ARP[1]	224 (75.2%)	265 (88.9%)	189 (63.4%)	

Note: 1: the ARP—Method of Accident Reduction Potential.
2: The number means the number of locations identified by both methods as upper 90% hazardous locations; the percent in the parenthesis stands for the corresponding percentage.

Table 20: Accumulated Similarity of Various Methods for All Classifications of Highways (δ = 0.95)

	Bayesian	Frequency	Rate	the ARP[1]
Bayesian	1	123 (82.0%)	52 (34.7%)	111 (74.0%)
Frequency	123 (82.0%)	1	77 (51.3%)	135 (90.0%)
Rate	52 (34.7%)	77 (51.3%)	1	74 (49.3%)
the ARP[1]	111 (74.0%)	135 (90.0%)	74 (49.3%)	1

Note: 1: the ARP—Method of Accident Reduction Potential.
2: The number means the number of locations identified by both methods as upper 95% hazardous locations; the percent in the parenthesis stands for the corresponding percentage.

Upon reviewing both the accumulated (see Tables 19~20) and disaggregated (see Appendix D) results of identification similarities of the HSID methods, the following conclusions can be obtained.

First, EB, ARP, and accident frequency methods yield the higher similarities among their identification results. Overall, around 82% of the hot spots identified by the EB method are the same as those flagged by the accident frequency method, and about 75% of them are same as the hazardous locations screened out by the ARP method. By contrast, accident rate method identified the total different hot spots. In most cases, more than 50% of the identified hot spots are different from those identified by the other three methods. It is indicated that, if the EB method is considered as the best HSID method as shown in the results of the previous tests and experiment design, many cautions should be given to the hot spots identified by the accident rate method.

Second, the relative high similarity between the EB method and ARP method indicates that many locations are expected to have not only higher expected crashes, but also higher accident reduction potentials. This makes it feasible to jointly use the criteria of crash number and accident reduction potential to screen the hazardous locations from the road network.

CONCLUSIONS AND RECOMMENDATIONS

Upon reviewing the previous results the following conclusions and recommendations are made:

1. The EB method outperforms the other three HSID methods under the site consistency test. The road sections identified by the EB method possess relatively small numbers of crash counts in Period 1 ("current" period), whereas these sections show very high numbers of crash counts in Period 2 ("future" period). The highest crash count difference in the two periods demonstrates that the EB method is efficient in identifying truly hazardous sections, which show serious safety problems in the future period. On the contrary, the accident rate method identifies dangerous sections having both the smallest number of crash counts in Period 2 and the fewest crash count difference in the two periods. The EB method is followed by the accident frequency method and the ARP method.
2. The EB method is superior to the other three methods in terms of method consistency. The same hot spots in both periods identified by the EB method make up 56% of all the top 10% sections and 47.3% of the top 5% sections. The EB method is followed by the accident frequency method, the ARP method, and the accident rate method.
3. Compared with the method consistency test, the total ranking differences test shows a larger advantage associated with the EB method. In both cases of $\delta = 0.90$ and $\delta = 0.95$, the EB method reduces the total ranking differences by about 50% compared with the accident frequency method, 80% compared with the ARP method, and more than 100% compared with the accident rate method. The results show that the accident rate method is not reliable and much caution should be taken to use this method even though it maybe shows good performance in other aspects. On the contrary, the EB method possesses the highest degree of dependability.
4. In the false identification test, there is a very slight difference between the accident frequency method and the EB method. Compared to these two methods, the accident rate method yields more false negatives and false positives, increasing the three percentages by more than 100%. The ARP method performed in between these three methods. Considering the 3-year crash history used, it is expected the EB would outperform the accident frequency method when a longer crash history is used.
5. In the false true Poisson mean differences test, there are small differences among the total false TPMs associated with the accident frequency method, the EB method and the ARP method. However, the ARP method yields the lowest average TPM difference per erroneous identification, indicating that each false identification resulting from the ARP method has the slightest consequences. Conversely, the false identification caused by the accident rate method has the most serious consequences.
6. Overall, on the basis of Arizona 3-year road section crash data, the EB method exhibits the best performance under most of the five evaluation tests. By contrast, the accident rate method performs badly in all the five tests, the accident

frequency method and the ARP method perform between the two methods, and the former one performs slightly better than the latter. Therefore, it is strongly recommended that EB be used when conducting HSID within Arizona in the future.

7. The EB method identifies the largest number of same hot spots as those identified by the accident frequency method. Overall, more than half of the hot spots identified by the accident rate method are different from those resulting from the other three HSID methods.

CHAPTER VI - HSID IN CURRENT ALGSP MODEL AND RECOMMENDED SOFTWARE CHANGES

Due to limited time and resource constraints and the extensive number of candidate sites requiring evaluation, it is impractical for agencies to examine all sites in detail. As a result, much emphasis is placed on the stage of hazardous site identification. On the basis of the previous literature review on HSID methods and thorough statistical analyses, some software changes are identified which can improve the ability of ALGSP model to accurately identify hazardous locations. These recommended changes, which include data requirements, analytical method enhancements, crash analysis period changes, etc., are provided following the description of the logic of HSID in current ALGSP model.

HSID IN CURRENT ALGSP MODEL

This section aims to familiarize the readers with the mechanics and procedures of HSID within the ALGSP model. According to the final report prepared by Carey (2001), hot spot identification is one main component of the current ALGSP model and it can be shown in the dashed rectangle shown in Figure 9.

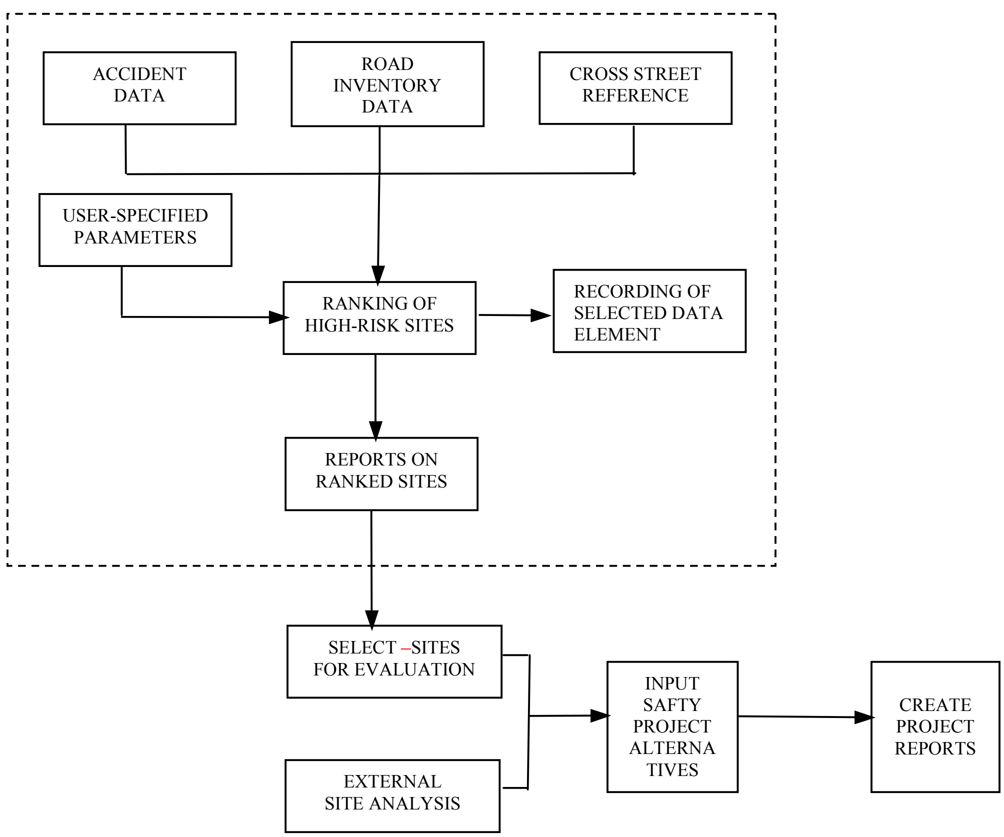


Figure 9: **Key Steps of ALGSP Model**

HSID is conducted based on two sub-procedures. First, the crashes occurring on the large amount of candidate road sections are divided into different subgroups with respect to users' specified selection parameters. Second, the road sites where each subgroup of crashes occur are examined to screen out the corresponding hot spots by using the four alternative weighting methods provided by the ALGSP model.

The allowable user-specified selection parameters in the ALGSP model include:

- **Jurisdiction:** This selection parameter allows users to specify the jurisdiction level of analysis, for example, county level or city/town level. When conducting county level analysis, the option is given to users to specify whether to include the sublevels (i.e., cities and towns) in the county.
- **Period of Analysis:** This option allows users to specify the start and end dates of crash analysis. In this model, all annualized data outputs are converted to whole-year values.
- **Alcohol Involvement:** Sometimes the involvement of alcohol relevant crashes would lead to the misleading information regarding remedial treatments selections. This option allows users to include or exclude the alcohol and drug-related crashes from the analysis in terms of studies of different purposes.
- **Location Reference:** The model provides four options for identifying a crash site, such as route-specific, junction-specific, junction–specific (junction), and junction–specific (intersection), all of which rely on cross streets to identify the reference point, rather than the milepost method. The cross street reference method suffers to some difficulties, for example, the lack of uniform geographical coordinates for site positioning, the infeasibility of the floating segment length, and so on.
- **Distance (Radius):** This selection parameter allows users to select the aggregation distance in terms of the specified location reference method. Any distance between 0 feet and 1 mile is optional, and an unlimited distance option is also available.

The four weighting methods used to prioritize hazardous locations within each subgroup are outlined as follows:

- **Total incidence method:** This method is used to rank sites for prioritization according to the number of crashes recorded.
- **Total fatalities method:** This method is used to rank sites for prioritization according to the number of fatal crashes recorded.
- **Total fatalities and injuries method:** This method is used to rank sites for prioritization according to the number of fatal crashes recorded.
- **Severity-weighted method:** This method is similar to the total fatalities and injuries method, except that the relative difference of severity in terms of crash costs is also considered.

RECOMMENDED SOFTWARE CHANGES

The allowable user-specified selection parameters provide much convenience to the software users to conduct analysis on the crash data, and the four weighing methods employed to rank sites for prioritization are straightforward.

However, based on the comprehensive literature review and comparison of alternate HSID methods, it is known that the algorithms for conducting HSID in the current ALGSP model allow some room for improvement. The corresponding recommended software changes to enhance its identification capability are outlined as follows.

Incorporating the Functional Classification as an Additional User Selection Parameter

The current selection parameters allow the software users to include or exclude the alcohol and drug-related crashes, to focus on the intersection-related crashes or the crashes on the road segments, to specify the crash history for analysis, and to specify the jurisdiction level. These parameters help the users to classify the crash data into alternate groups and are thus very convenient for doing studies of different purposes. However, considering the design criteria and level of service vary significantly according to the function of the highway facility, it is strongly recommended to embed the function classification of road entities as an additional user selection parameter. It is expected that grouping the crash data in this way would decrease much variation among the crash data and then increase the identification accuracy. In addition, the classification of crash data in this way can help the local governments (which have some problems of getting the required crash information) to perform the EB analysis of crash data (see the next two subsections).

Data Interface Improvement

It is suggested that additional data be incorporated into data files included with the current ALGSP model. For example, traffic counts are an optional data element in the software and are recommended after priority ranking has been established. From the former analyses it is known that exposure information (e.g., traffic counts, AADT, VMT, etc.) – perhaps the most important factor influencing road safety – is very important for developing accurate safety performance functions of various road entities. It also serves as the basis of EB implementation. Other valuable information required by sophisticated application includes shoulder width, width of lane, posted speed limits, number of lanes, etc.

A noteworthy point is that obtaining the somewhat detailed data may be problematic for local governments currently. In this case, the Bayesian correction based on the sample moments method is suggested to be performed to obviate the RTM biases, in which only the recorded crash number of the similar sites is required. The more common traits shared by the similar sites, the more accurate the identification result will be. Therefore, implementation of this type of the EB analysis requires at least that the software can group the road sites in terms of functional classification, and by various subcategories within functional classification if possible.

Exploring the Relationship between Exposure and Safety as Employed in the ALGSP

The exploration of the relationship between exposure and safety is recommended due to its significance in the implementation of Bayesian techniques. The relationship will be used to determine the expected crash counts of the reference population and it is likely to be non-linear across the different functional classifications of road sites. Although the Arizona road segment SPFs within this report can be used as an optional input for the ALGSP, they would need to be updated at some future time (they represent year 2000 data). These models should be updated regularly (e.g. every five years) by using more current data when conducting HSID. In addition, the relationships of traffic exposure and safety of Arizona intersections are not explored in this report and would need to be developed.

Incorporation of the EB Techniques to Calculate the Expected Crash Number

The experiment design demonstrates that under a range of practical conditions, the EB method offers around 50% reductions in the percentages of false positives and false negatives compared to CI and SR methods. By using SPFs of Arizona road sections, the EB method outperforms the accident frequency method, accident reduction potential method, and the accident rate method in most of the five evaluation tests based on 3-year crash data by showing the best site and method consistency, possessing the least total ranking differences, and yielding the least false identifications. Therefore, considering the great advantages associated with the EB method which can ensure that subsequent resources are directed to the truly hazardous locations, it is strongly recommended that the EB techniques be embedded in the current ALGSP model even though the additional resources are required to support data needs as well as the software modifications.

As mentioned previously, there are two kinds of EB techniques: the multivariate regression-based EB technique and the sample moments-based EB technique. Both are proved to obviate the RTM biases to some extent (the regression method is more accurate but requires more data). For the traffic agencies which have the detailed crash information available, it is suggested to use the multivariate regression method, whereas for the local governments which have problems collecting the required crash data, the method of sample moments is suggested. The corresponding flowchart of conducting EB analysis is shown in Figure 10.

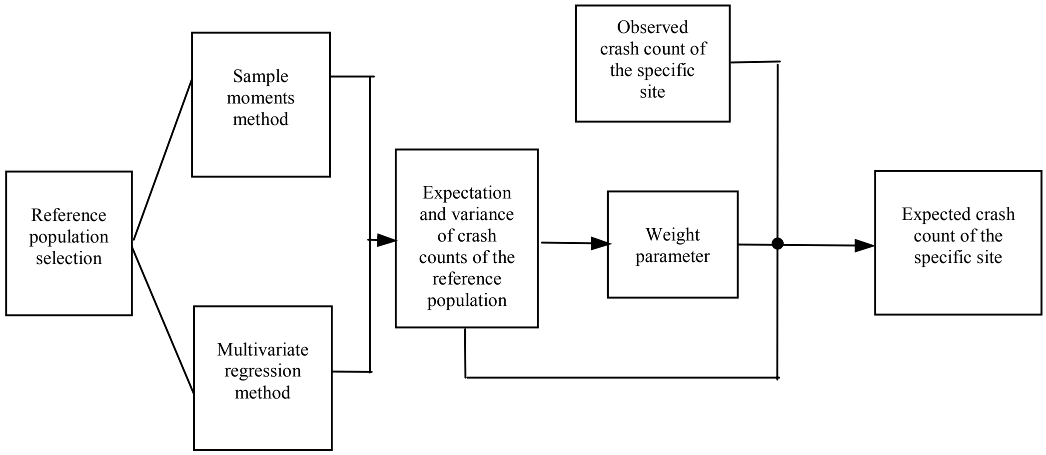


Figure 10: The Flowchart of Conducting EB Analysis

Incorporation of Accident Reduction Potential Method

Currently one of the disputes about road safety issue remaining unresolved is regarding which kind of sites should be identified as the hot spots, the sites with large expected crash counts or sites with high accident reduction potential. The practice to identify sites possessing large numbers of crashes on the road network is based on the assumption that the application of subsequent remedial treatments on the identified sites is to decrease the expected crash counts and/or severity by some fixed proportion (Hauer, et al., 2002). This assumption is echoed in the general usage of accident reduction factors (ARFs) to calculate the safety benefit of countermeasures, which is also the logic employed by ALGSP. Whereas the underlying assumption of employing the ARP value to screen out the hazardous locations is that the implementation of remedial treatments is to decrease the excess of the expected accident count and/or severity over what is 'normal' at similar sites (Hauer, et al., 2002). It follows that only the excess is reducible and it is more reasonable from the practical point of view.

Up to now there is lack of consensus on which criterion in general (or in some cases) is better than another for judging HSID methods. Hence, the hot spots resulting from the sole usage of either of the standards thus might be subject to the potential waste of investment. In addition, the results in chapter V indicate that each false identification resulting from the ARP method has the slightest consequences than those brought by the other three HSID methods such as the accident frequency method, the accident rate method, and EB method. Therefore, it is recommended that in addition to the incorporation of EB techniques to compute the expected crash count at selected sites (e.g. intersections or segments), the accident reduction potential should also be incorporated into the ALGSP as an additional weighting method to prioritize hazardous sites. It is expected that the remedial treatments applied to the sites showing both the high number of expected crash counts and the high ARP value would yield higher benefits.

The computation of the accident reduction potential is based on the expected crash number of specific sites and the associated reference population. Hence, similar to the EB technique, there are also two types of the ARP methods, or the method of sample moments and the multivariate regression method. The corresponding flowchart for calculating the accident reduction potential is illustrated in Figure 11.

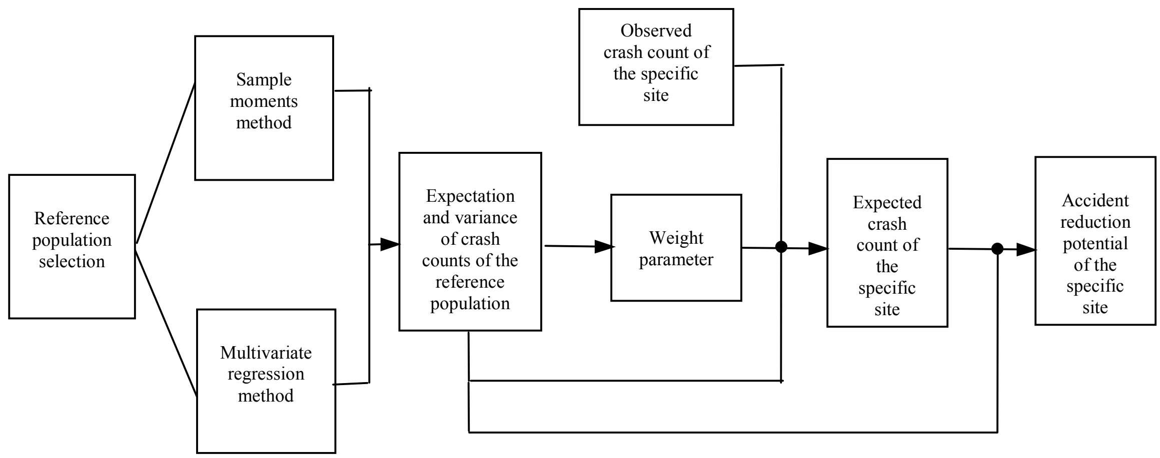


Figure 11: The Flowchart of Computing Accident Reduction Potential

Incorporation of the EB Techniques to Calculate the Expected Crash Costs

The severity-weighted method in the current ALGSP model offers software users the difference of severity in terms of crash costs. However, the crash costs are computed simply as the production of the observed number of crashes of various severities and the crash cost estimates of different severity levels, and thus is subject to the RTM bias. Considering the consistent advantages associated with the EB techniques, it is recommended that the EB techniques be used to compute the expected crash costs in the ALGSP in the future. Again, the method of sample moments can be used to compute the number of crashes of different levels of severity when the crash information is limited, and the multivariate regression method is suggested if the data requirement is satisfied. A noteworthy point is that simultaneity should be considered in some cases when developing the corresponding SPFs (Quyang, et al., 2002; Ladron, et al., 2004) in the sense that the crashes of different severities might share common characteristics. However, this correction is not mandatory and may be beyond the general level of capability at most DOTs and agencies wishing to implement the software.

Recommended Period of Analysis for Software Users

This recommendation is unrelated to the ALGSP model, but it is outlined herein due to the significant impacts of crash analysis on the accuracy of results. The experiment results show that 3-year crash histories represent the largest portion of the “best” study period of crash history, and 3 through 6 years make up almost 90% of all the optimum t-years. Hence, as the trade-off between the long and short histories, if there is no significant physical change in the location under securitization, it is suggested that the most recent 6-years of crash records is sufficient for the ALGSP model to capture the majority of the beneficial effect of the crash history. By contrast, 3-years of crash history data represent the ‘shortest’ period of time that should be used. Crash histories of 1 and 2 years provide relatively little benefit in most situations. Plus, it is important to note that Hazard Elimination Projects (HES) eligibility guidelines require that data from a period of at least 3 years be included in the site analysis.

While much of this research is focused on improving the ALGSP, it is worth noting that the ALGSP already incorporates many useful algorithms and information that will aid local jurisdictions in identifying hot spots. These include but are not limited to the inclusion of numerous crash modification factors associated with engineering countermeasures, the ability to upload crash data, the inclusion of costs of numerous countermeasures, and the ability to conduct cost-benefit analyses. The addition of Bayesian techniques and other recommended changes within the ALGSP software represents a small portion of the capabilities of the software and the enormous effort invested to date in the software.

REFERENCES

Brown, B., C. Farley, and M. Forgues. Identification of Dangerous Highway Locations: Results of Community Health Department Study in Quebec. In Transportation Research Record 1376, TRB, National Research Council, Washington D.C., 1992.

Carey, J. Arizona Local Government Safety Project Analysis Model. Final Report 504, Arizona Dept. of Transportation, Phoenix, AZ., 2001.

Davis, G. A. and S. Yang. Bayesian Identification of High-Risk Intersections for Older Drivers via Gibbs Sampling. In Transportation Research Record 1746, TRB, National Research Council, Washington D.C., 2001.

Deacon, J. A., C. V. Zegeer, and R. C. Deen. Identification of Hazardous Rural Highway Locations. In Transportation Research Record 543, TRB, National Research Council, Washington, D.C., 1975.

Flak, M. A., and J. C. Barbaresso. Use of Computerized Roadway Information System in Safety Analyses. In Transportation Research Record 844, TRB, National Research Council, Washington D.C., 1982.

Hakkert, A. S. and D. Mahalel. Estimating the Number of Accidents at Intersections from a Known Traffic Flow on the Approaches. Accident Analysis and Prevention, Vol.10, No.1, 1978.

Hauer, E. and B. N. Persaud. Problem of Identifying Hazardous Locations Using Accident Data. In Transportation research Record 975, TRB, National Research Council, Washington D.C., 1984.

Hauer, E., J. C. N. Ng, and J. Lovell. Estimation of Safety at Signalized Intersections. In Transportation Research Record 1185, TRB, National Research Council, Washington D.C., 1988.

Hauer, E. Empirical Bayes' Approach to the Estimation of Unsafe: The Multivariate Regression Approach. Accident Analysis and Prevention, Vol. 24, No.5, Oct. 1992.

Hauer, E., K. Quaye, and Z. Liu. On the use of Accident or Conviction Counts to Trigger Action. In Transportation Research Record 1401, TRB, National Research Council, Washington D.C., 1993.

Hauer, E. Identification of Sites with Promise. In Transportation Research Record 1542, TRB, National Research Council, Washington D.C., 1996.

Hauer, E. Observational Before-After Studies in Road Safety. PERGAMON, 1997.

Hauer, E., D. W. Harwood, F.M. Council, and M. S. Griffith. Estimating Safety by the Empirical Bayes Method: A Tutorial. In Transportation Research Record 1784, TRB, National Research Council, Washington, D.C., 2002.

Hauer, E., J. Kononov, B. Allery, and M. S. Griffith. Screening the Road Network for Sites with Promise. In Transportation Research Record 1784, TRB, National Research Council, Washington, D.C., 2002.

Higle J. L. and J. M. Witkowski. Bayesian Identification of Hazardous Locations. In Transportation Research Record 1185, TRB, National Research Council, Washington, D.C., 1988.

Higle J. L. and M. B. Hecht, A comparison of Techniques for the Identification of Hazardous Locations. In Transportation Research Record 1238, TRB, National Research Council, Washington D.C., 1989.

Kelton, W. D., R. P. Sadowski, and D.T. Sturrock. Simulation with Arena. The McGraw-Hill Companies, Inc., NY., 2003.

Kononov, J. and B. N. Jason. Diagnostic Methodology for the Detection of Safety Problems at Intersections. In Transportation Research Record 1784, TRB, National Research Council, Washington, D.C., 2002.

Ladron, F., S. Washington, and J. Oh. Forecasting Crashes at the Planning Level. In Transportation research Record 1897, TRB, National Research Council, Washington D.C., 2004.

Laughland, J. C., L. E. Haefner, J. W. Hall, and D. R. Clough. NCHRP Report 162: Methods for Evaluating Highway Safety Improvements. TRB, National Research Council, Washington D.C., 1975.

Mahal, D., A. S. Hakkert, and J. N. Phrasker. A system for the allocation of Safety Resources on a Road Network. Accident Analysis and Prevention, Vol. 14, No.1, 1982.

Mak, K. K., T. Chira-Chavala, and B. A. Hilger. Automated Analysis of High-Accident Locations. In Transportation Research Record 1068, TRB, National Research Council, Washington D.C., 1985.

May, J. F. A Determination of Accident Prone Location. Traffic Engineering, Vol. 34, No.5, Feb. 1964.

McGuigan, D. R. D. The use of Relationships between Road Accidents and Traffic Flow in ‘Black-Spot’ Identification. Traffic Engineering and Control, Aug.-Sept., 1981.

McGuigan, D. R. D. Nonjunction Accident Rates and their Use in ‘Black-Spot’ Identification. Traffic Engineering and Control, 1982.

Morin, D. A.. Application of Statistical Concepts to Accident Data. In Highway Research Record 188, HRB, National Research Council, Washington, D.C., 1967.

NCHRP (National Cooperative Highway Research Program). Safety Management Systems, A Synthesis of Highway Practice. Transportation Research Board, National Academies, Washington D.C., 2003.

Norden, N., J. Orlansky, and H. Jacobs. Application of Statistical Quality-Control Techniques to Analysis of Highway Accident Data. Bulletin 117, HRB, National Council Washington D.C., 1956.

Persaud, B. N. and E. Hauer. Comparison of Two Methods for Debiasing Before-and-After Accident Studies. In Transportation Research Record 975, TRB, National Research Council, Washington D.C., 1984.

Persaud, B. N. Estimating Accident Potential of Ontario Road Sections. In Transportation Research Record 1327, TRB, National Research Council, Washington D.C., 1991.

Persaud, B., C. Lyon, and T. Nguyen. Empirical Bayes Procedure for Ranking Sites for Safety Investigation by Potential for Safety Improvement. In Transportation Research Record 1665, TRB, National Research Council, Washington D.C., 1999.

Quyang Y., V. Shankar, and T. Yamamoto. Modeling the Simultaneity in Injury Causation in Multivehicle Collisions. In Transportation research Record 1784, TRB, National Research Council, Washington D.C., 2002.

Renshaw, D. L., and E. C. Charter. Identification of High Hazard Locations in Baltimore County Road-Rating Project. In Transportation Research Record 753, TRB, National Research Council, Washington D.C., 1980.

SAS Users Manual. SAS Institute Inc., Cary, NC., 1998.

Stokes, R. W., and M. I. Mutabazi. Rate-Quality Control Method of Identifying Hazardous Road Locations. In Transportation Research Record 1542, TRB, National Research Council, Washington D.C., 1996.

Tamburri, T. N., and R. N. Smith. The safety Index: Method of evaluating and Rating Safety Benefits. In Highway Research Record 332, HRB, National Research Council, Washington D.C., 1970.

Tarko, A. P., K. C. Sinha, and O. Farooq. Methodology for Identifying Highway Safety Problem Areas. In Transportation Research Record 1542, TRB, National Research Council, Washington D.C., 1996.

Washington, S., M. G. Karlaftis, and F. L. Mannering, Statistical and Econometric Methods for Transportation Data Analysis, Chapman & Hall, 2003.

Wright, C. C., C. R. Abbess and D. F. Jarrett. Estimating the Regression-to-Mean Effect Associated with Road Accident Black Spot Treatment: Towards a More Realistic Approach. Accident Analysis and Prevention, Vol.20, No.3, 1988.

APPENDIX A: REAL ARIZONA CRASH DATA USED FOR THE DEVELOPMENT OF SIMULATED CRASH DATA

In the Section III of this report, it is known that real Arizona crash data are required for the development of simulated crash data used in the experiment design. The six datasets, which represent the crash counts from intersections in six counties of Arizona, are shown in this appendix. Tables 21~26 present the frequency and cumulative percentage of observed crash count from each county. Figures 12~17 illustrate the empirical cumulative distribution of each dataset. It is noticeable that the six distributions follow into three kinds of shapes (i.e., exponential, linear, and "s" shapes) and the crash counts show the two levels of heterogeneity.

Table 21: Observed Data from Apache (E1)

x	1	2	3	4	5	7	8	9	11	12	14	15	19
n(x)	10	13	10	14	17	20	13	11	11	6	7	4	6
CDF	7	16.2	23.2	33.1	45.1	59.2	68.3	76.1	83.8	88	93	95.8	100

Note: x=observed crash count; n(x) =number of sites with count x; CDF=cumulative % of data observed with crash count ≥ x.

Table 22: Observed Data from Gila (E2)

x	2	4	5	6	8	9	11	13	15	16	17	18	19
n(x)	2	4	3	2	4	1	5	4	4	2	4	4	2
CDF	2	6	8	10	14	15	21	25	28	29	33	37	39
x	21	23	24	26	27	28	29	31	32	33	34	35	36
n(x)	3	3	2	3	1	3	3	2	1	1	2	1	2
CDF	42	45	47	50	51	54	57	59	60	61	63	64	66
x	37	38	39	40	41	42	46	48	49	52	53	54	55
n(x)	2	3	3	3	2	1	4	3	1	2	2	1	2
CDF	68	69	72	75	77	78	82	85	86	88	90	91	93
x	56	57	62	66	72								
n(x)	1	1	2	2	1								
CDF	94	95	97	99	100								

Note: x=observed crash count; n(x) =number of sites with count x; CDF=cumulative % of data observed with crash count ≥ x.

Table 23: Observed Data from Graham (L1)

x	1	2	3	4	5	6	7	8	9	10	11	14	15
n(x)	5	4	3	8	7	4	5	6	3	9	5	4	5
CDF	7.0	12.7	16.9	28.2	38.0	43.7	50.7	59.2	63.4	76.1	83.1	88.7	95.8
x	16												
n(x)	3												
CDF	100.0												

Note: x=observed crash count; n(x) =number of sites with count x; CDF=cumulative % of data observed with crash count ≥ x.

Table 24: Observed Data from Lapaz (L2)

x	4	5	7	8	9	12	14	15	16	17	18	19	21
n(x)	1	2	1	1	2	3	1	1	2	4	1	2	3
CDF	1	3	4	5	7	10	11	12	14	18	19	21	24
x	23	24	25	26	27	29	30	31	32	34	35	36	37
n(x)	1	2	7	3	2	1	1	2	6	4	5	1	2
CDF	25	27	34	37	39	40	41	43	49	53	58	59	61
x	38	43	45	47	49	51	52	55	57	59	61	62	64
n(x)	2	7	3	2	2	1	1	2	1	2	1	1	3
CDF	63	70	73	75	77	78	79	81	82	84	85	86	89
x	68	69	70										
n(x)	1	1	1										
CDF	90	91	92										

Note: x=observed crash count; n(x) =number of sites with count x; CDF=cumulative % of data observed with crash count ≥ x.

Table 25: Observed Data from Pima (S1)

x	1	2	3	4	5	6	7	8	9	10	11	13	14
n(x)	3	3	1	3	15	19	22	20	24	20	8	6	10
CDF	1.8	3.6	4.2	6	15	26.3	40	51.5	65.9	77.8	82.6	86.2	92.2
x	15	17	18										
n(x)	7	3	3										
CDF	96.4	98.2	100										

Note: x=observed crash count; n(x) =number of sites with count x; CDF=cumulative % of data observed with crash count ≥ x.

Table 26: Observed Data from Santacruz (S2)

x	1	2	4	5	6	7	8	9	10	11	13	14	15
n(x)	1	2	1	2	1	2	1	1	5	4	5	6	8
CDF	0.9	2.6	3.4	5.1	6	7.7	8.5	9.4	13.7	17.1	21.3	26.5	33.3
x	16	17	18	21	23	24	25	27	28	29	31	33	34
n(x)	7	5	3	3	4	6	4	4	2	3	2	4	3
CDF	39.3	43.6	46.2	48.7	52.1	57.3	60.7	64.1	65.8	68.4	70.1	73.5	76.1
x	35	36	38	40	41	43	44	46	48	52	57	61	
n(x)	5	3	3	4	2	3	1	2	2	1	1	1	
CDF	80.3	82.9	85.5	88.9	90.6	93.2	94	95.7	97.4	98.3	99.1	100	

Note: x=observed crash count; n(x) =number of sites with count x; CDF=cumulative % of data observed with crash count ≥ x.

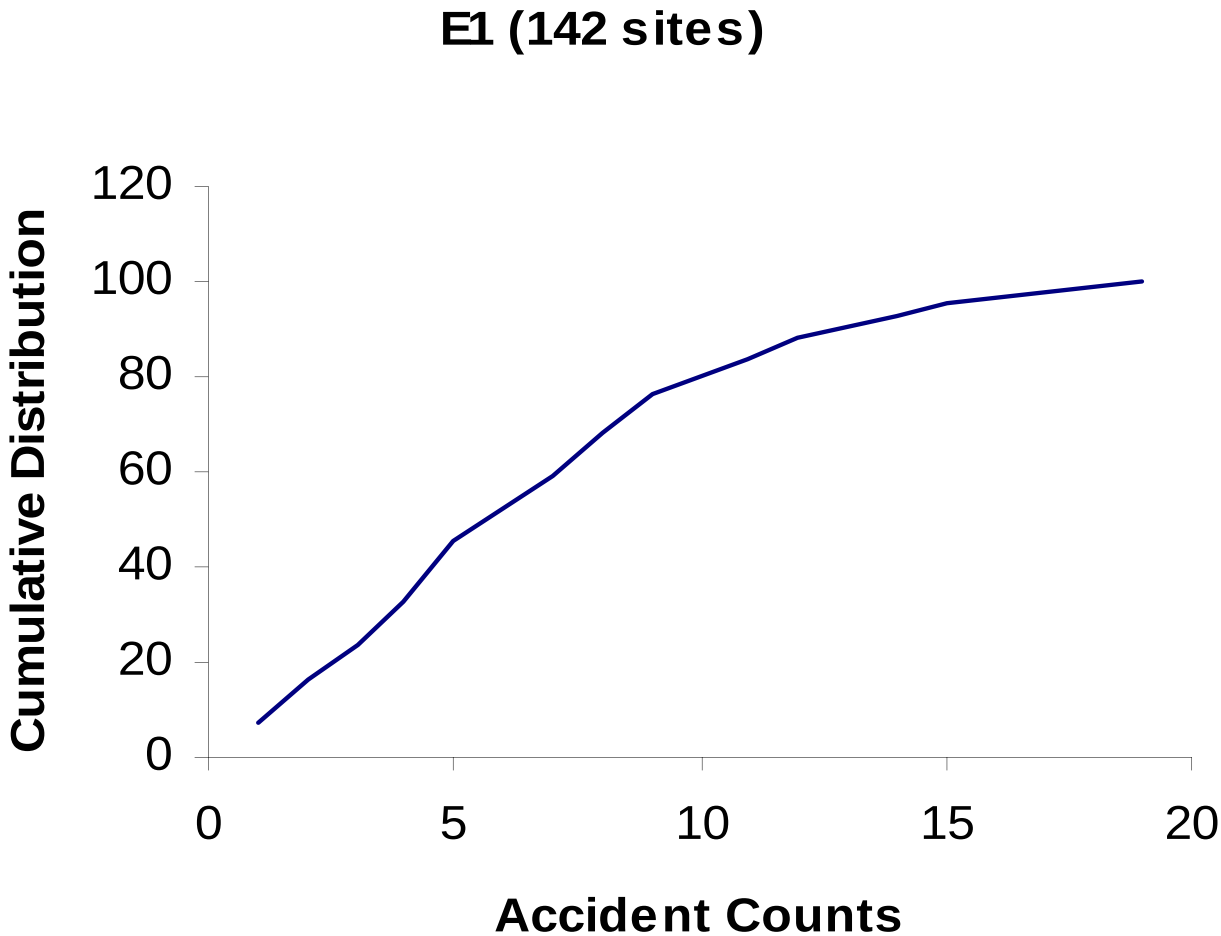


Figure 12: Empirical Cumulative Distribution of Dataset One (E1)

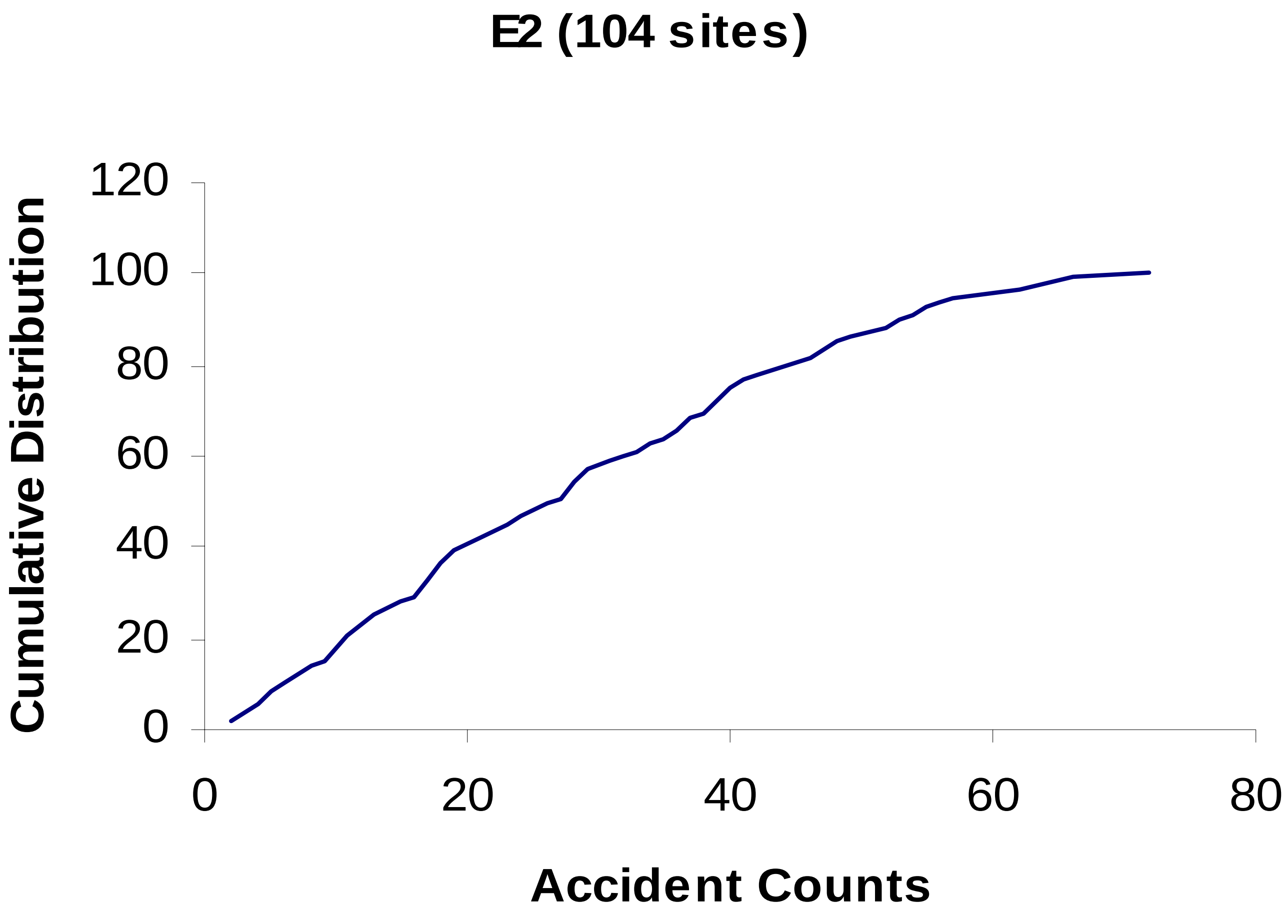


Figure 13: Empirical Cumulative Distribution of Dataset Two (E2)

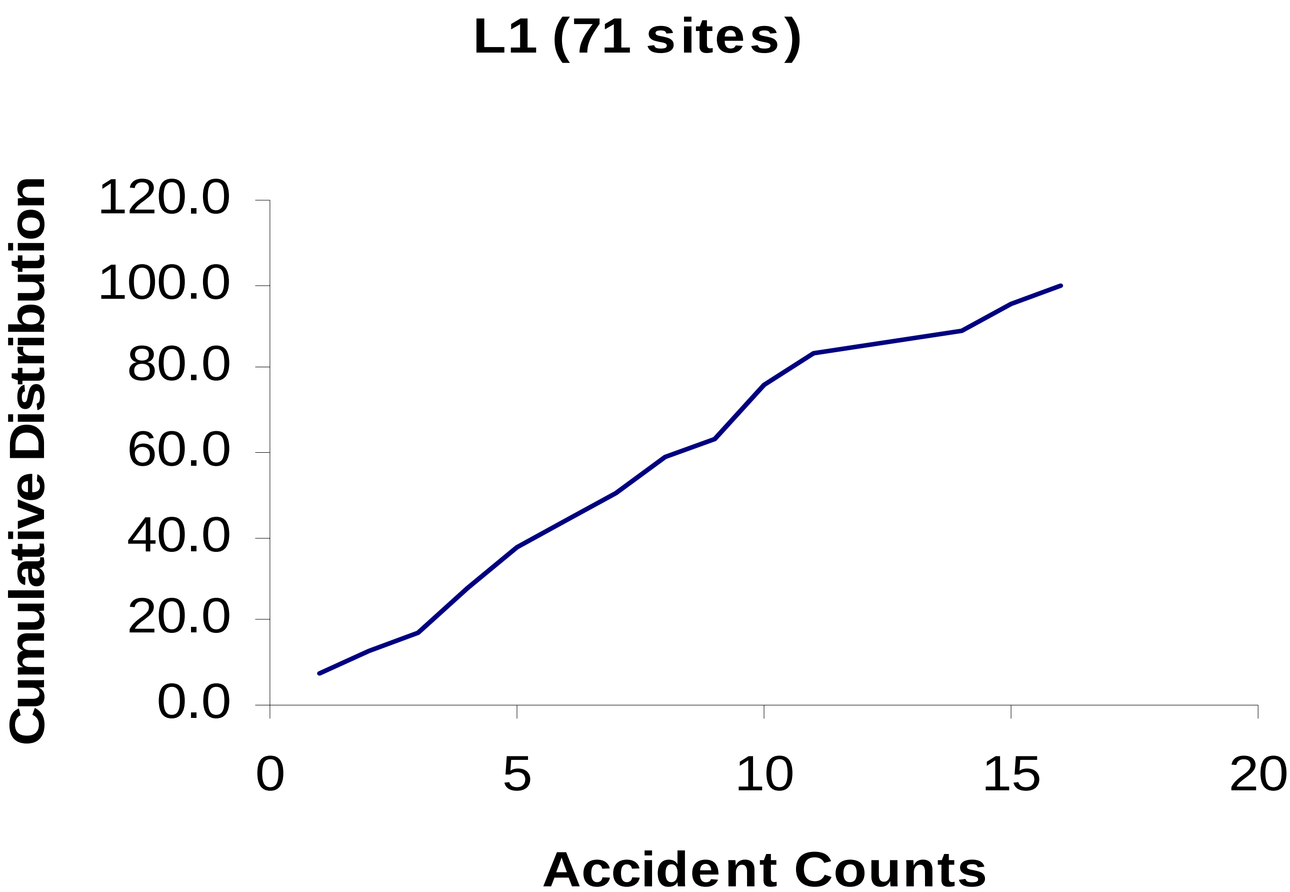


Figure 14: Empirical Cumulative Distribution of Dataset Three (L1)

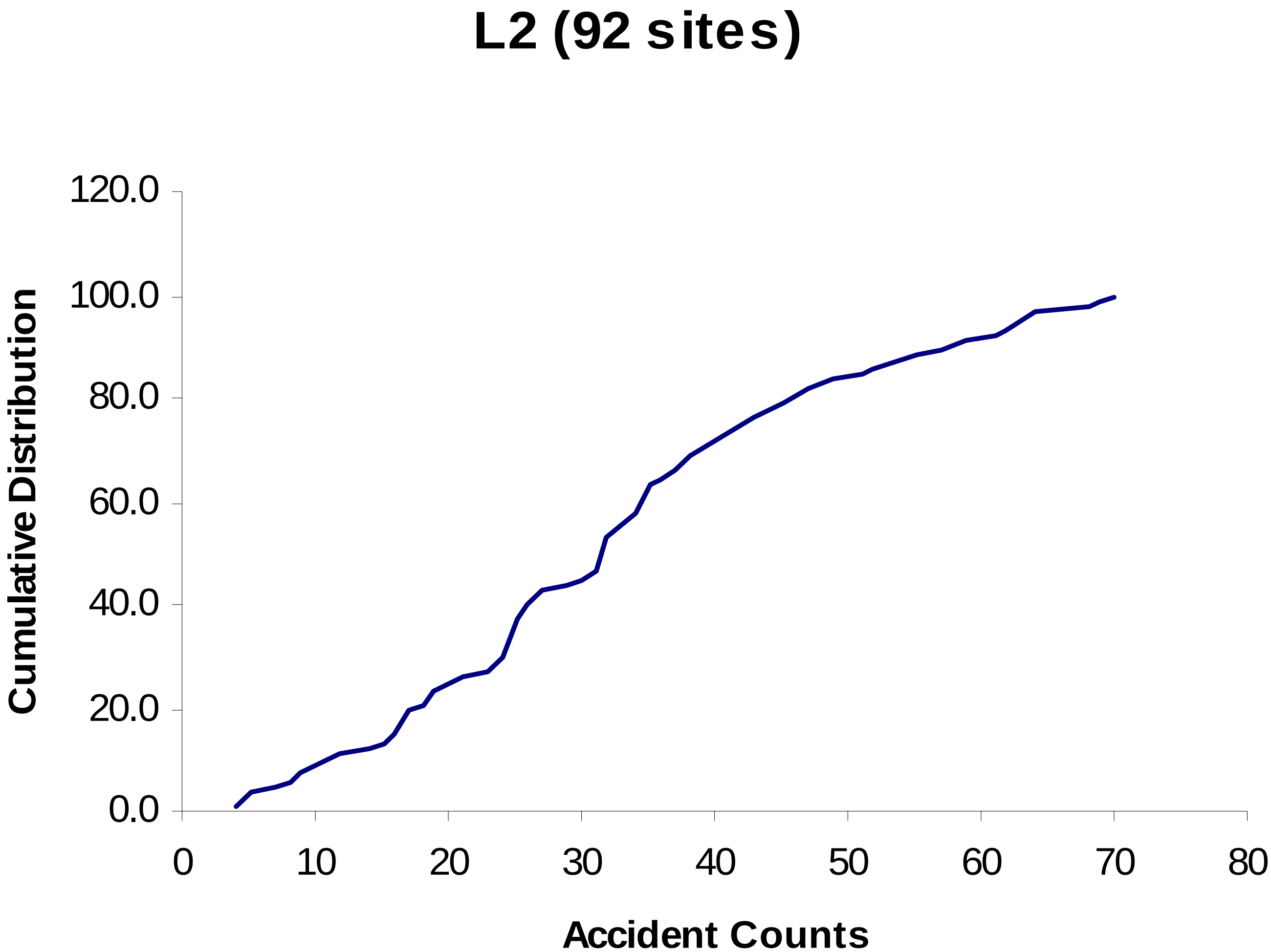


Figure 15: Empirical Cumulative Distribution of Dataset Four (L2)

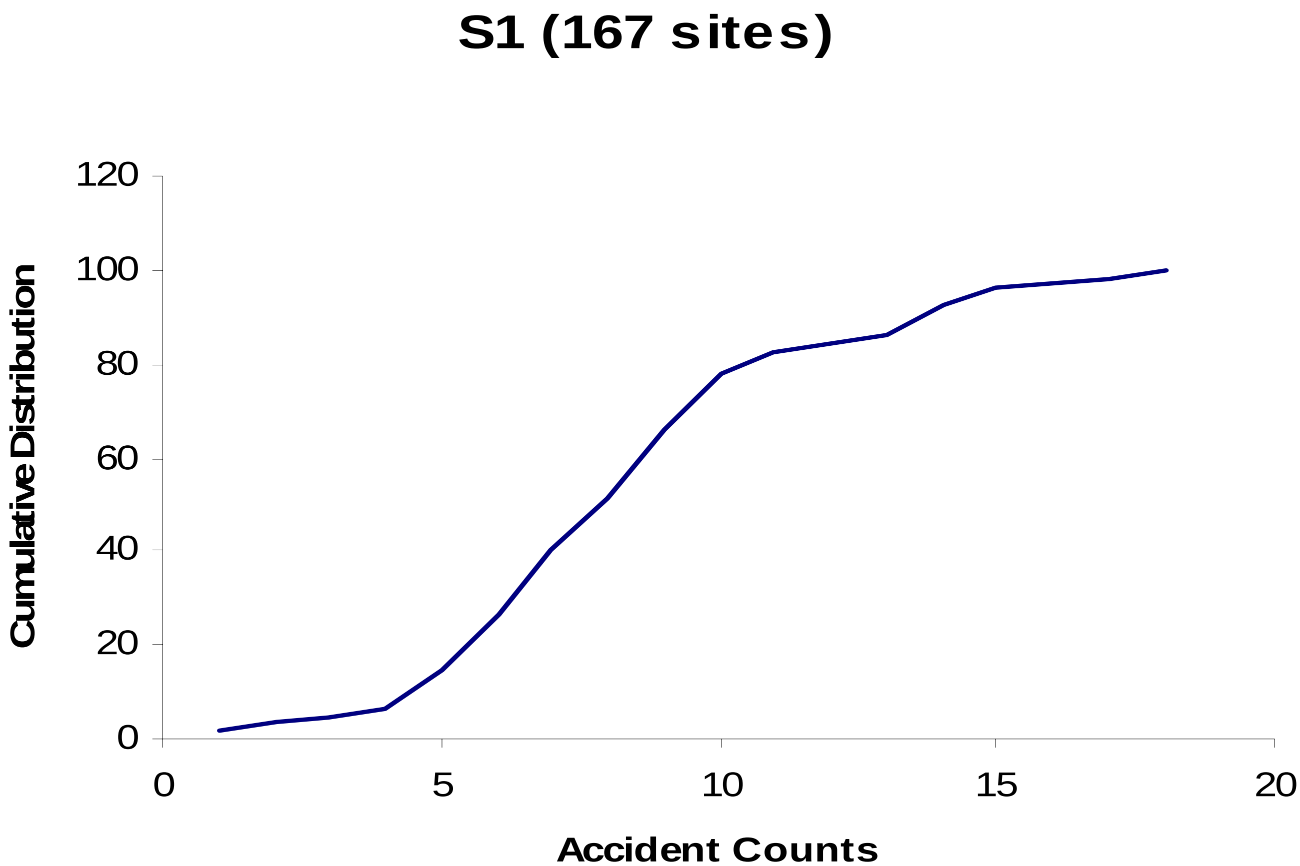


Figure 16: Empirical Cumulative Distribution of Dataset Five (S1)

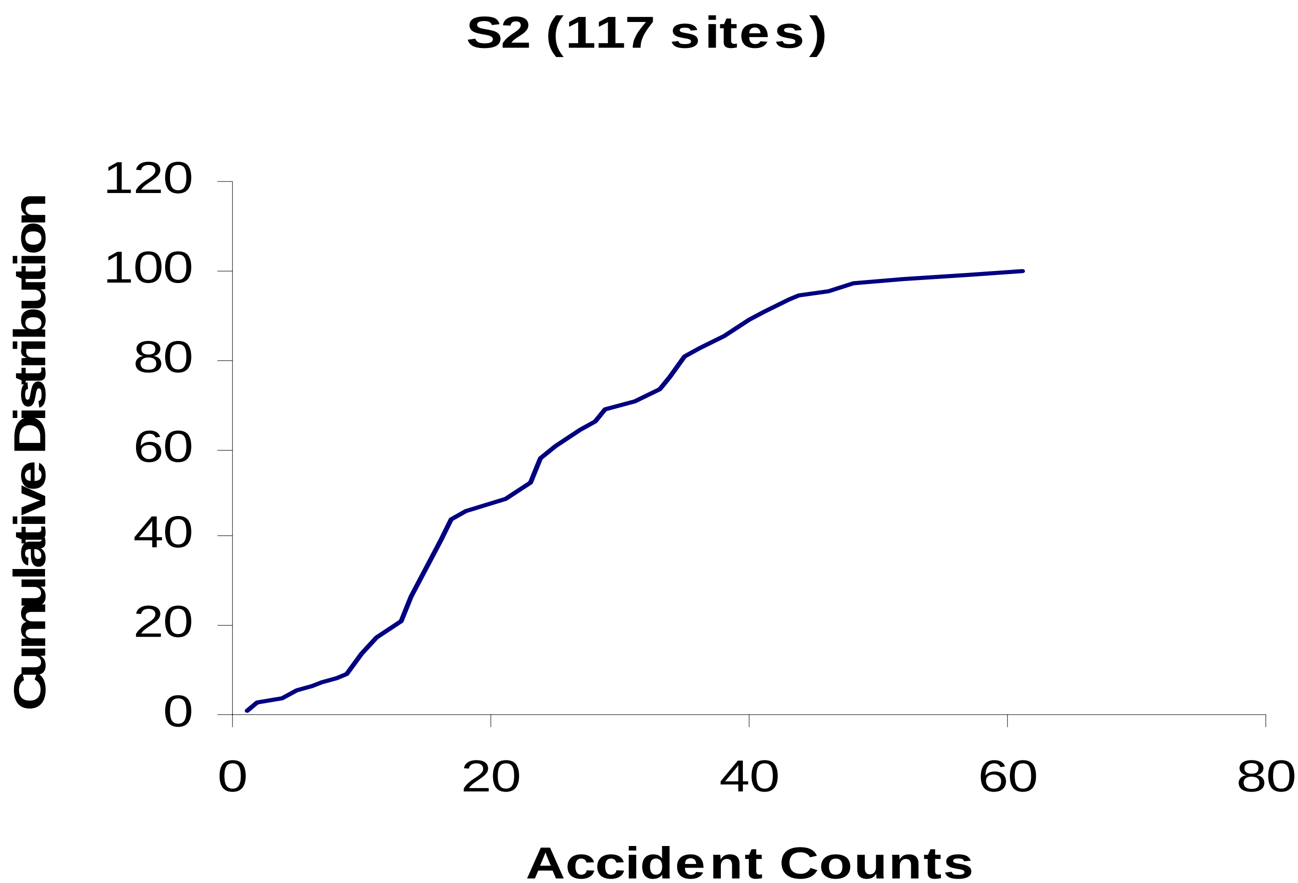


Figure 17: Empirical Cumulative Distribution of Dataset Six (S2)

APPENDIX B: THE IDENTIFICATION ERROR RATES ASSOCIATED WITH VARIOUS HSID METHODS, CONFIDENCE LEVELS, AND GROUPS

It is known from chapter III that the identification error rate is employed as the indicator to explore the best study duration for crash data analysis. Various "t" years are plotted against the associated identification error rate and the "knee" of the curve would be identified as the optimal study period. This appendix presents the three kinds of identification error rates (FN, FP, and FI) of the different "t" years across the three HSID methods, three confidence levels and three groups, which are shown in Tables 27~53.

Table 27: The Identification Error Rates of SR Method for Group 1 (δ = 0.90)

		1	2	3	4	5	6	7	8	9	10
E1	**FN**	3.80	2.85	2.32	2.13	1.83	1.73	1.67	1.70	1.67	1.33
	FP	34.20	25.67	20.88	19.14	16.50	15.60	15.00	15.33	15.00	12.00
	FI	6.84	5.13	4.18	3.83	3.30	3.12	3.00	3.07	3.00	2.40
E2	**FN**	2.17	1.46	1.28	1.10	0.89	0.78	0.58	0.52	0.50	0.56
	FP	19.50	13.11	11.50	9.86	8.00	7.00	5.25	4.67	4.50	5.00
	FI	3.90	2.62	2.30	1.97	1.60	1.40	1.05	0.93	0.90	1.00
L1	**FN**	4.20	3.26	2.75	2.49	2.22	1.93	1.92	1.74	1.61	1.67
	FP	37.80	29.33	24.75	22.43	20.00	17.40	17.25	15.67	14.50	15.00
	FI	7.56	5.87	4.95	4.49	4.00	3.48	3.45	3.13	2.90	3.00
L2	**FN**	2.54	2.00	1.63	1.59	1.46	1.31	1.28	1.15	1.06	1.11
	FP	22.90	18.00	14.63	14.29	13.17	11.80	11.50	10.33	9.50	10.00
	FI	4.58	3.60	2.93	2.86	2.63	2.36	2.30	2.07	1.90	2.00
S1	**FN**	6.36	5.43	4.88	4.68	4.43	4.09	3.94	3.85	3.56	3.33
	FP	57.20	48.89	43.88	42.14	39.83	36.80	35.50	34.67	32.00	30.00
	FI	11.44	9.78	8.78	8.43	7.97	7.36	7.10	6.93	6.40	6.00
S2	**FN**	2.79	2.31	2.04	1.86	1.76	1.64	1.58	1.37	1.44	1.44
	FP	25.10	20.78	18.38	16.71	15.83	14.80	14.25	12.33	13.00	13.00
	FI	5.02	4.16	3.68	3.34	3.17	2.96	2.85	2.47	2.60	2.60

Table 28: The Identification Error Rates of ER Method for Group 1 (δ = 0.90)

		1	2	3	4	5	6	7	8	9	10
E1	FN	2.47	1.73	1.53	1.51	1.44	1.31	1.31	1.19	1.22	1.22
	FP	22.20	15.56	13.75	13.57	13.00	11.80	11.75	10.67	11.00	11.00
	FI	4.44	3.11	2.75	2.71	2.60	2.36	2.35	2.13	2.20	2.20
E2	FN	1.19	0.83	0.67	0.59	0.54	0.60	0.47	0.48	0.50	0.56
	FP	10.70	7.44	6.00	5.29	4.83	5.40	4.25	4.33	4.50	5.00
	FI	2.14	1.49	1.20	1.06	0.97	1.08	0.85	0.87	0.90	1.00
L1	FN	3.03	2.40	2.13	2.00	1.85	1.87	1.78	1.78	1.72	1.67
	FP	27.30	21.56	19.13	18.00	16.67	16.80	16.00	16.00	15.50	15.00
	FI	5.46	4.31	3.83	3.60	3.33	3.36	3.20	3.20	3.10	3.00
L2	FN	1.87	1.41	1.36	1.29	1.28	1.20	1.08	1.04	1.06	1.11
	FP	16.80	12.67	12.25	11.57	11.50	10.80	9.75	9.33	9.50	10.00
	FI	3.36	2.53	2.45	2.31	2.30	2.16	1.95	1.87	1.90	2.00
S1	FN	5.39	4.75	4.33	4.11	4.02	3.91	3.92	3.81	3.61	3.33
	FP	48.50	42.78	39.00	37.00	36.17	35.20	35.25	34.33	32.50	30.00
	FI	9.70	8.56	7.80	7.40	7.23	7.04	7.05	6.87	6.50	6.00
S2	FN	2.08	1.81	1.65	1.59	1.56	1.51	1.08	1.41	1.33	1.44
	FP	18.70	16.33	14.88	14.29	14.00	13.60	9.75	12.67	12.00	13.00
	FI	3.74	3.27	2.98	2.86	2.80	2.72	1.95	2.53	2.40	2.60

Table 29: The Identification Error Rates of CI Method for Group 1 (δ = 0.90)

		1	2	3	4	5	6	7	8	9	10
E1	FN	2.70	2.50	2.02	1.57	1.21	1.14	1.03	1.15	1.00	0.89
	FP	62.06	35.51	28.06	27.03	24.84	22.94	24.02	21.90	20.59	25.49
	FI	7.47	5.87	4.68	4.17	3.62	3.36	3.38	3.27	3.00	3.40
E2	FN	1.81	1.15	1.08	0.67	0.60	0.61	0.51	0.37	0.51	0.56
	FP	25.00	18.23	13.96	11.71	9.46	8.65	8.11	7.21	6.31	8.11
	FI	4.22	2.99	2.51	1.90	1.58	1.50	1.35	1.13	1.15	1.40
L1	FN	2.28	2.91	2.56	2.28	1.81	1.55	1.27	1.15	1.11	1.11
	FP	104.29	46.67	33.16	29.88	26.70	24.29	24.23	24.15	23.47	22.45
	FI	8.71	6.86	5.56	4.99	4.25	3.78	3.53	3.40	3.30	3.20
L2	FN	1.91	1.93	1.72	1.71	1.74	1.55	1.37	1.31	1.40	1.01
	FP	37.61	16.10	14.37	14.29	14.49	12.90	11.45	10.90	11.68	8.41
	FI	5.28	3.44	3.08	3.06	3.10	2.76	2.45	2.33	2.50	1.80
S1	FN	0.43	1.24	1.10	3.45	3.10	2.98	2.65	2.52	2.30	2.08
	FP	832.73	248.25	228.21	63.95	58.91	57.21	55.23	50.78	49.42	58.14
	FI	9.59	9.89	9.05	8.66	7.90	7.64	7.18	6.67	6.35	6.90
S2	FN	2.26	2.06	2.10	2.04	1.71	1.71	1.53	1.48	1.34	1.56
	FP	36.41	24.94	18.01	16.81	15.20	15.29	14.22	13.07	12.75	14.71
	FI	5.40	4.24	3.73	3.54	3.08	3.10	2.83	2.67	2.50	2.90

Table 30: The Identification Error Rates of SR Method for Group 1 (δ = 0.95)

		1	2	3	4	5	6	7	8	9	10
E1	**FN**	2.11	1.64	1.34	1.01	1.05	0.86	0.89	0.74	0.84	0.63
	FP	40.00	31.11	25.50	19.14	20.00	16.40	17.00	14.00	16.00	12.00
	FI	4.00	3.11	2.55	1.91	2.00	1.64	1.70	1.40	1.60	1.20
E2	**FN**	1.33	1.10	0.95	0.90	0.79	0.72	0.58	0.56	0.63	0.63
	FP	25.20	20.89	18.00	17.14	15.00	13.60	11.00	10.67	12.00	12.00
	FI	2.52	2.09	1.80	1.71	1.50	1.36	1.10	1.07	1.20	1.20
L1	**FN**	2.38	1.98	1.72	1.53	1.35	1.22	1.13	0.91	0.84	0.74
	FP	45.20	37.56	32.75	29.14	25.67	23.20	21.50	17.33	16.00	14.00
	FI	4.52	3.76	3.28	2.91	2.57	2.32	2.15	1.73	1.60	1.40
L2	**FN**	1.52	1.10	0.96	0.93	0.88	0.80	0.82	0.77	0.79	0.63
	FP	28.80	20.89	18.25	17.71	16.67	15.20	15.50	14.67	15.00	12.00
	FI	2.88	2.09	1.83	1.77	1.67	1.52	1.55	1.47	1.50	1.20
S1	**FN**	3.48	3.11	2.88	2.63	2.58	2.40	2.29	2.28	2.11	2.32
	FP	66.20	59.11	54.75	50.00	49.00	45.60	43.50	43.33	40.00	44.00
	FI	6.62	5.91	5.48	5.00	4.90	4.56	4.35	4.33	4.00	4.40
S2	**FN**	1.25	0.90	0.66	0.59	0.47	0.40	0.26	0.25	0.26	0.21
	FP	23.80	17.11	12.50	11.14	9.00	7.60	5.00	4.67	5.00	4.00
	FI	2.38	1.71	1.25	1.11	0.90	0.76	0.50	0.47	0.50	0.40

Table 31: The Identification Error Rates of EB Method for Group 1 (δ = 0.95)

		1	2	3	4	5	6	7	8	9	10
E1	**FN**	1.39	1.02	0.82	0.66	0.65	0.55	0.55	0.56	0.53	0.53
	FP	26.40	19.33	15.50	12.57	12.33	10.40	10.50	10.67	10.00	10.00
	FI	2.64	1.93	1.55	1.26	1.23	1.04	1.05	1.07	1.00	1.00
E2	**FN**	0.97	0.84	0.71	0.69	0.70	0.69	0.66	0.70	0.58	0.63
	FP	18.40	16.00	13.50	13.14	13.33	13.20	12.50	13.33	11.00	12.00
	FI	1.84	1.60	1.35	1.31	1.33	1.32	1.25	1.33	1.10	1.20
L1	**FN**	1.78	1.43	1.25	1.07	1.05	0.95	0.82	0.81	0.79	0.74
	FP	33.80	27.11	23.75	20.29	20.00	18.00	15.50	15.33	15.00	14.00
	FI	3.38	2.71	2.38	2.03	2.00	1.80	1.55	1.53	1.50	1.40
L2	**FN**	1.07	0.88	0.87	0.80	0.79	0.80	0.79	0.74	0.68	0.63
	FP	20.40	16.67	16.50	15.14	15.00	15.20	15.00	14.00	13.00	12.00
	FI	2.04	1.67	1.65	1.51	1.50	1.52	1.50	1.40	1.30	1.20
S1	**FN**	3.07	2.70	2.54	2.42	2.37	2.21	2.21	2.25	2.26	2.32
	FP	58.40	51.33	48.25	46.00	45.00	42.00	42.00	42.67	43.00	44.00
	FI	5.84	5.13	4.83	4.60	4.50	4.20	4.20	4.27	4.30	4.40
S2	**FN**	0.75	0.63	0.55	0.53	0.46	0.46	0.47	0.49	0.47	0.42
	FP	14.20	12.00	10.50	10.00	8.67	8.80	9.00	9.33	9.00	8.00
	FI	1.42	1.20	1.05	1.00	0.87	0.88	0.90	0.93	0.90	0.80

Table 32: The Identification Error Rates of CI Method for Group 1 (δ = 0.95)

		1	2	3	4	5	6	7	8	9	10
E1	**FN**	1.58	1.41	1.36	1.67	1.66	1.38	1.32	1.15	1.14	0.97
	FP	82.22	50.65	38.60	24.00	22.00	20.00	16.67	19.11	17.33	16.00
	FI	5.43	4.33	3.69	3.34	3.18	2.78	2.48	2.50	2.35	2.10
E2	**FN**	1.30	1.09	0.96	0.66	0.63	0.54	0.43	0.47	0.38	0.43
	FP	35.93	20.81	15.32	14.08	13.15	11.27	11.97	12.21	11.27	11.27
	FI	3.39	2.43	1.98	1.61	1.52	1.30	1.25	1.30	1.15	1.20
L1	**FN**	1.56	1.42	1.69	1.75	1.46	1.32	1.17	1.17	1.28	1.49
	FP	109.61	71.50	50.00	35.83	30.95	26.67	28.17	24.34	23.02	25.40
	FI	5.72	4.64	4.33	3.90	3.32	2.92	2.88	2.63	2.65	3.00
L2	**FN**	1.51	1.23	1.19	1.12	1.06	0.90	0.89	0.79	0.65	0.54
	FP	35.58	19.73	15.07	11.02	7.86	8.29	5.36	6.67	7.14	5.71
	FI	3.64	2.46	2.14	1.81	1.53	1.42	1.20	1.20	1.10	0.90
S1	**FN**	0.19	0.29	0.68	1.39	1.24	1.06	0.83	0.76	0.78	0.73
	FP	1226.00	497.98	239.38	115.92	109.05	102.29	100.71	98.10	98.57	94.29
	FI	6.32	5.77	5.45	5.40	5.02	4.60	4.33	4.17	4.20	4.00
S2	**FN**	1.19	0.81	0.70	0.76	0.76	0.83	0.93	0.85	0.91	0.85
	FP	41.60	25.84	20.83	15.83	13.81	10.75	7.94	5.29	3.97	4.76
	FI	3.36	2.21	1.85	1.66	1.55	1.44	1.38	1.13	1.10	1.10

Table 33: The Identification Error Rates of SR Method for Group 1 (δ = 0.99)

		1	2	3	4	5	6	7	8	9	10
E1	**FN**	0.52	0.46	0.40	0.36	0.35	0.28	0.28	0.30	0.30	0.30
	FP	51.00	45.56	40.00	35.71	35.00	28.00	27.50	30.00	30.00	30.00
	FI	1.02	0.91	0.80	0.71	0.70	0.56	0.55	0.60	0.60	0.60
E2	**FN**	0.26	0.18	0.14	0.14	0.12	0.12	0.10	0.10	0.10	0.10
	FP	26.00	17.78	13.75	14.29	11.67	12.00	10.00	10.00	10.00	10.00
	FI	0.52	0.36	0.28	0.29	0.23	0.24	0.20	0.20	0.20	0.20
L1	**FN**	0.65	0.54	0.51	0.40	0.44	0.42	0.40	0.47	0.40	0.40
	FP	64.00	53.33	50.00	40.00	43.33	42.00	40.00	46.67	40.00	40.00
	FI	1.28	1.07	1.00	0.80	0.87	0.84	0.80	0.93	0.80	0.80
L2	**FN**	0.39	0.26	0.21	0.13	0.13	0.16	0.15	0.17	0.15	0.10
	FP	39.00	25.56	21.25	12.86	13.33	16.00	15.00	16.67	15.00	10.00
	FI	0.78	0.51	0.43	0.26	0.27	0.32	0.30	0.33	0.30	0.20
S1	**FN**	0.79	0.74	0.67	0.58	0.51	0.48	0.43	0.44	0.40	0.40
	FP	78.00	73.33	66.25	57.14	50.00	48.00	42.50	43.33	40.00	40.00
	FI	1.56	1.47	1.33	1.14	1.00	0.96	0.85	0.87	0.80	0.80
S2	**FN**	0.23	0.17	0.15	0.16	0.15	0.12	0.13	0.10	0.10	0.10
	FP	23.00	16.67	15.00	15.71	15.00	12.00	12.50	10.00	10.00	10.00
	FI	0.46	0.33	0.30	0.31	0.30	0.24	0.25	0.20	0.20	0.20

Table 34: The Identification Error Rates of EB Method for Group 1 (δ = 0.99)

		1	2	3	4	5	6	7	8	9	10
E1	FN	0.37	0.29	0.25	0.22	0.22	0.22	0.23	0.24	0.20	0.20
	FP	37.00	28.89	25.00	21.43	21.67	22.00	22.50	23.33	20.00	20.00
	FI	0.74	0.58	0.50	0.43	0.43	0.44	0.45	0.47	0.40	0.40
E2	FN	0.15	0.22	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10
	FP	15.00	22.22	10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00
	FI	0.30	0.44	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20
L1	FN	0.51	0.45	0.40	0.39	0.40	0.38	0.38	0.44	0.40	0.40
	FP	50.00	44.44	40.00	38.57	40.00	38.00	37.50	43.33	40.00	40.00
	FI	1.00	0.89	0.80	0.77	0.80	0.76	0.75	0.87	0.80	0.80
L2	FN	0.26	0.20	0.21	0.14	0.12	0.10	0.15	0.17	0.15	0.10
	FP	26.00	20.00	21.25	14.29	11.67	10.00	15.00	16.67	15.00	10.00
	FI	0.52	0.40	0.43	0.29	0.23	0.20	0.30	0.33	0.30	0.20
S1	FN	0.72	0.61	0.53	0.51	0.44	0.44	0.45	0.47	0.45	0.40
	FP	71.00	60.00	52.50	50.00	43.33	44.00	45.00	46.67	45.00	40.00
	FI	1.42	1.20	1.05	1.00	0.87	0.88	0.90	0.93	0.90	0.80
S2	FN	0.11	0.12	0.10	0.12	0.12	0.12	0.08	0.10	0.10	0.10
	FP	11.00	12.22	10.00	11.43	11.67	12.00	7.50	10.00	10.00	10.00
	FI	0.22	0.24	0.20	0.23	0.23	0.24	0.15	0.20	0.20	0.20

Table 35: The Identification Error Rates of CI Method for Group 1 (δ = 0.99)

		1	2	3	4	5	6	7	8	9	10
E1	**FN**	0.65	0.75	0.89	0.82	0.74	0.60	0.59	0.51	0.57	0.51
	FP	118.99	72.19	48.56	41.80	40.74	35.56	36.11	25.93	18.52	22.22
	FI	2.52	2.23	2.13	1.93	1.82	1.54	1.55	1.20	1.05	1.10
E2	**FN**	0.33	0.20	0.24	0.26	0.24	0.25	0.15	0.20	0.20	0.20
	FP	55.67	42.41	34.66	27.67	21.01	14.40	15.22	13.04	13.04	8.70
	FI	1.45	1.10	1.00	0.89	0.72	0.60	0.50	0.50	0.50	0.40
L1	**FN**	0.17	0.45	0.47	0.46	0.37	0.30	0.25	0.65	0.77	1.24
	FP	402.00	128.43	91.07	79.59	79.76	74.29	73.21	38.98	28.89	9.68
	FI	2.18	1.90	1.74	1.57	1.48	1.34	1.28	1.40	1.40	1.50
L2	**FN**	0.44	0.45	0.41	0.38	0.39	0.76	0.56	0.51	0.36	0.31
	FP	50.52	27.86	20.11	14.63	11.35	7.26	5.77	1.28	1.92	0.00
	FI	1.40	1.07	0.86	0.71	0.65	0.92	0.70	0.53	0.40	0.30
S1	**FN**	0.00	0.06	0.14	0.09	0.07	0.06	0.13	0.17	0.30	0.20
	FP	998.72	922.22	362.16	311.43	293.33	316.00	215.38	133.33	109.09	118.18
	FI	1.89	1.90	1.81	1.64	1.53	1.64	1.53	1.37	1.50	1.50
S2	**FN**	0.54	0.43	0.40	0.47	0.46	0.45	0.51	0.65	0.51	0.31
	FP	38.67	23.94	21.03	14.37	12.42	10.08	8.57	3.70	3.70	3.70
	FI	1.40	0.99	0.90	0.81	0.77	0.70	0.73	0.73	0.60	0.40

Table 36: The Identification Error Rates of SR Method for Group 2 (δ = 0.90)

		1	2	3	4	5	6	7	8	9	10
E1	**FN**	3.42	2.77	2.63	2.40	2.15	2.04	1.83	1.70	1.39	1.56
	FP	30.80	24.89	23.63	21.57	19.33	18.40	16.50	15.33	12.50	14.00
	FI	6.16	4.98	4.73	4.31	3.87	3.68	3.30	3.07	2.50	2.80
E2	**FN**	1.99	1.35	1.14	1.03	0.80	0.78	0.67	0.67	0.67	0.67
	FP	17.90	12.11	10.25	9.29	7.17	7.00	6.00	6.00	6.00	6.00
	FI	3.58	2.42	2.05	1.86	1.43	1.40	1.20	1.20	1.20	1.20
L1	**FN**	4.68	3.36	2.86	2.59	2.26	2.09	1.89	1.89	1.83	1.78
	FP	42.10	30.22	25.75	23.29	20.33	18.80	17.00	17.00	16.50	16.00
	FI	8.42	6.04	5.15	4.66	4.07	3.76	3.40	3.40	3.30	3.20
L2	**FN**	2.66	1.86	1.49	1.46	1.43	1.33	1.17	1.11	1.11	0.78
	FP	23.90	16.78	13.38	13.14	12.83	12.00	10.50	10.00	10.00	7.00
	FI	4.78	3.36	2.68	2.63	2.57	2.40	2.10	2.00	2.00	1.40
S1	**FN**	6.51	5.51	5.29	5.05	4.41	4.36	4.14	3.78	3.72	3.22
	FP	58.60	49.56	47.63	45.43	39.67	39.20	37.25	34.00	33.50	29.00
	FI	11.72	9.91	9.53	9.09	7.93	7.84	7.45	6.80	6.70	5.80
S2	**FN**	2.89	2.07	1.93	1.81	1.63	1.44	1.36	1.19	1.22	1.22
	FP	26.00	18.67	17.38	16.29	14.67	13.00	12.25	10.67	11.00	11.00
	FI	5.20	3.73	3.48	3.26	2.93	2.60	2.45	2.13	2.20	2.20

Table 37: The Identification Error Rates of EB Method for Group 2 (δ = 0.90)

		1	2	3	4	5	6	7	8	9	10
E1	**FN**	2.42	1.93	1.60	1.43	1.30	1.22	1.14	1.04	1.17	1.11
	FP	21.80	17.33	14.38	12.86	11.67	11.00	10.25	9.33	10.50	10.00
	FI	4.36	3.47	2.88	2.57	2.33	2.20	2.05	1.87	2.10	2.00
E2	**FN**	1.07	0.80	0.67	0.60	0.52	0.49	0.39	0.37	0.44	0.44
	FP	9.60	7.22	6.00	5.43	4.67	4.40	3.50	3.33	4.00	4.00
	FI	1.92	1.44	1.20	1.09	0.93	0.88	0.70	0.67	0.80	0.80
L1	**FN**	2.84	2.17	1.76	1.70	1.54	1.44	1.36	1.30	1.28	1.11
	FP	25.60	19.56	15.88	15.29	13.83	13.00	12.25	11.67	11.50	10.00
	FI	5.12	3.91	3.18	3.06	2.77	2.60	2.45	2.33	2.30	2.00
L2	**FN**	1.53	1.32	1.08	1.03	1.02	0.96	0.94	0.89	0.83	0.78
	FP	13.80	11.89	9.75	9.29	9.17	8.60	8.50	8.00	7.50	7.00
	FI	2.76	2.38	1.95	1.86	1.83	1.72	1.70	1.60	1.50	1.40
S1	**FN**	5.31	4.47	4.08	3.79	3.54	3.49	3.44	3.41	3.22	3.11
	FP	47.80	40.22	36.75	34.14	31.83	31.40	31.00	30.67	29.00	28.00
	FI	9.56	8.04	7.35	6.83	6.37	6.28	6.20	6.13	5.80	5.60
S2	**FN**	1.81	1.35	1.25	1.13	1.11	1.04	1.00	1.00	0.89	1.11
	FP	16.30	12.11	11.25	10.14	10.00	9.40	9.00	9.00	8.00	10.00
	FI	3.26	2.42	2.25	2.03	2.00	1.88	1.80	1.80	1.60	2.00

Table 38: The Identification Error Rates of CI Method for Group 2 (δ = 0.90)

		1	2	3	4	5	6	7	8	9	10
E1	**FN**	2.50	2.97	2.41	2.18	1.65	1.45	1.31	1.34	1.00	0.89
	FP	58.43	31.26	29.90	29.27	28.43	27.65	25.98	25.16	23.04	20.59
	FI	6.85	5.86	5.21	4.94	4.38	4.12	3.83	3.77	3.25	2.90
E2	**FN**	1.80	1.75	1.42	1.32	1.18	1.26	1.10	0.97	0.96	1.01
	FP	23.50	12.50	10.92	9.14	7.66	7.57	6.31	5.71	5.86	4.50
	FI	4.07	2.93	2.48	2.19	1.90	1.96	1.68	1.50	1.50	1.40
L1	**FN**	2.17	2.60	2.74	2.34	1.90	1.91	1.69	1.55	1.44	1.44
	FP	108.57	53.41	33.42	31.63	28.40	25.10	23.47	24.83	21.43	20.41
	FI	8.87	6.99	5.75	5.21	4.50	4.18	3.83	3.83	3.40	3.30
L2	**FN**	1.97	1.80	1.60	1.44	1.33	1.25	1.20	1.08	1.18	0.90
	FP	35.06	21.71	16.00	13.08	12.77	11.96	11.45	9.97	10.75	8.41
	FI	5.09	3.87	3.14	2.69	2.55	2.40	2.30	2.03	2.20	1.70
S1	**FN**	0.68	1.29	2.04	3.12	3.01	2.89	2.74	2.37	2.19	2.19
	FP	631.65	240.63	140.65	70.78	60.27	56.98	54.07	50.78	45.93	38.37
	FI	10.65	9.67	9.54	8.44	7.93	7.54	7.15	6.53	5.95	5.30
S2	**FN**	2.26	1.88	1.97	2.05	1.78	1.71	1.61	1.60	1.50	1.34
	FP	35.49	22.35	18.49	15.69	12.91	10.59	10.54	10.13	8.33	9.80
	FI	5.29	3.86	3.64	3.44	2.92	2.62	2.53	2.47	2.20	2.20

Table 39: The Identification Error Rates of SR Method for Group 2 (δ = 0.95)

		1	2	3	4	5	6	7	8	9	10
E1	**FN**	1.91	1.56	1.29	1.10	1.00	0.97	0.82	0.70	0.68	0.74
	FP	36.20	29.56	24.50	20.86	19.00	18.40	15.50	13.33	13.00	14.00
	FI	3.62	2.96	2.45	2.09	1.90	1.84	1.55	1.33	1.30	1.40
E2	**FN**	1.28	1.02	0.83	0.68	0.61	0.57	0.50	0.46	0.42	0.42
	FP	24.40	19.33	15.75	12.86	11.67	10.80	9.50	8.67	8.00	8.00
	FI	2.44	1.93	1.58	1.29	1.17	1.08	0.95	0.87	0.80	0.80
L1	**FN**	2.36	1.92	1.61	1.47	1.28	1.31	1.11	1.05	0.95	0.84
	FP	44.80	36.44	30.50	28.00	24.33	24.80	21.00	20.00	18.00	16.00
	FI	4.48	3.64	3.05	2.80	2.43	2.48	2.10	2.00	1.80	1.60
L2	**FN**	1.43	1.06	0.96	0.81	0.79	0.67	0.61	0.56	0.58	0.53
	FP	27.20	20.22	18.25	15.43	15.00	12.80	11.50	10.67	11.00	10.00
	FI	2.72	2.02	1.83	1.54	1.50	1.28	1.15	1.07	1.10	1.00
S1	**FN**	3.52	3.12	2.75	2.56	2.49	2.32	2.08	2.00	1.89	1.89
	FP	66.80	59.33	52.25	48.57	47.33	44.00	39.50	38.00	36.00	36.00
	FI	6.68	5.93	5.23	4.86	4.73	4.40	3.95	3.80	3.60	3.60
S2	**FN**	1.38	0.96	0.76	0.74	1.09	0.57	0.47	0.49	0.58	0.53
	FP	26.20	18.22	14.50	14.00	20.67	10.80	9.00	9.33	11.00	10.00
	FI	2.62	1.82	1.45	1.40	2.07	1.08	0.90	0.93	1.10	1.00

Table 40: The Identification Error Rates of EB Method for Group 2 (δ = 0.95)

		1	2	3	4	5	6	7	8	9	10
E1	**FN**	1.34	0.99	0.87	0.75	0.70	0.65	0.58	0.60	0.58	0.53
	FP	25.40	18.89	16.50	14.29	13.33	12.40	11.00	11.33	11.00	10.00
	FI	2.54	1.89	1.65	1.43	1.33	1.24	1.10	1.13	1.10	1.00
E2	**FN**	0.86	0.65	0.51	0.44	0.37	0.38	0.37	0.35	0.32	0.32
	FP	16.40	12.44	9.75	8.29	7.00	7.20	7.00	6.67	6.00	6.00
	FI	1.64	1.24	0.98	0.83	0.70	0.72	0.70	0.67	0.60	0.60
L1	**FN**	1.74	1.30	1.01	0.99	0.91	0.88	0.82	0.81	0.84	0.74
	FP	33.00	24.67	19.25	18.86	17.33	16.80	15.50	15.33	16.00	14.00
	FI	3.30	2.47	1.93	1.89	1.73	1.68	1.55	1.53	1.60	1.40
L2	**FN**	0.91	0.71	0.68	0.62	0.56	0.51	0.45	0.46	0.42	0.42
	FP	17.20	13.56	13.00	11.71	10.67	9.60	8.50	8.67	8.00	8.00
	FI	1.72	1.36	1.30	1.17	1.07	0.96	0.85	0.87	0.80	0.80
S1	**FN**	2.92	2.40	2.07	1.94	1.79	1.81	1.71	1.65	1.63	1.68
	FP	55.40	45.56	39.25	36.86	34.00	34.40	32.50	31.33	31.00	32.00
	FI	5.54	4.56	3.93	3.69	3.40	3.44	3.25	3.13	3.10	3.20
S2	**FN**	0.73	0.48	0.39	0.35	0.26	0.23	0.21	0.28	0.26	0.21
	FP	13.80	9.11	7.50	6.57	5.00	4.40	4.00	5.33	5.00	4.00
	FI	1.38	0.91	0.75	0.66	0.50	0.44	0.40	0.53	0.50	0.40

Table 41: The Identification Error Rates of CI Method for Group 2 (δ = 0.95)

		1	2	3	4	5	6	7	8	9	10
E1	**FN**	1.70	1.28	1.64	2.05	1.86	1.86	1.49	1.59	1.46	1.19
	FP	71.98	48.08	28.33	20.76	17.56	13.87	11.33	9.33	7.33	4.00
	FI	5.31	3.86	3.44	3.46	3.03	2.76	2.23	2.17	1.90	1.40
E2	**FN**	1.26	1.09	0.94	0.83	0.84	0.80	0.83	0.83	0.65	0.65
	FP	30.12	22.06	19.01	16.46	16.67	16.34	17.96	18.78	18.31	16.90
	FI	3.12	2.53	2.23	1.93	1.97	1.90	2.05	2.10	1.90	1.80
L1	**FN**	1.37	1.25	0.92	1.29	1.46	1.43	1.39	1.14	0.96	0.96
	FP	128.53	72.71	61.96	37.95	27.51	25.08	23.02	21.16	19.05	17.46
	FI	5.69	4.53	3.73	3.33	3.10	2.92	2.75	2.40	2.10	2.00
L2	**FN**	1.50	1.07	0.86	0.68	0.45	0.47	0.27	0.32	0.32	0.32
	FP	31.21	19.90	14.44	10.49	10.71	8.57	7.50	8.57	7.86	5.71
	FI	3.32	2.32	1.78	1.36	1.17	1.04	0.78	0.90	0.85	0.70
S1	**FN**	0.19	0.46	0.77	1.29	1.23	0.89	0.88	0.83	0.83	0.73
	FP	1282.00	450.51	235.63	112.24	106.19	101.71	92.86	93.33	87.14	85.71
	FI	6.60	5.41	5.46	5.17	4.90	4.42	4.10	4.07	3.85	3.70
S2	**FN**	1.23	0.75	0.65	0.48	0.57	0.51	0.67	0.57	0.48	0.32
	FP	38.95	28.63	21.21	17.56	14.12	13.04	7.54	8.47	7.14	7.94
	FI	3.24	2.31	1.80	1.44	1.37	1.26	1.10	1.07	0.90	0.80

Table 42: The Identification Error Rates of SR Method for Group 2 (δ = 0.99)

		1	2	3	4	5	6	7	8	9	10
E1	FN	0.62	0.47	0.43	0.33	0.22	0.20	0.23	0.20	0.15	0.20
	FP	61.00	46.67	42.50	32.86	21.67	20.00	22.50	20.00	15.00	20.00
	FI	1.22	0.93	0.85	0.66	0.43	0.40	0.45	0.40	0.30	0.40
E2	FN	0.26	0.19	0.19	0.19	0.20	0.16	0.18	0.17	0.20	0.20
	FP	26.00	18.89	18.75	18.57	20.00	16.00	17.50	16.67	20.00	20.00
	FI	0.52	0.38	0.38	0.37	0.40	0.32	0.35	0.33	0.40	0.40
L1	FN	0.61	0.52	0.43	0.43	0.42	0.42	0.33	0.27	0.20	0.20
	FP	60.00	51.11	42.50	42.86	41.67	42.00	32.50	26.67	20.00	20.00
	FI	1.20	1.02	0.85	0.86	0.83	0.84	0.65	0.53	0.40	0.40
L2	FN	0.36	0.28	0.20	0.14	0.15	0.16	0.10	0.10	0.05	0.00
	FP	36.00	27.78	20.00	14.29	15.00	16.00	10.00	10.00	5.00	0.00
	FI	0.72	0.56	0.40	0.29	0.30	0.32	0.20	0.20	0.10	0.00
S1	FN	0.79	0.80	0.72	0.72	0.71	0.73	0.66	0.61	0.56	0.51
	FP	78.00	78.89	71.25	71.43	70.00	72.00	65.00	60.00	55.00	50.00
	FI	1.56	1.58	1.43	1.43	1.40	1.44	1.30	1.20	1.10	1.00
S2	FN	0.27	0.16	0.10	0.09	0.08	0.06	0.03	0.07	0.05	0.10
	FP	27.00	15.56	10.00	8.57	8.33	6.00	2.50	6.67	5.00	10.00
	FI	0.54	0.31	0.20	0.17	0.17	0.12	0.05	0.13	0.10	0.20

Table 43: The Identification Error Rates of EB Method for Group 2 (δ = 0.99)

		1	2	3	4	5	6	7	8	9	10
E1	**FN**	0.43	0.28	0.21	0.17	0.12	0.10	0.08	0.10	0.10	0.10
	FP	43.00	27.78	21.25	17.14	11.67	10.00	7.50	10.00	10.00	10.00
	FI	0.86	0.56	0.43	0.34	0.23	0.20	0.15	0.20	0.20	0.20
E2	**FN**	0.19	0.17	0.19	0.17	0.15	0.18	0.15	0.20	0.20	0.20
	FP	19.00	16.67	18.75	17.14	15.00	18.00	15.00	20.00	20.00	20.00
	FI	0.38	0.33	0.38	0.34	0.30	0.36	0.30	0.40	0.40	0.40
L1	**FN**	0.46	0.39	0.33	0.29	0.25	0.22	0.25	0.17	0.20	0.10
	FP	46.00	38.89	32.50	28.57	25.00	22.00	25.00	16.67	20.00	10.00
	FI	0.92	0.78	0.65	0.57	0.50	0.44	0.50	0.33	0.40	0.20
L2	**FN**	0.21	0.15	0.09	0.09	0.10	0.08	0.05	0.00	0.00	0.00
	FP	21.00	14.44	8.75	8.57	10.00	8.00	5.00	0.00	0.00	0.00
	FI	0.42	0.29	0.18	0.17	0.20	0.16	0.10	0.00	0.00	0.00
S1	**FN**	0.72	0.70	0.59	0.55	0.57	0.55	0.56	0.54	0.56	0.51
	FP	71.00	68.89	58.75	54.29	56.67	54.00	55.00	53.33	55.00	50.00
	FI	1.42	1.38	1.18	1.09	1.13	1.08	1.10	1.07	1.10	1.00
S2	**FN**	0.14	0.08	0.08	0.03	0.02	0.00	0.00	0.00	0.00	0.00
	FP	14.00	7.78	7.50	2.86	1.67	0.00	0.00	0.00	0.00	0.00
	FI	0.28	0.16	0.15	0.06	0.03	0.00	0.00	0.00	0.00	0.00

Table 44: The Identification Error Rates of CI Method for Group 2 (δ = 0.99)

		1	2	3	4	5	6	7	8	9	10
E1	**FN**	0.68	0.49	0.67	0.53	0.51	0.45	0.41	0.41	0.41	0.31
	FP	104.12	62.01	33.33	29.63	27.78	22.96	25.93	22.22	22.22	22.22
	FI	2.44	1.71	1.55	1.31	1.25	1.06	1.10	1.00	1.00	0.90
E2	**FN**	0.41	0.26	0.22	0.10	0.10	0.10	0.05	0.03	0.05	0.10
	FP	59.61	45.08	31.87	28.57	26.09	20.87	15.22	15.94	10.87	13.04
	FI	1.61	1.22	0.94	0.76	0.70	0.58	0.40	0.40	0.30	0.40
L1	**FN**	0.15	0.17	0.46	0.39	0.34	0.30	0.33	0.20	0.20	0.10
	FP	443.18	240.32	101.79	98.98	92.86	92.86	94.64	100.00	107.14	92.86
	FI	2.10	1.82	1.88	1.77	1.63	1.60	1.65	1.60	1.70	1.40
L2	**FN**	0.46	0.42	0.40	0.41	0.39	0.47	0.51	0.48	0.51	0.51
	FP	54.40	25.37	18.18	17.37	11.33	6.92	7.69	6.41	5.77	3.85
	FI	1.50	0.98	0.81	0.81	0.67	0.64	0.70	0.63	0.65	0.60
S1	**FN**	0.00	0.11	0.21	0.23	0.22	0.28	0.53	0.44	0.40	0.30
	FP	932.46	752.38	327.50	294.29	263.33	190.32	100.00	75.76	90.91	90.91
	FI	1.92	1.87	1.85	1.70	1.53	1.46	1.63	1.27	1.40	1.30
S2	**FN**	0.42	0.33	0.26	0.18	0.19	0.18	0.21	0.14	0.15	0.31
	FP	37.78	19.16	14.58	11.31	13.19	9.76	9.09	4.00	1.96	0.00
	FI	1.26	0.78	0.60	0.44	0.50	0.42	0.43	0.23	0.20	0.30

Table 45: The Identification Error Rates of SR Method for Group 3 (δ = 0.90)

		1	2	3	4	5	6	7	8	9	10
E1	**FN**	3.43	2.79	2.15	2.03	1.63	1.53	1.42	1.33	1.22	1.11
	FP	30.90	25.11	19.38	18.29	14.67	13.80	12.75	12.00	11.00	10.00
	FI	6.18	5.02	3.88	3.66	2.93	2.76	2.55	2.40	2.20	2.00
E2	**FN**	2.10	1.48	1.32	1.08	1.02	0.93	0.81	0.74	0.67	0.78
	FP	18.90	13.33	11.88	9.71	9.17	8.40	7.25	6.67	6.00	7.00
	FI	3.78	2.67	2.38	1.94	1.83	1.68	1.45	1.33	1.20	1.40
L1	**FN**	4.44	3.65	2.97	2.63	2.54	2.33	2.22	1.96	1.83	2.00
	FP	40.00	32.89	26.75	23.71	22.83	21.00	20.00	17.67	16.50	18.00
	FI	8.00	6.58	5.35	4.74	4.57	4.20	4.00	3.53	3.30	3.60
L2	**FN**	2.44	1.62	1.50	1.19	1.20	1.09	1.06	0.85	0.94	0.78
	FP	22.00	14.56	13.50	10.71	10.83	9.80	9.50	7.67	8.50	7.00
	FI	4.40	2.91	2.70	2.14	2.17	1.96	1.90	1.53	1.70	1.40
S1	**FN**	6.71	5.77	5.11	4.52	4.33	4.11	3.83	3.74	3.89	3.78
	FP	60.40	51.89	46.00	40.71	39.00	37.00	34.50	33.67	35.00	34.00
	FI	12.08	10.38	9.20	8.14	7.80	7.40	6.90	6.73	7.00	6.80
S2	**FN**	2.50	1.83	1.53	1.33	1.11	0.93	0.83	0.85	0.94	1.00
	FP	22.50	16.44	13.75	12.00	10.00	8.40	7.50	7.67	8.50	9.00
	FI	4.50	3.29	2.75	2.40	2.00	1.68	1.50	1.53	1.70	1.80

Table 46: The Identification Error Rates of EB Method for Group 3 (δ = 0.90)

		1	2	3	4	5	6	7	8	9	10
E1	**FN**	2.32	1.77	1.49	1.22	1.11	1.04	0.97	1.00	0.89	1.00
	FP	20.90	15.89	13.38	11.00	10.00	9.40	8.75	9.00	8.00	9.00
	FI	4.18	3.18	2.68	2.20	2.00	1.88	1.75	1.80	1.60	1.80
E2	**FN**	1.18	0.86	0.69	0.63	0.54	0.51	0.50	0.44	0.44	0.33
	FP	10.60	7.78	6.25	5.71	4.83	4.60	4.50	4.00	4.00	3.00
	FI	2.12	1.56	1.25	1.14	0.97	0.92	0.90	0.80	0.80	0.60
L1	**FN**	3.08	2.32	1.94	1.63	1.50	1.36	1.31	1.19	1.17	1.22
	FP	27.70	20.89	17.50	14.71	13.50	12.20	11.75	10.67	10.50	11.00
	FI	5.54	4.18	3.50	2.94	2.70	2.44	2.35	2.13	2.10	2.20
L2	**FN**	1.49	1.12	1.04	0.97	0.89	0.89	0.83	0.74	0.67	0.78
	FP	13.40	10.11	9.38	8.71	8.00	8.00	7.50	6.67	6.00	7.00
	FI	2.68	2.02	1.88	1.74	1.60	1.60	1.50	1.33	1.20	1.40
S1	**FN**	5.38	4.33	3.88	3.63	3.37	3.24	3.08	3.00	3.06	3.00
	FP	48.40	39.00	34.88	32.71	30.33	29.20	27.75	27.00	27.50	27.00
	FI	9.68	7.80	6.98	6.54	6.07	5.84	5.55	5.40	5.50	5.40
S2	**FN**	1.60	1.28	1.01	0.87	0.85	0.76	0.75	0.74	0.78	0.78
	FP	14.40	11.56	9.13	7.86	7.67	6.80	6.75	6.67	7.00	7.00
	FI	2.88	2.31	1.83	1.57	1.53	1.36	1.35	1.33	1.40	1.40

Table 47: The Identification Error Rates of CI Method for Group 3 (δ = 0.90)

		1	2	3	4	5	6	7	8	9	10
E1	**FN**	2.27	2.51	2.13	1.51	1.37	1.11	0.92	1.04	0.72	0.67
	FP	68.00	33.55	26.72	26.05	22.88	22.75	22.06	19.28	18.63	21.57
	FI	7.20	5.68	4.64	4.01	3.57	3.32	3.08	2.90	2.55	2.80
E2	**FN**	1.72	1.21	0.93	0.76	0.43	0.38	0.42	0.30	0.45	0.34
	FP	24.62	15.23	12.16	11.84	9.91	10.45	10.14	9.01	7.21	6.31
	FI	4.10	2.74	2.18	1.99	1.48	1.50	1.50	1.27	1.20	1.00
L1	**FN**	2.17	2.29	2.72	2.38	2.24	2.15	1.94	1.70	1.77	1.77
	FP	105.56	64.50	36.22	32.51	28.23	26.12	23.21	20.07	17.86	23.47
	FI	8.68	7.18	6.00	5.33	4.78	4.50	4.03	3.50	3.35	3.90
L2	**FN**	1.79	1.80	1.50	1.10	0.95	0.87	0.87	0.82	0.73	0.78
	FP	36.88	15.78	14.02	12.95	11.68	11.03	10.51	11.21	10.75	12.15
	FI	5.05	3.30	2.84	2.37	2.10	1.96	1.90	1.93	1.80	2.00
S1	**FN**	0.51	1.15	2.09	3.55	3.15	2.91	2.76	2.48	2.63	2.52
	FP	849.09	244.76	134.18	59.80	56.01	53.02	47.09	46.12	47.09	44.19
	FI	9.84	9.68	9.24	8.39	7.70	7.22	6.58	6.23	6.45	6.10
S2	**FN**	1.97	1.57	1.49	1.29	1.15	0.96	0.86	0.93	1.00	1.22
	FP	32.50	21.45	13.65	12.75	9.48	8.63	7.60	8.50	7.35	7.84
	FI	4.78	3.47	2.71	2.46	2.00	1.74	1.55	1.70	1.65	1.90

Table 48: The Identification Error Rates of SR Method for Group 3 (δ = 0.95)

		1	2	3	4	5	6	7	8	9	10
E1	FN	2.26	1.65	1.58	1.43	1.39	1.22	1.16	1.16	1.26	1.26
	FP	43.00	31.33	30.00	27.14	26.33	23.20	22.00	22.00	24.00	24.00
	FI	4.30	3.13	3.00	2.71	2.63	2.32	2.20	2.20	2.40	2.40
E2	FN	1.39	1.06	0.91	0.87	0.77	0.65	0.63	0.60	0.53	0.53
	FP	26.40	20.22	17.25	16.57	14.67	12.40	12.00	11.33	10.00	10.00
	FI	2.64	2.02	1.73	1.66	1.47	1.24	1.20	1.13	1.00	1.00
L1	FN	2.47	2.06	1.84	1.55	1.39	1.41	1.29	1.23	1.21	1.05
	FP	47.00	39.11	35.00	29.43	26.33	26.80	24.50	23.33	23.00	20.00
	FI	4.70	3.91	3.50	2.94	2.63	2.68	2.45	2.33	2.30	2.00
L2	FN	1.35	1.09	0.93	0.72	0.72	0.67	0.58	0.53	0.53	0.53
	FP	25.60	20.67	17.75	13.71	13.67	12.80	11.00	10.00	10.00	10.00
	FI	2.56	2.07	1.78	1.37	1.37	1.28	1.10	1.00	1.00	1.00
S1	FN	3.45	3.05	2.64	2.63	2.30	2.23	2.16	2.11	1.95	2.00
	FP	65.60	58.00	50.25	50.00	43.67	42.40	41.00	40.00	37.00	38.00
	FI	6.56	5.80	5.03	5.00	4.37	4.24	4.10	4.00	3.70	3.80
S2	FN	1.29	1.03	0.86	0.77	0.72	0.61	0.61	0.53	0.53	0.42
	FP	24.60	19.56	16.25	14.57	13.67	11.60	11.50	10.00	10.00	8.00
	FI	2.46	1.96	1.63	1.46	1.37	1.16	1.15	1.00	1.00	0.80

Table 49: The Identification Error Rates of EB Method for Group 3 (δ = 0.95)

		1	2	3	4	5	6	7	8	9	10
E1	FN	1.52	1.19	1.14	1.07	1.05	0.93	1.05	1.09	1.05	1.16
	FP	28.80	22.67	21.75	20.29	20.00	17.60	20.00	20.67	20.00	22.00
	FI	2.88	2.27	2.18	2.03	2.00	1.76	2.00	2.07	2.00	2.20
E2	FN	0.85	0.68	0.61	0.60	0.49	0.48	0.45	0.46	0.37	0.42
	FP	16.20	12.89	11.50	11.43	9.33	9.20	8.50	8.67	7.00	8.00
	FI	1.62	1.29	1.15	1.14	0.93	0.92	0.85	0.87	0.70	0.80
L1	FN	1.77	1.36	1.17	1.07	1.07	0.99	0.95	0.91	0.84	0.84
	FP	33.60	25.78	22.25	20.29	20.33	18.80	18.00	17.33	16.00	16.00
	FI	3.36	2.58	2.23	2.03	2.03	1.88	1.80	1.73	1.60	1.60
L2	FN	0.88	0.75	0.66	0.59	0.56	0.57	0.53	0.53	0.53	0.53
	FP	16.80	14.22	12.50	11.14	10.67	10.80	10.00	10.00	10.00	10.00
	FI	1.68	1.42	1.25	1.11	1.07	1.08	1.00	1.00	1.00	1.00
S1	FN	2.88	2.29	2.05	1.86	1.72	1.75	1.76	1.65	1.74	1.58
	FP	54.80	43.56	39.00	35.43	32.67	33.20	33.50	31.33	33.00	30.00
	FI	5.48	4.36	3.90	3.54	3.27	3.32	3.35	3.13	3.30	3.00
S2	FN	0.76	0.54	0.43	0.41	0.40	0.38	0.37	0.39	0.32	0.32
	FP	14.40	10.22	8.25	7.71	7.67	7.20	7.00	7.33	6.00	6.00
	FI	1.44	1.02	0.83	0.77	0.77	0.72	0.70	0.73	0.60	0.60

Table 50: The Identification Error Rates of CI Method for Group 3 (δ = 0.95)

		1	2	3	4	5	6	7	8	9	10
E1	**FN**	1.36	1.54	1.31	1.30	1.07	1.31	1.43	1.41	1.35	1.30
	FP	96.06	56.16	43.26	32.36	27.18	19.72	13.67	13.78	14.00	13.33
	FI	5.20	4.54	3.73	3.27	2.77	2.62	2.35	2.33	2.30	2.20
E2	**FN**	1.41	1.06	0.77	0.72	0.74	0.60	0.65	0.50	0.48	0.32
	FP	33.85	21.42	17.25	16.30	15.49	15.49	14.79	13.62	12.68	14.08
	FI	3.52	2.47	1.94	1.83	1.78	1.66	1.65	1.43	1.35	1.30
L1	**FN**	1.23	1.48	1.10	1.19	1.43	1.56	1.47	1.32	1.12	0.85
	FP	129.85	73.67	62.08	48.03	33.43	22.54	19.84	20.63	15.87	15.87
	FI	5.41	4.80	4.04	3.57	3.27	2.88	2.63	2.53	2.05	1.80
L2	**FN**	1.47	1.17	1.03	1.04	0.95	0.97	0.81	0.75	0.75	0.75
	FP	29.81	19.23	15.11	11.32	7.86	6.29	6.07	5.71	5.71	5.71
	FI	3.24	2.37	1.98	1.76	1.43	1.34	1.18	1.10	1.10	1.10
S1	**FN**	0.24	0.46	0.98	1.21	1.05	0.91	0.78	0.66	0.52	0.31
	FP	1246.00	488.89	173.56	114.69	104.29	100.00	93.57	88.57	90.00	88.57
	FI	6.47	5.83	5.46	5.19	4.67	4.38	4.03	3.73	3.65	3.40
S2	**FN**	1.10	0.83	0.63	0.56	0.64	0.61	0.56	0.75	0.59	0.53
	FP	42.70	31.11	30.68	24.94	21.30	21.55	14.34	11.11	10.32	7.94
	FI	3.32	2.50	2.29	1.90	1.80	1.80	1.40	1.40	1.20	1.00

Table 51: The Identification Error Rates of SR Method for Group 3 (δ = 0.99)

		1	2	3	4	5	6	7	8	9	10
E1	**FN**	0.49	0.38	0.32	0.22	0.24	0.20	0.13	0.10	0.10	0.00
	FP	49.00	37.78	31.25	21.43	23.33	20.00	12.50	10.00	10.00	0.00
	FI	0.98	0.76	0.63	0.43	0.47	0.40	0.25	0.20	0.20	0.00
E2	**FN**	0.26	0.19	0.16	0.16	0.12	0.16	0.15	0.13	0.10	0.20
	FP	26.00	18.89	16.25	15.71	11.67	16.00	15.00	13.33	10.00	20.00
	FI	0.52	0.38	0.33	0.31	0.23	0.32	0.30	0.27	0.20	0.40
L1	**FN**	0.60	0.45	0.39	0.35	0.32	0.26	0.30	0.24	0.25	0.30
	FP	59.00	44.44	38.75	34.29	31.67	26.00	30.00	23.33	25.00	30.00
	FI	1.18	0.89	0.78	0.69	0.63	0.52	0.60	0.47	0.50	0.60
L2	**FN**	0.35	0.31	0.27	0.26	0.25	0.24	0.23	0.27	0.25	0.20
	FP	35.00	31.11	26.25	25.71	25.00	24.00	22.50	26.67	25.00	20.00
	FI	0.70	0.62	0.53	0.51	0.50	0.48	0.45	0.53	0.50	0.40
S1	**FN**	0.86	0.76	0.73	0.69	0.64	0.69	0.68	0.67	0.56	0.61
	FP	85.00	75.56	72.50	68.57	63.33	68.00	67.50	66.67	55.00	60.00
	FI	1.70	1.51	1.45	1.37	1.27	1.36	1.35	1.33	1.10	1.20
S2	**FN**	0.26	0.17	0.14	0.14	0.10	0.08	0.08	0.00	0.00	0.10
	FP	26.00	16.67	13.75	14.29	10.00	8.00	7.50	0.00	0.00	10.00
	FI	0.52	0.33	0.28	0.29	0.20	0.16	0.15	0.00	0.00	0.20

Table 52: The Identification Error Rates of EB Method for Group 3 (δ = 0.99)

		1	2	3	4	5	6	7	8	9	10
E1	FN	0.33	0.24	0.16	0.14	0.13	0.10	0.03	0.00	0.00	0.00
	FP	33.00	23.33	16.25	14.29	13.33	10.00	2.50	0.00	0.00	0.00
	FI	0.66	0.47	0.33	0.29	0.27	0.20	0.05	0.00	0.00	0.00
E2	FN	0.17	0.16	0.11	0.12	0.10	0.14	0.10	0.10	0.10	0.10
	FP	17.00	15.56	11.25	11.43	10.00	14.00	10.00	10.00	10.00	10.00
	FI	0.34	0.31	0.23	0.23	0.20	0.28	0.20	0.20	0.20	0.20
L1	FN	0.45	0.34	0.29	0.27	0.25	0.26	0.25	0.27	0.25	0.20
	FP	45.00	33.33	28.75	27.14	25.00	26.00	25.00	26.67	25.00	20.00
	FI	0.90	0.67	0.58	0.54	0.50	0.52	0.50	0.53	0.50	0.40
L2	FN	0.23	0.19	0.18	0.17	0.15	0.12	0.15	0.17	0.10	0.10
	FP	23.00	18.89	17.50	17.14	15.00	12.00	15.00	16.67	10.00	10.00
	FI	0.46	0.38	0.35	0.34	0.30	0.24	0.30	0.33	0.20	0.20
S1	FN	0.77	0.68	0.59	0.59	0.56	0.55	0.56	0.54	0.45	0.51
	FP	76.00	67.78	58.75	58.57	55.00	54.00	55.00	53.33	45.00	50.00
	FI	1.52	1.36	1.18	1.17	1.10	1.08	1.10	1.07	0.90	1.00
S2	FN	0.11	0.09	0.03	0.01	0.03	0.02	0.00	0.00	0.00	0.00
	FP	11.00	8.89	2.50	1.43	3.33	2.00	0.00	0.00	0.00	0.00
	FI	0.22	0.18	0.05	0.03	0.07	0.04	0.00	0.00	0.00	0.00

Table 53: The Identification Error Rates of CI Method for Group 3 (δ = 0.99)

		1	2	3	4	5	6	7	8	9	10
E1	**FN**	0.56	0.65	0.59	0.73	0.70	0.62	0.62	0.62	0.51	0.51
	FP	120.78	83.04	61.93	39.78	32.72	30.37	23.15	22.22	20.37	14.81
	FI	2.41	2.21	1.94	1.74	1.57	1.42	1.23	1.20	1.05	0.90
E2	**FN**	0.45	0.32	0.29	0.29	0.31	0.31	0.26	0.17	0.20	0.20
	FP	55.94	39.90	29.21	19.88	16.67	9.57	10.87	14.49	10.87	4.35
	FI	1.57	1.17	0.94	0.74	0.68	0.52	0.50	0.50	0.45	0.30
L1	**FN**	0.12	0.19	0.34	0.29	0.27	0.24	0.23	0.24	0.05	0.10
	FP	454.76	208.97	96.43	84.69	97.62	97.14	91.07	76.19	85.71	85.71
	FI	2.03	2.00	1.69	1.47	1.63	1.60	1.50	1.30	1.25	1.30
L2	**FN**	0.41	0.36	0.35	0.34	0.39	0.43	0.49	0.41	0.46	0.41
	FP	51.58	30.81	22.28	17.37	16.99	11.54	11.54	11.54	9.62	7.69
	FI	1.38	1.03	0.85	0.74	0.82	0.72	0.78	0.70	0.70	0.60
S1	**FN**	0.00	0.07	0.28	0.29	0.22	0.24	0.53	0.54	0.40	0.20
	FP	967.32	922.22	335.00	305.71	360.00	312.00	118.18	78.79	104.55	81.82
	FI	1.91	1.91	1.95	1.81	2.02	1.80	1.83	1.40	1.55	1.10
S2	**FN**	0.45	0.38	0.35	0.34	0.36	0.33	0.28	0.24	0.15	0.20
	FP	38.67	30.99	25.52	22.02	22.22	24.17	26.04	23.61	20.83	16.67
	FI	1.31	1.10	0.95	0.86	0.88	0.90	0.90	0.80	0.65	0.60

APPENDIX C: SAFETY PERFORMANCE FUNCTIONS OF VARIOUS FUNCTIONAL CLASSIFICATIONS OF ARIZONA ROAD SEGMENTS

Due to the existence of overdispersion of crashes in the nine functional classifications of road segments, the NB regression model is employed to develop the SPFs. Two different model forms are tried, and the model form resulting in less prediction errors is selected. The major results of SPFs, including model form, estimated parameter values, associated standard errors, t-statistics, and over-dispersion parameters are shown in Tables 54~62 respectively. Figures 18~26 demonstrate the visual relationships between predicted crashes per km per year and AADT.

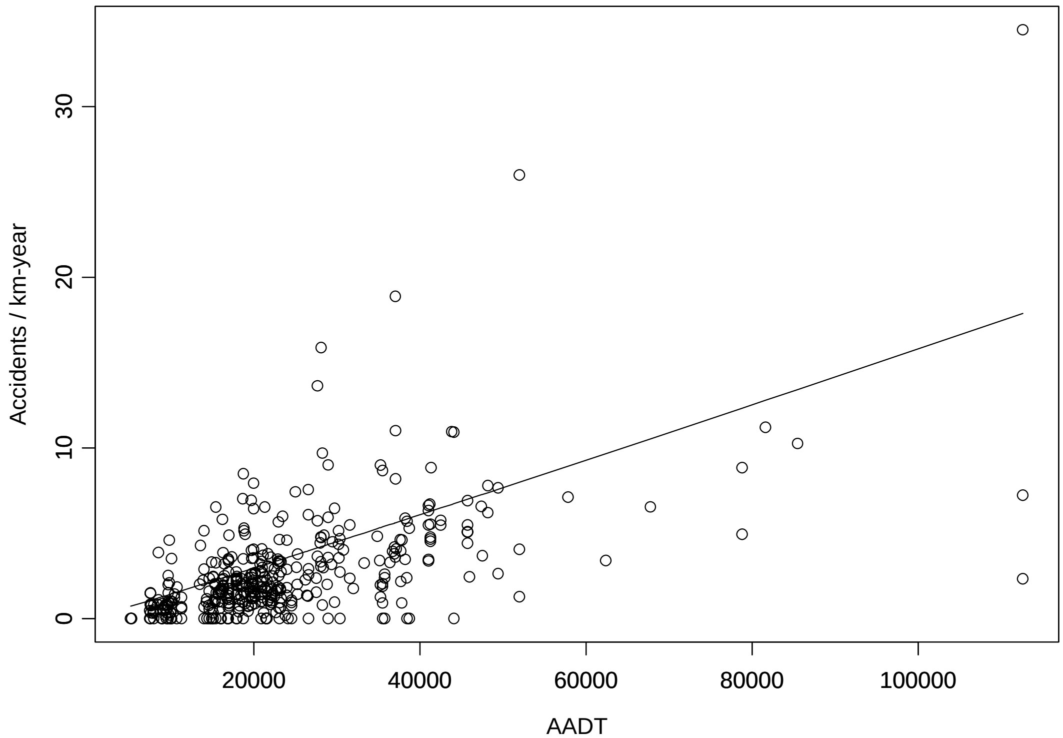


Figure 18: Relation of AADT and Crashes/year-km for Rural Interstate Principle Arterials (Functional Code: 1, year: 2000)

Table 54: Estimation Results for SPF of Rural Interstate Principle Arterials (Functional Code: 1)

Model[1]: λ=exp{-9.506+0.258×SL+1.043×ln(AADT)}			
Parameter	**Estimate**	**Std Error**	**T-Statistics**
Intercept	-9.5064694	0.82539379	-11.51750
SL[2]	0.2582805	0.01422997	18.15046
Ln(AADT)	1.0427151	0.07924371	13.15833
Over-dispersion Parameter: 3.183979, R_p-squared= 0.5396, G^2=1230.6			

Note: 1: λ—crash number in road segment during 1 year.
2: SL—road section length (km).

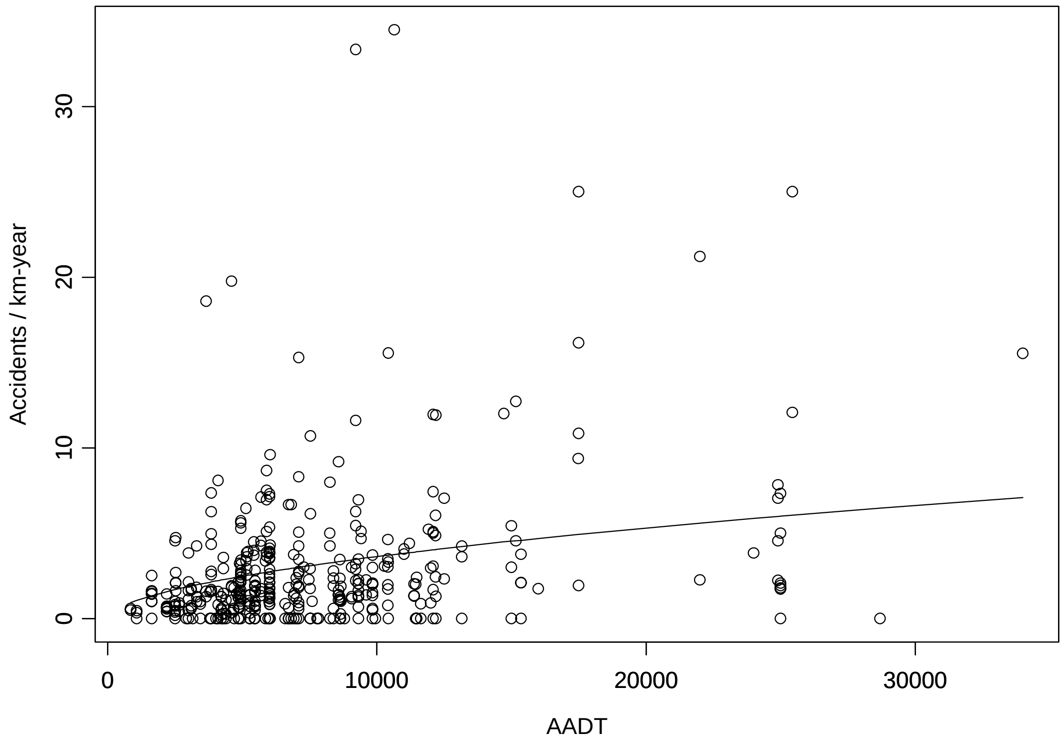


Figure 19: Relation of AADT and Crashes/year-km for Rural Other Principle Arterials (Functional Code: 2, year: 2000)

Table 55: Estimation Results for SPF of Rural Other Principle Arterials (Functional Code: 2)

Model[1]: λ=exp{-3.959+0.191×SL+0.549×ln(AADT)}			
Parameter	**Estimate**	**Std Error**	**T-Statistics**
Intercept	-3.9588146	0.76381057	-5.182980
SL[2]	0.1912418	0.01317196	14.518864
Ln(AADT)	0.5493370	0.08342578	6.584739
Over-dispersion Parameter: 4.557303, R_p-squared= 0.3947, G^2=1873.3			

Note: 1: λ—crash number in road segment during 1 year.
2: SL—road section length (km).

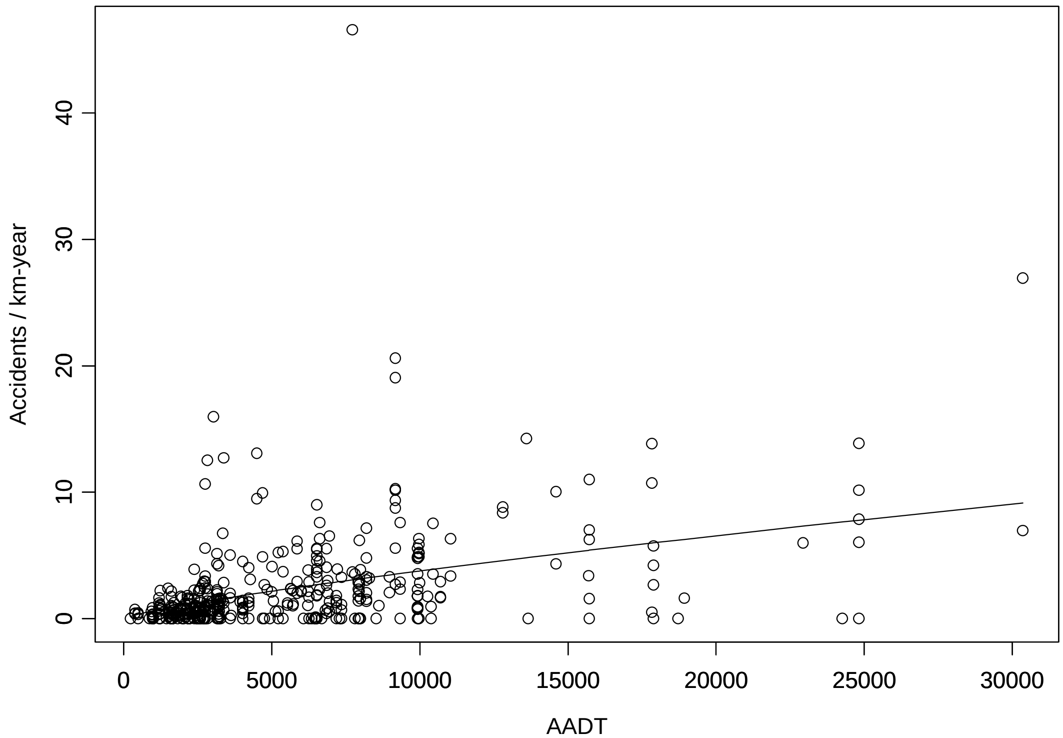


Figure 20: Relation of AADT and Crashes/year-km for Rural Minor Arterials (Functional Code: 6, year: 2000)

Table 56: Estimation Results for SPF of Rural Minor Arterials (Functional Code: 6)

Model[1]: λ=exp{-6.263+0.230×SL+0.799×ln(AADT)}			
Parameter	**Estimate**	**Std Error**	**T-Statistics**
Intercept	-6.2631910	0.54049300	-11.58792
SL[2]	0.2297051	0.01259435	18.23874
Ln(AADT)	0.7995184	0.05990074	13.34739
Over-dispersion Parameter: 3.222105, R_p-squared= 0.6319, G^2=1380.6			

Note: 1: λ—crash number in road segment during 1 year.

2: SL—road section length (km).

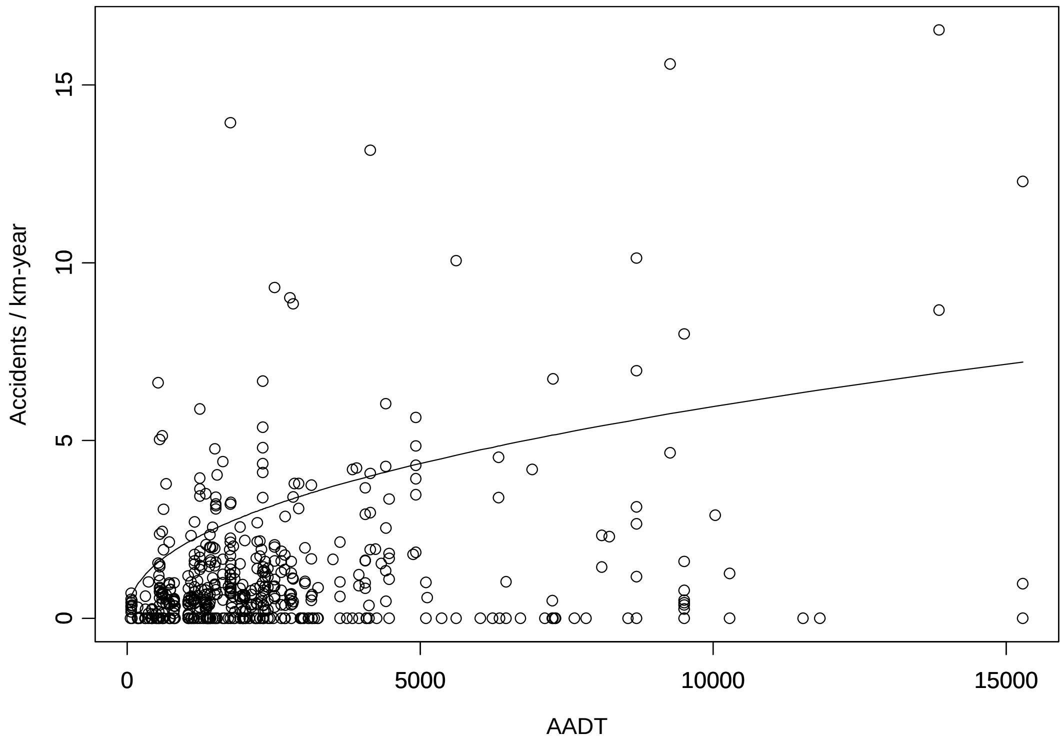


Figure 21: Relation of AADT and Crashes/year-km for Rural Major Collectors (Functional Code: 7, year: 2000)

Table 57: Estimation Results for SPF of Rural Major Collectors (Functional Code: 7)

Model[1]: $\lambda=\exp\{-2.574+0.195\times SL+0.352\times \ln(AADT)\}$			
Parameter	**Estimate**	**Std Error**	**T-Statistics**
Intercept	-2.5742562	0.48346848	-5.324559
SL[2]	0.1954693	0.01156483	16.902053
Ln(AADT)	0.3518093	0.06069124	5.796707
Over-dispersion Parameter: 4.113486, R_p-squared= 0.4026, G^2=2058.6			

Note: 1: λ—crash number in road segment during 1 year.

2: SL—road section length (km).

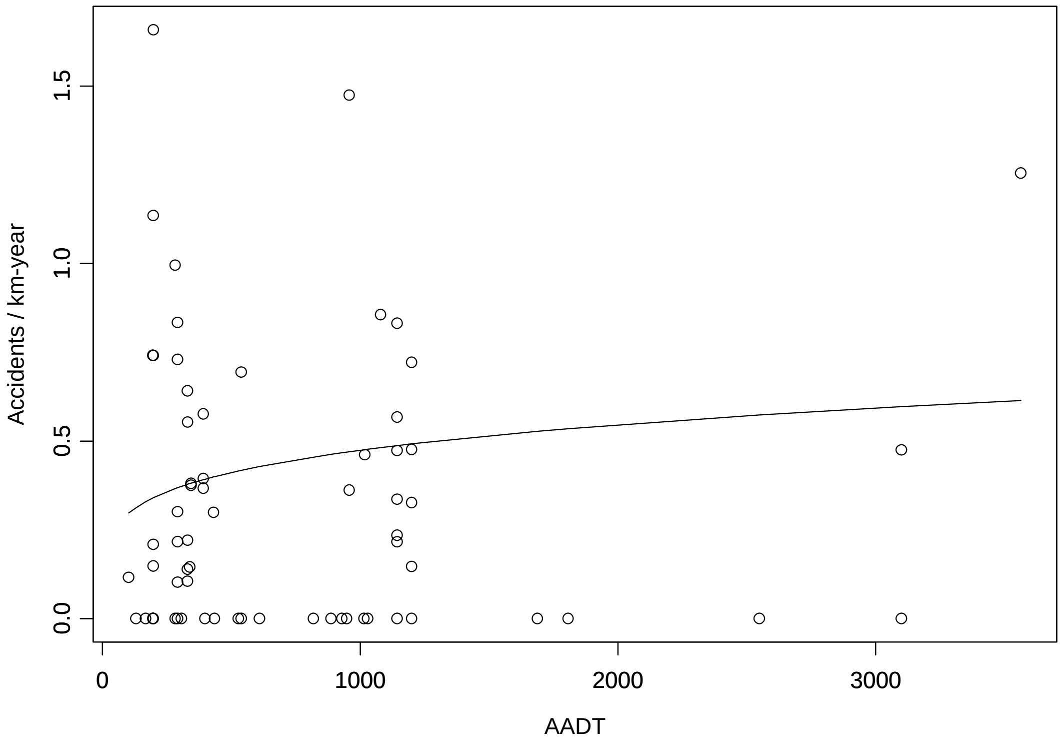


Figure 22: Relation of AADT and Crashes/year-km for Rural Minor Collectors (Functional Code: 8, year: 2000)

Table 58: Estimation Results for SPF of Rural Minor Collectors (Functional Code: 8)

Model[1]: λ=exp{-2.376+0.220×SL+0.204×ln(AADT)}			
Parameter	**Estimate**	**Std Error**	**T-Statistics**
Intercept	-2.3757517	1.11444736	-2.131776
SL[2]	0.2201045	0.03297558	6.674771
Ln(AADT)	0.2040924	0.16920269	1.206201
Over-dispersion Parameter: 1.753917, R_p-squared= 0.3950, G^2=141.5			

Note: 1: λ—crash number in road segment during 1 year.
2: SL—road section length (km).

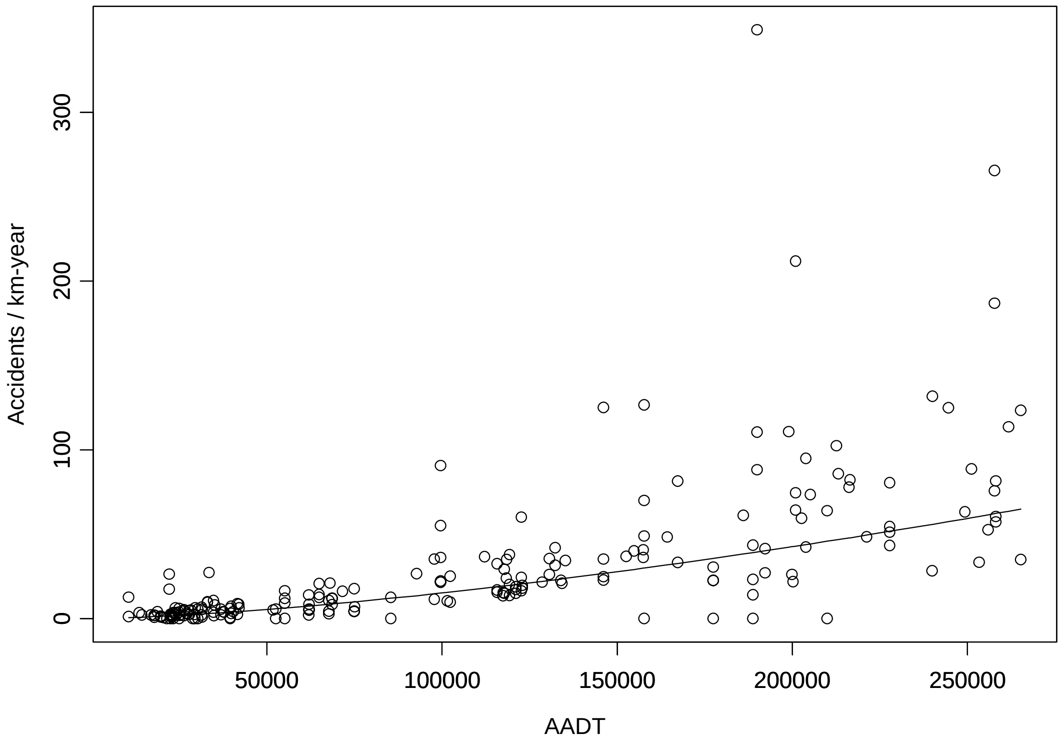


Figure 23: Relation of AADT and Crashes/year-km for Urban Interstate Principle Arterials (Functional Code: 11, year: 2000)

Table 59: Estimation Results for SPF of Urban Interstate Principle Arterials (Functional Code: 11)

Model[1]: λ=exp{-15.063+0.641×SL+1.489×ln(AADT)}			
Parameter	**Estimate**	**Std Error**	**T-Statistics**
Intercept	-15.0635312	1.26037041	-11.95167
SL[2]	0.6413353	0.05782069	11.09180
Ln(AADT)	1.4892735	0.10328018	14.41974
Over-dispersion Parameter: 9.826755, R_p-squared= 0.7196, G^2=1739.5			

Note: 1: λ—crash number in road segment during 1 year.
2: SL—road section length (km).

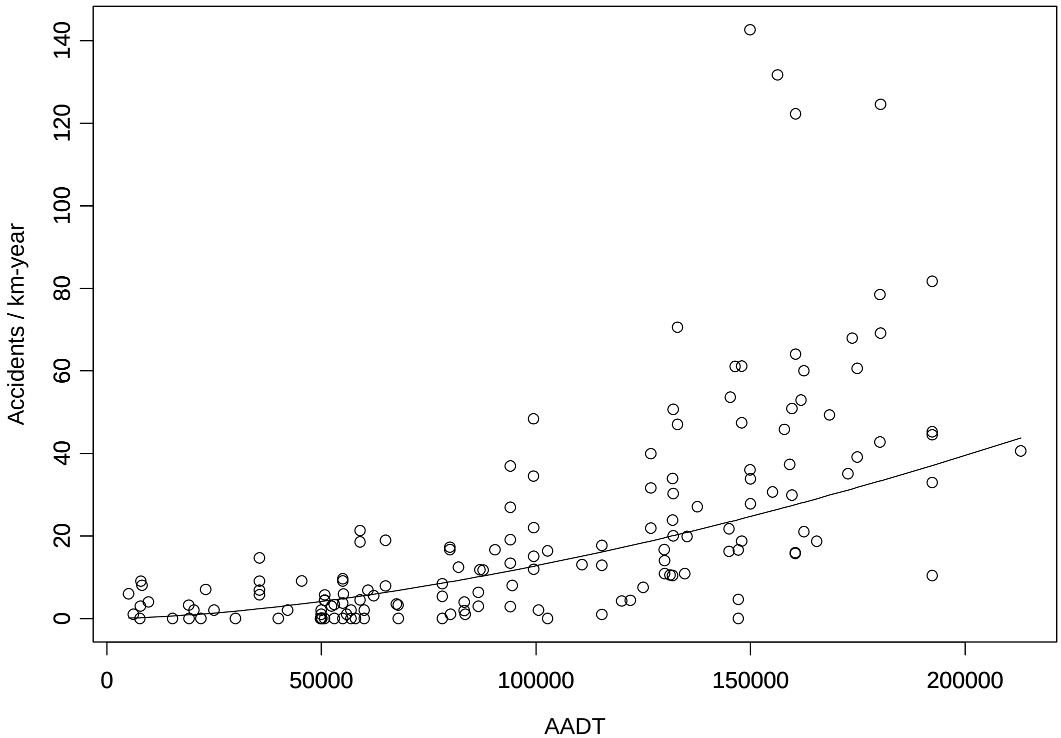


Figure 24: Relation of AADT and Crashes/year-km for Urban Freeways (Functional Code: 12, year: 2000)

Table 60: Estimation Results for SPF of Urban Freeways (Functional Code: 12)

Model[1]: $\lambda=\exp\{-16.322+0.115\times SL+1.629\times \ln(AADT)\}$			
Parameter	**Estimate**	**Std Error**	**T-Statistics**
Intercept	-16.3219233	2.86360896	-5.699774
SL[2]	0.1154657	0.01346027	8.578257
Ln(AADT)	1.6289176	0.24278373	6.709336
Over-dispersion Parameter: 19.71501, R_p-squared= 0.3374, G^2=1946.3			

Note: 1: λ—crash number in road segment during 1 year.
2: SL—road section length (km).

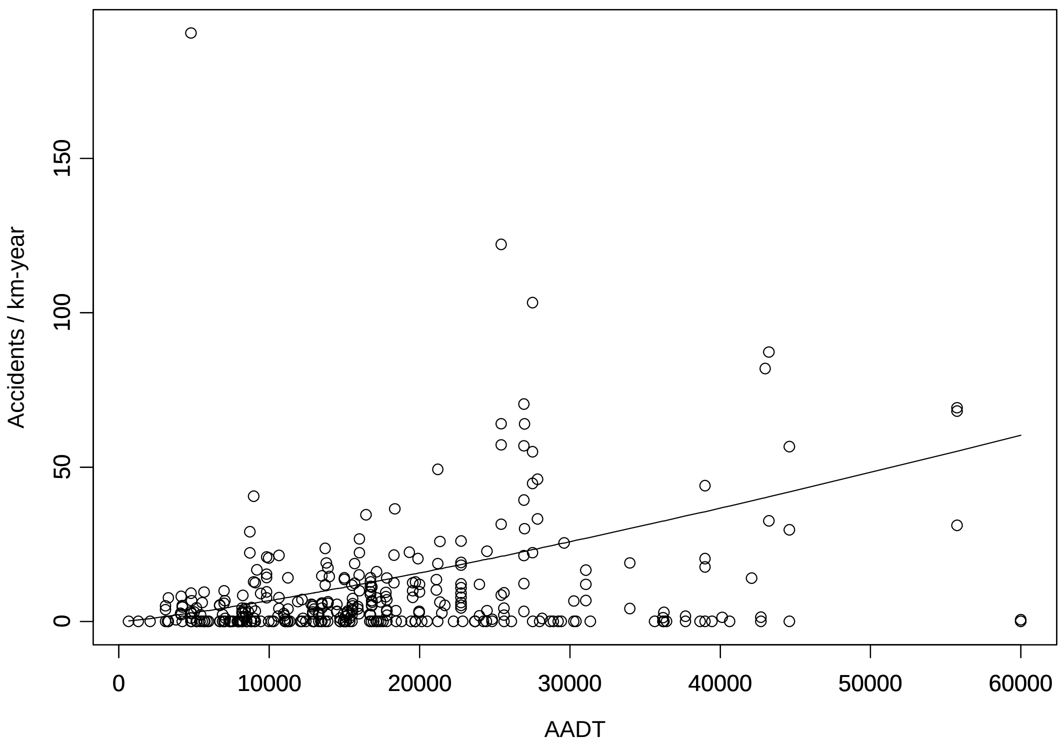


Figure 25: Relation of AADT and Crashes/year-km for Urban Other Principle Arterials (Functional Code: 14, year: 2000)

Table 61: Estimation Results for SPF of Urban Other Principle Arterials (Functional Code: 14)

Model[1]: λ=exp{-10.192+0.825×SL+1.124×ln(AADT)}			
Parameter	**Estimate**	**Std Error**	**T-Statistics**
Intercept	-10.1915082	1.29736797	-7.855526
SL[2]	0.8251083	0.09225009	8.944255
Ln(AADT)	1.1243125	0.12909345	8.709292
Over-dispersion Parameter: 10.31901, R_p-squared= 0.5672, G^2=3646.5			

Note: 1: λ—crash number in road segment during 1 year.

2: SL—road section length (km).

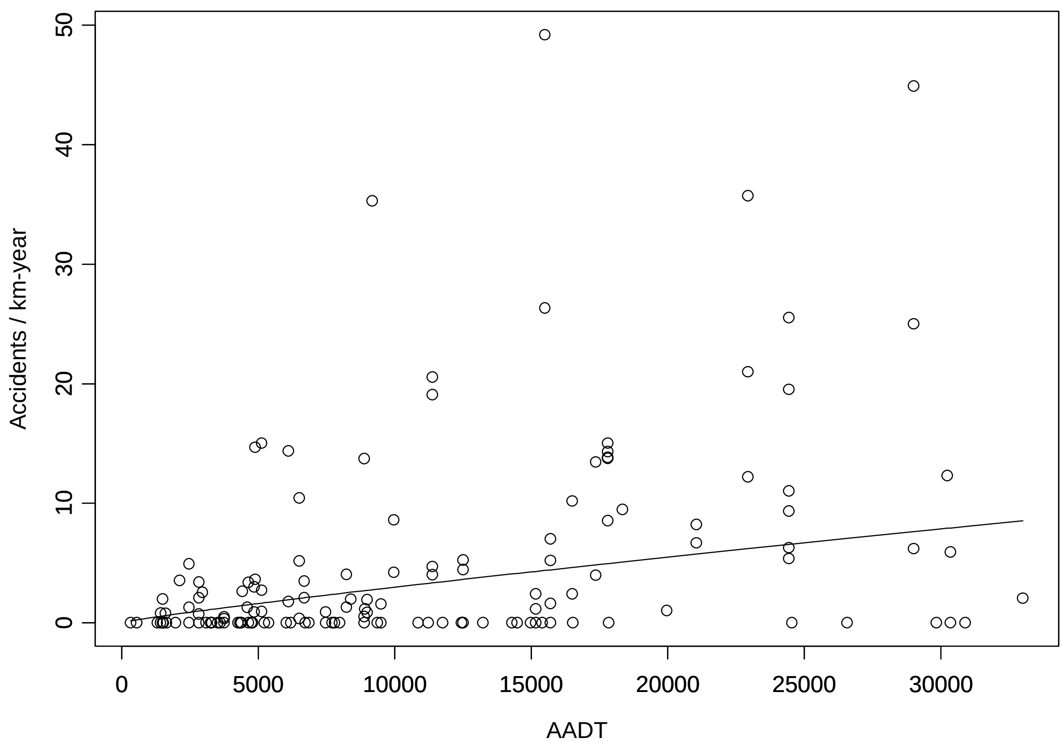


Figure 26: Relation of AADT and Crashes/year-km for Urban Minor Arterials (Functional Code: 16, year: 2000)

Table 62: Estimation Results for SPF of Urban Minor Arterials (Functional Code: 16)

Model[1]: $\lambda=\exp\{-7.755+0.703\times SL+0.864\times \ln(AADT)\}$			
Parameter	**Estimate**	**Std Error**	**T-Statistics**
Intercept	-7.7549163	1.49116881	-5.200562
SL[2]	0.7030970	0.08656798	8.121907
Ln(AADT)	0.8642005	0.15184554	5.691313
Over-dispersion Parameter: 4.939187, R_p-squared= 0.4711, G^2=681.4			

Note: 1: λ—crash number in road segment during 1 year.
2: SL—road section length (km).

APPENDIX D: COMPARISON TESTS RESULTS AND SIMILARITY OF ALTERNATIVE HSID METHODS FOR VARIOUS CLASSIFICATIONS OF HIGHWAY SECTIONS

In addition to the experiment design based on the simulated crash data, real Arizona crash data are also used in chapter V of this report to compare the performances of alternative HSID methods. Five evaluation tests are conducted, which include the site consistency test, method consistency test, total ranking differences test, false identification test, and false true Poisson means difference test. Similarity of identification results of these HSID methods are explored as well. This appendix presents the comparison tests results and similarity of alternative HSID methods for each of the nine classifications of highway sections, which are shown in Tables 63~125.

Table 63: Similarity of Identification Results (δ = 0.90) of Various Methods (Functional Code: 1)

	Bayesian	Frequency	Rate	the ARP[1]
Bayesian		33 (82.5%)	19 (47.5%)	19 (47.5%)
Frequency	33 (82.5%)		26 (65%)	26 (65%)
Rate	19 (47.5%)	26 (65%)		38 (95%)
the ARP1	19 (47.5%)	26 (65%)	38 (95%)	

Note: 1: the ARP—Method of Accident Reduction Potential.
2: The number means the number of locations identified by both methods as upper 90% hazardous locations; the percent in the parenthesis stands for the corresponding percentage.

Table 64: Similarity of Identification Results (δ = 0.95) of Various Methods (Functional Code: 1)

	Bayesian	Frequency	Rate	the ARP[1]
Bayesian		18 (90%)	12 (60%)	7 (35%)
Frequency	18 (90%)		14 (70%)	9 (45%)
Rate	12 (60%)	14 (70%)		13 (65%)
the ARP1	7 (35%)	9 (45%)	13 (65%)	

Note: 1: the ARP—Method of Accident Reduction Potential.
2: The number means the number of locations identified by both methods as upper 95% hazardous locations; the percent in the parenthesis stands for the corresponding percentage.

Table 65: Results of Site Consistency Test of Various Methods (Functional Code: 1)

Method	δ = 0.90		δ = 0.95	
	2000 Crashes	2001-2002 Crashes	2000 Crashes	2001-2002 Crashes
Frequency	384	724	249	418
Rate	341	541	204	259
Bayesian	377	751	248	421
the ARP[1]	347	573	236	354

Note: 1: the ARP—Method of Accident Reduction Potential.

Table 66: Results of Method Consistency Test of Various Methods (Functional Code: 1)

Method	δ = 0.90	δ = 0.95
Bayesian	26 (65%)	9 (45%)
Frequency	21 (52.5%)	9 (45%)
Rate	20 (50%)	4 (20%)
the ARP[1]	16 (40%)	6 (30%)

Note: 1: the ARP—Method of Accident Reduction Potential.

2: The number means the number of locations identified by methods in both periods, the percent in the parenthesis stands for the corresponding percentage.

Table 67: Results of Total Ranking Differences Test of Various Methods (Functional Code: 1)

Method	δ = 0.90	δ = 0.95
Bayesian	2042	1547
Frequency	2674	2000
Rate	3781	2522
the ARP[1]	4739	3038

Note: 1: the ARP—Method of Accident Reduction Potential.

Table 68: Results of False Identification Test of Various Methods (Functional Code: 1)

Method		$\delta = 0.90$	$\delta = 0.95$
Bayesian	FN	15 (2.1%)	13 (1.8%)
	FP	15 (19.2%)	13 (34.2%)
	FI	30 (3.7%)	26 (26%)
Frequency	FN	19 (2.6%)	13 (1.8%)
	FP	19 (24.4%)	13 (34.2%)
	FI	38 (4.7%)	26 (26%)
Rate	FN	46 (6.3%)	25 (3.3%)
	FP	46 (59.0%)	25 (65.8%)
	FI	92 (11.4%)	50 (6.2%)
the ARP[1]	FN	38 (5.2%)	18 (2.3%)
	FP	38 (48.7%)	18 (47.4%)
	FI	76 (9.4%)	36 (4.5%)

Note: 1: the ARP—Method of Accident Reduction Potential.
2: The number means the number of false identifications; the percent in the parenthesis stands for the corresponding percentage.

Table 69: Results of False True Poisson Means Differences Test of Various Methods (Functional Code: 1)

Method		$\delta = 0.90$	$\delta = 0.95$
Bayesian	FN	29.9	47.5
	FP	52.5	48.1
	FI	82.4	95.6
Frequency	FN	46.9	48.9
	FP	57.7	40.3
	FI	104.6	89.2
the ARP[1]	FN	91.8	65.5
	FP	182.2	112.5
	FI	274.0	178.0
Rate	FN	151.7	102.5
	FP	151.0	98.4
	FI	302.7	200.9

Note: 1: the ARP—Method of Accident Reduction Potential.

Table 70: Similarity of Identification Results (δ = 0.90) of Various Methods (Functional Code: 2)

	Bayesian	Frequency	Rate	the ARP[1]
Bayesian		34(77%)	21 (47.7%)	32 (72.7%)
Frequency	34(77%)		30 (68.2%)	42 (95.5%)
Rate	21 (47.7%)	30 (68.2%)		32 (72.7%)
the ARP1	32 (72.7%)	42 (95.5%)	32 (72.7%)	

Note: 1: the ARP—Method of Accident Reduction Potential.

2: The number means the number of locations identified by both methods as upper 90% hazardous locations; the percent in the parenthesis stands for the corresponding percentage.

Table 71: Similarity of Identification Results (δ = 0.95) of Various Methods (Functional Code: 2)

	Bayesian	Frequency	Rate	the ARP[1]
Bayesian		19 (86.4%)	8 (36.4%)	19 (86.4%)
Frequency	19 (86.4%)		10 (45.5%)	22 (100%)
Rate	8 (36.4%)	10 (45.5%)		10 (45.5%)
the ARP1	19 (86.4%)	22 (100%)	10 (45.5%)	

Note: 1: the ARP—Method of Accident Reduction Potential.

2: The number means the number of locations identified by both methods as upper 95% hazardous locations; the percent in the parenthesis stands for the corresponding percentage.

Table 72: Results of Site Consistency Test of Various Methods (Functional Code: 2)

Method	δ = 0.90		δ = 0.95	
	2000 Crashes	2001-2002 Crashes	2000 Crashes	2001-2002 Crashes
Frequency	520	955	361	694
Rate	448	784	263	508
Bayesian	500	976	354	731
the ARP[1]	517	937	361	694

Note: 1: the ARP—Method of Accident Reduction Potential.

Table 73: Results of Method Consistency Test of Various Methods (Functional Code: 2)

Method	$\delta = 0.90$	$\delta = 0.95$
Bayesian	23 (52.3%)	14 (63.6%)
Frequency	20 (45.5%)	11 (50%)
Rate	19 (43.2%)	7 (31.9%)
the ARP[1]	18 (40.9%)	10 (45.5%)

Note: 1: the ARP—Method of Accident Reduction Potential.
2: The number means the number of locations identified by methods in both periods; the percent in the parenthesis stands for the corresponding percentage.

Table 74: Results of Total Ranking Differences Test of Various Methods (Functional Code: 2)

Method	$\delta = 0.90$	$\delta = 0.95$
Bayesian	3705	1963
Frequency	4683	2715
Rate	7304	4432
the ARP[1]	5165	3395

Note: 1: the ARP—Method of Accident Reduction Potential.

Table 75: Results of False Identification Test of Various Methods (Functional Code: 2)

Method		$\delta = 0.90$	$\delta = 0.95$
Bayesian	**FN**	25 (3.1%)	10 (1.2%)
	FP	25 (29.1%)	10 (23.8%)
	FI	50 (5.7%)	20 (2.3%)
Frequency	**FN**	24 (3.0%)	11 (1.3%)
	FP	24 (27.9%)	11 (26.2%)
	FI	48 (5.4%)	22 (2.5%)
Rate	**FN**	49 (6.2%)	29 (3.5%)
	FP	49 (57.0%)	29 (69.0%)
	FI	98 (11.1%)	58 (6.6%)
the ARP[1]	**FN**	26 (3.3%)	12 (1.4%)
	FP	26 (30.2%)	12 (28.6%)
	FI	52 (5.9%)	24 (2.7%)

Note: 1: the ARP—Method of Accident Reduction Potential.
2: The number means the number of false identifications; the percent in the parenthesis stands for the corresponding percentage.

Table 76: Results of False True Poisson Means Differences Test of Various Methods (Functional Code: 2)

Method		$\delta = 0.90$	$\delta = 0.95$
Bayesian	**FN**	91.1	54.6
	FP	90.2	65.7
	FI	181.3	120.3
Frequency	**FN**	85.6	52.1
	FP	99.0	68.5
	FI	184.6	120.6
the ARP[1]	**FN**	81.3	64.5
	FP	100.1	69.1
	FI	181.4	133.6
Rate	**FN**	253.4	194.3
	FP	199.9	167.2
	FI	453.3	361.5

Note: 1: the ARP—Method of Accident Reduction Potential.

Table 77: Similarity of Identification Results (δ = 0.90) of Various Methods (Functional Code: 6)

	Bayesian	Frequency	Rate	the ARP[1]
Bayesian		38 (86.4%)	15 (34.1%)	31 (70.5%)
Frequency	38 (86.4%)		20 (45.5%)	37 (84.1%)
Rate	15 (34.1%)	20 (45.5%)		23 (52.3%)
the ARP1	31 (70.5%)	37 (84.1%)	23 (52.3%)	

Note: 1: the ARP—Method of Accident Reduction Potential.
2: The number means the number of locations identified by both methods as upper 90% hazardous locations; the percent in the parenthesis stands for the corresponding percentage.

Table 78: Similarity of Identification Results (δ = 0.95) of Various Methods (Functional Code: 6)

	Bayesian	Frequency	Rate	the ARP[1]
Bayesian		14 (63.6%)	3 (13.6%)	13 (59.1%)
Frequency	14 (63.6%)		11 (50%)	21(95.5%)
Rate	3 (13.6%)	11 (50%)		11 (50%)
the ARP1	13 (59.1%)	21(95.5%)	11 (50%)	

Note: 1: the ARP—Method of Accident Reduction Potential.
2: The number means the number of locations identified by both methods as upper 95% hazardous locations; the percent in the parenthesis stands for the corresponding percentage.

Table 79: Results of Site Consistency Test of Various Methods (Functional Code: 6)

Method	δ = 0.90		δ = 0.95	
	2000 Crashes	2001-2002 Crashes	2000 Crashes	2001-2002 Crashes
Frequency	471	718	319	399
Rate	334	340	219	181
Bayesian	452	723	283	431
the ARP[1]	464	659	318	408

Note: 1: the ARP—Method of Accident Reduction Potential.

Table 80: Results of Method Consistency Test of Various Methods (Functional Code: 6)

Method	δ = 0.90	δ = 0.95
Bayesian	23 (52.3%)	10 (45.5%)
Frequency	21 (47.7%)	7 (31.8%)
Rate	13 (29.5%)	3 (13.6%)
the ARP[1]	18 (40.9%)	7 (31.8%)

Note: 1: the ARP—Method of Accident Reduction Potential.
2: The number means the number of locations identified by methods in both periods; the percent in the parenthesis stands for the corresponding percentage.

Table 81: Results of Total Ranking Differences Test of Various Methods (Functional Code: 6)

Method	δ = 0.90	δ = 0.95
Bayesian	2118	970
Frequency	5040	2050
Rate	6904	4380
the ARP[1]	6900	2651

Note: 1: the ARP—Method of Accident Reduction Potential.

Table 82: Results of False Identification Test of Various Methods (Functional Code: 6)

Method		δ = 0.90	δ = 0.95
Bayesian	**FN**	24 (3.1%)	17 (2.0%)
	FP	24 (27.9%)	17 (40.5%)
	FI	48 (5.5%)	34 (3.9%)
Frequency	**FN**	23 (2.9%)	17 (2.0%)
	FP	23 (26.7%)	17 (40.5%)
	FI	46 (5.3%)	34 (3.9%)
Rate	**FN**	52 (6.6%)	31 (3.7%)
	FP	52 (60.5%)	31 (73.8%)
	FI	104 (11.9%)	62 (7.1%)
the ARP[1]	**FN**	31 (3.9%)	16 (1.9%)
	FP	31 (36.0%)	16 (38.1%)
	FI	62 (7.1%)	32 (3.7%)

Note: 1: the ARP—Method of Accident Reduction Potential.
2: The number means the number of false identifications; the percent in the parenthesis stands for the corresponding percentage.

Table 83: Results of False True Poisson Means Differences Test of Various Methods (Functional Code: 6)

Method		δ = 0.90	δ = 0.95
Bayesian	**FN**	91.8	57.4
	FP	61.2	76.4
	FI	153.0	133.8
Frequency	**FN**	71.7	43.0
	FP	60.8	67.4
	FI	132.5	110.4
the ARP[1]	**FN**	85.4	85.0
	FP	126.6	40.8
	FI	212.0	125.9
Rate	**FN**	215.0	161.6
	FP	190.5	121.7
	FI	405.5	283.4

Note: 1: the ARP—Method of Accident Reduction Potential.

Table 84: Similarity of Identification Results (δ = 0.90) of Various Methods (Functional Code: 7)

	Bayesian	Frequency	Rate	the ARP[1]
Bayesian		54 (85.7%)	23 (36.5%)	54 (85.7%)
Frequency	54 (85.7%)		32 (50.8%)	61 (98.4%)
Rate	23 (36.5%)	32 (50.8%)		32 (50.8%)
the ARP1	54 (85.7%)	61 (98.4%)	32 (50.8%)	

Note: 1: the ARP—Method of Accident Reduction Potential.
2: The number means the number of locations identified by both methods as upper 90% hazardous locations; the percent in the parenthesis stands for the corresponding percentage.

Table 85: Similarity of Identification Results (δ = 0.95) of Various Methods (Functional Code: 7)

	Bayesian	Frequency	Rate	the ARP[1]
Bayesian		25 (78.1%)	6 (18.8%)	25 (78.1%)
Frequency	25 (78.1%)		12 (37.5%)	32 (100%)
Rate	6 (18.8%)	12 (38.7%)		12 (37.5%)
the ARP1	25 (78.1%)	32 (100%)	12 (37.5%)	

Note: 1: the ARP—Method of Accident Reduction Potential.
2: The number means the number of locations identified by both methods as upper 95% hazardous locations; the percent in the parenthesis stands for the corresponding percentage.

Table 86: Results of Site Consistency Test of Various Methods (Functional Code: 7)

Method	δ = 0.90		δ = 0.95	
	2000 Crashes	2001-2002 Crashes	2000 Crashes	2001-2002 Crashes
Frequency	349	607	240	409
Rate	225	326	126	165
Bayesian	337	622	230	438
the ARP[1]	349	602	240	409

Note: 1: the ARP—Method of Accident Reduction Potential.

Table 87: Results of Method Consistency Test of Various Methods (Functional Code: 7)

Method	$\delta = 0.90$	$\delta = 0.95$
Bayesian	34 (54.0%)	12 (37.5%)
Frequency	29 (46.0%)	12 (37.5%)
Rate	30 (47.6%)	13 (40.6%)
the ARP[1]	29 (46.0%)	13 (40.6%)

Note: 1: the ARP—Method of Accident Reduction Potential.
2: The number means the number of locations identified by methods in both periods; the percent in the parenthesis stands for the corresponding percentage.

Table 88: Results of Total Ranking Differences Test of Various Methods (Functional Code: 7)

Method	$\delta = 0.90$	$\delta = 0.95$
Bayesian	5107	2054
Frequency	10764	5016
Rate	8979	5081
the ARP[1]	8276	5246

Note: 1: the ARP—Method of Accident Reduction Potential.

Table 89: Results of False Identification Test of Various Methods (Functional Code: 7)

Method		$\delta = 0.90$	$\delta = 0.95$
Bayesian	**FN**	33 (2.9%)	21 (1.8%)
	FP	33 (26.6%)	21 (33.9%)
	FI	66 (5.3%)	42 (3.3%)
Frequency	**FN**	35 (3.1%)	20 (1.7%)
	FP	35 (28.2%)	20 (32.3%)
	FI	35 (5.6%)	40 (3.2%)
Rate	**FN**	81 (7.2%)	49 (4.1%)
	FP	81 (65.3%)	49 (79.0%)
	FI	160 (12.9%)	98 (7.8%)
the ARP[1]	**FN**	35 (3.1%)	19 (1.6%)
	FP	35 (28.2%)	19 (30.8%)
	FI	35 (5.6%)	38 (3.0%)

Note: 1: the ARP—Method of Accident Reduction Potential.
2: The number means the number of false identifications; the percent in the parenthesis stands for the corresponding percentage.

Table 90: Results of False True Poisson Means Differences Test of Various Methods (Functional Code: 7)

Method		$\delta = 0.90$	$\delta = 0.95$
Bayesian	FN	77.2	51.0
	FP	54.0	58.9
	FI	131.2	109.9
Frequency	FN	59.4	54.6
	FP	66.2	59.3
	FI	125.6	113.9
the ARP[1]	FN	70.0	61.3
	FP	72.0	73.5
	FI	142.0	134.8
Rate	FN	252.4	186.4
	FP	139.6	146.4
	FI	392.0	332.8

Note: 1: the ARP—Method of Accident Reduction Potential.

Table 91: Similarity of Identification Results (δ = 0.90) of Various Methods (Functional Code: 8)

	Bayesian	Frequency	Rate	the ARP[1]
Bayesian		5(50%)	2 (20%)	5 (50%)
Frequency	5(50%)		7 (70%)	10 (100%)
Rate	2 (20%)	7 (70%)		7 (70%)
the ARP1	5(50%)	10 (100%)	7 (70%)	

Note: 1: the ARP—Method of Accident Reduction Potential.
2: The number means the number of locations identified by both methods as upper 90% hazardous locations; the percent in the parenthesis stands for the corresponding percentage.

Table 92: Similarity of Identification Results (δ = 0.95) of Various Methods (Functional Code: 8)

	Bayesian	Frequency	Rate	the ARP[1]
Bayesian		2 (40%)	0 (0%)	2 (40%)
Frequency	2 (40%)		3 (60%)	5 (10%)
Rate	0 (0%)	3(60%)		3 (60%)
the ARP1	2 (40%)	5 (10%)	3 (60%)	

Note: 1: the ARP—Method of Accident Reduction Potential.
2: The number means the number of locations identified by both methods as upper 95% hazardous locations; the percent in the parenthesis stands for the corresponding percentage.

Table 93: Results of Site Consistency Test of Various Methods (Functional Code: 8)

Method	δ = 0.90		δ = 0.95	
	2000 Crashes	2001-2002 Crashes	2000 Crashes	2001-2002 Crashes
Frequency	10.5	8.9	6.5	4.9
Rate	9.5	8.2	5.3	3.5
Bayesian	8.8	12.8	4.9	6.9
the ARP[1]	10.5	8.9	6.5	4.9

Note: 1: the ARP—Method of Accident Reduction Potential.

Table 94: Results of Method Consistency Test of Various Methods (Functional Code: 8)

Method	$\delta = 0.90$	$\delta = 0.95$
Bayesian	2 (20%)	1 (20%)
Frequency	2 (20%)	1 (20%)
Rate	3 (30%)	1 (20%)
the ARP[1]	2 (20%)	0 (0%)

Note: 1: the ARP—Method of Accident Reduction Potential.
2: The number means the number of locations identified by methods in both periods; the percent in the parenthesis stands for the corresponding percentage.

Table 95: Results of Total Ranking Differences Test of Various Methods (Functional Code: 8)

Method	$\delta = 0.90$	$\delta = 0.95$
Bayesian	223	89
Frequency	437	147
Rate	475	274
the ARP[1]	592	320

Note: 1: the ARP—Method of Accident Reduction Potential.

Table 96: Results of False Identification Test of Various Methods (Functional Code: 8)

Method		$\delta = 0.90$	$\delta = 0.95$
Bayesian	**FN**	9 (4.9%)	5 (2.6%)
	FP	9 (50.0%)	5 (62.5%)
	FI	18 (9%)	10 (5%)
Frequency	**FN**	9 (4.9%)	4 (2.1%)
	FP	9 (50.0%)	4 (50%)
	FI	18 (9%)	8 (4%)
Rate	**FN**	14 (7.7%)	7 (3.6%)
	FP	14 (77.8%)	49 (87.5%)
	FI	28 (14%)	98 (7 %)
the ARP[1]	**FN**	9 (4.9%)	5 (2.6%)
	FP	9 (50.0%)	5 (62.5%)
	FI	18 (9%)	10 (5%)

Note: 1: the ARP—Method of Accident Reduction Potential.
2: The number means the number of false identifications; the percent in the parenthesis stands for the corresponding percentage.

Table 97: Results of False True Poisson Means Differences Test of Various Methods (Functional Code: 8)

Method		$\delta = 0.90$	$\delta = 0.95$
Bayesian	**FN**	4.5	4.2
	FP	4.0	3.1
	FI	8.5	7.3
Frequency	**FN**	3.9	3.2
	FP	3.1	3.4
	FI	7.0	6.6
the ARP[1]	**FN**	4.3	3.4
	FP	5.0	4.8
	FI	9.3	8.2
Rate	**FN**	11.7	8.2
	FP	4.9	5.9
	FI	16.6	14.1

Note: 1: the ARP—Method of Accident Reduction Potential.

Table 98: Similarity of Identification Results (δ = 0.90) of Various Methods (Functional Code: 11)

	Bayesian	Frequency	Rate	the ARP[1]
Bayesian		19 (90.5%)	13 (61.9%)	17 (80.9%)
Frequency	19 (90.5%)		15 (71.4%)	19 (90.5%)
Rate	13 (61.9%)	15 (71.4%)		16 (76.2%)
the ARP1	17 (80.9%)	19 (90.5%)	16 (76.2%)	

Note: 1: the ARP—Method of Accident Reduction Potential.
2: The number means the number of locations identified by both methods as upper 90% hazardous locations; the percent in the parenthesis stands for the corresponding percentage.

Table 99: Similarity of Identification Results (δ = 0.95) of Various Methods (Functional Code: 11)

	Bayesian	Frequency	Rate	the ARP[1]
Bayesian		11 (100%)	6 (54.5%)	10 (90.9%)
Frequency	11 (100%)		6 (54.5%)	0 (0%)
Rate	6 (54.5%)	6 (54.5%)		6 (54.5%)
the ARP1	10 (90.9%)	0 (0%)	6 (54.5%)	

Note: 1: the ARP—Method of Accident Reduction Potential.
2: The number means the number of locations identified by both methods as upper 95% hazardous locations; the percent in the parenthesis stands for the corresponding percentage.

Table 100: Results of Site Consistency Test of Various Methods (Functional Code: 11)

Method	δ = 0.90		δ = 0.95	
	2000 Crashes	2001-2002 Crashes	2000 Crashes	2001-2002 Crashes
Frequency	2774	6706	1868	4589
Rate	2433	5606	1438	3505
Bayesian	2758	6694	1868	4589
the ARP[1]	1868	4589	1865	4837

Note: 1: the ARP—Method of Accident Reduction Potential.

Table 101: Results of Method Consistency Test of Various Methods (Functional Code: 11)

Method	$\delta = 0.90$	$\delta = 0.95$
Bayesian	17 (80.9%)	7 (63.6%)
Frequency	15 (71.4%)	7 (63.6%)
Rate	12 (57.2%)	8 (72.7%)
the ARP[1]	15 (71.4%)	7 (63.6%)

Note: 1: the ARP—Method of Accident Reduction Potential.
2: The number means the number of locations identified by methods in both periods; the percent in the parenthesis stands for the corresponding percentage.

Table 102: Results of Total Ranking Differences Test of Various Methods (Functional Code: 11)

Method	$\delta = 0.90$	$\delta = 0.95$
Bayesian	99	30
Frequency	112	39
Rate	440	147
the ARP[1]	143	65

Note: 1: the ARP—Method of Accident Reduction Potential.

Table 103: Results of False Identification Test of Various Methods (Functional Code: 11)

Method		$\delta = 0.90$	$\delta = 0.95$
Bayesian	**FN**	4 (1.1%)	4 (1.1%)
	FP	4 (10%)	4 (10%)
	FI	8 (1.9%)	8 (1.9%)
Frequency	**FN**	6 (1.6%)	4 (1.1%)
	FP	6 (15%)	4 (10%)
	FI	12 (2.9%)	8 (1.9%)
Rate	**FN**	17 (4.5%)	12 (3.0%)
	FP	17 (42.5%)	12 (60%)
	FI	34 (8.2%)	24 (5.8 %)
the ARP[1]	**FN**	9 (2.4%)	4 (1.1%)
	FP	9 (22.5%)	4 (10%)
	FI	18 (4.3%)	8 (1.9%)

Note: 1: the ARP—Method of Accident Reduction Potential.
2: The number means the number of false identifications; the percent in the parenthesis stands for the corresponding percentage.

Table 104: Results of False True Poisson Means Differences Test of Various Methods (Functional Code: 11)

Method		$\delta = 0.90$	$\delta = 0.95$
Bayesian	**FN**	27.2	158.6
	FP	44.1	133.6
	FI	71.3	292.2
Frequency	**FN**	95.1	128.6
	FP	83.1	184.2
	FI	178.2	312.8
the ARP[1]	**FN**	101.0	213.4
	FP	148.5	102.6
	FI	249.5	316.0
Rate	**FN**	1402.9	1108.2
	FP	516.1	449.5
	FI	1919.0	1557.7

Note: 1: the ARP—Method of Accident Reduction Potential.

Table 105: Similarity of Identification Results (δ = 0.90) of Various Methods (Functional Code: 12)

	Bayesian	Frequency	Rate	the ARP[1]
Bayesian		15 (88.2%)	10 (58.8%)	15 (88.2%)
Frequency	15 (88.2%)		10 (58.8%)	16 (94.1%)
Rate	10 (58.8%)	10 (58.8%)		10 (58.8%)
the ARP1	15 (88.2%)	16 (94.1%)	10 (58.8%)	

Note: 1: the ARP—Method of Accident Reduction Potential.
2: The number means the number of locations identified by both methods as upper 90% hazardous locations; the percent in the parenthesis stands for the corresponding percentage.

Table 106: Similarity of Identification Results (δ = 0.95) of Various Methods (Functional Code: 12)

	Bayesian	Frequency	Rate	the ARP[1]
Bayesian		8 (88.9%)	4 (44.4%)	8 (88.9%)
Frequency	8 (88.9%)		5 (55.6%)	9 (100%)
Rate	4 (44.4%)	5 (55.6%)		5 (55.6%)
the ARP1	8 (88.9%)	9 (100%)	5 (55.6%)	

Note: 1: the ARP—Method of Accident Reduction Potential.
2: The number means the number of locations identified by both methods as upper 95% hazardous locations; the percent in the parenthesis stands for the corresponding percentage.

Table 107: Results of Site Consistency Test of Various Methods (Functional Code: 12)

Method	δ = 0.90		δ = 0.95	
	2000 Crashes	2001-2002 Crashes	2000 Crashes	2001-2002 Crashes
Frequency	1352	3437	888	2337
Rate	1064	2602	662	1840
Bayesian	1337	3452	878	2353
the ARP[1]	1352	3356	888	2337

Note: 1: the ARP—Method of Accident Reduction Potential.

Table 108: Results of Method Consistency Test of Various Methods (Functional Code: 12)

Method	$\delta = 0.90$	$\delta = 0.95$
Bayesian	12 (70.6%)	8 (88.9%)
Frequency	12 (70.6%)	7 (77.8%)
Rate	10 (58.8%)	7 (77.8%)
the ARP[1]	11 (64.7%)	6 (66.7%)

Note: 1: the ARP—Method of Accident Reduction Potential.
2: The number means the number of locations identified by methods in both periods; the percent in the parenthesis stands for the corresponding percentage.

Table 109: Results of Total Ranking Differences Test of Various Methods (Functional Code: 12)

Method	$\delta = 0.90$	$\delta = 0.95$
Bayesian	122	29
Frequency	135	30
Rate	277	73
the ARP[1]	236	37

Note: 1: the ARP—Method of Accident Reduction Potential.

Table 110: Results of False Identification Test of Various Methods (Functional Code: 12)

Method		$\delta = 0.90$	$\delta = 0.95$
Bayesian	**FN**	9 (3.0%)	1 (0.3%)
	FP	9 (28.1%)	1 (6.3%)
	FI	18 (5.5%)	2 (0.6%)
Frequency	**FN**	5 (1.7%)	2 (0.6%)
	FP	5 (15.6%)	2 (12.5%)
	FI	10 (3.0%)	4(1.2%)
Rate	**FN**	14 (4.7%)	11 (3.5%)
	FP	14 (43.8%)	11 (68.8%)
	FI	28 (8.5%)	22 (6.7 %)
the ARP[1]	**FN**	7 (2.3%)	4 (1.3%)
	FP	7 (21.9%)	4 (25%)
	FI	14 (4.2%)	8 (2.4%)

Note: 1: the ARP—Method of Accident Reduction Potential.
2: The number means the number of false identifications; the percent in the parenthesis stands for the corresponding percentage.

Table 111: Results of False True Poisson Means Differences Test of Various Methods (Functional Code: 12)

Method		$\delta = 0.90$	$\delta = 0.95$
Bayesian	FN	171.7	29.3
	FP	268.6	54.8
	FI	440.3	84.1
Frequency	FN	88.3	61.3
	FP	183.7	80.7
	FI	272.0	142.0
the ARP[1]	FN	114.3	87.5
	FP	173.5	102.0
	FI	287.9	189.5
Rate	FN	622.7	428.6
	FP	272.4	274.4
	FI	895.1	703.0

Note: 1: the ARP—Method of Accident Reduction Potential.

Table 112: Similarity of Identification Results (δ = 0.90) of Various Methods (Functional Code: 14)

	Bayesian	Frequency	Rate	the ARP[1]
Bayesian		35 (81.4%)	20 (46.5%)	37 (86.0%)
Frequency	35 (81.4%)		27 (62.8%)	40 (93.0%)
Rate	20 (46.5%)	27 (62.8%)		24 (55.8%)
the ARP1	37 (86.0%)	40 (93.0%)	24 (55.8%)	

Note: 1: the ARP—Method of Accident Reduction Potential.

2: The number means the number of locations identified by both methods as upper 90% hazardous locations; the percent in the parenthesis stands for the corresponding percentage.

Table 113: Similarity of Identification Results (δ = 0.95) of Various Methods (Functional Code: 14)

	Bayesian	Frequency	Rate	the ARP[1]
Bayesian		18 (85.7%)	10 (47.6%)	19 (90.5%)
Frequency	18 (85.7%)		13 (61.9%)	19 (90.5%)
Rate	10 (47.6%)	13 (61.9%)		11 (52.4%)
the ARP1	19 (90.5%)	19 (90.5%)	11 (52.4%)	

Note: 1: the ARP—Method of Accident Reduction Potential.

2: The number means the number of locations identified by both methods as upper 95% hazardous locations; the percent in the parenthesis stands for the corresponding percentage.

Table 114: Results of Site Consistency Test of Various Methods (Functional Code: 14)

Method	δ = 0.90		δ = 0.95	
	2000 Crashes	2001-2002 Crashes	2000 Crashes	2001-2002 Crashes
Frequency	2020	4896	1445	2833
Rate	1731	3970	1160	1870
Bayesian	1964	4727	1271	2863
the ARP[1]	1214	4899	1434	2856

Note: 1: the ARP—Method of Accident Reduction Potential.

Table 115: Results of Method Consistency Test of Various Methods (Functional Code: 14)

Method	$\delta = 0.90$	$\delta = 0.95$
Bayesian	21 (48.9%)	8 (38.1%)
Frequency	21 (48.9%)	7 (33.3%)
Rate	17 (39.5%)	5 (23.8%)
the ARP[1]	22 (51.2%)	8 (38.1%)

Note: 1: the ARP—Method of Accident Reduction Potential.
2: The number means the number of locations identified by methods in both periods, the percent in the parenthesis stands for the corresponding percentage.

Table 116: Results of Total Ranking Differences Test of Various Methods (Functional Code: 14)

Method	$\delta = 0.90$	$\delta = 0.95$
Bayesian	2296	2925
Frequency	5030	1253
Rate	5577	3492
the ARP[1]	5899	3585

Note: 1: the ARP—Method of Accident Reduction Potential.

Table 117: Results of False Identification Test of Various Methods (Functional Code: 14)

Method		$\delta = 0.90$	$\delta = 0.95$
Bayesian	**FN**	25 (3.2%)	16 (2.0%)
	FP	25 (29.8%)	16 (40%)
	FI	50 (5.8%)	32 (3.7%)
Frequency	**FN**	22 (2.8%)	14 (1.7%)
	FP	22 (26.2%)	14 (35%)
	FI	44 (5.1%)	28 (3.3%)
Rate	**FN**	42 (5.4%)	24 (2.9%)
	FP	42 (50%)	24 (60%)
	FI	84 (9.8%)	48 (5.6%)
the ARP[1]	**FN**	21 (2.7%)	13 (1.6%)
	FP	21 (25%)	13 (32.5%)
	FI	42 (4.9%)	26 (3.0%)

Note: 1: the ARP—Method of Accident Reduction Potential.
2: The number means the number of false identifications; the percent in the parenthesis stands for the corresponding percentage.

Table 118: Results of False True Poisson Means Differences Test of Various Methods (Functional Code: 14)

Method		$\delta = 0.90$	$\delta = 0.95$
Bayesian	**FN**	533.2	575.3
	FP	570.9	609.1
	FI	1104.2	1184.4
Frequency	**FN**	482.5	360.9
	FP	601.5	529.1
	FI	1084.0	890.0
the ARP[1]	**FN**	383.8	479.9
	FP	524.8	570.3
	FI	908.6	1050.2
Rate	**FN**	1010.7	851.7
	FP	897.1	920.5
	FI	1907.8	1772.2

Note: 1: the ARP—Method of Accident Reduction Potential.

Table 119: Similarity of Identification Results ($\delta = 0.90$) of Various Methods (Functional Code: 16)

	Bayesian	Frequency	Rate	the ARP[1]
Bayesian		13 (81.3%)	7 (43.8%)	14 (87.5%)
Frequency	13 (81.3%)		10 (62.5%)	13 (81.3%)
Rate	7 (43.8%)	10 (62.5%)		7 (43.8%)
the ARP[1]	14 (87.5%)	13 (81.3%)	7 (43.8%)	

Note: 1: the ARP—Method of Accident Reduction Potential.
2: The number means the number of locations identified by both methods as upper 90% hazardous locations; the percent in the parenthesis stands for the corresponding percentage.

Table 120: Similarity of Identification Results ($\delta = 0.95$) of Various Methods (Functional Code: 16)

	Bayesian	Frequency	Rate	the ARP[1]
Bayesian		8 (100%)	3 (37.5%)	8 (100%)
Frequency	8 (100%)		3 (37.5%)	8 (100%)
Rate	3 (37.5%)	3 (37.5%)		3 (37.5%)
the ARP1	8 (100%)	8 (100%)	3 (37.5%)	

Note: 1: the ARP—Method of Accident Reduction Potential.
2: The number means the number of locations identified by both methods as upper 95% hazardous locations; the percent in the parenthesis stands for the corresponding percentage.

Table 121: Results of Site Consistency Test of Various Methods (Functional Code: 16)

Method	$\delta = 0.90$		$\delta = 0.95$	
	2000 Crashes	2001-2002 Crashes	2000 Crashes	2001-2002 Crashes
Frequency	395	1170	262	921
Rate	313	935	180	294
Bayesian	389	1246	262	921
the ARP[1]	392	1275	262	921

Note: 1: the ARP—Method of Accident Reduction Potential.

Table 122: Results of Method Consistency Test of Various Methods (Functional Code: 16)

Method	$\delta = 0.90$	$\delta = 0.95$
Bayesian	9 (56.3%)	2 (25%)
Frequency	7 (43.8%)	2 (25%)
Rate	7 (43.8%)	1 (12.5%)
the ARP[1]	8 (50%)	2 (25%)

Note: 1: the ARP—Method of Accident Reduction Potential.
2: The number means the number of locations identified by methods in both periods; the percent in the parenthesis stands for the corresponding percentage.

Table 123: Results of Total Ranking Differences Test of Various Methods (Functional Code: 16)

Method	$\delta = 0.90$	$\delta = 0.95$
Bayesian	463	298
Frequency	727	435
Rate	1132	811
the ARP[1]	651	450

Note: 1: the ARP—Method of Accident Reduction Potential.

Table 124: Results of False Identification Test of Various Methods (Functional Code: 16)

Method		$\delta = 0.90$	$\delta = 0.95$
Bayesian	FN	9 (3.0%)	7 (2.2%)
	FP	9 (30%)	7 (50%)
	FI	18 (5.5%)	14 (4.3%)
Frequency	FN	9 (3.0%)	6 (1.9%)
	FP	9 (30%)	6 (42.9%)
	FI	18 (5.5%)	12 (3.7%)
Rate	FN	18 (6.0%)	11 (3.5%)
	FP	18 (60%)	11 (78.6%)
	FI	36 (11.0%)	22 (6.7%)
the ARP[1]	FN	8 (2.7%)	7 (2.2%)
	FP	8 (26.7%)	7 (50%)
	FI	16 (4.9%)	14 (4.3%)

Note: 1: the ARP—Method of Accident Reduction Potential.
2: The number means the number of false identifications; the percent in the parenthesis stands for the corresponding percentage.

Table 125: Results of False True Poisson Means Differences Test of Various Methods (Functional Code: 16)

Method		$\delta = 0.90$	$\delta = 0.95$
Bayesian	**FN**	114.9	119.3
	FP	125.8	127.8
	FI	240.7	247.1
Frequency	**FN**	108.0	106.3
	FP	123.9	135.4
	FI	231.9	241.7
the ARP[1]	**FN**	115.6	111.5
	FP	132.9	127.9
	FI	248.5	239.4
Rate	**FN**	244.2	173.9
	FP	177.9	154.3
	FI	422.1	328.2

Note: 1: the ARP—Method of Accident Reduction Potential.