1-1-2010

Exploration Of Theoretical And Application Issues In Using Fully Bayesian Methods For Road Safety Analysis

Bo Lan
Ryerson University

EXPLORATION OF THEORETICAL AND APPLICATION ISSUES IN USING FULLY BAYESIAN METHODS FOR ROAD SAFETY ANALYSIS

By

Bo Lan

Master of Engineering, University of Waterloo, Canada, 2004

Master of Applied Science, Chongqing Jiaotong University, China, 1991

A dissertation
Presented to Ryerson University
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
in the Program of
Civil Engineering

Toronto, Ontario, Canada, 2010

Author's Declaration

I hereby declare that I am the sole author of this thesis.

I authorize Ryerson University to lend this thesis to other institutions or individuals for the purpose of scholarly research.

Signature ____________________________________

I further authorize Ryerson University to reproduce this thesis by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

Signature ____________________________________

Exploration of Theoretical and Application Issues in using Fully Bayesian Methods for Road Safety Analysis

Bo Lan 2010

Doctor of Philosophy
in Civil Engineering
Ryerson University
Toronto, Ontario, Canada

ABSTRACT

The Fully Bayesian (FB) approach to road safety analysis has been available for some time, but it is largely unevaluated and untested. This study is trying to bridge the gap by conducting a thorough evaluation of FB method for black spots identification and treatment effect analysis.

First, an evaluation is conducted on the univariate FB versus the empirical Bayesian (EB) method for single level severity data through the development of various models, and multivariate FB versus univariate FB for multilevel severity data, as well as the performance of various ranking and evaluation criteria for black spots identification. It is confirmed that the FB method is superior to the EB with respect to key ranking criteria (expected rank, mode rank and median rank of posterior PM, etc.). The multivariate FB method is better than univariate FB for the multilevel severity crashes.

Then a test of the FB before-after method for treatment effect analysis is performed. Two FB testing frameworks were employed. First the univariate before-after fully Bayesian (FB) method was examined using three simulated datasets. Then multivariate Poisson log

normal (MVPLN), univariate Poisson log normal (PLN) and PG (Poisson gamma) models were evaluated using two groups of California unsignalized intersections. Hypothetical treatment sites were selected from these datasets such that a significant effect would be estimated by the naive before-after method that does not account for regression to the mean. This study confirmed that FB methods can indeed provide valid results, in that they correctly estimate a treatment effect of zero at these hypothetical treatment sites after accounting for regression to the mean.

Finally the EB and the validated FB before after methods were applied to evaluation of two treatments: the conversion of rural intersections from unsignalized to signalized control; and the conversion of road segments from a four-lane to a three-lane cross-section with two-way left turn lanes (also known as road diets). The result indicates that both FB and EB method can provide comparable treatment effect estimates. This would suggest it is still appropriate to conduct treatment effect analysis using the EB method for univairate crash data, but that it is essential in so doing to account for temporal trends in crash frequency.

ACKNOWLEDGEMENT

First and foremost, I am extremely grateful to my advisor, Dr. Bhagwant Persaud, for his guidance and great support throughout my graduate study. I especially thank him for his endless encouragement, concern, and patience in my study and life.

My sincere thanks are due to my committee members, Dr. Frank F. Saccomanno, Dr. Saeed Zolfaghari, Dr. Mike Chapman and Dr. Arnold Yuan for their time and insightful suggestions.

I wish to thank Dr. Grace Luk, professor and graduate program director in the Department of Civil Engineering, for her encouragement and kind help during the hard time of my study. I also wish to thank Craig Lyon for his time and useful comments on my research.

Finally, I would like to thank my parents, Yuanlan Chen and Jiyue Lan for their constant encouragement and support. I also thank my brother, Yi Lan, and my sisters, Hua Lan, Xue Lan and Meng Lan for their emotional support during the difficult times of this long journey.

Of all people, I am most thankful to my husband, Yong Wang, for his support and contributions. I also thank my two lovely daughters Hongfan Wang and Michelle Hongyu Lan. Without their encouragement and support, I would not have been able to complete the study.

DEDICATION

To My Grandparents,

Chen, Sanming and Jiang, Jingzhen

Table of Contents

List of Tables

List of Figures

List of Symbols and Acronyms

The following notations are used in this dissertation:

AADT	Annual average daily traffic
AR(1)	First order autoregressive model
BCI	Bayesian credit interval
CR	Crash reduction
CRR	Crash reduction rate
EB	Empirical Bayesian
FB	Fully Bayesian
GEE	Generalized estimating equation
HB	Hierarchical Bayesian
LL	Log likelihood
MCMC	Markov chain Monte Carlo
MVPLN	Multivariate Poisson lognormal
P-AR (1)	Poisson AR (1) model
pdf	Probability density function
PDO	Property damage only
PG	Poisson gamma
PLN	Poisson lognormal
PM	Posterior Poisson mean
PPSI	Pseudo potential safety improvement
PSI	Potential safety improvement
P_{worst}	The probability of a site being the highest ranked hot spot
RTM	Regression-to-mean
SPF	Safety Performance Functions
λ	The expected crash frequency at an entity had a specific treatment NOT been implemented

π	The expected crash frequency in the after a specific treatment has been implemented
μ	The expected crash frequency at similar sites
y	observed crash frequency

CHAPTER 1 INTRODUCTION

1.1 OVERVIEW OF ROAD SAFETY

1.1.1 The Status of Road Safety

Road traffic crashes constitute one of the world's largest public health and injury problems. According to the World Health Organization (2009), 1.2 million people are killed on the world's roads each year, and as many as 50 million others are injured. The injuries due to road collisions are identified as one of the leading epidemics of our time and this epidemic of road traffic injuries in most regions of the world is still increasing. The World Health Organization predicts that road traffic injury will rise to become the fifth leading cause of death by 2030. In the United States, ninety-four percent of all transportation fatalities occur on highways. More than 41,000 Americans are killed each year in motor vehicle crashes, and three million are injured. More productive years of life are lost due to road collisions than any other disease, more than heart disease and cancer combined (US Department of Transport, 2001). In Canada, about 200,000 people were injured and 3,000 people were killed in 2006 (Transport Canada, 2005). The human and economic consequences of motor vehicle crashes are unaffordable and unacceptable. The majority of motor vehicle crashes are predictable and preventable; the carnage is unnecessary.

1.1.2 The Need for Road Safety Study

The above facts suggest a significant need for continuing research to improve road safety through reduction of the harm (deaths, injuries, and property damage) that results from road crashes. Road safety management exclusively deals with road traffic crashes with regard to ways to reduce the number of crashes and their consequences.

The analytical aspects of safety management of a road network can be simply divided into three basic parts: hazardous site identification (or network ranking), implementation of treatment for identified hazardous sites, and analysis of treatment results (Hauer, 1997; Geurts and Wets, 2003). Hazardous site identification and treatment effect analysis play important roles in the improvement of road safety and provide guidance on road planning, design, maintenance, construction and operation. Both procedures will be investigated and discussed in this thesis using fully Bayesian methods. The current approaches for both procedures and the objective of this thesis will be briefly introduced in the rest of Chapter 1.

1.2 CURRENT METHODS FOR HAZARDOUS SITE IDENTIFICATION

Hazardous site identification is also called hot spot identification, black spot identification, priority investigation location flagging, identification of sites with promise, network screening, or network ranking. In this thesis, these terms are used without distinction. Network ranking is the first step in the site safety improvement process. A black spot can be defined as a site where the accident risk or a safety indicator is "unacceptably high" and safety countermeasures are most warranted. The product of network screening is a list of sites that are ranked by priority for the conduct of detailed engineering studies (Hauer et al., 2004). It is important that the process for identifying sites requiring safety investigation be efficient because resources can be wasted on sites that are incorrectly identified as unsafe and sites that are truly unsafe can remain untreated if they are not properly identified in this process (Persaud et al., 1999).

There are several methods currently applied for network screening: conventional or naïve methods, the Empirical Bayes (EB) method and Fully Bayesian or hierarchical Bayesian method (denoted as FB or HB, respectively). Details are discussed below:

1.2.1 Conventional Methods (Naïve Methods)

Conventional methods (or naïve methods) are used to rank sites utilizing accident counts and/or rates, often in a statistical quality control framework. Top ranked sites are identified as sites with promise for further examination and possible treatment. Typically, resources are invested to improve correctable sites from the top down until allocated funds are expended.

The naïve methods are now known to have difficulties in identifying deviant sites because of potential bias due to the regression-to-the-mean (RTM) phenomenon in which sites with a randomly high accident count can be wrongly identified as being hazardous and vice versa (Persaud et al., 1999; Hauer et al., 2004; Hauer, 1996; Elvik, 2008a; Elvik, 2008b; Brüde and Larson et al., 1988). In the area of traffic safety, the RTM effect may be explained as follows.

Sites with accident counts above (or below) the expected frequency at sites with similar traits in a short period will in the following short period have accident counts which on average, are closer to the expected number of accidents at sites with similar traits. That is, when data is not enough, i.e. at a particular location which only spans a few years, a naïve statistical analysis that relies on information from a short period data at that site fails to capture the true long-term behavior of that site. The estimated long-term crash frequency obtained over a few years can be excessively influenced by a single year with an unusually high (or low) number of crashes. This is known as the RTM effect. In other words, the crash counts have regressed to the mean (see Davis, 1976; Bland and Altman, 1994a and 1994b, for a general description and examples).

A. Ranking criteria for the conventional method

The ranking criteria associated with the conventional method are the upper tail crash frequency, upper tail crash rates, or upper tail crash frequency combining upper tail crash rates (Elvik, 2008a). The details are described below.

a. Upper Tail Crash Frequency

Black spots are identified based on the total crash counts in the whole period (normally 3 or 4 years), where the recorded number of crashes belongs to the upper percentage (i.e., 1%, 2.5% or 5%) or the top ranking of the entire population distribution.

b. Upper Tail Crash Rates

The procedure is the same as above, but crash rates are used as criteria instead. The crash rate is the accident count divided by traffic volume or entering vehicle volume, usually in units such as crashes per thousand or million vehicles.

The use of crash rates makes an implicit assumption that crashes are linear to exposure. The possible nonlinearity of the relation, which many investigators have confirmed with developed safety performance functions (SPFs) (Hauer, 1992; Hauer, 1997; Persaud et al., 1999; Lord and Persaud, 2002; Miao and Lord, 2007; Persaud and Nguyen, 1998; Persaud et al., 2002; Sayed and Rodrigez, 1999; Turner-Fairbank Highway Research Center, 1999; Brüde and Larson, 1988 etc.), is the primary argument against using crash rates as a criterion.

c. Upper Crash Frequency Combined with Upper Crash Rates

Sites that record a number of accidents greater than upper critical percentage values (i.e. 2.5%, 1%, or 5%) in a population of sites and have higher-than-average accident rates are classified as hazardous sites. The average accident rate refers to the overall average for the whole population of sites.

All of the above criteria directly use observed crash counts or crash rates to identify hot spots. It implies that this method cannot address the RTM problem and the results are not reliable.

B. Ranking criteria involving crash severity

Instead of ranking locations based on only one severity crashes, the identification of hazardous locations can be done in terms of the total risk when crash counts at different levels of severity are available, which is defined as the product of the crash frequency and its consequences (usually in terms of weight). It is obvious that different levels of severity crashes contribute different safety levels to the site. For example, in Belgium (Geurts, 2003) the safety of the site can be expressed as:

$$P = X + 3*Y + 5*Z \qquad (1\text{-}1)$$

where

X = total number of light injuries,

Y = total number of serious injuries, and

Z = total number of deadly injuries.

C. Application of the conventional method

Despite the drawbacks mentioned above, conventional methods are currently still used to identify black spots in many jurisdictions due to the ease of application (Elvik, 2008b; SØrenthen, 2007; Geurts, 2004). Specifically, Elvik conducted a detailed survey with the black spot identification method in Europe. The survey results are shown in Table 1-1 (Elvik, 2008b). It is seen that in many countries, recorded crash counts are directly used for ranking despite the fact that methods based on this random variable fail to address the RTM problem.

Table 1-1 Overview of definitions of hazardous road locations in selected European countries

Country	Reference to population of sites	Sliding window applied	Reference to normal level of safety	Recorded or expected number of crashes	Crash severity considered	Length of identification period
Austria	No	Yes, 250m	Yes, by means of critical values for crash rate	Recorded, minimum critical value 3-function of traffic volume	No	3 years
Denmark	Yes, detailed categorization of roadway elements	Yes, for road sections-variable length	Yes, by means of crash prediction models	Recorded, based on statistical test-minimum four crashes	No	5 years
Flanders	No	Yes, 100m	No	Recorded, weighted by severity	Yes, by means of weights	3 years
Germany	No	No, crash maps inspected	No	Recorded, minimum values 3 or 5	Yes, by different critical values	1 year (all crashes) or 3 years (injury crashes)
Hungary	No	Yes, 100 or 1000m	No	Recorded, minimum 4	No	3 years
Norway	Not when identifying black spots	Yes, 100m (spot) or 1000m (section)	Yes, by means of normal crash rates for roadway elements	Recorded higher than normal by statistical test, minimum values 4 (spots) or 10 (sections)	Yes, by estimating crash costs and potential savings	5 years
Portugal	Yes, for one definition; no for the other	Yes, for one definition; no for the other	Yes, for one definition; no for the other	Recorded in one definition (minimum 5), expected in the other	Yes in one definition (by severity weighting); no in other	1 or 5 years
Switzerland	Yes, open roads and junctions	No, fixed sections of variable length	Yes	Recorded, a set of critical values	Yes, by different critical values	2 years

1.2.2 Empirical Bayesian Method

To overcome the drawbacks of the conventional techniques, the empirical Bayes (EB) approach was originally developed for before-after studies to evaluate the effects of road safety treatments. Hauer (1980) was among the first researchers to indicate how the EB method eliminates the effects of RTM in road crash data**.** Since then, the EB approach has been suggested, examined and widely explored by several researchers (Brüde and Larson et al., 1988, Persaud et al., 1999; Persaud and Lyon, 2007; Hauer and Persaud, 1983; Hauer, 1992; Hauer et al., 2004; Hauer et al., 2002; Saccomanno et al., 2001; Cheng and Washington 2005a, 2008; Hauer, 1996; Elvik, 1997; Elvik, 2008a; Elvik, 2008b, 2008c). The EB approach is now a primary method for both treatment effect analysis and black spot identification.

A. Basics of the EB approach

According to the EB method, the best estimate of expected crashes for a specific site is obtained by combining two sources of information: (1) the crash record (y) for a specific entity or site (intersection, road section, etc.), and (2) expected crashes (μ) for similar sites, which is obtained from a crash prediction model or a safety performance function. In this way, it can be seen that the EB procedure essentially aims to smooth out the random fluctuations in crash data by specifying the safety of a site as an estimate of its long-term mean (λ) instead of its short-term count. The expected crashes for a specific site can be estimated as:

$$\lambda_i = E(\mu_i|y_i) = \alpha_i\mu_i + (1-\alpha_i)y_i \qquad (1-2)$$

where

λ_i = expected crash counts in n years at site i,

μ_i= expected crashes in n years at similar sites, estimated from safety performance functions (SPFs),

y_i = observed crash counts in n years at site i, and

α_i = the weight given to the estimated expected crashes for similar entities and estimated from the mean and variance of the SPFs estimate.

For a negative binomial (NB) model where the expected number of accidents is gamma distributed with shape parameter k, and the recorded number of crashes x_i for each entity is Poisson distributed, α_i can be calculated as:

$$\alpha_i = \frac{1/k}{1/k + \mu_i} \qquad (1-3)$$

$$1 - \alpha_i = \frac{\mu_i}{1/k + \mu_i} \qquad (1-4)$$

where k = the over-dispersion parameter of the NB model and is estimated from the SPF calibration process with the use of maximum likelihood estimation for the required reference group. The density function is:

$$P[Y = y_i] = \frac{\Gamma(y_i + k)}{\Gamma(y_i + 1)\Gamma(k)} \left(\frac{\mu_i}{\mu_i + k}\right)^{y_i} \left(\frac{k}{\mu_i + k}\right)^{k} \qquad \mu_i, k > 0,\ y_i = 0, 1, 2, \ldots \qquad (1-5)$$

Generalized linear modeling is used to estimate the required reference group SPF by using, e.g., the software package SAS (SAS Institute 1998) and assuming a negative binomial (NB) error distribution. The NB dispersion parameter, k, is also estimated by SAS.

Annual SPF multipliers are calibrated to account for temporal effects on safety due to variations in weather, demography, crash reporting and so on. After applying the multipliers, the estimated λ_i can be directly used to identify hot spots or to derive other criteria, such as potential for safety improvement (PSI) to detect hot spots.

B. Ranking criteria

a. Expected crashes λ_i

For each site, the EB estimate of the expected number of crashes is obtained by combining the observed crash counts with an estimate of the normal number of crashes from an SPF as mentioned in Equation 1-2. Sites with the highest estimates of λ_i are classified as hazardous. In other words, black spots are identified based on the expected crashes (λ_i) in the whole data group, where λ_i belongs to the highest percentage (i.e., 1%, 2.5% or 5%) or the top ranking subgroup of the whole population distribution.

b. Potential for safety improvement (PSI)

The PSI was originally introduced by McGuigan (1981) as the difference between the observed crash count of a site and expected crashes for similar sites estimated from SPFs, denoted as PPSI:

$$PPSI_i = y_i - \mu_i \qquad (1\text{-}6)$$

This ranking criterion is based on the achievable benefits due to potential highway engineering improvements. μ_i represents what might be normally expected on the basis of traffic volume alone or similar sites, and may not be reduced by highway engineering treatments. This method seems reasonable as it reflects the belief that any road or intersection which is open to traffic will have a certain level of risk.

Because the suggested accident count is included in Equation1-6, it would be difficult for application due to random fluctuations in counts where, as is often the case, a relatively short accident history is used. To overcome the limitation, Persaud (1999) proposed that the EB estimated expected crashes λ_i rather than observed crashes y_i for each site should be used. In this way, the PSI for each site is calculated and sites can be ranked to identify hazardous sites. The revised PSI is:

$$PSI_i = \lambda_i - \mu_i = \alpha_i\mu_i + (1-\alpha_i)y_i - \mu_i = (1-\alpha_i)y_i - (1-\alpha_i)\mu_i \qquad (1\text{-}7)$$

By comparing Equations 1-2 and 1-7, it can be found that higher crash counts y_i will influence the priority for further investigation for a particular site for both the EB expected crashes and EB PSI methods, but it is not the sole factor. On the other hand, the value of μ_i has a different impact on the selection of hot spots for both ranking criteria. Larger values of μ_i will increase the value of expected crashes λ_i, thus increasing the probability of the site being identified as a hot spot for the first criterion; for the PSI criterion, larger values of μ_i will decrease the corresponding probability because the PSI is diminished with the increase of μ_i. Thus, sites ranked as unsafe by the PSI method could indeed have no safety issue because of a low μ_i and unsafe sites (a largeλ_i) might not be ranked to the top list because of a large μ_i.

Different weights for crashes of different severity levels can be introduced in both criteria.

C. Evaluation criteria

There are several evaluation criteria to identify the performance of the ranking methods. Persaud and Lyon (1999) developed two criteria to quantify the performance of the methods: observed crashes in the following period, and the difference between observed crashes and predicted crashes for similar sites (estimations from SPFs) in the subsequent period. Another criterion borrowed from epidemiology (Elvik, 2008a) is usually used to conduct evaluations: the percentage of correct positives, also called sensitivity (Elvik, 2008a), which is the percentage of safe sites that are correctly claimed; and the percentage of correct negatives, also called specificity (Elvik, 2008a), which is the percentage of unsafe sites that are correctly claimed. Besides these criteria, Cheng and Washington (2008) developed three new evaluation criteria: the method consistency, total rank differences and Poisson mean differences tests. These three tests are designed to evaluate a method's performance by measuring the consistency in terms of number of the same hot spots identified, the sum of total rank differences of hazardous road sections identified and the sum of Poisson mean differences of black spots recognized across two periods. The evaluation criteria are described below:

- Criterion 1: Sum of observed crashes in the succeeding time period

$$C_1 = Y_j^{i+1} = \sum_{k=n-n\alpha}^{n} Y_{j,k}^{i+1} \qquad (1\text{-}8)$$

where

Y_j^{i+1} =sum of observed crash counts in the second time period i+1 for ranking method j,

n = total number of sites,

α = the percentage of top ranked high risk sites, and

$\sum_{k=n-n\alpha}^{n} Y_{j,k}^{i+1}$ = observed crash counts at top ranked nα sites by method j during second time period i+1.

This criterion rests on the premise that a site identified as high risk during period 1 should also reveal inferior safety performance in a subsequent period 2, given that no significant changes have occurred at the site and that the site is, in fact, high risk. It simply requires a comparison of the sum of observed crashes at the ranked high-risk sites (identified by method 1 during time period i) during the succeeding time period i + 1 to crashes which occur at the same number of high-risk sites (in time period i + 1) identified by other possible ranking criteria. The method which provides the most crashes in period i+1 at the top ranked sites is the best.

- Criterion 2: Sum of differences between observed and predicted crashes at similar sites

$$C_2 = PPSI_j^{i+1} = \sum_{k=n\alpha}^{n}(Y_{j,k}^{i+1} - SPF\ \text{estimate}_{j,k}^{i+1}) = \sum_{k=n\alpha}^{n}(Y_{j,k}^{i+1} - \mu_{j,k}^{i+1}) \qquad (1\text{-}9)$$

where

$PPSI_j^{i+1}$ =Pseudo potential safety improvement for top high risk sites ranked by method j during the second time period i+1.

Similar to the first criterion, this evaluation measure aims to determine whether the high risk sites in the first period are also high risk in the second period, except that it is based on differences between observed and predicted crashes at similar sites (estimated from SPFs) rather than crash counts. The differences between observed and predicted crashes at similar sites are somewhat like the PSI. Likewise, the sum of differences (observed crashes minus estimations from SPFs) at ranked high-risk sites during the subsequent time period i + 1 identified by method 1 (during previous time period i) are compared with those from other possible ranking criteria. A greater sum of differences means a better method, which indicates that there is more room to improve safety. It should be noted that predicted crashes for similar sites are estimated from SPFs.

- Criterion 3: Sensitivity and specificity

Data for the second time period i+1 are usually used to assess whether the hazardous sites identified in the first time period are true or false positives. The idea is that true positives will persist in having a bad safety record, whereas false positives will regress toward a more normal safety record. There are also some false negatives (i.e., sites not detected in the first time period, but which are detected in the second time period). Usually two measures are used to evaluate the ranking criteria:

$$C_{31} = \text{Sensitivity} = \frac{\text{number of correct positives}}{\text{total number of positives}} \qquad (1-10-\text{a})$$

$$C_{32} = \text{Specificity} = \frac{\text{number of correct negatives}}{\text{total number of negatives}} \qquad (1-10-\text{b})$$

where

number of correct positives = number of sites that continue to belong to the top ranked nα in the second period,

number of false positives = number of sites that drop out of the top ranked list (nα) in time periods i+1,

number of correct negatives = number of sites that do not belong to the top ranked list (nα) in both the time periods i and i+1,

number of false negatives = number of new sites that enter the list (nα) in the time period i+1,

total number of positives = the number of correct (true) positives + the number of false negatives, and

total number of negatives = the number of correct negatives + the number of false positives.

In statistics, the terms false positive, which refers to type I errors, and false negative, which is associated with type II errors, are used to describe possible errors made in a statistical decision process. Table 1-2 presents the concepts of these terms more clearly.

Table 1-2 Concept of False Positive and False Negative

		1st time period i	
		High risk	**Not High risk**
2nd time period i+1	**High risk**	True (correct) Positive	False Negative **Type II error**
	Not High risk	False Positive **Type I error**	True Negative

This evaluation criterion is borrowed from epidemiology (Deeks, 2001; Rothman and Greenland, 1998). The criterion employs a number of correct positives, or complementarily false positives, and correct negatives, or complementarily false negatives, to assess the performances of various ranking criteria. It can be seen that a larger evaluation measure, which means more consistency in the next period, results in a better method.

- Criterion 4: Method Consistency

$$C_4 = \{k^i_{n-n\alpha,}k^i_{n-n\alpha+1}, \dots, k^i_n\}_j \cap \{k^{i+1}_{n-n\alpha,}k^{i+1}_{n-n\alpha+1}, \dots, k^{i+1}_n\}_j \qquad (1\text{-}11)$$

where

$\{k^i_{n-n\alpha,}k^i_{n-n\alpha+1}, \dots, k^i_n\}_j$= top ranked nα high risk sites by method j during first time period i, and

$\{k^{i+1}_{n-n\alpha,}k^{i+1}_{n-n\alpha+1}, \dots, k^{i+1}_n\}_j$= top ranked nα high risk sites by method j during time period i+1.

This test is designed to evaluate the performance of a method by measuring the number of the same hot spots identified in both periods. This criterion is simply used to identify the intersection of the top nα ranked sites identified in time period i and the subsequent time period *i* + 1 from various ranking criteria. It can be found that a better method means more intersections of top ranked sites. The method yielding the largest intersection of sites is said to be the most consistent. This criterion is similar to criterion 3 and can be seen as another one of its forms.

- Criterion 5: Total Rank Differences

This test is built on the method consistency test, and takes into account the rankings of safety performance of road sections in the two periods. The sum of total rank differences between the ranks of the hazardous road sections identified in the first period and ranks identified in the second period for the same group sites is used to reflect the performance of consistent rankings of sites across periods. This criterion is used to reflect the performance of consistent rankings of sites across periods.

$$C_5 = \sum_{k=n-n\alpha}^{n} (\mathrm{Rank}(k_j^{i+1}) - \mathrm{Rank}(k_j^{i})) \qquad (1-12)$$

where

n = total number of sites,

$\mathrm{Rank}(k_j^{i})$= the rank order for site k by method j during time period i, and

$\mathrm{Rank}(k_j^{i+1})$ = the ranked order for site k (identified in time period i) by method j during subsequent time period i+1.

It can be seen that a smaller total rank difference means more consistency in the ranking method. However, this criterion has a problem in that it cannot differentiate the volatile changes of identified sites in the second period. For example, if sites 10 and 15 have rankings of 1^{st} and 8^{th} in period 1, but 6^{th} and 3^{rd} in period 2, ranked by method j, from Equation 1-12, $C_5 = 0$, so that a consistent conclusion can be falsely drawn with this criterion. To avoid the situation, thus the absolute values of difference should be used in calculating the sum of rank differences of the two periods.

- Criterion 6: Poisson Mean Differences

One major problem with these rank related evaluation criteria, as pointed out by Cheng and Washington (2008), is that each false identification is weighted equally. For example, if a site with a total Poisson mean (TPM) of 16.8 is wrongly selected for treatment instead of one with

16.9, the error is really rather small, whereas if a site with a TPM of 6.9 is mistakenly selected instead of one with 16.9, the error is much more significant. The Poisson mean differences associated with the two false identifications are 0.1 and 10, respectively which are relatively large differences, whereas sensitivity and specificity differences are the same with these false identifications. Poisson mean differences are proposed by Chen and Washington (2005b) to obviate this drawback. This criterion can be expressed as:

$$C_6 = \sum_{k=n-n\alpha}^{n} \left| PM_{j,k}^{i+1} - PM_{j,k}^{i} \right| \qquad (1-13)$$

They suggested that a smaller value of this criterion is desirable. From ranks 3 through to 6, an underlying assumption is that there is homogeneity across the two periods. However, this may not be the case in real applications. Thus, we believe that the sum of the Poisson means in the succeeding time period might be a better criterion.

D. Applications of the EB method and limitations

The EB method was extensively examined by researchers (Persaud, 1999; Saccomanno et al., 2001; Elvik, 2008a, 2008c; Cheng and Washington, 2005a, 2008). It was confirmed that the EB method can provide promising results. Hauer (1997) and Hauer et al. (2002) presented an excellent illustration on how to clearly implement the EB method with a step by step procedure. Now the EB method is widely applied for road safety studies. The Highway Safety Manual (to be published in 2010 by AASHTO) employs the EB method as a standard method for road safety analyses. However, there are still some limitations of the EB method:

- It requires a large sample size of data to develop SPFs. This can be costly or otherwise impractical.
- There is no flexibility to define underlying distributions for the observed crashes in that only an NB distribution for the observed crashes can be assumed, however, there may be other distributions which are more suitable for the data set, but which cannot be implemented with the EB method
- Only point estimations of expected crashes are available. This implies that the EB method does not consider the uncertainty of the obtained data

- It is difficult to select the function form of SPFs. The chi-square test is commonly used to evaluate the fit of the model. However, it does not take into consideration the penalty of an overparameterized model
- It cannot handle multivariate correlation distributions. This is even worse when crash data with different levels of severity are available for network ranking. Intuitively, there should be some correlation between crashes of different severity levels, and disregarding this correlation may lead to biased results
- It is difficult to incorporate spatial correlations and/or time series correlations, and
- The ranking procedure can be time consuming and costly. To implement the EB method, the first step is to calibrate SPFs. When different severity crash data are available, the whole EB procedure needs to apply for each severity of crash individually. In other words, the whole procedure needs to be performed again and again. This is time consuming and costly.

With these issues, the application of the EB method can be problematic and a new method, the fully or Hierarchical Bayes (FB or HB, used interchangeably) method was proposed to overcome these limitations. The current FB ranking method and associated ranking criteria will be investigated in detail in Chapter 5.

1.3 TREATMENT EFFECT ANALYSIS

Following black spot identification, the diagnosis of safety issues and the development of potential remedies, there is implementation of countermeasures to improve the safety performance of some identified black spots. After that, it is prudent to conduct a treatment effect analysis to determine if implemented countermeasures improve road safety, and to provide feedback information for a road safety management system. The treatment effect can be quantified in terms of the number or percentage of crashes reduced.

1.3.1 Measurement of Treatment Effect

- Crash Reduction

$$CR = \lambda_A - E(Y_A) \qquad (1-14)$$

where

CR = crash reduction in terms of number of crashes reduced in the period after implementation,

λ_A = the expected crashes for the whole treatment group without treatment in after period, and

$E(Y_A)$ = the expected crashes for the whole treatment group with treatment in after period.

- Crash Reduction Rate

$$CRR = \frac{\lambda_A - E(Y_A)}{\lambda_A} \qquad (1-15)$$

where

CRR = crash reduction rate in terms of percentage of crashes reduced in the after period.

In order to obtain CR and CRR, there are two tasks: estimate the expected crashes for the whole treatment group without treatment in after period λ_A; and estimate the expected crashes for the whole treatment group with treatment in after period $E(Y_A)$. Normally, $E(Y_A)$ is estimated to be the observed crashes in the after period (Hauer, 1997). Thus the critical issue for the treatment effect analysis becomes how to estimate λ_A. Currently, there are three methods to estimate λ_A: naïve before-after treatment study, comparison group before-after study, and the EB before-after safety study (Hauer, 1997). These are described below.

1.3.2 Method for Treatment Effect Analysis

Before-after studies, also called longitudinal studies, are commonly used methods to evaluate the safety effects of a single treatment or a combination of treatments in highway safety (Hauer, 1997). This type of study is deemed superior to cross-sectional studies because they can exclude time-invariant unobserved individual differences, and can account for temporal order of events. The fundamental difference between cross-sectional and before-after studies is that cross-sectional studies take place at a single point in time while before-after studies involve a series of measurements taken over a period of time. Both are observational studies.

Before-after studies can be grouped into three types: the simple (naïve) before-after study, the before-after study with comparison group (also called before after C-G method), and the before-after study using the EB method (using a "reference" group similar in concept to a comparison). The selection of the method is usually governed by the availability of data, such as crashes, traffic flow, etc., and can also be influenced by the amount of available data (or sample size). Here, the term "after" means the safety status after the implementation of a treatment; correspondingly, the term "before" refers to the status before the implementation of a treatment.

A. Naïve before-after study (simple before-after study)

This approach assumes that all of the observed changes in crashes are due to treatment. It does not account for the temporal effects on safety due to variations in weather, demography, crash reporting and so on. Thus, the expected number of accidents in the after periods with or without treatment has the form:

$$E(Y_A) = \sum_{i=1}^{n} Y_{Ai} \qquad (1-16)$$

$$\lambda_A = \sum_{i=1}^{n} Y_{Bi} \frac{t_{Ai}}{t_{Bi}} \qquad (1-17)$$

Crash reduction and crash reduction rate are calculated as:

$$CR = \lambda_A - E(Y_A) = \sum_{i=1}^{n} Y_{Bi} \frac{t_{Ai}}{t_{Bi}} - \sum_{i=1}^{n} Y_{Ai} \qquad (1-18)$$

$$CRR = \frac{\lambda_A - E(Y_A)}{\lambda_A} = \frac{\sum_{i=1}^{n} Y_{Bi} \frac{t_{Ai}}{t_{Bi}} - \sum_{i=1}^{n} Y_{Ai}}{\sum_{i=1}^{n} Y_{Bi} \frac{t_{Ai}}{t_{Bi}}} \qquad (1-19)$$

where

Y_{Bi}= observed crash counts at site i in a before period of t_{Bi} years, and

Y_{Ai} = the observed crash counts at site i in an after period of t_{Ai} years.

Naïve before-after methods, like naïve methods for black spot identification, are still appealing in that they are easy to apply. Although widely used, they are inevitably likely to have errors in that they fail to address the RTM problem. That is, a randomly large number of crashes for a site during a before period is normally followed by a reduced number of crashes during a similar after period, even if no countermeasures have been implemented (while the opposite applies in the case of a randomly small number of crashes). In the latter part of the thesis, an example will be given. In that example, naïve results show a large crash change in the before- after period at high crash sites even if there is no treatment implemented.

B. Comparison group before-after method

This method employs a comparison group to estimate the expected number of accidents in the after period for treatment sites (λ_A) had treatment not been implemented.

Assume that:

C_{Bi}= crashes observed at comparison site(s) which correspond to treatment site i in the before period of t_{Bi} years, and

C_{Ai} = crashes observed at comparison site(s) which correspond to treatment site i in the after period of t_{Ai} years.

Then,
$$\lambda_A = \sum_{i=1}^{n} Y_{Bi} \frac{C_{Ai}}{C_{Bi}} \qquad (1-20)$$

Compared with the naïve method, this method can account for unrelated effects (Persaud and Lyon, 2007), or non-scheme effects (Hirst et al., 2004), or confounding factors (Elvik, 2002), such as time and traffic trends, but will not account for RTM unless the comparison group is similar to the treatment group in all of the possible factors that could influence safety. Persaud and Lyon (2007) reported that there are immense practical difficulties in achieving this ideal. Moreover, it is difficult to test the necessary assumption where the comparison group is unaffected by the treatment. In addition, this method will not control for changes in safety which results from changes in traffic volume at the treatment sites that might result from the treatment itself.

It should be noted that normally, the sample size of the comparison group is relatively small; again, this method cannot be used to conduct a treatment analysis by itself. The detailed information of this method can be obtained from Hauer (1997), Persaud and Lyon (2007), and Hirst et al. (2004).

C. EB before-after method

a. Procedure

In the EB approach (Hauer, 1992; Hauer, 1997; Hauer and Harwood, 2002; Persaud and Nguyen, 1998; Harkey et al., 2008), the change in safety for a given crash type at a location is given by:

$$\lambda - \pi \qquad (1\text{-}21)$$

where λ is the expected number of crashes that would have occurred in the after period without treatment and π is expected number of crashes that occurred in the after period with treatment, which is normally estimated to be the number of reported crashes in the after period.

In estimating λ, the effects of the regression to the mean and changes in traffic volume are explicitly accounted for using SPFs which relate crashes to traffic flow and other relevant factors. Annual SPF multipliers are calibrated to account for the temporal effects on safety due to variations in weather, demography, crash reporting and so on.

In the EB procedure for the treatment effect analysis, the SPFs is used to first estimate the number of crashes that would be expected in each year of the before period at locations with traffic volumes and other characteristics similar to the one being analyzed. The procedure to obtain an estimate of λ is the same as Equation 1-2.

A factor is then applied to π to account for the length of the after period and differences in traffic volumes between the before and after periods. This factor is the sum of the annual SPF predictions for the after period divided by P, the sum of these predictions for the before period. The result, after applying this factor, is an estimate of λ. The procedure also produces an estimate of the variance of λ.

The estimate of λ is then summed over all sites in a treatment group of interest (to obtain λ_{sum}) and compared with the count of crashes during the after period in that group (π_{sum}). The variance of λ is also summed over all sections in the treatment group.

The index of effectiveness (θ) is estimated as:

$$\theta = (\pi_{sum}/\lambda_{sum})/\{1 + [Var(\lambda_{sum})/\lambda_{sum}{}^2]\} \quad (1-22)$$

The standard deviation of θ is given by:

$$Stddev(\theta) = \left(\frac{\theta^2[Var(\pi_{sum})/\pi_{sum}{}^2] + [Var(\lambda_{sum})/\lambda_{sum}{}^2]}{[1 + \frac{Var(\lambda_{sum})}{\lambda_{sum}{}^2}]^2}\right)^{0.5} \quad (1-23)$$

The percent change in crashes is in fact 100(1−θ); thus a value of θ = 0.80 with a standard deviation of 0.10 indicates a 20 percent reduction in crashes with a standard deviation of 10%.

b. Applications and limitations

Like the EB application for network ranking, the EB method for the treatment effect analysis was extensively evaluated and found to provide promising results (Persaud and Lyon 2007; Hauer 1992; Hauer 1997; Hauer and Harwood 2002; Elvik 2008c; Brüde and Larson 1988, etc.). It is now widely used for the treatment effect analysis (Persaud 1988; Persaud and Hauer 1997; Persaud and Nguyen 1998; Persaud et al. 2001; Persaud et al. 2003; Persaud 2005; Harkey et al. 2008) because it can address RTM problems with step by step procedures for implementation. However, there are still some limitations of this method similar to those identified for network ranking. The new FB method for the treatment effect analysis is explored, evaluated and discussed in Chapter 6.

1.4 MOTIVATION AND RESEARCH OBJECTIVES

Due to the limitations of the currently popular method of EB for road safety analyses, it is necessary to explore the development of a new method which can overcome its drawbacks. The FB method has been recently introduced to road safety analyses (Miaou and Lord 2003; Bossche et al. 2003; Brijs et al. 2004; Carriquiry and Pawlovich 2005; Miaou and Song 2005; Pawlovich et al. 2005; Lord 2006; Aul 2006; Miranda-Moreno and Fu 2007; Lan et al. 2009; Persaud et al. 2010; Park and Lord 2007; Ma and Kockelman 2006; Ma and Kockelman 2008; El-Basyouny and Sayed 2009; Aguero-Valverde and Jovanis 2009; Lan and Persaud 2010). However, few have applied the FB in either network ranking or a treatment effect analysis mainly due to the lack of systemic evaluations. Moreover, in the above studies, normally only one function form of expected crashes was investigated. Furthermore, no evaluation of the FB method was conducted for both network ranking and treatment effect analysis. All of these issues will be discussed, explored and investigated in detail in the latter part of this thesis.

To this end, the objectives of this dissertation are to:

- Explore various FB models with correlated data. Various FB models will be proposed and discussed with correlated data that might occur in road safety studies, including data with time series, spatial, temporal spatial, multivariate (with or without temporal correlation and/or with or without spatial correlations).

- Investigate the model selection criteria and to find possible best model criteria. Different model selection criteria, e.g., log likelihood (LL), Akaike information criterion (denoted as AIC, see Akaike, 1973 and Bozdogan, 2000), Bayes information criterion or the Schwarz criterion (denoted as BIC, see Schwarz, 1978) and deviance information criterion (defined as DIC, see Spiegelhalter et al., 2002), will be investigated.

- Develop a proper approach to conduct a thorough evaluation of FB for black spots identification. Two categories of data, single severity data and multilevel severity data, will be used to explore the performance of the FB methods using various ranking and evaluation criteria.

- Investigate ranking and evaluation criteria to identify the possible criteria. Specifically, the posterior mode rank of the decision parameter (the Poisson mean in this study) is proposed as a ranking criterion and evaluated for hotspot identification. The other seven ranking criteria including posterior Poisson mean, posterior expected rank, posterior median rank, posterior probability of being the worst, raw data, posterior PSI and posterior PPSI will be also evaluated and compared, and the most robust criteria will be identified. Moreover, in addition to sensitivity and specificity, the sum of crash counts and sum of PPSI in the evaluation period, the sum of the Poisson mean and sum of the PSI in the second time period are proposed and employed as two new criteria to evaluate the performance of the ranking methods and various ranking criteria.

- Design a method to test FB for treatment effect analysis. Two FB testing frameworks will be employed. First the univariate before-after FB method will be examined using three simulated datasets. Then multivariate and univariate FB methods will be evaluated and compared using two groups of untreated California unsignalized intersections (one with high crash counts and another one with low crash counts). The test will be performed for hypothetical treatment sites that have significant naive treatment effects due to regression to the mean. The FB method will be validated if it accounts for regression to the mean and estimates a treatment effect of zero at these hypothetical treatment sites.

- Explore and compare the performance of the EB and FB approaches for network ranking and treatment effect analysis and to identify the advantages of FB method over the EB method. Both FB and EB methods will be evaluated and compared based on their performance in terms of various evaluation criteria for black spots identification using single severity crash data. For treatment effect analysis, EB and FB methods will be compared with the application to evaluate two treatments: the conversion of rural intersections from unsignalized to signalized control; and the conversion of road segments from a four-lane to a three-lane cross-section with two-way left turn lanes (also known as road diets).

1.5 ORGANIZATION OF DISSERTATION

This dissertation is composed of eight chapters.

Chapter 2 presents the basics of the FB method. Two approaches, univariate FB (Poisson-gamma (PG) and Poisson-log normal (PLN) and multivariate FB (multivariate Poisson-log normal, denoted as MVPLN) are discussed in terms of marginal and posterior distributions. The procedure to obtain FB posterior distributions is described. Then, the basics of EB and FB are compared. Finally, the advantages of FB are discussed in terms of the ability to handle correlation longitudinally (time series correlation) and spatially (spatial correlation), and correlations between different severities / different types of crashes, flexibility for model selection either in terms of function forms or underlined distributions of dependent variables (PG, PLN, mixture distributions), and capability to provide rich inference information.

Chapter 3 discusses FB models with correlated data. Various univariate FB methods to deal with each correlated data case (temporal, spatial, temporal combined with spatial correlations) are discussed in detail. Multivariate FB models with different severity or types of crash counts with a combination of spatial and/or temporal correlated data are also studied.

Chapter 4 indicates that the objective in model selection is to use as parsimonious a model as possible while ensuring that reliable results are obtained. The competing models should be compared based on a trade-off between the fit of the data to the model and the corresponding complexity of the model. Popular model selection criteria, including maximum likelihood, Bayes factors (Burnham and Anderson, 2004), AIC, BIC and DIC, are discussed in detail. The advantages and disadvantages of these methods are highlighted and issues with model selection are summarized.

Chapter 5 presents the evaluation of FB for network ranking using 5 severity levels of data (fatal, incapacitating-injury, non-incapacitating injury, minor injury and property damage only (PDO) crashes) and 1 severity level of crash (total crashes) data for California four legged unsignalized intersections with 2 lanes on major roads. The evaluations are performed by using data from one level of severity (total injured) to compare the EB results with univariate FB results, and 5 severity levels of data to compare the results obtained from univariate FB models with multivariate Poisson-log normal (MVPLN) models. Eight ranking criteria (posterior expected rank, mode rank, median rank, the probability of being the worst, Poisson mean, PSI, PPSI and raw data) and 5 evaluation criteria (sum of Poisson mean, sensitivity and specificity, sum of the PSI, sum of the PPSI and sum of crash counts in evaluation period) are explored in this study. The evaluation methods include the EB, univariate PG and univariate PLN with four function forms of the expected crashes, and univariate Poisson autoregressive first order (P AR (1) model) models for data from a single level of severity and MVPLN AR (1) and P AR (1) for multilevel severity crash data. In addition, a sensitivity analysis of ranking criteria and a data history sensitivity analysis are conducted with data for different ranking periods (first 3 years vs. first 6 years). The results are compared with each other and the best method and most robust ranking criteria are identified and discussed for each of the cases (multilevel of severity cases and one type of crash case).

Chapter 6 first presents the evaluation of the univariate FB using simulated data for a before-after study, then performs the evaluation for both univariate FB and MVPLN using different types of crash data from California unsignalized intersections. For the latter case, two data sets, one with relatively high crash counts and the other with lower crash counts, are selected for evaluation, respectively. The objective of the evaluation for different types of crashes is to determine if MVPLN is superior to univariate FB for each of these cases.

Chapter 7 provides a comparison and discussion of the pros and cons of the two Bayesian approaches, EB and FB methods, based on, and illustrated with, empirical applications. These applications pertain to the evaluation of two treatments: the conversion of rural intersections from unsignalized to signalized control, and the conversion of road segments from a four-lane to a three-lane cross-section with two-way left turn lanes (road diets). In each case, the numerical results from the two approaches are compared.

Lastly, Chapter 8 concludes this dissertation with a short summary of this research and some suggestions for future studies.

CHAPTER 2 FULLY BAYESIAN METHOD FOR ROAD SAFETY ANALYSIS

2.1 INTRODUCTION OF THE BAYES METHOD

In Bayesian models, the likelihood of the observed data y given parameters θ, denoted by $p(y|\theta)$ or $L(\theta|y)$, is used to modify the prior beliefs $\pi(\theta)$ with updated knowledge to obtain a posterior density of θ, $\pi(\theta|y)$. Thus, according to Bayes' theorem,

$$p(y,\theta) = p(y|\theta)\pi(\theta) = L(\theta|y)\pi(\theta) = \pi(\theta|y)m(y) \qquad (2-1)$$

where

y = observed data,

θ = parameters,

$\mathrm{p}(\mathrm{y},\theta) =$ joint probability of observed data y and parameters θ,

$\mathrm{p}(\mathrm{y}|\theta) =$ conditional distribution or likelihood function of θ given fixed data y,

$\pi(\theta) =$ prior distribution of parameters, can be informative or non − informative ,

$\mathrm{m}(\mathrm{y}) =$ marginal distribution of the data y , $\mathrm{m}(\mathrm{y}) = \int \mathrm{p}(\mathrm{y}|\theta)\pi(\theta)\mathrm{d}\theta$, and

$\pi(\theta|\mathrm{y}) =$ posterior distribution of parameters θ.

Therefore, the posterior density can be written as:

$$\pi(\theta|y) = \frac{p(y|\theta)\pi(\theta)}{m(y)} \qquad (2-2)$$

Equation 2-2 gives a general solution for updating prior probabilities into posterior probabilities. However, the actual calculation can be laborious. It can be seen that $\mathrm{m}(\mathrm{y})$ is constant with respect to θ, and has the role of a normalizing constant. Generally, there are two approaches to obtaining the posterior distribution:

- by finding $m(y) = \int p(y|\theta)\pi(\theta)d\theta$. This approach, however, is normally impossible in road safety studies unless it has a conjugate prior for the likelihood,

- by the Markov chain Monte Carlo (MCMC) method. From Equation 2-2, ignoring the normalizing constant, Bayes' formula is often written in a proportional form:

$$\pi(\theta|y) \propto p(y|\theta)\pi(\theta). \qquad (2-3)$$

Comparing Equations 2-1 and 2-3, we can find that posterior distribution is proportional to the joint distribution. For this reason, we can derive all the estimates and even draw random samples from the posterior density by MCMC simulation techniques (Brooks, 1998; Gelfand, 1990) without knowing the constant $m(y)$. In fact, MCMC is now frequently used to estimate the posterior distribution $\pi(\theta|y)$, and $m(y)$ can be ignored in many calculations.

2.2 WINBUGS AND ITS SAMPLING METHODS

The software WinBUGS (Lunn et al., 2000; Spiegelhalter et al., 2003; Cowles, 2004) is commonly used to implement the FB method. This section summarizes the basics of how this is achieved.

2.2.1 The Gibbs Sampler

An MCMC algorithm, which is known as Gibbs sampling (Congdon, 2003; Cowles, 2004), is used to construct the transition kernels for its Markov chain samplers. It uses a fixed sequence of Gibbs transition kernels each of which updates a different component of the state vector, as follows. Given starting values (initial values), the Gibbs sampler proceeds by systematically updating each variable in turn, via a single Gibbs update, as follows:

Specify an initial value: $\theta_1^{(0)}, \theta_2^{(0)}, \ldots, , \theta_p^{(0)}$

Repeat for $j = 1, 2, \ldots, m$

sample $\theta_1^{(j)} \sim p\left(\theta_1 \middle| \theta_2^{(j-1)}, \ldots, , \theta_p^{(j-1)}, y\right)$

.
.
.

sample $\theta_k^{(j)} \sim p\left(\theta_k \middle| \theta_1^{(j)}, \theta_2^{(j)}, \ldots, , \theta_{k-1}^{(j)}, \theta_{k+1}^{(j-1)}, \theta_p^{(j-1)}, y\right)$

.
.
.

sample $\theta_p^{(j)} \sim p\left(\theta_p \middle| \theta_1^{(j)}, \theta_2^{(j)}, \dots, \theta_{p-1}^{(j-1)}, y\right)$

Return the values $\left\{\theta_1^{(j)}, \theta_2^{(j)}, \dots, , \theta_p^{(j)}, \quad j = 1, 2, \dots, m \right\}$

It should be noted the first L iterations must be disregarded to ensure convergence of the chains due to the effects of the initial values. The first disregarded iteration is also called burn-in iteration (Lunn et al., 2000; Spiegelhalter et al., 2003). After convergence, when the sample size m is large enough (m-L), then the mean of the parameters is:

$$\theta_k = \frac{1}{m-L} \sum_{j=L+1}^{m} \theta_k^{(j)} \qquad k = 1, 2, \dots, p \qquad (2-4)$$

From the above, it can be seen that, conceptually, the Gibbs transition is fairly straightforward. Ideally, the conditional distribution $\mathrm{p}\left(\theta_k \middle| \theta_1^{(j)}, \theta_2^{(j)}, \dots, \theta_{k-1}^{(j)}, \theta_{k+1}^{(j-1)}, \dots, \theta_p^{(j-1)}, y\right)$ will be in the form of a standard distribution which allows efficient random variate generation, and a suitable prior specification often ensures that this is the case.

2.2.2 The Metropolis-Hastings Algorithm

The Gibbs sampler owes some of its success and popularity to the fact that in many statistical models, the complete conditional posterior distributions $\mathrm{p}\left(\theta_k \middle| \theta_1^{(j)}, \dots, \theta_{k-1}^{(j)}, \theta_{k+1}^{(j-1)}, \dots, \theta_p^{(j-1)}, y\right)$ take the form of some well-known distributions, such as having conjugate priors, i.e. Poisson-gamma models, which allow efficient random variate generation. However, there remain many important applications where this is not the case, which require alternative MCMC schemes. In such cases, the Metropolis-Hastings algorithm (Chib and Greenberg, 1995; Brooks, 1998) is used to draw posterior samples for parameters estimation in WinBUGs. This powerful algorithm provides a general approach for producing a correlated sequence of draws from the target density that may be difficult to sample by a classical independence method. The Metropolis-Hastings algorithm generates a sequence of draws as depicted below:

Step 1: Start with any initial value θ_0 satisfying f (θ_0)>0.

Step 2: Using current θ_i value, sample a candidate point θ^* from some distribution $q(\theta^*|\theta_i)$, which is the probability of returning a value of θ^* given a previous value of θ_i. This distribution is also referred to as the proposal or candidate-generating distribution.

Step 3: Calculate

$$\alpha_{\theta^*|\theta_i} = min\left\{1, \frac{\pi(\theta^*)q(\theta_i|\theta^*)}{\pi(\theta_i)q(\theta^*|\theta_i)}\right\}$$

where

$q(\theta^*|\theta_i)$ = the proposal distribution, which is the probability of returning a value of θ^*given a previous value of θ_i.,

$q(\theta_i|\theta^*)$ = the proposal distribution, the probability of returning a value of θ_i given a previous value of θ^*, and

π = stationary limiting distribution which is the same as the distribution that we wish to simulate.

It should be noted that the choice of the proposal distribution $q(\theta^*|\theta_i)$ and $q(\theta_i|\theta^*)$ is essentially arbitrary, subject to the condition that the resulting chain is aperiodic and irreducible, has a stationary distribution π, and in practice, generally selected so that observations may be generated with reasonable ease.

Step 4: generate a random variable u from uniform distribution U[0,1]

Step 5: if $u < \alpha_{\theta^*|\theta_i}$, accept the proposal value θ_j for θ_{i+1} , $\theta_{i+1} = \theta_j$

Otherwise, reject the proposal value θ^*, $\theta_{i+1} = \theta_i$

Step 6: go to step 2, repeat until a large enough sample has been generated. Finally, return for θ_0 , $\theta_1, \dots, \theta_m$.

The one-dimensional Gibbs sampler is a special case of Metropolis-Hastings, where the proposal distribution $q(\theta^*|\theta_i) = \pi(\theta^*)$. For them, $\alpha_{\theta^*|\theta_i} = 1$. It means that the proposal θ^* is always accepted.

2.3 HIERARCHICAL BAYES METHOD

The hierarchical Bayes (HB) method is widely used in Bayesian analysis. It is a powerful tool for expressing rich statistical models that more fully reflect a given problem than a simpler model could otherwise.

Given data y and parameters θ, a simple Bayesian analysis starts with a prior probability (prior) $\pi(\theta)$ and likelihood $\pi(y|\theta)$ *or* $L(\theta|y)$ to compute a posterior probability $\pi(\theta|y) \propto p(y|\theta)\pi(\theta)$. Often, the prior on θ depends, in turn, on other parameters φ (called hyper parameter) that is not mentioned in the likelihood. So, the prior $\pi(\theta)$ must be replaced by a prior $\pi(\theta|\varphi)$, and a prior $\pi(\varphi)$ on the newly introduced parameters φ is required, which results in a posterior probability $\pi(\theta|y)$ shown below.

$$\pi(\theta,\varphi,y) = \pi(\theta,y|\varphi)\pi(\varphi) = p(y|\theta,\varphi)\pi(\theta,\varphi) \qquad (2-5)$$

Since

$$\pi(\theta,y|\varphi)\pi(\varphi) = \pi(\theta|y,\varphi)m(y|\varphi)\pi(\varphi) \qquad (2-6)$$

$$p(y|\theta,\varphi)\pi(\theta,\varphi) = p(y|\theta,\varphi)\pi(\theta|\varphi)\pi(\varphi) \qquad (2-7)$$

Plugging Equations 2-6 and 2-7 into Equation 2-5,

$$\pi(\theta|y,\varphi)m(y|\varphi)\,\pi(\varphi) = p(y|\theta,\varphi)\pi(\theta|\varphi)\pi(\varphi)$$

The posterior distribution is:

$$\pi(\theta|y,\varphi) = \frac{p(y|\theta,\varphi)\pi(\theta|\varphi)}{m(y|\varphi)} = \frac{p(y|\theta)\pi(\theta|\varphi)}{m(y|\varphi)} \qquad (2-8)$$

So,

$$m(y|\varphi) = \int p(y|\theta,\varphi)p(\theta|\varphi)d\theta \qquad (2-9)$$

$$\pi(\theta|y,\varphi) \propto p(y|\theta)\pi(\theta|\varphi) \qquad (2-10)$$

This is the simplest example of an HB model. The process may be repeated; for example, the parameters φ may depend in turn, on additional parameters ψ, which will require its own prior. Eventually, the process must terminate with priors that do not depend on any other unmentioned parameters. It is quite common that almost all the Bayes methods involve HB study, and this is also quite popular in road safety studies, i.e. crashes Y~ Poisson (θ), and θ ~ gamma (α, β). In this study, HB and FB will be used interchangeably. Five HB models, which are PG, PLN, P AR(1), MVPLN and MVPLN AR (1), are investigated which will be described in detail in the latter part of this thesis.

2.4 FULL BAYESIAN (FB) MODELS

In road safety studies, sometimes only one dependent variable or several uncorrelated dependent variables is/are introduced into the analysis procedure, e.g., a treatment effect analysis based on total crashes. In this case, a univariate FB model is appropriate for providing promising results (Lan et al. 2009, Persaud et al. 2010). Actually, it is confirmed in this study that a univariate FB method works well even with multiple uncorrelated dependent variables (Lan and Persaud 2010), for example, for different types of crashes. While in other situations, this may require a multivariate correlated random effects model for promising results, e.g., network ranking based on multilevel severity crashes, intuitively, these different severity crashes are inherently correlated. In this latter case, a univariate FB model might no longer be proper for the study since it fails to capture the inherent relationship between these correlated dependent variables. Under this situation, a multivariate FB must be performed. The univariate and multivariate FB models involved in this study are introduced in the next sub-sections.

2.4.1 Univariate FB Models

Crash counts $Y_{i,t}$ at site i in year t are typically assumed Poisson distributed with a mean $\lambda_{i,t}$,

$$Y_{i,t} \sim \text{Pois}\left(\lambda_{i,t}\right) \qquad (2-11)$$

where

$$\lambda_{i,t} = \; exp\;(\boldsymbol{\beta}.\boldsymbol{X} + \varepsilon_i) \qquad (2-12)$$

$$\mu_{i,t} = exp\ (\boldsymbol{\beta}.\boldsymbol{X}) \qquad (2-13)$$

and

$Y_{i,t}$ = observed number of crashes at site i in year t,

$\lambda_{i,t}$= expected number of crashes at site i in year t,

$\mu_{i,t}$ = expected number of crashes at similar sites in year t,

$\boldsymbol{X} = (1, X_1, X_2 \dots, X_M)'$, $\boldsymbol{X}$ could be the logarithm of the covariates (logarithm of traffic volumes) or just covariates and M is the number of covariates,

$\boldsymbol{\beta} =$ vector of coefficients, $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_M)$, and

e^{ε_i} = random effect at site i.

Crash counts at a given site are inherently discrete, positive numbers, and often are small, as in the case of fatal and injury accidents. In an EB study, the NB distribution is regarded as an effective model and is the only available distribution applied for overdispersed count data. FB, however, has more flexibility to choose the distributions of the crash counts. Generally, there are two popular HB models (PG and PLN models) with respect to different distributions of random effects e^{ε_i} in road safety analyses.

$Y_{i,t} \sim$ Pois $(\lambda_{i,t})$ is rewritten as $Y_{i,t} \sim$ Pois $(e^{\varepsilon_i}\mu_{i,t})$. $y_{i,t}$ is said to follow PG or PLN distribution with regard to the distributions that e^{ε_i} follows:

A. Poisson-gamma models ($Y_{i,t} \sim$ Pois $(e^{\varepsilon_i}\mu_{i,t})$ and $e^{\varepsilon_i} \sim Gamma\ (\alpha, \beta)$)

The FB model is called a PG model if $Y_{i,t} \sim$ Pois $(e^{\varepsilon_i}\mu_{i,t})$ and $e^{\varepsilon_i} \sim Gamma\ (\alpha, \beta)$. It can be seen that this model introduces a gamma distributed multiplicative random effect. The posterior distribution of e^{ε_i} and marginal distribution of crash counts $Y_{i,t}$ can be expressed or derived as follows:

a. Posterior distribution of $\lambda_{i,t}$

$$\lambda_{i,t} = e^{\varepsilon_i}\mu_{i,t} \qquad (2-14)$$

Since $e^{\varepsilon_i} \sim$ Gamma (α, β)

Then $e^{\varepsilon_i}\mu_{i,t} \sim$ Gamma $(\alpha, \mu_{i,t}\beta)$

From Equation 2-10,

$$\pi(\lambda_{i,t}|y_{i,t}, \alpha, \mu_{i,t}\beta) \propto p(y_{i,t}|\lambda_{i,t})\pi(\lambda_{i,t}|\alpha, \mu_{i,t}\beta) \qquad (2-15)$$

$Y \sim$ Pois (λ), the Poisson likelihood is:

$$p(y_{i,t}|\lambda_{i,t}) = \frac{{\lambda_{i,t}}^{y_{i,t}} e^{-\lambda_{i,t}}}{y_{i,t}!} \qquad (2-16)$$

Prior gamma distribution $\lambda_{i,t} \sim Gamma\ (\alpha, \mu_{i,t}\beta)$,

$$\pi(\lambda_{i,t}|\alpha, \mu_{i,t}\beta) = \frac{{\lambda_{i,t}}^{\alpha-1} e^{-\lambda_{i,t}/\mu_{i,t}\beta}}{(\mu_{i,t}\beta)^{\alpha}\Gamma(\alpha)} \qquad (2-17)$$

Plugging Equations 2-16 and 2-17 into Equation 2-15, and the posterior distribution is:

$$\pi(\lambda_{i,t}|y, \alpha, \mu_{i,t}\beta) \propto \frac{{\lambda_{i,t}}^{y_{i,t}+\alpha-1} e^{-\lambda_{i,t}(1+\frac{1}{\mu_{i,t}\beta})}}{(\mu_{i,t}\beta)^{\alpha}\Gamma(\alpha)}$$

$$\propto \ {\lambda_{i,t}}^{y_{i,t}+\alpha-1} e^{-\lambda_{i,t}(1+\frac{1}{\mu_{i,t}\beta})} \qquad (2-18)$$

It can be found that the posterior distribution of the PG FB model has the same form of prior distribution (Equation 2-17). The posterior distribution is also gamma distributed, such that $\pi(\lambda_{i,t}|y, \alpha, \mu_{i,t}\beta) \sim Gamma\left(y_{i,t} + \alpha, 1 + \frac{1}{\mu_{i,t}\beta}\right)$. The prior gamma distribution and the posterior are then called conjugate distributions, and the prior gamma is called a conjugate prior for the likelihood Poisson distribution (Gelman et al., 2003). A conjugate prior is an algebraic convenience; otherwise, a difficult numerical integration may be necessary. Furthermore, conjugate priors may provide intuition, by more transparently showing how a likelihood function updates a distribution.

b. Marginal distribution $m(y|\alpha, \beta)$:

For simplicity of notation, we omit the subscript. Apply Bayes' theorem, the joint distribution of (y, λ) has the probability density function (abbreviated as pdf):

$$P(y, \lambda|\alpha, \mu\beta) = P(y|\lambda)\pi(\lambda|\alpha, \mu\beta) \qquad (2-19)$$

The marginal distribution is:

$$m(y|\alpha, \mu\beta) = \int P(y, \lambda|\alpha, \mu\beta)\, d\lambda = \int P(y|\lambda)\pi(\lambda|\alpha, \mu\beta) d\lambda \qquad (2-20)$$

The marginal distribution of crash counts y is:

$$m(y|\alpha, \mu\beta) = \int \frac{\lambda^y e^{-\lambda}}{y!} \times \frac{\lambda^{\alpha-1} e^{-\frac{\lambda}{\mu\beta}}}{(\mu\beta)^\alpha \Gamma(\alpha)} d\lambda$$

$$m(y|\alpha, \beta) = \frac{1}{y!\,(\mu\beta)^\alpha \Gamma(\alpha)} \int \lambda^{y+\alpha-1} e^{-(1+\frac{1}{\mu\beta})\lambda}\, d\lambda \qquad (2-21)$$

Assume that $-\left(1 + \frac{1}{\mu\beta}\right)\lambda = -t$

$$d\lambda = \frac{dt}{1 + \frac{1}{\mu\beta}}$$

$$m(y|\alpha, \beta) = \frac{1}{y!\,(\mu\beta)^\alpha \Gamma(\alpha)} \int \left(\frac{t}{1 + 1/(\mu\beta)}\right)^{y+\alpha-1} \times \frac{e^{-t}}{1 + 1/(\mu\beta)} dt$$

$$= \frac{1}{y!\,(\mu\beta)^\alpha \Gamma(\alpha)} \left(\frac{1}{1 + \frac{1}{\mu\beta}}\right)^{y+\alpha} \int t^{y+\alpha-1} e^{-t} dt$$

$$= \frac{1}{y!\,(\mu\beta)^\alpha \Gamma(\alpha)} \left(\frac{1}{1 + \frac{1}{\mu\beta}}\right)^{y+\alpha} \Gamma(y+\alpha)$$

$$m(y|\alpha, \beta) = \frac{\Gamma(y+\alpha)}{y!\,\Gamma(\alpha)} \left(\frac{1}{\mu\beta + 1}\right)^{\alpha} \left(\frac{\mu\beta}{\mu\beta + 1}\right)^{y} \qquad (2-22)$$

It can be seen that Equation 2-22 is a Polya density function with parameters $and\ \mu\beta$. When random effects follow $e^{\varepsilon_i} \sim Gamma\ (\alpha, \beta)$, then $y \sim Polya\ (\alpha, \mu\beta)$. A special case is $= 1/\alpha$, when Equation 2-22 becomes:

$$m\left(y|\alpha, \frac{1}{\alpha}\right) = \frac{\Gamma(y+\alpha)}{y!\,\Gamma(\alpha)}\left(\frac{\alpha}{\mu+\alpha}\right)^{\alpha}\left(\frac{\mu}{\mu+\alpha}\right)^{y} \qquad (2-23)$$

Equation 2-23 is a density function of NB distribution. When the crash count y follows the Poisson distribution with its own mean $\lambda(y \sim Poisson\ (\lambda))$ and λ follows a gamma distribution with shape parameter α and scale parameter μ/α ($\lambda \sim Gamma\ (\alpha, \mu/\alpha)$, then crash counts marginally follow the NB distribution with mean μ and dispersion parameter α: $y \sim NB(\mu, \alpha)$. The estimators of the expected Poisson mean $E(\lambda)$ and variance $Var(\lambda)$ are:

$$\lambda \sim Gamma\ (\alpha, \mu/\alpha)$$

$$E(\lambda) = \mu$$

$$Var(\lambda) = \alpha(\mu/\alpha)^2 = \mu^2/\alpha$$

The estimators of expected crash counts $E(y)$ and variance $Var(y)$ are:

$$y \sim \text{NB}\ (\mu, \alpha)$$

$$E(y) = \mu$$

$$Var(y) = \mu + \mu^2/\alpha$$

B. Poisson-log normal models ($y_{i,t} \sim \text{Pois}\ (e^{\varepsilon_i}\mu_{i,t})$ and $\varepsilon_i \sim \text{Normal}\ (0, \sigma^2)$)

When $\varepsilon_i \sim \text{Normal}\ (0, \sigma^2)$, then $e^{\varepsilon_i} \sim \text{log normal}\ (0, \sigma^2)$, the FB model is said to be a PLN model. Similarly, the posterior and marginal distribution can be shown as follows:

- Posterior distribution of $\lambda_{i,t}$ of PLN model

The subscript is omitted for a simple expression as before. If $e^{\varepsilon_i} \sim \text{log normal}\ (0, \sigma^2)$, then $\lambda \sim \text{log normal}\ (\ln\mu, \sigma^2)$,

Prior Log-Normal distribution $\lambda \sim \text{log normal}\ (\ln\mu, \sigma^2)$:

$$\pi(\lambda|ln\mu,\sigma^2) = \frac{1}{\lambda\sqrt{2\pi}\sigma}e^{-\frac{(ln\lambda-ln\mu)^2}{2\sigma^2}} \qquad \lambda > 0 \qquad (2-24)$$

Plugging Equations 2-16 and 2-24 into Equation 2-15,

$$\begin{aligned}
\pi(\lambda|y,\sigma^2) &\propto p(y|\lambda)\pi(\lambda|\alpha,\mu\beta) \\
&\propto \frac{\lambda^y e^{-\lambda}}{y!} \times \frac{1}{\lambda\sqrt{2\pi}\sigma}e^{-\frac{(ln\lambda-ln\mu)^2}{2\sigma^2}} \\
&\propto \frac{\lambda^{y-1}e^{-\lambda}}{\sqrt{2\pi}\sigma}e^{-\frac{(ln\lambda-ln\mu)^2}{2\sigma^2}} \qquad (2-25)
\end{aligned}$$

It has been found that the prior log normal distribution and the posterior are not conjugate distributions. In this case, the estimators of the expected Poisson mean $E(\lambda)$ and variance $Var(\lambda)$ cannot have a closed form. Both can be estimated through the MCMC method.

- Marginal distribution $\mathrm{m}(\mathrm{y}|\alpha,\beta)$ of PLN model

As defined in Equation 2-19, the marginal distribution of crash counts y is:

$$m(y|\sigma^2) = \int \frac{\lambda^y e^{-\lambda}}{y!} \times \frac{1}{\lambda\sqrt{2\pi}\sigma}e^{-\frac{(ln\lambda-ln\mu)^2}{2\sigma^2}}d\lambda$$

$$m(y|\sigma^2) = \frac{1}{y!\sqrt{2\pi}\sigma}\int \lambda^{y-1}e^{-\frac{\lambda+(ln\lambda-ln\mu)^2}{2\sigma^2}}d\lambda \qquad (2-26)$$

It can be seen that there is also no closed form of the integral in Equation 2-26. The estimators of expected crash counts $E(y)$ and variance $Var(y)$ cannot be expressed in a closed form either.

2.4.2 Multivariate FB models

Practically, the analysed dependent variables (crash counts) may not be a single level and/or single type. In fact, crash data are normally collected at different severity levels (i.e. fatal, injured, PDO, etc.) and pertain to different types (e.g., total, read end, right angle and left turn). Intuitively, collisions at different severity levels are correlated while crashes for different types may or may not be correlated. In such cases, a univariate FB model is unable to capture the underlying correlation that might occur between these different severity levels and/or different types of crashes. For this reason, it is natural to believe that a multivariate FB approach might be a better approach for safety analyses based on crash types and severities. The following sections will introduce two multivariate FB models: multivariate Poisson (MVP) and MVPLN models.

A. Multivariate Poisson (MVP) models

MVP models can be presented in a different way (Tsionas, 1999, 2001; Ma and Kockelman, 2006; Karlis and Meligkotsidou, 2005; Brijs et al., 2006). For ease of implementation, the following assumption is made for MVP distributions, as used by Tsionas (1999, 2001).

Let crash counts $\boldsymbol{Y_i} = Y_i^1, Y_i^2, \dots, Y_i^L$ be described as L types or injury severity levels of multivariate crash records at location i (where i=1, 2…N). Suppose that $\mathbf{X_i} = \mathrm{X}_\mathrm{i}^1, \mathrm{X}_\mathrm{i}^2, \dots, \mathrm{X}_\mathrm{i}^\mathrm{L}$ are independent Poisson variables at site i with parameters $\theta_\mathrm{i}^1, \theta_\mathrm{i}^2, \dots, \theta_\mathrm{i}^\mathrm{L}$, and Δ_i follows a Poisson distribution with parameter ζ_i, independently of $\mathrm{X}_\mathrm{i}^1, \mathrm{X}_\mathrm{i}^2, \dots, \mathrm{X}_\mathrm{i}^\mathrm{L}$. Define

$$\begin{aligned} Y_i^1 &= \mathrm{X}_\mathrm{i}^1 + \Delta_\mathrm{i} \\ Y_i^2 &= \mathrm{X}_\mathrm{i}^2 + \Delta_\mathrm{i} \qquad (2-27) \\ &\vdots \\ Y_i^\mathrm{L} &= \mathrm{X}_\mathrm{i}^\mathrm{L} + \Delta_\mathrm{i} \end{aligned}$$

Then, the variables $Y_i^1, Y_i^2, \dots, Y_i^L$ (where i=1, 2,...,N) are said to follow the MVP distribution. It can be seen that $Y_i^1, Y_i^2 \dots, Y_i^L$ marginally follow Poisson distributions with means $\theta_\mathrm{i}^1 + \zeta_\mathrm{i}, \theta_\mathrm{i}^2 + \zeta_\mathrm{i}, \dots, \theta_\mathrm{i}^\mathrm{L} + \zeta_\mathrm{i}$. Thus MVP cannot model data with overdispersion.

The correlation coefficients between Y_i^m and Y_i^k is $R_{Y_i^\mathrm{m}, Y_i^\mathrm{k}}$,

$$R_{Y_i^\mathrm{m}, Y_i^\mathrm{k}} = \frac{\mathrm{Cov}(Y_i^\mathrm{m}, Y_i^\mathrm{k})}{\sigma_{Y_i^\mathrm{m}} \sigma_{Y_i^\mathrm{k}}} \qquad (2-28)$$

where

$\mathrm{Cov}(Y_i^m, Y_i^k)$ = the covariance between Y_i^m and Y_i^k, and

$\sigma_{Y_i^m}, \sigma_{Y_i^k}$ = standard devations

Since Δ_i is independent of X_i. Δ_i and X_i are Poisson distributed.

$$\begin{aligned}\mathrm{Cov}(Y_i^m, Y_i^k) &= \mathrm{Cov}(X_i^m + \Delta_i, X_i^k + \Delta_i) \\ &= \mathrm{Cov}(X_i^m, X_i^k) + \mathrm{Cov}(X_i^m, \Delta_i) + \mathrm{Cov}(X_i^k, \Delta_i) + \mathrm{Cov}(\Delta_i, \Delta_i) \\ &= \mathrm{Var}(\Delta_i, \Delta_i)\end{aligned}$$

$$\mathrm{Cov}(Y_i^m, Y_i^k) = \zeta_i \qquad (2-29)$$

$$\begin{aligned}\mathrm{Var}(Y_i^m) &= \mathrm{Var}(X_i^m + \Delta_i) \\ &= \mathrm{Var}(X_i^m) + \mathrm{Var}(\Delta_i) + 2\mathrm{Cov}(X_i^m, \Delta_i) \\ &= \theta_i^m + \zeta_i\end{aligned}$$

$$\mathrm{Var}(Y_i^m) = \theta_i^m + \zeta_i$$

Similarly,

$$\mathrm{Var}(Y_i^k) = \theta_i^k + \zeta_i$$

Thus

$$R_{Y_i^m, Y_i^k} = \frac{\mathrm{Cov}(Y_i^m, Y_i^k)}{\sigma_{Y_i^m}\, \sigma_{Y_i^k}} = \frac{\zeta_i}{\sqrt{(\theta_i^m + \zeta_i)(\theta_i^k + \zeta_i)}} \qquad (2-30)$$

It can be seen that the correlation coefficient $R_{Y_i^m, Y_i^k}$ is a non negative number. This could be critical for the analyzed data. In fact, the data do not always meet this requirement. This method has been investigated by several researchers in road safety studies. For example, Ma and Kockelman (2006) applied an MVP regression approach to assess the effects of covariates on collision counts at different severity levels, Brijs et al. (2006) employed MVP to identify and rank sites according to their total expected cost to society by using accident data from 23,184 accident locations in Flanders (Belgium). However, one must keep in mind that MVP models do not support negative covariances (or negative correlation coefficients) between random variables. The covariance (or correlation coefficient) in the MVP setting is always positive, as shown in Equations 2-29 and 2-30. Moreover, the MVP model has Poisson marginal distributions and thus cannot model overdispersion. Furthermore, crash data are found to be significantly overdispersed relative to the mean, and using the Poisson regression models may overstate or understate the likelihood of crashes (Maher and Summersgill, 1996).

The above two drawbacks, especially the second one, greatly limit the application of MVP. As a matter of fact, MVP has been rarely applied in road safety studies. For this reason, MVP models will not be further explored in this study. To overcome the shortcomings of MVP, a more flexible and powerful multivariate FB model, MVPLN, has been suggested in road safety studies and will be briefly introduced below.

B. Multivariate Poisson-log normal (MVPLN) models

Let crash counts $\boldsymbol{Y_i} = Y_i^1\ , Y_i^2, \ldots, Y_i^L$ be L types or injury severity levels of multivariate crash records at location i. M is defined to be the number of covariates and $\boldsymbol{X} = (1, X_1, X_2, \ldots, X_M)'$. Let $\boldsymbol{\beta^k} = \left(\beta_0^k, \beta_1^k, \beta_2^k, \ldots, \beta_M^k\right)$ be the (M+1) dimensional regression coefficients for crash type or severity k. $\boldsymbol{\varepsilon_i} = (\varepsilon_i^1\ , \varepsilon_i^2, \ldots, \varepsilon_i^L)'$ are the unobserved random errors for crash type or severity 1, type or severity 2,..., type or severity L, respectively.

Each type or severity of crash is assumed to be independently Poisson distributed. That is,

$$y_{i,t}^k | \boldsymbol{\beta^k}, \boldsymbol{X}, \varepsilon_i^k \sim \text{Pois}\left(\lambda_{i,t}^k\right)$$

where

$$\lambda_{i,t}^k = \ exp\ (\boldsymbol{\beta^k}.\boldsymbol{X} + \varepsilon_i^k) \qquad (2-31)$$

$$\mu_{i,t}^k = \exp\ (\boldsymbol{\beta^k}.\boldsymbol{X}) \qquad (2-32)$$

where

$\lambda_{i,t}^k$= the expected crashes of type/severity level k at location *i* in year *t*, and

$\mu_{i,t}^k$= the expected crashes of type/severity *k* at similar sites in year *t*.

Assume that $\mathrm{\varepsilon_i^L}|\Sigma \sim \mathrm{N_L}(0, \Sigma)$, where Σ is an unrestricted $\mathrm{L} \times \mathrm{L}$ covariance matrix between different severity/type of crashes; thus the correlation between different types or severity of crashes is built through the multivariate normal distribution of ε_i. The details of MVPLN is described in Chapter 3.

2.5 COMPARISON OF THE BASICS OF FB AND EB

From the above sections and Chapter 1, it is seen that both EB and FB combine prior and current information to derive an estimate for the expected safety of a site that is being evaluated. In the context of crash analyses, the prior information is the expected accident frequency from a group of similar sites and the current information is the site specific observed accident frequency. EB and FB are not different types of studies. They are indeed two related approaches to combining prior and current information. However, there are still some differences in the two approaches, which can be summarized as follows.

In the EB approach, prior information comes from using a reference group of sites similar to those under evaluation to calculate a sample mean and variance or from a calibrated safety performance function (SPF) that relates the crash frequency of the reference sites to their characteristics. The SPFs are developed by the maximum likelihood method. The inference of the parameters of SPFs is based on the likelihood of the data alone. The point estimates of the expected mean and the variance are then combined with the site specific crash count to obtain an improved estimate of a site's long-term expected crash frequency. Basically, there are two steps to conduct an EB study: the first step of the procedure (SPF development) uses a classical approach (maximum likelihood) to develop SPFs by using observed data while the second step (the estimation of the expected crashes) employs a Bayesian approach, which combines the observed crashes with the SPF estimates to obtain estimates of expected crashes. It can be seen that the observed data are used twice in the EB procedure: once in the development of SPFs, and in another time, to estimate site specific long-term expected crash frequency. Hence, this method is called the empirical Bayesian method. It should be noted, however, in EB studies, the NB distribution of the dependent variable (crash counts) is usually assumed, and a large sample of the reference group is required for developing SPF.

In the Full Bayes approach, the likelihood of the observed data y given parameters θ, denoted as $p(y|\theta)\ or\ L(\theta|y)$, is used to modify prior beliefs $\pi(\theta)$ with updated knowledge to obtain posterior density $\pi(\theta|y)$. The procedure to combine prior information with site specific crash frequency is integrated. Unlike the EB procedure, the observed crash data are used only once with the FB procedure. Instead of a point estimate of the expected mean and its variance, a distribution of likely values of expected crashes is generated. Thus rich information can be obtained from the posterior distributions. The FB or HB methods offer a number of potential advantages and are summarized below:

- the small sample properties of FB models, which may allow the estimation of valid crash models with smaller sample sizes,
- the ability to include prior knowledge in the coefficient values in the modeling along with the data collected,
- the ability to include spatial correlation between sites in the model formulation,
- the ability to handle correlation longitudinally (time series correlation),
- the ability to deal with correlations between different severities / different types of crashes,
- the capability to provide rich information of the inference such as the posterior distribution of outcomes,
- the ability to specify very complex model forms,
- the use of an integrated procedure for all outcomes, avoiding the need for independent SPFs, and
- the flexibility for model selection either in terms of function forms or underlying, more complex distributions of dependent variables (PG, PLN, and mixture distributions), through an HB structure.

With regards to the last bulleted point, the FB method can accommodate distributions such as hierarchical PG, PLN and even mixture models (Lord, 2006; Miranda-Moreno and Fu, 2007; Lan et al., 2009; Persaud et al., 2010; Maher and Mountain, 2009; Park and Lord, 2009), while the EB approach relies on the assumption of an NB distribution of crash counts in using an NB dispersion parameter directly in the estimation process.

CHAPTER 3 FB MODELS WITH CORRELATED DATA

Observed crash counts are usually correlated, for example, along a corridor or across years. For the former case, there might be a spatial correlation between the sites while a time series correlation might exist for the latter case. Furthermore, crashes of different types/severities at the same site might be correlated. When this is the case, then a multivariate approach that was described in Chapter 2 is necessary. The situation becomes extremely complicated when all of these conditions converge, for instance, when there are observed crash counts of different types/severities along a corridor across several years. In this case, spatial correlation, time series correlation and a multivariate approach might need to be combined together. This might induce underestimation or overestimation of the posterior distributions of crash frequency if these correlations are disregarded. However, it is impossible to conduct this sort of analysis by using the empirical Bayes approach (EB), because EB itself is a univariate approach. Because of MCMC simulation techniques, this problem can be solved with the FB method.

Similar to Chapter 2, FB models with correlated data will be introduced in two separate ways: univariate FB models with correlated data and multivariate FB with correlated data.

3.1 UNIVARIATE FB MODELS WITH CORRELATED DATA

3.1.1 Univariate FB Models with Time Series Data

Accident counts on an entity often exhibit time trends due to temporal changes in factors such as road conditions, traffic flow, weather, the economy, accident-reporting practices, advances in vehicular technologies, and design standard improvements. Successive observations are likely to be dependent. It therefore stands to reason that FB models which accommodate these trends should provide better estimates of safety than traditional models in the identification of hazardous entities and in the evaluation of treatments applied to those entities since both tasks require the use of time series accident data. In an EB framework, this time correlation might be conducted by applying a time multiplier as described in Chapter 1 and a generalized estimating equation (GEE) procedure (Zeger and Liang, 1986; Liang and Zeger, 1986; Lord and Persaud, 2000). Lord and Persaud (2000) applied the GEE procedure to analyse motor-vehicle accidents at 868 four-legged signalized intersections in Toronto for 1990 through to 1995. They assumed that the crash counts follow an NB distribution. The results demonstrated that not accounting for temporal correlation does not affect the coefficient estimates, but that the variance of these estimates is considerably underestimated. For the EB approach, this means that the estimated SPFs would be the same with or without the GEE procedure; thus the final results of either a network ranking or treatment effect analysis should be the same. This result could be caused by the limitations of EB as mentioned in Chapter 2 as well as the GEE procedure itself. Another reason might be that the crash database used has only 5 years of observations that may not exhibit strong time correlation. The GEE method for other data might provide better results, but applications of GEE are minimal. For FB, it is much more flexible to work with time series data. The details are as follows.

A. Time multiplier model

Suppose $\mathrm{Y_{i,t}} \sim \text{Pois}\left(\mu_{i,t} e^{\varepsilon_i}\right) \quad i = 1, 2, \ldots, n$

$$\lambda_{i,t} = \mu_{i,t} e^{\varepsilon_i}$$

$$\log \mu_{i,t} = \beta_{0,t} + \beta_1 X_{i,t,1} + \cdots \beta_M X_{i,t,M} \qquad (3-1)$$

where

$Y_{i,t}$ = observed number of crashes in time t at site i,

$\lambda_{i,t}$= expected number of crashes in time t at site i,

$\mu_{i,t}$ = expected number of crashes in time t at the sites similar to site i,

$X_{i,t,1}, X_{i,t,2} \ldots, X_{i,t,M}$ = Covariates in time t at site i and M is the number of covariates,

$\beta_{0,t}$= Time varying intercept,

$\beta_1, \ldots, \beta_M =$ coefficients that correspond to covariates $X_{i,t,1}, X_{i,t,2} \ldots, X_{i,t,M}$, and

e^{ε_i} = random effects at site i , which follows gamma distribution or log normal distribution as described in Chapter 2.

The time varying intercept $\beta_{0,t}$is used to account for temporal variations in crash occurrences. It is similar, in principle, to the time multiplier in an EB study. It has been demonstrated that this model can provide promising predicted results (Persaud et al., 2010; Lan and Persaud, 2010).

B. Time varying coefficient model

$$\log \mu_{i,t} = \beta_{0,t} + \beta_{1,t} X_{i,t,1} + \cdots \beta_{M,t} X_{i,t,M} \qquad (3-2)$$

where

$\beta_{1,t}, \ldots, \beta_{M,t} =$ Time varying coefficients that correspond to covariates $X_{i,t,1}, \ldots, X_{i,t,M}$

This model can be seen an extension of the time multiplier model. However, it does not mean that this model is superior to the time varying model. In fact, it has been confirmed that the time multiplier model is better than this model probably due to overparameterizing (Lan and Perssaud, 2010).

C. Time trend model

A time trend model might be employed to deal with time series data. It can be seen as an alternative way to consider the relationship among time series data. The model can be described as:

$$\log \mu_{i,t} = \beta_0 + \beta_1 X_{i,t,1} + \cdots \beta_M X_{i,t,M} + \gamma t \qquad (3-3)$$

In a comparison of Equation 3-3 with Equation 3-1, it can be found that Equation 3-3 has an extra term γt to account for a time trend and that all the coefficients are fixed. This model was found to probably provide better results than other previous models (Lan and Persaud, 2010).

The above three models are simple ways to deal with time series data. However, they do not really introduce time series correlation between successive time periods. The following section introduces models that are able to explain inherent temporal correlations.

D. Autoregressive (AR) FB models

One major class of models for time series data is the autoregressive (AR) model (Chib and Greenberg1994; Congdon, 2001 and 2003), and the first order autoregressive process AR (1) is the simplest model to describe dependence in the values of an outcome variable over successive time points. An AR (1) model in random error can be employed to account for time dependence in road safety studies, which means that values of random effects at time t depend upon their immediate predecessor. A model that allows for AR (1) dependence in the random errors might be ideal to reproduce the dynamic features of time series crash data.

The AR (1) model can be written as:

$$Y_{i,t} \sim \text{Pois}\left(\mu_{i,t} e^{\omega_{i,t}}\right) \qquad i = 1, 2, \dots, n$$

and

$$\omega_{i,t} = r\omega_{i,t-1} + u_{i,t} \qquad t \geq 2 \qquad (3-4)$$

$$\lambda_{i,t} = \mu_{i,t} e^{\omega_{i,t}}$$

$$\mu_{i,t} = exp\ (\boldsymbol{\beta}.\boldsymbol{X})$$

where

$\mathbf{X} = (X_{i,t,1}, X_{i,t,2} \dots, X_{i,t,M})'$, and M is the number of covariates at time t,

$\boldsymbol{\beta} =$ Vector of coefficients, $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_M)$,

$e^{\omega_{i,t}} =$ Random effects at time t ,

$u_{i,t} =$ Unstructured white noise, $u_{i,t} \sim N(0, \sigma^2)$, $t \geq 1$, and

$r =$ Correlation coefficient of random effects between successive time periods.

The first point of $\omega_{i,t}$ is usually modeled as:

$$\omega_{i,1} = r\omega_{i,0} + u_{i,1} \qquad (3-5)$$

$\omega_{i,0}$ is a latent data point, typically modelled as a fixed effect or an unknown parameter, and the diffuse uninformed prior can be a normal distribution with a large variance, i.e., $\omega_{i,0} \sim N(0,1000)$.

It should be noted that when $|r| < 1$, the AR process is stationary, otherwise the process is nonstationary (Congdon, 2001,2003; Chib and Greenberg, 1994). It should also be noted that a classical statistical approach with the AR (1) rests on a stationarity assumption. However, for the FB approach, this restriction can be relaxed, which means that some trends may be seen in the data.

- Stationary AR (1) process

For this, simply confine $r \in (-1,1)$, i.e., $r \sim \text{Uniform}\ (-1,1)$, and the above models can be used to model a stationary process. Alternatively, the first point of $\omega_{i,t}$ can be simply set to $\omega_{i,1} \sim N(0, \sigma^2/(1-r^2))$ (Haque et al., 2010; Congdon, 2001).

- Nonstationary AR (1) process

On the other hand, for a nonstationary process, it can be modeled by defining $r \sim N(0,1)$ such that r does not belong to $(-1,1)$, and the variance of the first point $\omega_{i,1}$ can be relaxed.

For a road safety study, the Poisson mean usually transforms to its logarithm in the modeling process. The stationarity of AR (1) should be met. However, in this study, the assumptions for prior of stationarity or non-stationarity in the AR (1) process have been investigated and the results will be presented in Chapter 5.

E. AR (1) with time trend FB model

The above mentioned non-stationary process can be transformed to a stationary process by adding a trend variable into the model. Alternatively, one can consider that a trend analysis only accounts for a broadscale time series pattern in a long period, while an AR (1) model with random effects explains fine-scale autocorrelation between successive time periods.

The functional form of the proposed model is similar to the above AR (1) model except that a time trend variable is included in the functional form of $\mu_{i,t}$. In other words, this model includes one more covariate of time, t. It should be noted, however, that the literature contains no instances where this model was used for road safety analysis. This model, and the previous AR (1) model, are explored for network ranking studies only; they are not applicable for treatment effect analyses since the treatment year typically needs to be excluded. The results are detailed in Chapter 5. It is anticipated that the AR (1) model with a stationarity assumption of random errors would be better if it represents a stationary AR process.

3.1.2 Univariate Spatial FB Models

Spatial FB models are required when analysed sites are along a corridor or within a road network because the sites along a certain corridor will affect each other, especially for those that are close to each other. Several adjacent sites, e.g.., signalized intersections along a certain corridor or within a road network, share a high percentage of the same or similar traffic. For example, signals within a road network are coordinated in most circumstances, and this coordination will promote the platooning of vehicles that cross the intersections. Furthermore, adjacent entities probably have similar types of land use and roadway design. In order to improve estimation in safety analysis, there is the need to look at the spatial relationship for adjacent sites along a corridor or within a road network rather than treat each intersection as an isolated entity.

Spatial correlation can be managed in several ways. For example, Abdel-Aty and Wang (2006) employed the GEE procedure (Zeger and Liang, 1986; Liang and Zeger, 1986; Lord and Persaud, 2000) to introduce spatial correlations by using the maximum likelihood method. They assumed that crash counts follow the NB distribution, analyzed the resulting regression coefficients, and concluded that the spatial model is better based on the cumulative residual plots from all of the developed models. It should be noted that they used an EB method.

A. Gamma distribution for the latent variable

Shaddick et al. (2007) used a gamma distribution for the latent variable. The fundamental difference is that instead of the mean of a Poisson distribution at a particular location being directly associated with the value of a latent variable at that location, the latent variables lie on the boundaries between the locations. The mean for a particular location is then modeled as a combination of the latent variables that lie on its boundaries; this combination induces correlation between the Poisson means as shown below. The appeal of this approach is the ease in working with a PG set-up in which exact expressions for expectations and variances are available.

Suppose

$$Y_i \sim \text{Pois}\left(\mu_i \theta_i^*\right) \qquad i = 1, 2, \dots, n$$

and

$\lambda_i = \mu_i \theta_i^*$

$\mu_i = \exp\left(\boldsymbol{\beta}.\boldsymbol{X}\right)$

$$\theta_i^* = w\theta_{i-1} + (1 - w)\theta_i \qquad (3-6)$$

where

Y_i = observed number of crashes at site i,

λ_i= expected number of crashes at site i,

μ_i = expected number of crashes at sites similar to site i,

$\boldsymbol{X} = (1,\ X_1,\ X_2 \dots,\ X_M)'$, and M is the number of covariates,

$\beta =$ vector of coefficients, $\boldsymbol{\beta} = (\ \beta_0, \beta_1, \dots, \beta_M)$,

θ_i^* = random effect at site i,

θ_{i-1}, θ_i = latent variable controls between (Y_{i-1}, Y_i) and (Y_i, Y_{i+1}), respectively, $\theta_0, \theta_1, \dots, \theta_n$ are considered to be independent and follow a gamma distribution, $\theta_i \sim \text{Gamma}\,(a, 1/a)$ for $i = 1,2, \dots, n$, and

w = the level of dependence between the latent variables, $w \in (0,1)$.

It can be seen that spatial correlation is included in the developed models through random effects at site i. However, this spatial model is only appropriate for spatial correlations along a corridor where the analyzed site has only two adjacent sites. This spatial method cannot model cases with more than two adjacent sites, such as road networks.

B. Log normal distribution for the latent variable

In most cases, the latent variables are treated as log normal distributions for easy presentation and implementation. Unlike the gamma distribution of latent variables, this method does not have any restriction; it works for both corridor and road network cases. The hierarchical Poisson models, however, have several forms. The two most popular forms are the conditional auto-regressive (CAR) models with or without site specific random effects (Aguero-Valverde and Jovanis, 2008: Lichstein et al, 2002; Lu et al., 2007).

a. Model 1--- only spatial correlated random effects included

$$Y_i \sim \text{Pois}\left(\mu_i e^{\psi_i}\right) \qquad i = 1, 2, \dots, n \qquad (3-7)$$

where

$\lambda_i = exp\ (\boldsymbol{\beta}.\boldsymbol{X} + \psi_i)$

$\mu_i = exp\ (\boldsymbol{\beta}.\boldsymbol{X})$

$$\log\lambda_i = \log\mu_i + \psi_i \qquad i = 1, 2, \dots, n \qquad (3-8)$$

$$\psi_i \sim N(\ \mu_{\psi_i}, \sigma^2_{\psi_i}) \qquad i = 1, 2, \dots, n \qquad (3-9)$$

There are two methods to obtain μ_{ψ_i} and $\sigma^2_{\psi_i}$,

- Method A (CAR model)

$$\mu_{\psi_i} |\ \psi_{j\neq i} = \rho \sum_j w_{ij}\ \psi_j \qquad (3-10)$$

and

$$\sigma^2_{\psi_i} = \text{constant}, \qquad \sigma^2_{\psi_i} \sim \text{Inv Gamma}\ (\alpha, \beta)$$

where

e^{ψ_i} = random effect at site i,

$e^{\psi_{j\neq i}}$ = all the random effects at sites j which are adjacent to site i,

μ_{ψ_i}= expected logarithm of random effect at site i,

w_{ij} = weight that determines the relative influence of location j on location i, typically defined in CAR models to decrease with increasing distance between i and j (e.g., $w_{ij} = {}^{1}\!/_{\text{Distance}_{ij}}$) and is zero if i and j are not adjacent; sometimes one can simply set $w_{ij} = 1$ if sites i and j are adjacent, otherwise $w_{ij} = 0$ (Lichstein et al, 2002; Congdon, 2003a, 2003b),

ρ = a parameter to be estimated that determines the direction (positive or negative) and magnitude of the spatial correlated effect.

- Method B (intrinsic CAR model)

$$\mu_{\psi_i} |\ \psi_{j\neq i} = {}^{\sum_j w_{ij}\ \psi_j}\!\Big/_{\sum_j w_{ij}} \qquad (3-11)$$

and

$$\sigma^2_{\psi_i} = {}^{\sigma^2_{\psi}}\!\Big/_{\sum_j w_{ij}} \qquad (3-12)$$

σ_{ψ}^{2} = Variance that controls the extra Poisson variation, typically set equal to some fixed value, or assigned a distribution itself, often a relatively vague inverse gamma distribution.

Several researchers have introduced similar models in their study (Lu et al., 2007; Aguero-Valverde and Jovanis, 2008; Congdon, 2003b). However, it should be noted that no study comparing methods A and B has been found in the literature, and there are no applications in road safety studies which use model 1 and include methods A and B.

b. Model 2--- Spatial Correlated Random Effects Combined with Site Specific Random Effects

$$Y_i \sim \text{Pois}\left(\mu_i e^{\varepsilon_i} e^{\psi_i}\right) \qquad i = 1, 2, \ldots, n \qquad (3-13)$$

where

e^{ε_i} = uncorrelated site specific random effects, which is the same as Equation (2-12), basically reflecting unmeasured differences among segments, and assumed to be independent and identically gamma or log normal distributed, $e^{\varepsilon_i} \overset{\text{iid}}{\sim} \text{Gamma}(\alpha, \beta)$ or $\varepsilon_i \overset{\text{iid}}{\sim} N(0, \sigma^2)$.

3.1.3 Univariate Spatial–Temporal FB Models

When spatial data are collected over time, a spatial-temporal statistical analysis can provide benefits which are not possible with only the spatial or temporal model. For example, a spatial–temporal FB model is a good option for investigation when an analysis is performed for a roadway corridor or a network that uses longitudinal data. The spatial-temporal FB models can be developed and have two basic forms with regard to different ways in dealing with time series data, as presented below.

A. Basic spatial–temporal FB models

This basic spatial–temporal model actually only considers spatial correlation and time variation. It does not include time series correlation in the model. As for the spatial correlation part, there are two models, with and without site specific random effects e^{ε_i} .

- Model A: with site specific random effects e^{ε_I}

$$Y_{i,t} \sim \text{Pois}\left(\mu_{i,t} e^{\varepsilon_i} e^{\psi_i}\right) \qquad i = 1, 2, \ldots, n \qquad (3-14)$$

- Model B: without site specific random effects e^{ε_I}

$$Y_{i,t} \sim \text{Pois}\left(\mu_{i,t} e^{\psi_i}\right) \qquad i = 1, 2, \ldots, n \qquad (3-15)$$

$\mu_{i,t}$ takes the same functional form of Equations 3-1, 3-2, and 3-3 to account for temporal variation in crash occurrence. Random effects e^{ψ_i} associated spatial correlations are obtained through the procedure presented above. The site specific random effects e^{ε_i} can have gamma or log normal distribution structures.

There has been only one known instance of a spatial-temporal model application in road safety studies. For this, Li et al. (2007) employed a GIS-based Bayesian crash rate model for an intra-city motor vehicle crash analysis. Their model can be rewritten as:

$$Y_{ijt} \sim \text{Pois}\left(E_{ijt}\mu_{ijt} e^{\varepsilon_{ijt}} e^{\psi_{ijt}}\right) \qquad i = 1, 2, \ldots, n \qquad (3-16)$$

$$\mu_{ijt} = \beta_{jt} + \alpha x$$

where

Y_{ijt} = crash counts at site i on road type j in time t.,

E_{ijt} = a factor which is proportional to the annual average daily vehicle miles traveled (VMT),

μ_{ijt} = expected crashes at site i on road type j in time t. It can be seen that μ_{ijt} is a multiplier model although only one covariate x, which corresponds to road type, is included,

$e^{\varepsilon_{ijt}}$ = site specific random effects at site i on road type j in time t and which is assumed to be $\varepsilon_{ijt} \overset{iid}{\sim} N(0, \sigma_t^2)$, and

$e^{\psi_{ijt}}$ = spatial correlated random effects at site i on road type j in time t. In other words, the adjacent spatial effects are correlated, but vary with time. It is assumed to follow the CAR model, which is defined in Equation (3-10).

It is evident that the developed model seems to count temporal effects repeatedly and might have some problems. The temporal effect was counted 3 times: in coefficient β_{jt}, in site specific random effects $e^{\varepsilon_{ijt}}$ and in spatial effects $e^{\psi_{ijt}}$. It should be noted that this was the only model explored in the Li et al. study (2007).

B. Spatial–temporal AR FB models

Fully spatial–temporal FB models account for correlations between successive time periods and among adjacent road segments or sites. Time series correlations can be properly accounted for by AR (1) random errors $e^{\omega_{i,t}}$. $\mu_{i,t}$ has two forms: one with time trend and another one without the trend. The method to obtain $\omega_{i,t}$, ψ_i is again, the same.

$$Y_{i,t} \sim \text{Pois}\left(\mu_{i,t} e^{\omega_{i,t}} e^{\psi_i}\right) \qquad i = 1, 2, \dots, n \qquad (3-17)$$

It should be noted that there is no site specific random effect e^{ε_i} included in this model in that the initial time series random effects $e^{\omega_{i,1}}$ already implicitly include site specific random effects. The univariate spatial–temporal FB models can have different expressions with various combinations of time series correlation random effects and spatial random effects as mentioned before.

3.2 Multivariate Poisson Log normal FB Models

3.2.1 Case where random errors $\boldsymbol{\varepsilon_i}$ follow multivariate normal distributions

Basic case:

From Chapter 2,

$Y_{ik}|\boldsymbol{\beta_k}, \boldsymbol{X}, \varepsilon_{ik} \sim$ Pois (λ_{ik}), denoted f $(y_{ik}|\boldsymbol{\beta_k}, \boldsymbol{X}, \varepsilon_{ik})$

$$\lambda_{ik} = e^{\varepsilon_{ik}} \mu_{ik}$$

$$log\lambda_{ik} = \beta_{0k} + \beta_{1k} X_{i1} + \cdots \beta_{Mk} X_{iM} + \varepsilon_{ik} \qquad (3-18)$$

$$\mu_{ik} = exp\,(\boldsymbol{\beta_k}.\boldsymbol{X}) \qquad (3-19)$$

$$i = 1,2, \dots, n \qquad k = 1, 2, \dots, L$$

where

$\boldsymbol{\beta_k} = (\beta_{0k}, \beta_{1k}, \dots, \beta_{Mk})$, are coefficients which correspond to type or severity k crash models, and

$\boldsymbol{X} = (1, X_{i1}, \dots, X_{iM})'$,covariates at location i.

$$p(y_{ik}|\lambda_{ik}) = \frac{{\lambda_{ik}}^{y_{ik}} e^{-\lambda_{ik}}}{y_{ik}!} \qquad (3-20)$$

Given that the random effects $(\varepsilon_{i1}, \varepsilon_{i2}, \dots, \varepsilon_{iL})'|\boldsymbol{\Sigma} \sim \mathrm{N_L}(0, \boldsymbol{\Sigma})$, denoted as $f(\boldsymbol{\varepsilon_i}|\boldsymbol{\Sigma})$, $\log\lambda_i$ independently follow an L-dimensional normal distribution $\log\boldsymbol{\lambda}_i|\boldsymbol{\Sigma} \sim \mathrm{N_L}(\boldsymbol{\mu_i}, \boldsymbol{\Sigma})$, the probability density function of the L-dimensional log normal distribution is

$$p(\boldsymbol{\lambda_i}|\boldsymbol{\mu_i}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{L/2}|\boldsymbol{\Sigma}|^{1/2}\prod_{k=1}^{L}\lambda_{ik}} \mathrm{e}^{-\frac{1}{2}(\boldsymbol{log\lambda_i} - \boldsymbol{log\mu_i})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{log\lambda_i} - \mathbf{log}\boldsymbol{\mu_i})} \qquad (3-21)$$

where

$\boldsymbol{\varepsilon_i} = (\varepsilon_{i1}, \varepsilon_{i2}, \dots, \varepsilon_{ik}, \dots, \varepsilon_{iL})'$, and

Σ = an unrestricted $L \times L$ covariance matrix between different severity/types of crashes.

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1L} \\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{2L} \\ \dots & \dots & \dots & \dots \\ \sigma_{L1} & \sigma_{L2} & \dots & \sigma_{LL} \end{pmatrix}$$

$\boldsymbol{\mu_i} = (\mu_{i1}\ \ \mu_{i2}\ \ \dots \mu_{iL})'$

$\boldsymbol{\lambda_i} = (\lambda_{i1}\ \lambda_{i2} \dots\ \lambda_{iL})'$

$\boldsymbol{log\lambda_i} = (log\lambda_{i1}\ log\lambda_{i2}\ \dots\ log\lambda_{iL})'$

$\boldsymbol{log\mu_i} = (log\mu_{i1}\ log\mu_{i2} \dots \mathrm{log}\mu_{iL})'$

Uninformative prior distributions are usually specified when there is a lack of sufficient prior knowledge of the distributions for individual parameters. The most common priors for regression parameters $\beta = \beta_{01}, \ldots, \beta_{M1}, \ldots, \beta_{0L}, \ldots, \beta_{ML}$ are defined as the diffused normal distributions (with zero mean and large variance). $\beta_{jk} \overset{iid}{\sim} N(0, \sigma^2)$, $j = 0,1, \ldots, M$, $k = 1, 2, \ldots, L$. The joint distribution of $\beta_{01}, \ldots, \beta_{M1}, \ldots, \beta_{0L}, \ldots, \beta_{ML}$ is defined as f_β,

$$f_\beta = \frac{1}{(\sqrt{2\pi}\sigma)^{L\times M}} e^{-\frac{\beta_{01}{}^2 + \cdots + \beta_{M1}{}^2 + \ldots + \beta_{ML}{}^2}{2\sigma^2}} \quad (3-22)$$

A Wishart (R, p) prior is defined for Σ^{-1}, denoted as (f_Σ) where R is the scale matrix and p is the degrees-of-freedom parameter respectively. The hyper-prior parameters R and $p \geq L$ are known, usually assuming $p = L$ for vague priors (Tunaru, 2002). The parameterization of the Wishart probability density function (pdf) is

$$f_{\boldsymbol{\Sigma}} = p(\boldsymbol{\Sigma}^{-1}|R, p) = |R|^{\frac{p}{2}} |\boldsymbol{\Sigma}^{-1}|^{\frac{p-L-1}{2}} exp^{-0.5Tr(R\boldsymbol{\Sigma}^{-1})} \quad (3-23)$$

The conditional density of observed crashes y_i given $\boldsymbol{\mu_i}, \boldsymbol{\Sigma}$ is:

$$p(y_i|\boldsymbol{\mu_i}, \boldsymbol{\Sigma}) = \int \prod_{k=1}^{L} p(y_{ik}|\lambda_{ik}) p(\boldsymbol{\lambda_i}|\boldsymbol{\mu_i}, \boldsymbol{\Sigma}) d\lambda_{i1} \ldots d\lambda_{iL} \quad (3-24)$$

Since $\mu_{ik} = \exp(\beta_k . X)$, $p(y_i|\mu_i, \Sigma)$ can be rewritten as $p(y_i|\beta, X, \Sigma)$. According to the Bayes theory, the posterior joint pdf of parameters β and Σ is proportional to the product of prior and likelihood,

$$p(\beta, \boldsymbol{\Sigma}|y, X) \propto prior * Likelihood$$

$$p(\beta, \boldsymbol{\Sigma}|y, X) \propto f_{\boldsymbol{\Sigma}} f_\beta \prod_{i=1}^{n} p(y_i|\boldsymbol{\mu_i}, \boldsymbol{\Sigma})$$

From Equation 3-20,

$$p(\beta, \boldsymbol{\Sigma}|y, X) \propto f_{\boldsymbol{\Sigma}} f_\beta \prod_{i=1}^{n} \int \prod_{k=1}^{L} p(y_{ik}|\lambda_{ik}) p(\boldsymbol{\lambda_i}|\boldsymbol{\mu_i}, \boldsymbol{\Sigma}) d\lambda_{i1} \ldots d\lambda_{iL} \quad (3-25)$$

Plugging in $f_\Sigma, f_\beta, p(y_{ik}|\lambda_{ik})$, and $p(\lambda_{i1}\ \lambda_{i2} \ldots \lambda_{iL}|\mu_{i1}\ \mu_{i2} \ldots \mu_{iL}, \boldsymbol{\Sigma})$, we get

$$p(\beta,\mathbf{\Sigma}|y,\mathbf{X}) \propto e^{-\frac{{\beta_{01}}^2+\cdots+{\beta_{M1}}^2+\ldots+{\beta_{ML}}^2}{2\sigma^2}} \times |R|^{\frac{p}{2}}|\mathbf{\Sigma}^{-1}|^{\frac{p-L-1}{2}} e^{-0.5Tr(R\mathbf{\Sigma}^{-1})}$$

$$\times \prod_{i=1}^{n}\int\prod_{k=1}^{L} \frac{{\lambda_{ik}}^{y_{ik}}e^{-\lambda_{ik}}}{y_{ik}!} \times \frac{1}{|\mathbf{\Sigma}|^{\frac{1}{2}}\prod_{k=1}^{L}\lambda_{ik}} \mathrm{e}^{-\frac{1}{2}(\boldsymbol{log\lambda_i}-\boldsymbol{log\mu_i})'\mathbf{\Sigma}^{-1}(\boldsymbol{log\lambda_i}-\mathbf{log}\boldsymbol{\mu_i})}\ d\lambda_{i1}\ldots d\lambda_{iL} \qquad (3-26)$$

As well, the posterior pdf of parameters β and Σ can be obtained as follows:

$$p(\beta|y,\mathrm{X}) \propto f_\beta \prod_{i=1}^{n} p(y_i|\mu_i,\mathbf{\Sigma}) \qquad (3-27)$$

$$p(\mathbf{\Sigma}|y,\mathrm{X}) \propto f_\Sigma \prod_{i=1}^{n} p(y_i|\mu_i,\mathbf{\Sigma}) \qquad (3-28)$$

Since there is no standard density form for conditional posterior distributions of $\mathrm{p(\beta|y,X)}$, and $\mathrm{p(\Sigma|y,X)}$, they require the use of the Metropolis–Hastings (M-H) algorithm set in WinBUGs, as mentioned in Chapter 2.

A. Multivariate with longitudinal crash data

If the data are time series data (longitudinal data), as mentioned earlier, it might be necessary to address these effects in models. Two methods can be employed to deal with possible temporal variations as described below.

a. Method 1: Function for expected crashes that contains time effects

μ_{ik}, expected crashes of type k at sites which is similar to site i, can have model forms which introduce time effects, such as a time multiplier (Equation 3-1), time varying coefficients (Equation 3-2) or time trend (Equation 3-3). Then, the best model can be selected based on model selection criteria, such as DIC (Spiegelhalter et al., 2002). Temporal effects can be accounted for in this way. Again, the random effects are $\boldsymbol{\varepsilon_i}|\mathbf{\Sigma} \sim \mathrm{N_L}(0,\mathbf{\Sigma})$.

b. Method 2: Autoregressive FB models

Instead of using FB models that contain temporal effects, time series correlation models (mostly AR (1) model) can be directly employed to obtain the expected Poisson mean $\lambda_{ik,t}$, crashes at site i for type k crashes in year t.

$$Y_{ik,t}|\boldsymbol{\beta_k}, \boldsymbol{X}, \varepsilon_{ik} \sim \text{Pois}\left(\lambda_{ik,t}\right)$$

$$\lambda_{ik,t} = \mathrm{e}^{\omega_{ik,t}}\mathrm{e}^{\varepsilon_{ik}}\mu_{ik,t} \qquad (3-29)$$

$$\mu_{\mathrm{ik,t}} = \exp(\boldsymbol{\beta_k}.\boldsymbol{X})$$

where

$\mathrm{e}^{\varepsilon_{\mathrm{ik}}}$ =random effects at site i for type k crashes, following multiple normal distribution, and

$\mathrm{e}^{\omega_{\mathrm{ik,t}}}$ = time series random effects at site i for type k crashes in year t and similar to Equation 3-5,

$$\omega_{\mathrm{ik,t}} = \mathrm{r}\omega_{\mathrm{ik,t-1}} + \mathrm{u}_{\mathrm{ik,t}} \qquad \mathrm{t} \geq 2 \qquad (3-30)$$

$\mathrm{u}_{\mathrm{ik,t}}$ = unstructured white noise, $\mathrm{u}_{\mathrm{i,t}} \sim \mathrm{N}(0, \sigma^2)$, $\mathrm{t} \geq 1$

Similarly, for a stationary process, the first point of $\omega_{\mathrm{ik,t}}$ can be set as $\omega_{\mathrm{ik,1}} \sim \mathrm{N}(0, \sigma^2/(1 - \mathrm{r}^2)$.

B. Multivariate models with spatial correlated crash data

When multivariate crash data are along a corridor or within a road network, spatial multivariate FB models are required.

$$\lambda_{ik,t} = e^{\psi_{ik}}\mathrm{e}^{\varepsilon_{ik}}\mu_{ik,t} \qquad (3-31)$$

where,

Spatially correlated random effects $\mathrm{e}^{\psi_{\mathrm{ik}}}$ can be obtained by methods discussed previously, such as CAR or ICAR (Equations 3-7 to 3-12). Local specific random effects $\boldsymbol{\varepsilon_i} = (\varepsilon_{\mathrm{i1}}\ \varepsilon_{\mathrm{i2}} \ldots \varepsilon_{\mathrm{iL}})'$ follow multivariate normal distributions, $\boldsymbol{\varepsilon_i} \sim \mathrm{N_L}(0, \boldsymbol{\Sigma})$.

C. Multivariate models with spatial correlated longitudinal crash data

Most often, spatial correlated data are also longitudinal. For multivariate models with spatial correlated longitudinal crash data, the method is similar to the above MVPLN AR (1) model. The random effects can be divided into two parts: one is the spatial correlated random effect $e^{\psi_{ik}}$ for type k crashes; another is the site specific random effects, $e^{\varepsilon_{ik}}$, which basically reflects unmeasured differences among segments and are assumed to be correlated to the random effects of other types of crashes at the same location. The expected crashes at site i for type k crash in year t can be written as:

$$\lambda_{ik,t} = e^{\psi_{ik}} e^{\omega_{ik,t}} e^{\varepsilon_{ik}} \mu_{ik,t} \qquad (3-32)$$

where,

Time series random effects $e^{\omega_{ik,t}}$, spatially correlated random effects $e^{\psi_{ik}}$ and local specific random effects $\boldsymbol{\varepsilon_i} = (\varepsilon_{i1}\, \varepsilon_{i2} \ldots \varepsilon_{iL})'$ can be obtained from aforementioned methods.

3.2.2 Case where both coefficients and random errors $\boldsymbol{\varepsilon_i}$ follow multivariate normal distributions

This method is quite similar to the previous case, except that coefficients $\beta_{01},\ \beta_{11}, \ldots, \beta_{ML}$ no longer independently follow the diffused univariate normal distributions. Instead, similar to random errors $\boldsymbol{\varepsilon_i}$, regression coefficients which correspond to different types of crashes follow a multivariate normal distribution $\beta_{j1},\ \beta_{j2}, \ldots, \beta_{jL} \sim N_L\left(0, \boldsymbol{\Sigma_{\beta_j}}\right), j = 0,1,2, \ldots, M$ for a vague prior, and similarly, the prior of ${\boldsymbol{\Sigma_{\beta_j}}}^{-1}$ is set to follow the Wishart distribution.

This model can be seen as an extension of previous MVPLN models. It is investigated and compared with normal MVPLN in this study and the results are shown in Chapter 6.

3.3 SUMMARY

Various FB models for correlated data are introduced, proposed and documented in this chapter. Correlated data may include spatially, temporally and locally correlated crash data of various types or severities. However, due to the limitations of the data at hand, FB models which involve spatially correlated data are not explored further in this research.

In addition, for longitudinal data, it should be noted that AR (1) cannot be applied for the treatment effect analysis in that the countermeasure implementation year typically needs to be excluded, and thus there is a gap between the before and after treatment. However, the multiplier FB model, the FB model with time varying coefficients and time trend FB models can be investigated as alternative ways to account for time effects. For network ranking, the FB models which consider time correlation and multivariate correlation have been extensively explored and the results are shown in Chapter 5.

CHAPTER 4 BAYESIAN MODEL SELECTION

In Chapters 2 and 3 various FB models, were proposed and introduced. Intuitively, the results from each model might be different. The question then arises as to how the quality of a computational model should be evaluated and which model should be selected to draw conclusions about application. To answer this question, a model comparison is required for a diversity of features, including variable selection in regression, determination of the number of components in a mixture model, or the choice of parametric family. Currently, there are few road safety research studies which involve model selection. Most studies only rely on one model while other research uses one functional form of expected crashes, but with different distributions of random errors, such as PLN or PG and MVPLN FB models. Of course, these studies are not comprehensive and the best model for the data might not even be considered. Thus the results could be biased. In our recent publication (Lan et al., 2010), it can be seen that estimations from competing models are indeed very different.

As with frequentist analogues, Bayesian model comparison will not indicate which model is `true', but rather will reveal the preference for the model given the data and other information. These preferences can be used to choose a single `best' model or improve estimation via model averaging, in which expected values obtained from different models are weighted by their corresponding posterior probabilities (Congdon, 2001; Raftery, 1999). The latter case is beyond the scope of this study and will not be further investigated. Rather the focus of this research is on ways to identify the best among competing models. This chapter starts with the discussion of model selection criteria, followed by the introduction of popular model selection methods, and ends with a summary of issues for the model selection methods.

4.1 MODEL EVALUATION CRITERIA

Model evaluation criteria not only depend on descriptive adequacy which determines whether the model fits the observed data, but also complexity or simplicity, which determines whether the model’s description of the observed data is achieved in the simplest possible manner (also defined as generalizability, which implies whether the model provides a good predictor of future observations) (Myung et al., 2009). However, it should be noted that this is highly reliant on the knowledge, experience, and preferences of the modeller for model selection as to whether the theoretical construct of the model helps make sense of the observed data, and whether the components of the model, especially its parameters, are understandable. This can be challenging when quantified criteria of competing models are very close with each other.

4.1.1 Descriptive adequacy

The descriptive adequacy of a model is assessed by measuring how well it fits a set of empirical data, in other words, by testing goodness-of-fit. A number of goodness-of-fit (GOF) measures can be employed, including sum of squared errors, maximum likelihood, and chi-squared values. GOF measures are popular because they are relatively easy to compute and the measures are versatile. Perhaps most of all, a good fit is an almost irresistible piece of evidence in favour of the adequacy of a model. As Myung et al. (2009) point out, a model that appears to do just what one wants it to do, which is to mimic the process that generates data, is a very attractive model. This, however, does not necessarily mean that a better fit will result in a more accurate model, as frequentists tend to expect. In fact, when comparing competing models, the result may be that the selected model is not a good model after all.

GOF would be suitable for model evaluation and comparison were it not for the fact that data are noisy (measurement error). A data set contains the regularity that is presumed to reflect the phenomenon of interest plus noise. GOF does not distinguish between the two (Myung et al., 2009), and provides a single measure of a model's fit for both (i.e., GOF = fit to regularity + fit to noise). Thus, a good fit can be achieved for the wrong reasons, by fitting noise well instead of regularity. For this reason, GOF alone cannot be used as a criterion for model selection because of the potential to yield misleading information.

4.1.2 Complexity or simplicity

What allows a model to fit noisy data better than its competitors is that it is the most complex. What distinguishes a simple model from a complex one is the sensitivity of the model to parameter variation (Myung et al., 2009). A complex model with many parameters, because of its extra flexibility, tends to capture these spurious patterns more easily than a simple model with few parameters for a noisy data. Consequently, the complex model yields a better fit to the data, not because of its ability to more accurately approximate the underlying process, but rather because of its ability to capitalize on sampling errors. Therefore, choosing a model based solely on its fit, without appropriately filtering out the effects due to sampling errors, will result in choosing an overly complex model that poorly generalizes to other data from the same underlying process. A consequence of such practice is that the model may become more sophisticated as additional parameters or modifications of the model are introduced to account for newly found discrepancy which may be, in fact, sampling errors between a model's predictions and new observations, and the model's generalizability may be further decreased (Myung, 2000).

It can be seen that complexity affects not only model fit, but also the generalizability of a model and the variability in parameter estimation. It is thus necessary to take this reality into account when evaluating models. Normally, a simple model will generalize better to new data sets than a complex model and therefore will have a higher degree of predictive accuracy. In addition, the behaviour of a simple model is more tractable because parameter estimates will be more stable after repeated data fittings than those of complex models (Myung, 2000). Hence, this indicates as a rule of thumb in practice, that the simple model is always preferred to ensure high generalizability, provided that there are similar quantified criteria from competing models.

4.1.3 Generalizability

The goal of model selection is to identify one model, from a set of competing models, which best captures the regularities underlying the cognitive process of interest. Thus, in order to measure a model's generalizability, the model selection method must be sensitive to the properties of the model in addition to considering GOF. We know that simplicity and parsimony of models can improve model generalizability because a complex model with many parameters tends to capture these false patterns more easily than a simple model with few parameters for noisy data.

There are many examples in the literature in which model generalizability is addressed. For example, through a few simulation studies, Pitt et al. (2003) found that model selection criteria that consider model generalizability are superior to those only based on the GOF method; Liu and Aitkin (2008) investigated model selection criteria which consider model generalizability, such as Bayes factor (Kass and Raftery, 1995; Raftery, 1999), BIC (Raftery, 1999; Schwarz, 1978; Burnham, 2004) and DIC (Spiegelhalter et al., 2002; Berg et al., 2004); Myung et al. (2009) especially explained why generalizability is the preferred criterion for model selection and pointed out that good generalizability is achieved by trading off GOF with model complexity. Figure 4-1 (Pitt and Myung, 2002; Myung et al., 2009) gives an excellent presentation of such a trade-off. That is, one way of estimating the generalizability of a model is by appropriately discounting the model's GOF relative to its complexity. More details can be found in Myung et al. (2000, 2009), Pitt et al. (2003), Liu and Aitkin (2008), and Yu and Meyer (2006).

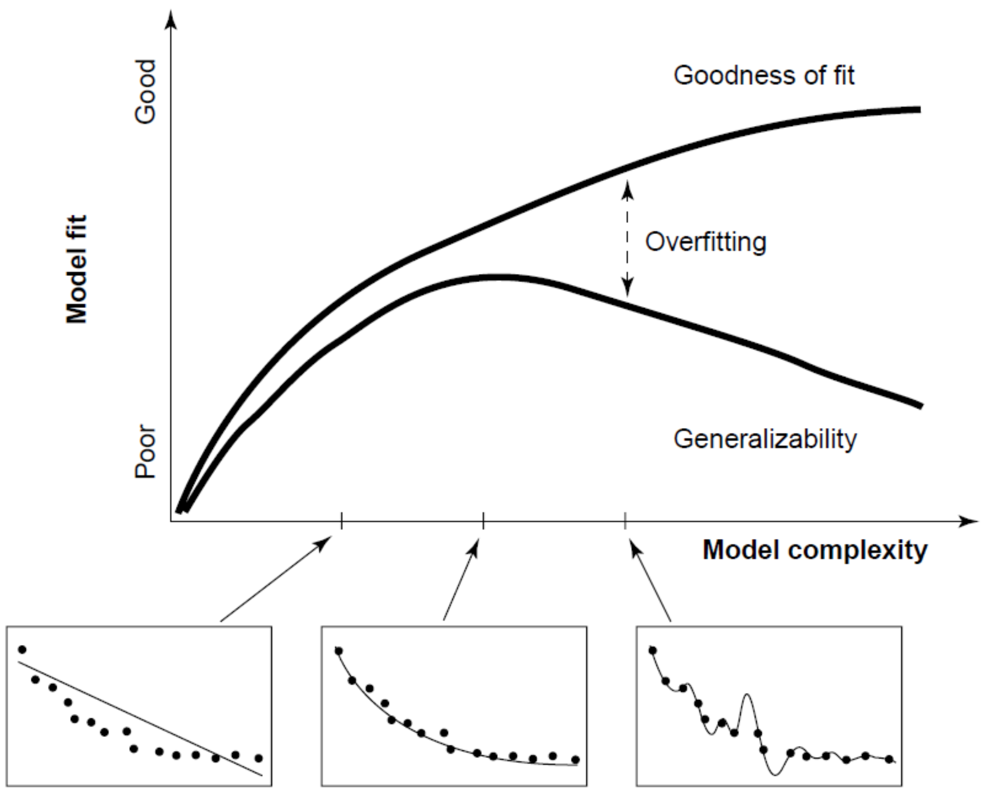


Figure 4-1 An Illustration of the Relationship between Goodness of Fit and Generalizability as a Function of Model Complexity

The *y* axis represents any fit index, where a larger value indicates a better fit (e.g., maximum likelihood). The three smaller graphs provide a concrete example of how fit improves as complexity increases. In the left graph, the model (line) is not complex enough to match the complexity of data (dots). The two are well matched in complexity in the middle graph, which is why this occurs at the peak of the generalizability function. In the right graph, the model is more complex than data, capturing micro variation due to random error.

4.2 MODEL SELECTION METHODS

Although each criterion mentioned above identifies a property of a model that can be evaluated on its own, in practice they are rarely independent of one another. Consideration of all three simultaneously is necessary to fully assess the adequacy of a model.

The core of model selection is that to avoid choosing unnecessarily complex models, a model should be selected based on its generalizability, rather than its GOF. Inference under models with too few parameters (variables) can be biased, while with models having too many parameters (variables), there may be poor precision or identification of effects that are, in fact, spurious. These considerations call for a balance between under- and overfitted models—the so-called model selection problem (Forster, 2000).

Model selection is realized by defining a selection criterion that makes an appropriate adjustment to its GOF by taking into account the contribution from model complexity (Myung, 2000). In a Bayesian framework, there are several different selection methods for choosing between competing models, such as Bayes factors, AIC (Akaike,1973; Bozdogan, 2000; Burnham and Anderson, 2002, 2004), BIC, DIC, marginal likelihood, etc. They differ from each other in terms of if and how such adjustments are made to best estimate a model's generalizability. Among them, DIC is the most popular in that it counts penalties for model overfitting and is readily available. (It is the default setting in software WinBUGs (Lunn et al., 2000; Spiegelhalter 2003; Cowles, 2004.) This is also the reason why DIC is the major criterion for model selection in this study, while other methods are used as alternative measures for comparison.

4.2.1 The Method of Maximum Likelihood

The maximum likelihood method is principally a method of parameter estimation, but extends straightforwardly to model selection. The objective is to choose the best of the best. That is, out of the maximum likelihood hypotheses in the competing models, the one that has the greatest likelihood or equivalently, the greatest log-likelihood is selected (Forster, 2000). This method, however, does not account for generalizability since it is in fact a typical model selection method based only on GOF. In fact, in the case of nested models, it can never favour anything less than the most complex of all the competing models. Thus, this method is rarely used to compare

Bayesian models. It is just reviewed here for completeness and studied in Chapters 5 and 7 for reference.

4.2.2 The Bayes Factor Method

A. Bayes factor

In Bayes' theorem, Bayesian inference is often described as a method which shows how belief is altered by data and the Bayes factor is the index through which the data "speak", as distinct from the purely subjective part of the equation (Goodman, 1999). The Bayes factor is actually a summary of the evidence provided by the data in favour of one scientific theory, represented by a Bayesian model, as opposed to another. It is a formal Bayesian model assessment method.

For example, as in Chapter 2, assume that a model class J can be specified for some observed data y by a likelihood function, denoted as $p(y|\theta)$ or $L(\theta|y)$, which gives the probability of observing y as a function of parameter θ. A Bayesian analysis begins with a prior density, $\pi(\theta)$ (the terms "density" and "distribution" are used interchangeably), which represents one's uncertainty about the true parameter before observing any data. Once the data are observed, Bayes' theorem can be applied to produce an updated posterior density, and $\pi(\theta|y)$, also written as $p(\theta|y)$ as follows:

$$\pi(\theta|y) = \frac{p(y|\theta)\pi(\theta)}{m(y)} = \frac{L(\theta|y)\pi(\theta)}{m(y)} \qquad (4-1)$$

where

$$m(y) = \int p(y|\theta)\pi(\theta)d\theta \qquad (4-2)$$

Similarly, we can apply Bayes' theorem to model selection by defining J competing models,$M_j, j = 1,2, \dots, J$, as having prior probabilities $p_j = P(M_j)$ of being a true model, where $\sum_{j=1}^{J} p_j = 1$. Letting θ_j be the parameter set in model j with prior $\pi(\theta_j)$, then, the posterior probabilities $P(M_j|y)$ of being a true model for model j after observing data y are:

$$\mathrm{P}(\mathrm{M_j}|\mathrm{y}) = \frac{\mathrm{P}(\mathrm{y}|\mathrm{M_j})\mathrm{P}(\mathrm{M_j})}{\sum_{\mathrm{j=1}}^{\mathrm{J}} \mathrm{P}(\mathrm{y}|\mathrm{M_j})\mathrm{P}(\mathrm{M_j})} = \frac{\mathrm{m_j}(\mathrm{y})\mathrm{P}(\mathrm{M_j})}{\sum_{\mathrm{j=1}}^{\mathrm{J}} \mathrm{m_j}(\mathrm{y})\mathrm{P}(\mathrm{M_j})} \qquad (4-3)$$

where

$\mathrm{P}(\mathrm{y}|\mathrm{M_j})$ = the marginal probability for model j, also denoted as $\mathrm{m_j}$ (y) in this thesis,

$\mathrm{P}(\mathrm{M_j}|\mathrm{y})$ = the posterior probability of being true model for model j, and

$\sum_{\mathrm{j=1}}^{\mathrm{J}} \mathrm{m_j}(\mathrm{y})\mathrm{P}(\mathrm{M_j})$ = averaged marginal density across competing models.

Plugging in $\mathrm{m}(\mathrm{y}) = \int \mathrm{p}(\mathrm{y}|\theta)\pi(\theta)\mathrm{d}\theta$, Equation 4-3 can be rewritten as

$$\mathrm{P}(\mathrm{M_j}|\mathrm{y}) == \frac{\int \mathrm{P}(\mathrm{y}|\theta_\mathrm{j})\pi(\theta_\mathrm{j})\mathrm{d}\theta_\mathrm{j}\, \mathrm{P}(\mathrm{M_j})}{\sum_{\mathrm{j=1}}^{\mathrm{J}} \mathrm{P}(\mathrm{M_j}) \int \mathrm{P}(\mathrm{y}|\theta_\mathrm{j})\pi(\theta_\mathrm{j})\mathrm{d}\theta_\mathrm{j}} \qquad (4-4)$$

Suppose there are two competing models i and j, then the posterior odds of being the true model is:

$$\frac{P(M_i|y)}{P(M_j|y)}$$

$$\frac{P(M_i|y)}{P(M_j|y)} = \frac{m_i(y)P(M_i)/\sum_{j=1}^{J} m_j(y)P(M_j)}{m_j(y)P(M_j)/\sum_{j=1}^{J} m_j(y)P(M_j)} = \frac{\mathrm{m}_i(y)}{\mathrm{m}_j(y)} \times \frac{P(M_i)}{P(M_j)} \qquad (4-5)$$

where

$\frac{\mathrm{m}_i(y)}{\mathrm{m}_j(y)} = \text{Bayes factor, denoted as } \mathrm{B}_{ij}$

$\frac{P(M_i)}{P(M_j)} = \text{prior odds}$

$\frac{P(M_i|y)}{P(M_j|y)} = \text{posterior odds}$

If one's belief in each competing model as being the true model is the same, that would suggest a non-informative prior, $\mathrm{P(M_i)} = \mathrm{P(M_j)}$. The posterior odds is equal to the Bayes factor in favour of model i, and this is why the Bayes factor is employed as a typical method for Bayesian model selection.

$$\frac{P(M_i|y)}{P(M_j|y)} = \frac{m_i(y)}{m_j(y)} \qquad (4-6)$$

In fact, Bayesian model comparison is a method of model selection based on Bayes factor (Burnham and Anderson, 2004). It has been shown that the Bayes factor method prefers a parsimonious model to a more complex one (Gelfand and Dey, 1994). Table 4-1 gives the reference values of the Bayes factor as well as that (in natural logarithm) for Bayesian model selection (Kass and Raftery, 1995).

Table 4-1 Reference values of Bayes factor for Bayesian model selection

$2\ln(\mathrm{B}_{ij})$	B_{ij}	Evidence against $\mathrm{H_0}$ ($\mathrm{H_0}$: model i is the true model)
0-2	1-3	Not worth more than a bare mention
2-6	3-20	Positive
6-10	20-150	Strong
>10	>150	Very strong

B. Marginal likelihood

From Equations 4-5 and 4-6, it can be seen that the core problem is how to calculate marginal likelihood (the integral of Equation 4-2) in order to obtain the Bayes factor. Marginal likelihood might be sometimes analytically available, for example, for exponential family distributions with conjugate priors (e.g., the PG model, where crashes follow the Poisson distribution and the Poisson mean follows a gamma distribution, implying that crashes marginally follow NB distribution as derived in Chapter 2). However, more often, computation in the models is intractable, requiring the implementation of MCMC numerical methods. The simplest apparent estimator of the marginal likelihood is the harmonic mean estimator.

To increase the efficiency of the model likelihood estimator, it is preferable to use samples from the posterior distribution. Newton and Raftery (1994) suggested the harmonic mean of the posterior sample likelihood as the estimator for marginal likelihood $m_j(y)$ under model j, that is, $m_j(y) = \int \mathrm{P}(\mathrm{y}|\theta_\mathrm{j})\pi(\theta_\mathrm{j})\mathrm{d}\theta_\mathrm{j}$ can be estimated as

$$m_j(y) = \frac{M}{\sum_{m=1}^{M} \frac{1}{P\left(y|\theta_j{}^{(m)}\right)}} \qquad (4-7)$$

where

$M =$ the total number of obtained MCMC iterations excluding burn in

$\mathrm{P}\left(\mathrm{y}|\theta_\mathrm{j}{}^{(\mathrm{m})}\right) =$ likelihood for draw m from a series of draws (draw1, draw 2, ..., draw M) from MCMC output for model j.

The harmonic mean can be easily obtained from the output of WinBUGs. However, it is worthwhile to mention that a possible problem with this approach is that it can be quite unstable because the inverse likelihood does not have finite variance (i.e., some of the likelihood terms in the sum might be near 0).

4.2.3 AIC

AIC is an abbreviation of the Akaike information criterion (Akaike, 1973 and Bozdogan, 2000). Akaike (1973) applied a correction of the estimation bias by penalizing extra parameters when the maximum likelihood estimations (MLEs) are used in estimating the expected log likelihood. AIC is aimed at solving the prediction problem, i.e., finding the model $\mathrm{M_j}$ that produces estimates of the density $\mathrm{P}(\hat{\theta}_\mathrm{j}|\mathrm{y})$ which is close, on average, to the true density. It was derived based on the information theory (Bozdogan, 2000; Burnham and Anderson, 2002, 2004) and it is one of the most popular penalizing approaches for Bayesian model selection, especially in econometrics. It adds a penalty factor (shown in Equation 4-8) that is proportional to the difference in the number of parameters between two models. Thus, the question in model selection, which is how much additional information a parameter must add to justify the "cost" of its inclusion, might be answered. The AIC can be written as follows:

$$\mathrm{AIC} = -2\log L(\hat{\theta}) + 2k \qquad (4-8)$$

$L(\hat{\theta})$ is the maximized likelihood function of the parameters in model, computed at a value θ that maximizes the probability of the data given the model; and k, which is the number of free parameters in the model, promotes model parsimony by penalizing models with increased model complexity (larger k).

A better model means a larger $L(\hat{\theta})$, which results in a smaller value of the AIC. Thus the model with a minimum AIC value is chosen as the best model to fit the data, that is, the model in the suite with the best overall statistical properties and parameter balance. Burnham and Anderson (2002) suggested that models with a difference in AIC < 2 are all plausible; values of 4-7 are considerably less so, while >10 means that the models are missing some important explanatory variables. Note that, per Akaike's rule of thumb, two models are essentially indistinguishable if the difference of their AICs is less than 2.

Note that the penalty term is relatively more important for small sample sizes, which increases the tendency to select simpler models. Since the penalty in the AIC does not increase with sample size, this method clearly favours larger models as the sample increases. This is especially so when the sample size increases (say, sample size N to infinity), in which case the AIC produces the same selection as the chi-square criterion. That is, similar to the chi-square method, it tends to favour overly complex models with large sample sizes (Busemeyer and Wang, 2000; Browne, 2000).

4.2.4 BIC

Closely related to the AIC method is Bayes information criterion (BIC) or the Schwarz criterion (Schwarz, 1978). This is another popular ranking model method which considers model generalizability. As AIC is very popular in econometrics, the BIC, on the other hand, is more popular in sociology (Weakliem, 2004).

BIC can be regarded as an approximation to the log Bayes factor. For well-behaved models and moderate to large sample sizes, BIC provides a useful approximation to the log Bayes factor (Wasserman, 2000).

$$BIC = -2logL(\hat{\theta}) + klog(n) \qquad (4-9)$$

where

$L(\hat{\theta}) =$ the maximized likelihood function,

$k =$ the number of free parameters in the model, and

$n =$ the sample size

It can be seen that AIC and BIC have the same form: −2 log L plus a penalty for each free parameter in the model. BIC differs from AIC only in the second term, which now depends on the sample size n. The penalty for each parameter is 2 for AIC, and log (n) for BIC. Clearly, as n increases, BIC favours simpler approximating models (that is, models with a smaller number of parameters k) than AIC.

Since BIC can be seen as a useful approximation to the log Bayes factor (Wasserman, 2000), the critical values of BIC differences for model comparison and selection can be taken from those for Bayes factors shown in Table 4-1.

Despite the superficial similarity between AIC and BIC, researchers believe that the latter is derived in a very different way and within a Bayesian framework (detailed derivation see Wasserman, 2000), while the former is based on the information theory (classical statistical methods) (Bozdogan, 2000). However, Burnham and Anderson (2004) proved that AIC can be justified as a Bayesian result by using a "savvy" prior on models, that is, a function of sample size and the number of model parameters, and BIC can be derived as a non-Bayesian result. More detailed discussion on the relationship between AIC and BIC can be found in Kuha (2004) Wasserman (2000), Weakliem (2004) and Burnham and Anderson (2004.

4.2.5 DIC

As noted, AIC and BIC methods trade off a measure of model adequacy, measured by the log-likelihood, against a measure of complexity, measured by the number of free parameters. Obviously, the calculation of AIC or BIC requires the specification of the number of free parameters. For a nonhierarchical Bayesian model with parameter θ, obtaining the number of free parameters is straightforward. However, for a complex hierarchical model, the specification of the dimensionality of the parameter space is rather arbitrary. (For details, see Yu and Meyer, 2006.)

DIC was introduced as a model selection method by Spiegelhalter et al.(2002). It soon became very popular because it can overcome the drawbacks of AIC and BIC. First, DIC is easily calculated from the samples generated by an MCMC simulation (indeed, DIC is automatically computed by WinBUGS 1.4). Second, there is no need for a number of free parameters. On the other hand, AIC and BIC require calculating the likelihood at its maximum with respect to θ, a result that is not readily available from an MCMC simulation, and the specification of a number of free parameters. Berg et al. (2004) extensively examined DIC through a simulation study and found that DIC clearly identifies the correct model out of eight different alternatives.

The deviance can be defined as $D(\theta) = -2\log\big(p(y|\theta)\big) + C$, where y are the data, θ are the unknown parameters of the model and $p(y|\theta)$ is the likelihood function. C is a constant that cancels out when compared with different models, and therefore, does not need to be known. The DIC of model j can be calculated as follows:

$$DIC_j = Dhat_j + 2pD_j = Dbar_j + pD_j = 2Dbar_j - Dhat_j \qquad (4-10)$$

where

$Dbar_j$ = the expected deviance for the jth model, $Dbar_j = E\big(D\big|y, \theta_j\big)$, given by the mean of the sampled deviances from MCMC simulations. This is a measure of how well the model fits the data in that a larger value means a worse fit,

$Dhat_j$ = the deviance at the posterior mean θ_j of the parameters for model j, and

pD_j = the effective number of parameters of the model, computed as the difference between $Dbar_j$ and $Dhat_j$, namely, $pD_j = Dbar_j - Dhat_j$. A larger value means more ease for the model to fit the data. This can be seen as a penalty term for increasing model complexity.

Models are penalized by the value of $Dbar_j$, which favors a good fit, but also (in common with AIC and BIC) by the effective number of parameters pD_j. Since $Dbar_j$ decreases as the number of parameters in a model increases, the pD_j term compensates for this effect by favoring models with a smaller number of parameters. Models with smaller DIC should be preferred to models with larger DIC. Differences of more than 10 in the value of the DIC might rule out the model with higher DIC values while differences between 5 and 10 are considered substantial. Attention should be paid to models when the differences are less than 5. In such cases, the expected deviance for the jth model ($Dbar_j$) combined with engineering judgment, might be used as a criterion and the model with a substantially lower $Dbar$ is favored.

From the definition of DIC, it can be seen that DIC is particularly suited to comparing Bayesian models when posterior distributions have been obtained using MCMC simulation. This can greatly reduce computation costs, especially for complicated hierarchical Bayesian models. It should be pointed out that because WinBUGS calculates DIC at the posterior mean, it requires the posterior mean to be a good estimate of the stochastic parameters. Therefore, it is important to check skewness and modality of the posterior distribution when using DIC. That is, it is only valid when the posterior distribution is approximately multivariate normal. Another issue is that DIC might not be adequate for missing data models. Celeux et al. (2006) extensively examined DIC for missing data models and found that DIC indeed favours complex models.

4.3 SUMMARY

The objective in model selection is to use a model that is as parsimonious as possible while ensuring that reliable results are obtained. A natural way to compare models is to use criterion based on a trade-off between the fit of the data to the model and the corresponding complexity of the model. Although model selection is very popular in other fields such as econometrics, sociology and psychology, as is evident from the cited references, there are few applications in road safety. In road safety analyses, most studies only rely on one model or functional form of expected crashes. Of course, that model might not be the best one, given the data. Thus the results might be biased. In our most recent published research (Lan et al., 2010), it can be seen that the estimations from competing models may indeed have large differences.

A few popular model selection criteria, such as AIC, BIC and DIC, are introduced in this chapter. Each of these methods has its own advantages and drawbacks, but DIC seems to have more advantages over other methods. Even so, it is suggested that all of these model selection criteria be calculated for model comparisons. If most of the criteria, if not all, favour the same model, one can be more confident that the final decision is not overly dependent on an assumed prior. If, however, the selected model is quite different for each criterion, one should select the best model based on experience and expertise. In that case DIC might be a very useful criterion.

It should be kept in mind that only the difference of the AIC, BIC, or DIC values between models is meaningful for model comparison. An individual AIC, BIC, or DIC value is meaningless because it varies with different data.

In addition, it should be emphasized that there is no way to find the true model in real cases, so an approximate will often need to suffice. Model specification is difficult because our knowledge about the phenomenon being modeled is rarely complete. That is, empirical data obtained from studying the phenomenon are limited, as they only provide partial information about its properties and the variables that influence them. With limited information, it is next to impossible to construct a “true” model.

Finally, some researchers, such as Burnham and Anderson (2004), Wasserman (2000) and Raftery (1999), argue for the advantages of model averaging over selecting a single model. However, model-averaged inference is not common, nor has there been much effort to evaluate it even in major publications on model selection. Model averaging might deserve more research, but it is not the objective of our study.

CHAPTER 5 EVALUATION OF THE FB METHOD FOR NETWORK RANKING

5.1 INTRODUCTION

This chapter begins with a literature review of actual applications of the FB method for hazardous site identification in road safety, and the FB ranking criteria that are used in these studies; this is followed by the the objectives, details and results of the evaluation study.

5.1.1 Literature Review

The first stage of road safety studies commonly involves a comparison of the decision parameters (i.e., Poisson mean or expected crashes) of the sites estimated from the accident numbers during some period for all sites. Then, potentially hazardous sites are determined, resulting in an ordered list. This list is constructed by ranking locations based on the ranking criteria from a promising method, e.g,, the posterior mean, which is similar to the EB method, the expected rank of the posterior distribution of decision parameters (Tunaru, 2002; Miaou and Song, 2005), or the probability that a site is the worst (Tunaru, 2002; Miaou and Song, 2005). The ranked sites are generally selected by working down the list until the allocated resources are exhausted for the detailed examination (i.e., the diagnosis and identification of potential treatments), and perhaps, for subsequent treatment of locations. Different list orderings may lead to different sets of locations being examined in detail. An inappropriate ordering of locations, therefore, could lead to a truly hazardous location not being examined and considered for treatment. Thus it is vital to properly select the method and ranking criteria for hot spot identification.

The first application of the FB method for road safety evaluation is probably the study of network ranking performed by Schluter et al. in 1997. They proposed a Bayesian hierarchical PG model to rank high risk sites for 35 intersection sites using criteria such as the posterior probability of selecting the worst site and the posterior mean of crashes.

Bossche et al. (2003) employed a Bayesian binomial hierarchical model to rank hazardous intersections for bicycles in a small university town in Belgium based only on traffic crash data. In other words, no covariates were included in their study. The authors used the posterior Poisson mean of crashes to rank the sites. They concluded that there is no such thing as "the" correct ranking because of the stochastic character of bicycle crashes and that, as a result, the estimated crashes are not deterministic. The difficulty with this study, first of all, is that the binomial hierarchical model used is questionable as it is now well known that Poisson hierarchical models are more favorable; secondly, their conclusion is open to question in that it is believed that correct ranking results do exist based on the evaluation of the performance, even if the crashes are not deterministic. This will be further discussed in this chapter.

Geurts et al. (2004) investigated the effects on identification and ranking of black spots based on four different weighting value combinations that correspond to light, serious and deadly injuries. The four weighting value combinations are: 1-1-1; 1-1-10; 1-3-5 and 1-10-10 for light, serious and deadly injuries, respectively. They concluded that weighting schemes greatly affect the ranking results, which is an obvious conclusion. Their ranking criterion is based on the posterior mean, again obtained from crash records only, which means that no covariates such as traffic volumes were considered in the ranking method.

Miranda-Moreno and Fu (2007) explored the differences of EB and FB through a simulation study. They used a PG model to generate random samples, then applied the EB method and FB PG model to calculate the posterior mean to rank the sites. They found that the FB estimators performed better than the EB estimators when working with data sets that have a small number of sites (observations) and which are characterized by an overall low mean accident frequency. Furthermore, when the data set is sufficiently large (e.g. over 300 sites), these two approaches yield practically the same results. However, it should be noted, that the FB and EB methods used indeed follow the same distribution (PG distribution) in their study, while this may not be the case for real data. Intuitively, the ranking results from their approach should be very close to that from the EB method. For a small sample case, the estimated SPFs for EB are not reliable resulting in errors in the estimation of expected crashes, while the FB method can carry that uncertainty to the final estimation. This is why, in principle, the FB estimator is better than the

EB estimator for a small sample case. For a large data set, SPFs are no longer a problem for the EB method, and the results are logically very close to that of FB. Another issue with Miranda-Moreno and Fu (2007) is that they only use one year of simulated data to rank sites, and thus time series correlation could not be addressed in their study.

Recently, Huang et al. (2009) conducted an evaluation of the FB method for hot spot ranking using crash records of 582 four-legged signalized intersections from 1997 to 2006 in Singapore. Three measurements, observed crashes, expected crashes from EB, and the Poisson mean from FB obtained from the last 3 years (2004 to 2006) of crash data were used to rank hot spots. Then, the average of the observed crash counts in the whole time period (1997 to 2006) is treated as a true mean to evaluate the FB method as well as the EB method. Sensitivity and specificity are used as evaluation criteria in their study. They concluded that the selected FB hierarchical models outperform the standard EB approach in correctly identifying hot spots. The major problem is probably with the true mean estimate, intuitively, since even ten years of crash count may be subject to regression to the mean.

The above mentioned studies employ a univariate approach for network ranking studies. For the multivariate count data, where different severity crash data (i.e., fatal, injured and PDO), which are potentially correlated, are used to rank hazardous sites, it is necessary to conduct ranking using a multivariate FB approach.

Tunaru (2002) employed a bivariate PLN model for black spot identification using two types of severity data: fatal or seriously injured crash records, and slightly injured crash data. Two ranking criteria, the probability that a site is the worst and the median of posterior distributions based on the Poisson mean, were used to rank sites in his study.

Miaou and Song (2005) employed a three-variate PLN model to develop a crash rate model for black spot identification using a two-lane rural Texas low-volume road data set, which includes 3 different severities of crash counts: fatal, incapacitating injury and non-incapacitating injury. They used the posterior mean and the posterior expected rank of crash rate and crash rate cost for network ranking.

Brijs et al. (2006, 2007) developed a three-variate Poisson distribution model for black spot identification using three crash severity types(no covariates are involved in the study): fatal, seriously injured and slightly injured. Due to the limitations of MVP models as mentioned in Chapter 2, the covariance (or correlation coefficient) in the MVP setting is always positive, and, furthermore, the MVP model cannot model overdispersed crash data.

From this literature review, it can be concluded work remains to be done on the evaluation of the FB method for hot spot identification, especially since the ranking criteria involved in previous studies were very limited and little attention has been paid to model selection. In addition only one function form of the expected crashes (Poisson mean) has been explored in these studies. The following section provides a brief introduction of the ranking criteria adopted in the reviewed studies.

5.1.2 FB Ranking Criteria

A. Posterior mean of decision parameter

The posterior mean of a decision parameter can be the Poisson mean after obtaining the data, or some other measures based on the Poisson mean. The Poisson mean of crashes is perhaps the most popular ranking criteria in the safety literature in that it is convenient to estimate during the model development procedure; indeed, it was used in almost all of the above applications for hazardous site identification. It has conceptually the same meaning the expected crashes from the EB method. It is the expected value of λ_i taken over the posterior distribution of λ_i given all data y, i.e., $p(\lambda_i|y)$. It can be calculated by:

$$E(\lambda_i|y) = \int \lambda_i \, p(\lambda_i|y) d\lambda_i \qquad (5-1)$$

As can be seen, the posterior mean is a point estimate of the mean number of accidents over a long time. Obviously, it does not take advantage of the full distribution of λ_i. Even so, this criterion continues to be popular because it is easy to obtain and is clearly understood.

B. Posterior expected rank of the decision parameter

The posterior expected rank of the decision parameter is based on the ranks of the mean parameter λ_i, which are the site specific parameters. The ranks $R_{\lambda_1}, \ldots, R_{\lambda_i}, \ldots, R_{\lambda_N}$ can be obtained by (Shen and Louis, 1998):

$$R_{\lambda_i} = \mathrm{rank}(\lambda_i) = \sum_{j=1}^{N} I(\lambda_i \geq \lambda_j) \qquad (5-2)$$

where

$N =$ the total number of sites to be ranked,

λ_i = Poisson mean of crashes for site i,

λ_j = Poisson mean of crashes for site j, and

I = the indicator function. $I(\lambda_i > \lambda_j) = 1$ if $\lambda_i > \lambda_j$, otherwise, $I(\lambda_i > \lambda_j) = 0$. It can be seen that the smallest has rank 1 and the greatest ranks correspond to the most hazardous site.

Based on this posterior distribution, the expected value of the true rank order of λ_i, indicated as $\overline{R}_{\lambda_i}$, can be obtained using the posterior distribution of R_{λ_i} given all data y. The posterior distribution of R_{λ_i} is denoted as $p(R_{\lambda_i}|y)$.

The posterior expectation of ranks has been widely recommended as a ranking criterion and is defined as:

$$\overline{R}_{\lambda_i} = E\left(R_{\lambda_i}|y\right) = \sum R_{\lambda_i}\, p(R_{\lambda_i}|y) \qquad (5-3)$$

This method is technically sound and the measure can be obtained from the posterior distributions of λ_i through the analysis of each iteration. For example, in WinBUGs, R_{λ_i} can be obtained through the analysis of an MCMC output. Shen and Louis (1998), and Miao and Song (2005) employed this criterion in their studies while Tunaru (2002) employed a median rank, which represents the expected rank, for easy of availability (i.e., it can be directly obtained in WinBUGs). Tunaru (2002), and Miao and Song (2005) mentioned that this criterion is optimal if the ranks of λ_i are of interest, whilst the posterior means $E(\lambda_i)$ are optimal estimates when the aim is to produce inference about λ_i. Furthermore, Laird and Louis (1989) employed the Gaussian model with some assumptions in their study and concluded that the posterior means can perform poorly. It is worthwhile to mention that no MCMC method was employed in their study. Since the MCMC method was not employed in the study and because of the assumptions, it is of interest to re-examine their conclusion using current MCMC methods. This, however, might be difficult for large samples, where the iterations of MCMC cannot be saved for an output analysis due to the limited available computer memory. With technological improvement in computers, this should not be an issue in the future. Alternatively, the posterior median can be used and is readily available in WinBUGs output.

C. **The probability that the site is the worst among all sites considered in terms of the decision parameter (P_{worst})**

The probability that the site is the worst among all sites considered in terms of the decision parameter (P_{worst}) represents the posterior probability that site i has the largest decision parameter value (i.e., Poisson mean) than any other site, given all data y. It can be expressed as:

$$P_{worst} = p\left(\lambda_i > \lambda_{j,} \text{ for all } j \neq i \middle| y\right) \qquad (5-4)$$

The procedure to calculate this criterion is similar to the posterior expected rank. The posterior distributions of λ_i is used to obtain P_{worst}, which can be calculated through each iteration of the MCMC process. Schluter et al. (1997), Brijs et al. (2006) and Tunaru (2002) provide detailed information on this criterion.

Currently, the above three criteria seem to be the most common for network ranking. There are also some less popular ranking criteria, such as the predictive distribution of accident frequency (Schluter et al., 1997), which will not be introduced here.

5.1.3 Objective of the Evaluation Study

Researchers have explored a few ranking criteria for hazardous site identification using the FB method. However, there is no evaluation study with regards to the ranking criteria and little research has been conducted on the evaluation of the ranking criteria themselves, and on the performance of the FB method and its variations, including comparisons of univariate EB versus FB, and multivariate FB versus univariate FB. The objective of this aspect of the research was to fill this void by conducting a thorough evaluation of the FB method for black spot identification. To completely evaluate the FB method, crash data on a single severity level were used for a univariate FB study for comparison with the EB method. Specifically, two data sets extracted from the same data with different data history were employed (one uses 6 years of data while another one uses 3 years of data) for the evaluation, including an investigation of the sensitivity of the ranking criteria. For the 6 years of data, 11 FB models are developed and compared, and the best model is used for the evaluation study. Then, an evaluation study is performed by using multilevel severity data where 5 levels of severity data are used for hot spot identification. Various ranking and evaluation criteria are proposed and employed for the study, details of which will be introduced in the next section.

The remainder of the chapter is structured as follows. Section 2 briefly introduces the data for this study. The employed ranking criteria are described in Section 3 while the evaluation criteria are explained in Section 4. Section 5 presents the approach and procedure for this evaluation study. An evaluation of the FB method with single severity data and a comparison with the EB method are illustrated in Section 6. Section 7 presents the evaluation study results for the multi level severity data. Finally, a brief summary can be found in Section 8.

5.2 DATA DESCRIPTION

The Highway Safety Information System (HSIS) provided all data used in this study. Geometry, traffic volume and crash data were acquired from the state of California (1993-2002) for 726 stop-controlled 4 legged intersections with 2 lanes on major roads that were selected to conduct this study. The last four years of data (1999-2002) were used to evaluate the ranking results from the FB and EB methods, while the preceding 3 years (1996-1998) and preceding 6 years were employed to rank the sites for the single level severity data (total crashes each year), respectively. Data composed of 5 severity levels of crash data were employed to conduct an FB analysis using both univariate and multivariate FB methods to determine if multivariate FB is superior to univariate FB. Detailed information will be presented in the latter part of this chapter.

Dataset 1: 726 stop-controlled 4 legged intersections with 2 lanes on major roads

In order to see the difference of the ranking results from the FB and EB methods, an identical data set was used to conduct FB and EB analyses. The severity data on a single level for this study is the total crashes each year. Various ranking and evaluation criteria were explored and used to conduct the evaluation of the FB method for comparison with the EB method, and for the sensitivity analysis of the ranking criteria.

Dataset 2: 436 intersections with high crash counts

It was of interest to investigate whether the multivariate FB method is superior to the univariate FB method for network ranking using multi levels of severity. Five crash severity levels were used, including Sev1: fatal (K), Sev2: incapacitating-injury (A), Sev3: non-incapacitating injury (B), Sev4: minor injury (C), and Sev5: PDO. Since the effect or cost of each severity of crash should be quite different for black spot identification, arbitrary weights were applied to these 5 crash severity levels: 5 to fatal, 4 to injury, 3 to non-incapacitating injury, 2 to minor injury and 1 to PDO, respectively. That is,

$$\text{weighted total crashes} = 5\mathrm{K} + 4\mathrm{A} + 3\mathrm{B} + 2\mathrm{C} + \mathrm{D} \qquad (5-5)$$

For example, if there is one crash for each of the 5 levels of severity, the combined weighted total crashes would be 15. Since the multivariate FB model is computationally demanding, as a result, the MCMC procedure is very slow when using WinBUGs. In addition, the computer memory is usually not enough for a large data set (i.e., there is an inadequate number of iterations obtained before WinBUGs freezes for the 726 sites) to obtain the MCMC output. To solve this problem, the weighted total crashes for each site in the first 6 years were first calculated. Then, the 436 sites where the weighted total crashes are greater than or equal to 8 were finally selected for further ranking studies.

Table 5-1 provides the summary information of these datasets.

Table 5-1 Summary Information of the Datasets

	Univariate approach: EB vs. FB	FB method: Multivariate FB vs univariate FB
Number of sites	726 sites	436 sites (weighted total crashes ≥ 8 in year 93-98)
Crash types	Total crashes	fatal (K), incapacitating-injury (A), non-incapacitating injury (B), minor injury (C), property damage only (PDO)
Ranking	Dataset 1:Year 1993-Year 1998 Dataset 2: year 1996-Year 1998	Year 1993-Year 1998
Evaluation	Year 1999-Year 2002 Posterior Poisson Mean was estimated by the model developed using 10 years' data	Year 1999-Year 2002 Posterior Poisson Mean was estimated by the model developed using 10 years' data

5.3 FB RANKING CRITERIA

The posterior Poisson mean (PM) of crash frequency and posterior mean of PSI were used as major ranking criteria for the comparison study of the FB and EB methods for hot spot identification in that these two criteria are available for both the FB and EB methods. For reference purposes, the raw crash count was also used as a ranking criterion. For the FB method, a sensitivity analysis of ranking criteria was exclusively conducted using the following eight ranking criteria: posterior expected, posterior median and posterior mode ranks of the posterior distribution of the Poisson mean, the probability that the site is the worst among all sites considered in terms of the Poisson mean (P_{worst}), PM, potential for safety improvement (PSI), and for reference purposes, observed crash counts and pseudo potential for safety improvement (PPSI) were also used as ranking criteria.

The evaluation of the ranking criteria is very important because a different ranking list can be obtained based on different ranking criteria even if the method is the same. For example, it has been shown that the ranked results would be different based on ranking criterion PM or PSI using the EB method (Elvik, 2008a). One of the objectives of this evaluation study is to provide an overview of the ranking results in terms of differences and similarities from all the ranking criteria, and identify promising ranking criteria. Details of each criterion are presented below.

A. Posterior Poisson mean of crash frequency (PM)

Based on the popularity of the posterior Poisson mean for ranking, and recognizing the fact that the number of crashes is not linear to traffic volume as is assumed in using crash rates as a ranking criterion, PM was selected as the major ranking criterion in this study.

B. Potential safety improvement (PSI)

Because of its availability for the EB method, PSI is also selected as one of the major criteria for exploration.

$$PSI_{\ i} = \lambda_i - \mu_i \qquad (5-6)$$

where

λ_i is the posterior Poisson mean of crash frequency, and

μ_i is the expected crashes at similar sites.

C. Posterior expected rank of the Poisson mean (expected rank)

The posterior expected rank of the Poisson mean was used to conduct a sensitivity analysis of different ranking criteria. Another purpose for using this ranking criterion is to conduct a comparison with the results from the PM. Intuitively, the results from the PM and expected rank might be the same, or at least very close, depending on the procedure used to obtain the expected rank. There are two methods to obtain the expected rank. i.e., if two chains with 8000 iterations are used to estimate these criteria, one method is to average the Poisson mean of two chains at each iteration to obtain a new chain with 8000 iterations; then the expected rank is calculated from the new chain (equivalent to one chain with average values of two chains for 8000 iterations). Another method is to combine two chains together to get one chain with 16000 iterations, and then the expected rank is calculated based on this combined chain. To maximally take advantage of the posterior MCMC output, the second method is used to obtain posterior expected rank, $\mathrm{P_{worst}}$ and mode rank (presented below) in our study. It should be noted that the difference between these two methods should be minor.

D. The probability that the site is the worst among all sites considered based on the PM ($\mathbf{P_{worst}}$)

The Poisson mean of crashes is used to calculate the probability of a site being the highest ranked hot spot. In other words, from a Bayesian MCMC output, the number of times for which each site has the largest PM is calculated. A larger number suggests more safety problems. Then, this number can be sorted from largest to smallest to rank black spots.

E. Median rank of the posterior distribution of the Poisson mean (median rank)

The median rank of the Poisson mean is used as a ranking criterion to compare the results with other ranking criteria, especially with the posterior expected rank. If the ranking results from the median rank are very close to those obtained by the expected rank, then the median rank can be used as a substitute for the expected rank of the posterior Poisson mean since it is readily available in WinBUGs output.

F. Mode rank of the posterior distribution of the Poisson mean (mode rank)

This ranking criterion is proposed since, intuitively, the mode rank of the posterior distribution of the Poisson mean is conceptually solid. It, thus, might be a better ranking criterion than median rank and expected rank. Mode rank means that each site is ranked by the most frequently occurring rank in the posterior distribution of Poisson mean. Similarly, it can be obtained by an analysis of data from the MCMC output, where a site can have a different rank order for each iteration.

G. Observed crash counts

This measure is only used for reference comparison. Due to the random variation of this measure, it usually provides biased results if sites with a high counts are identified as black spot.

H. Pseudo Potential safety improvement (PPSI)

Conceptually, PPSI is the same as PSI, but it is calculated as the difference between the observed crash counts (rather than the PM) and the expected crashes at similar sites. Similarly, this measure could provide biased results due to the RTM problem.

$$PPSI_{\ i} = \text{observed crash counts at site } i - \mu_i \qquad (5-7)$$

5.4 FB EVALUATION CRITERIA

The objective of this study is not only to use the FB method for network ranking, but more importantly, to evaluate the performance of the FB method for network ranking. The data in the first period are used to produce a ranked list of the hot spots, while data in the succeeding period are used to rank another list of the unsafe sites. Then, the results from these two ranked lists are compared and evaluated using various evaluation criteria. Properly selected evaluation criteria are vital for this evaluation study. The following three measures described in Chapter 1 are used to evaluate the performance of the FB method.

- **Criterion 1: sum of observed crashes in the succeeding time period**

Given the short period for evaluation, normally a few years, the RTM problem is a big issue which cannot be accounted. This criterion is, as mentioned, used for reference purposes.

- **Criterion 2: sum of differences between observed crashes and predicted crashes at similar sites in the next period (sum of the PPSI)**

This is the same as the ranking criteria PPSI. It is used for reference purposes, since there is no control for randomness in accident counts using this criterion.

- **Criterion 3: sensitivity and specificity**

This is the most popular evaluation criteria, especially in epidemiology. From Equations 1-9-a and 1-9-b, it can be seen that sensitivity is used to evaluate the ability of a specific method to correctly identify true hazardous sites, whilst specificity is used to measure the capability to identify safe sites. Ideally, a good ranking method should perform well in relation to both sensitivity and specificity. This means that it identifies as many of the truly hazardous sites as possible (sensitivity), while at the same time, not identifying a large number of sites that are truly not hazardous (specificity). Unfortunately, a trade-off must be made between sensitivity and specificity. Higher sensitivity means lower specificity, and vice versa. However, for the top ranked limited sites, specificity normally is large and so is not an issue. The critical measure is sensitivity for the top ranked sites as is shown later.

It should be noted that the assumption associated with these two evaluation criteria is homogeneity. In the two closed time periods, it is assumed that the road sections or sites are in the same or similar underlying operational states over these two time periods (similar traffic volume, driver population, pavement conditions, weather fluctuations, driving environment, traffic controls, etc.), and their expected safety performance remains virtually unaltered. Under this homogeneity assumption, a good ranking method will perfectly identify the same set of hot spots across two periods. However, in reality, homogeneity cannot be completely met. If the assumption is violated, these evaluation criteria should be used with caution.

Recognizing that these criteria fail to differentiate the disparity of the posterior Poisson mean or other decision parameters associated with false identification, we propose two other evaluation criteria: sum of the posterior Poisson mean and sum of the PSI in the second period, described as follows.

- **Criterion 4: sum of the posterior Poisson mean in the subsequent period**

The posterior Poisson mean is an estimate of the expected true mean after obtaining the data in the second period. This measurement can properly address the RTM problem and is a convenient criterion for ranking evaluation. Obviously, a larger sum of the PM in the second time period suggests a better method.

- **Criterion 5: sum of the PSI**

Similar to the ranking criterion PSI, the sum of the PSI in the second period is used as an evaluation criterion.

In all, it can be found that higher values of the above five evaluation criteria indicate a better ranking method. Eight ranking criteria and five evaluation criteria are explored in this study. However, it should be noted that the major criteria for ranking are PM, PSI and expected rank, mode rank, median rank and P_{worst}, while the key ones for evaluation are sensitivity and specificity, sum of the PM, and sum of the PSI in the succeeding period. It should be noted that the sum of the PSI and PM may not be used as evaluation criteria for the comparison of the FB and EB methods in that the estimates of these two measures may not be comparable due to the different modeling structure. A further investigation of these two criteria should be conducted in terms of the values in the evaluation period. Other criteria, including the sum of observed crashes, sum of differences between observed crashes, and predicted crashes for similar sites, are only used for reference.

5.5 THE APPROACH FOR NETWORK RANKING AND EVALUATION

This section presents detailed information with regard to the ranking criteria as well as evaluation criteria involved in this study.

For single level crash severity data cases, where a univariate approach was applied, 726 unsignalized four legged intersections were used to conduct this study. The ranking and evaluation results from the FB are compared with those from the current prevailing EB method by using two ranking criteria, PM and PSI, and various evaluation criteria as mentioned earlier. Moreover, the above mentioned eight ranking criteria are exclusively explored for the FB method. Furthermore, the sensitivity of the data history is studied for both the FB and EB methods. In other words, a sensitivity analysis with different periods of data history for ranking is conducted. To this end, 3 years of data (1996-1998) and six years of data history (1993-1998) are used to identify the most hazardous sites, respectively, while the second period (1999-2002) is used to evaluate the ranking results identified by both ranking periods using different methods and ranking criteria.

Generally, there are two ways to obtain the estimates of the expected crashes in the second period. One uses the second time period only (1999-2002) to develop FB models (or SPFs for the EB method) to estimate the true mean while the other one takes advantage of the whole time period data (1993-2002) to develop FB models. It is reasonable to believe that FB models developed using ten years of data can provide a better estimate of the true mean for 1999-2002. Thus, the second method to estimate the true mean was adopted.

Data for the two time frames described as follows were used to conduct a comprehensive evaluation study including a comparison with the EB method, a time sensitivity study and a ranking criteria sensitivity study.

Time Frame 6-4-10:

Six years of data (1993-1998) are used to develop the FB models (or calibrate the SPF for the EB method) and identify the hotspots. The years 1993-2002 are employed to develop the models (or calibrate the SPF for the EB method); then the developed models (or calibrated SPF) are used to estimate the true mean of the crashes in 1999-2002 for evaluation.

Time Frame 3-4-10: in this case, only three years of data (1996-1998) are used to rank sites.

For multi-level severity data, several crash severity types, such as fatal, injury and PDO, are used to identify the black spots. In such cases, a multivariate approach and/or a univariate approach can be performed on these data. In this study, both multivariate FB and univariate models were developed. The results are compared and evaluated based on different ranking criteria and evaluation criteria. One of the objectives of this study was to identify if there is an advantage in the multivariate approach (in our study, MVPLN AR (1) model) over the univariate Poisson AR (1) model. Since the safety effect of different severity crashes is quite different, five weights were given to the five severity levels of crash data as previously explained. Finally, 436 unsignalized intersections with weighted combined high crash counts were extracted from the sites used for a single level severity study.

5.6 FULL BAYESIAN METHOD WITH SINGLE SEVERITY DATA

The total crash count each year from 1993 to 2002 at each of the 726 unsignalized California four legged intersections was used to conduct this study. First, comprehensive FB model development was done for Time Frame 6-4-10. That is, six years of data (1993-1998) were used for ranking, and four years of data (1999-2002) combined with 10 years of data (1993-2002) for evaluation in that the true mean in the evaluation period was estimated from the model developed using ten years of data. Then, the best model was selected to conduct a comparison study with the EB method. The selected model was applied to Time Frame 3-4-10 and the multi-level severity data. More information is provided below.

5.6.1 Bayesian Model Framework

In road safety studies, it is normally assumed that the observed crash count $Y_{i,t}$ at site i in year t follows a Poisson distribution: $Y_{i,t} \sim \text{Pois}(\lambda_{i,t})$. The expected crash at intersection i in year t is $\lambda_{i,t}$, where $\lambda_{i,t} = \mu_{i,t} e^{\varepsilon_i}$. The term $\exp^{\varepsilon_i}$ represents random effects, which account for latent variables across the sites while $\mu_{i,t}$ is the expected crashes at similar sites. Normally, the basic form of the regression term $\mu_{i,t}$ is of the same form as SPFs used in EB studies. In this study, the basic form of FB models is a product form:

$$Y_{i,t} \sim \text{Pois}(\lambda_{i,t})$$

$$\lambda_{i,t} = \mu_{i,t} e^{\varepsilon_i}$$

$$\mu_{i,t} = e^{\beta_0} ML_{i,t}^{\beta_1} XST_{i,t}^{\beta_2} \qquad (5-8)$$

where

$\lambda_{i,t}$ = the expected crashes at location i in year *t*,

$\mu_{i,t}$ = the expected crashes at the sites similar to site *i* in year *t*,

$ML_{i,t}$ =AADT on the major road at intersection i in year *t*,

$XST_{i,t}$ =AADT on the minor road at intersection i in year t,

$\beta_0, \beta_1, \beta_2$ = fixed coefficients, and

e^{ε_i} = random effect at site i.

It can be seen that $\mu_{i,t}$ is the same form as that used in EB studies for intersections (Sayed and Rodrigez, 1999; Turner-Fairbank Highway Research Center, 1999; Persaud and Nguyen, 1998; Persaud et al., 2002). In this study, the form of $\mu_{i,t}$ has some variations, as described in Chapter 3.

Basically, there are three categories of FB models based on the different formats of random effects $\exp^{\varepsilon_i}$: PG models where $e^{\varepsilon_i} \sim \text{gamma}(\varphi, 1/\varphi)$; PLN models where $\exp^{\varepsilon_i} \sim \text{log normal}$ $(0, \sigma^2)$ and Poisson AR models where random effects have an AR format. In this research, 11 Bayesian models, including 4 PG, 4 PLN, and 3 Poisson AR(1) models, were developed using six years of data (1993-2002) as described below.

A. Poisson Gamma models

The random effect is $e^{\varepsilon_i} \sim \text{gamma}(\varphi, 1/\varphi)$ and $\varphi \sim \text{gamma}(0.01, 0.01)$. There are four PG models based on the different forms of $\mu_{i,t}$ in this study.

- Model 1: Original PG model

$\mu_{i,t}$ has the same form as Equation 5-8. This is the basic form of the PG model and does not account for time effects. This model is defined as PG_6yrs for data in 1993-1998 and PG_10yrs for 10 years' data.

- Model 2: Time multiplier PG model

Similarly, this model is denoted as PG_M_6yrs and PG_M_10yrs hereafter. $\mu_{i,t}$ is the same as Equation 5-8. However, instead of a fixed β_0, a time varying coefficient β_t, which is used to account for temporal variations in crash occurrences, is introduced into this model. It is similar in principle to the time multiplier in an EB study. It has been demonstrated that this model is promising (Persaud et al., 2010).

- Model 3: Time trend PG model

$$\mu_{i,t} = e^{\beta_0} ML_{i,t}^{\beta_1} XST_{i,t}^{\beta_2}\ e^{\beta_3 \times t} \qquad (5-9)$$

A potential time trend $\beta_3 \times t$ in the observed crash series is included in this model as an alternative way to deal with temporal variation. This model is described as PG_T_6yrs and PG_T_10yrs respectively in the study for 6 years (1993-1998) ranking data and the data (1993-2002) used to estimate the true mean for evaluation.

- Model 4: Time varying coefficient PG model

Similar to Model 2, aside from the time varying multiplier $\beta_{0,t}$, the other coefficients are also relaxed to be all time varying $\beta_{1,t}$ and $\beta_{2,t}$. It is expected that this model could be comparable to Model 2. For convenience, this model is called PG_VC_6yrs and PG_VC_10yrs, respectively.

B. Poisson Log Normal models

For the PLN models, the random effects are $\varepsilon_i \sim \mathrm{N}(0, \sigma^2)$ and $\sigma^2 \sim$ Inverse Gamma $(0.001, 0.001)$, such that the prior expected value of the precision parameter $1/\sigma^2$ is 1, and the prior variance of the precision is set at 1000. Accordingly, there are four PLN models in our study, based on the four different forms of $\mu_{i,t}$ which are the same as those for the PG models.

The four PLN models are:

- Model 5: Regular PLN model

(Denoted as PLN_6yrs and PLN_10yrs, respectively)

- Model 6: Time multiplier PLN model

(PLN_M_6yrs and PLN_M_10yrs)

- Model 7: Time trend PLN model

(Defined as PLN_T_6yrs and PLN_T_10yrs)

- Model 8: Time varying coefficients PLN model

(PLN_VC_6yrs and PLN_VC_10yrs, respectively)

C. Poisson AR (1) models

The last category in this study is the Poisson AR model, where the random effect ε_i is the AR model of the order 1. As mentioned in Chapter 3, AR (1) random errors can be employed to count time dependence in road safety studies, meaning that values of random effects at time t depends upon their immediate predecessor. A model which allows for AR (1) dependence in random errors might be an ideal model to reproduce the dynamic features of time series crash data.

Two AR (1) forms of ε_i are explored: one has a stationary form and the other does not. In addition, a time trend Poisson stationary AR (1) is also developed to see if there is a broad scale trend in the data. Details are as follows.

- Model 9: Stationary Poisson AR (1) model (denoted as P_AR(1)_6yrs and P_AR(1)_10yrs respectively)

The AR (1) model can be written as

$$Y_{i,t} \sim \text{Pois}\left(\mu_{i,t} e^{\varepsilon_{i,t}}\right) \qquad i = 1, 2, \ldots, n \qquad (5-10)$$

and

$$\varepsilon_{i,t} = r\varepsilon_{i,t-1} + u_{i,t} \qquad t \geq 2 \qquad (5-11)$$

where

$\mu_{i,t}$ is the same as Equation 5-8,

$u_{i,t}$ = unstructured white noise, $u_{i,t} \sim N(0, \sigma^2)$, $t \geq 1$, and

r = correlation coefficient of random effects between successive time periods.

To ensure a stationary process, $r \in (-1,1)$ is confined by setting $r = 2\rho - 1$ and $\rho \sim \text{Beta}\ (1,1)$ (Beta (1,1) is the same as Uniform [0,1]) and the first point of $\omega_{i,t}$ is simply set as $\omega_{i,1} \sim N(0, \sigma^2/(1 - r^2)$ in the study.

- Model 10: Non-Stationary Poisson AR (1) Model

This model is the same as Equations 5-10 and 5-11, but the correlation coefficient r does not necessarily belong to $(-1,1)$ and can be set: $r \sim N(0,1)$. This model was used for comparison with Model 9. It was expected that the results from these two models should be very close. This model is denoted as P_NAR(1)_6yrs and P_NAR(1)_10yrs for convenience.

- Model 11: Stationary Poisson AR (1) trend model

As mentioned in Chapter 3, this model can be seen as an alternative for the non-stationary process by adding a trend variable for transformation into a stationary process. Alternatively, it can be considered that a trend analysis only accounts for a broad scale time series pattern for a long period of time while AR (1) in the random effects explains fine-scale autocorrelation between successive time periods.

The model form is the same as Model 9, but the expected crashes at similar sites $\mu_{i,t}$ takes the form of Equation 5-9. Similarly, the model is abbreviated as P_AR (1)_T_6yrs and P_AR (1)_T_10yrs hereafter.

It should be noted that the Poisson mean usually transforms to its logarithm in the modeling process in road safety studies. The stationary process of AR (1) should be met, thus it is anticipated that the AR (1) model with a stationary assumption of random errors would be better.

All of the prior distributions for all coefficients $(\alpha, \beta_1, \beta_2, \beta_3)$ are assumed to be non-informative N(0,1000) to reflect the lack of precise knowledge of the coefficient values. The posterior distributions were calibrated by MCMC methods using all the data from 1993 to 1998, and 1993 to 2002, respectively.

5.6.2 Bayesian Model Selection

The years 1993-1998 and 1993- 2002 were used to develop the above 11 FB models for ranking and an evaluation analysis, respectively. The dependent variable is the total crashes each year and the independent variables are annual average daily traffic volume (AADT) on major and minor roads. The summary of the data is tabulated in Table 5-2. It can be seen that the AADT on major roads increases by a small amount in the evaluation method. Since traffic volume increases systematically for almost all intersections across years, it should not be a problem to use sensitivity and specificity as evaluation criteria. However, the sum of the PM might be a better evaluation criterion since a few intersections did not follow this pattern.

Table 5-2 Summary data for 726 California Unsignalized Intersections

year 1993- year 1998				
	Mean	Standard Deviation	Maximum	Minimum
Total Crashes /site. year	1.34	1.77	13	0
$AADT_{Major}$	8209	4239	29732	2917
$AADT_{Minor}$	652	860	7800	100
year 1993 - year 2002				
	Mean	Standard Deviation	Maximum	Minimum
Total Crashes /site. year	1.40	1.89	18	0
$AADT_{Major}$	8526	4459	29732	2900
$AADT_{Minor}$	653	863	7800	100

As explained in Chapter 4, the objective in model selection is to use as parsimonious a model as possible while ensuring that reliable results are predicted. Model choice is better based on penalized measures of fit than unmodified likelihood and deviances (Spiegelhalter et al., 2003; Congdon, 2001 and 2003).

The results relating to model selection criteria are listed in Tables 5-3 to 5-8. AIC, BIC and DIC were selected and calculated for major model selection criteria. In addition, the marginal likelihood was estimated by the harmonic mean from the MCMC output. It can be seen that the marginal likelihood is almost the same as the corresponding posterior mean of log likelihood (LL) in Tables 5-3 and 5-4. This is because of the large sample in the MCMC output (at least 10,000 samples in our study) indicating that the distribution of the output might be indeed close to a normal distribution. Hence, for this reason, marginal LL will not be listed in other tables. Since LL does not introduce penalties for including extra parameters, it is used only for reference purposes.

WinBUGS 1.4 was used for the model development. Two parallel chains were run for both scenarios of initial status to obtain the posterior distributions of the coefficients, LL and other decision parameters, such as Poisson mean of crashes, PSI, etc. Convergence was monitored by Gelman-Rubin convergence diagnostic plots and historical plots (Spiegelhalter et al., 2003; Cowles, 2005; Brooks and Gelman, 1998) set in WinBUGs. The results in terms of parameters and model selection criteria from all the above eleven models are presented in Tables 5-3 to 5-8. The parameter estimates from the SPFs developed for the EB study, using the maximum likelihood approach, are also listed in Tables 5-7 and 5-8.

The meanings of the symbols in the following tables and hereafter are:

$\beta_0, \beta_1, \beta_2, \ \beta_3$, are coefficients that correspond to the intercept, AADT on major roads, AADT on minor roads, and time trend as shown in Equation 5-9,

r is the time series correlation coefficient,

BCI is the Bayesian credit interval,

K means number of parameters in the model,

LL is the log likelihood of the developed model, and

Mar. LL is the marginal log likelihood estimated from the harmonic mean of likelihood from the MCMC output (see Equation 4-7).

For the PG models (Tables 5-3 and 5-4), PG and PG_T are comparable and better than the other two models (time multiplier model and varying coefficients) in terms of the lower values of AIC, BIC and DIC. PG_T_6yrs is deemed to be the optimal model in this group in consideration of the time series of the crash data and given that it also has comparable values of AIC, LL and to PG_6yrs. For all four developed PG models that use 6 years of data, LL is almost the same, which indicates that the fitting is not improved with the extra parameters. For 10 years of data, similarly, model PG_T_10 yrs is the best model. The value of LL for the more complex model PG_M is greater (although not by much) than PG and PG_T which means that the extra parameters somewhat improve the fit, but the scale is very limited. However, the LL of PG_VC_10yrs is almost the same as that of PG_M_10yrs.

It can be seen that the differences of BIC among these four models are much larger than other criteria in that BIC has a large penalty $\mathrm{K} \times \log(\mathrm{N})$ for the extra parameters in a large sample size (i.e., when the sample size is greater than 7, then the penalty of BIC is greater than that of AIC). The difference is even larger for 10 years of data because the sample size is bigger.

For the PLN group (Tables 5-5 and 5-6), not surprisingly, the values of all the model selection criteria follow a similar pattern as PG models. For the sake of the values of the model selection criteria, and in consideration of the time series of the crash data, PLN_T_6yrs and PLN_T_10yrs are the best for both data groups, respectively, as expected. Also, it can be seen that the parameters of the PLN group are generally comparable to those from the PG group.

Table 5-3 Parameter Estimation and Model Diagnostics (PG models)
(year 1993-year 1998)

	PG_6yrs			**PG_T_6yrs**			**PG_M_6yrs**			**PG_VC_6yrs**		
	mean	95% BCI		mean	95% BCI		mean	95% BCI		mean	95% BCI	
Constant	-7.83	-8.54	-6.97	-7.71	-8.17	-7.19	-8.15	-8.45	-7.83	-7.36	-9.27	-5.91
β_0							-8.19	-8.50	-7.86	-8.08	-9.49	-6.82
							-8.15	-8.46	-7.83	-7.84	-9.52	-5.66
							-8.22	-8.53	-7.89	-8.19	-9.56	-6.59
							-8.17	-8.48	-7.84	-7.34	-8.61	-6.11
							-8.21	-8.52	-7.87	-7.91	-9.32	-6.42
$AAdt_{major}$	0.58	0.49	0.66	0.58	0.51	0.62	0.61	0.56	0.65	0.51	0.34	0.70
β_1										0.61	0.46	0.75
										0.60	0.37	0.80
										0.59	0.41	0.73
										0.57	0.44	0.69
										0.56	0.39	0.73
$AAdt_{minor}$	0.47	0.41	0.53	0.46	0.41	0.51	0.49	0.43	0.53	0.50	0.43	0.58
β_2										0.47	0.38	0.55
										0.45	0.37	0.52
										0.51	0.42	0.59
										0.41	0.32	0.49
										0.51	0.43	0.58
Trend β_3				-0.01	-0.02	0.01						
No. of Parameters: K	4			5			9			19		
Log likelihood: LL	-5632			-5632			-5633			-5634		
Mar. LL	-5632			-5632			-5633			-5633		
AIC	11271			11272			11282			11304		
BIC	11290			11298			11333			11419		
DIC	11746			11748			11753			11764		

Table 5-4 Parameter Estimation and Model Diagnostics (PG models)
(year 1993-year 2002)

	PG_10yrs			PG_T_10yrs			PG_M_10yrs			PG_VC_10yrs		
	mean	95% BCI		mean	95% BCI		mean	95% BCI		mean	95% BCI	
β_0	-8.05	-8.39	-7.74	-7.97	-8.27	-7.70	-7.73	-8.13	-6.94	-6.84	-8.18	-5.34
							-7.76	-8.17	-6.98	-7.79	-9.66	-6.54
							-7.73	-8.13	-6.95	-7.57	-9.09	-5.99
							-7.79	-8.20	-7.01	-7.89	-9.66	-6.43
							-7.74	-8.15	-6.96	-7.01	-8.68	-5.17
							-7.78	-8.19	-7.00	-7.88	-9.59	-6.20
							-7.77	-8.17	-6.98	-7.71	-9.49	-6.01
							-7.74	-8.15	-6.95	-7.56	-9.18	-5.87
							-7.65	-8.06	-6.86	-8.27	-9.79	-6.60
							-7.69	-8.09	-6.90	-9.33	-10.49	-7.71
β_1	0.60	0.57	0.63	0.59	0.55	0.62	0.57	0.50	0.61	0.46	0.29	0.60
										0.58	0.44	0.77
										0.58	0.39	0.72
										0.56	0.41	0.75
										0.54	0.32	0.71
										0.56	0.37	0.74
										0.57	0.39	0.75
										0.54	0.36	0.73
										0.61	0.42	0.76
										0.75	0.57	0.87
β_2	0.48	0.44	0.55	0.47	0.43	0.53	0.47	0.43	0.53	0.49	0.40	0.58
										0.46	0.37	0.54
										0.44	0.36	0.51
										0.50	0.41	0.58
										0.40	0.33	0.48
										0.51	0.43	0.59
										0.46	0.39	0.54
										0.48	0.40	0.56
										0.51	0.44	0.60
										0.47	0.39	0.54
β_3				0.01	0.00	0.01						
K	4			5			13			31		
LL	-9524			-9522			-9518			-9517		
Mar. LL	-9523			-9522			-9518			-9517		
AIC	19054			19052			19060			19094		
BIC	19075			19080			19143			19301		
DIC	19602			19600			19601			19611		

Table 5-5 Parameter Estimation and Model Diagnostics (PLN models)
(year 1993-year 1998)

	PLN_6yrs			**PLN_T_6yrs**			**PLN_M_6yrs**			**PLN_VC_6yrs**		
	mean	95% BCI		mean	95% BCI		mean	95% BCI		mean	95% BCI	
β_0	-8.43	-9.45	-7.24	-7.89	-9.01	-6.93	-7.87	-8.92	-6.90	-8.42	-9.66	-7.47
							-7.91	-8.96	-6.94	-9.09	-10.18	-8.16
							-7.87	-8.92	-6.91	-8.84	-9.92	-7.50
							-7.94	-9.00	-6.97	-9.35	-10.71	-8.24
							-7.88	-8.94	-6.91	-8.30	-9.39	-7.10
							-7.92	-8.98	-6.95	-8.95	-10.25	-7.11
β_1	0.61	0.49	0.71	0.56	0.46	0.67	0.56	0.45	0.68	0.60	0.50	0.72
										0.69	0.59	0.80
										0.69	0.54	0.81
										0.70	0.58	0.84
										0.65	0.51	0.77
										0.64	0.46	0.77
β_2	0.49	0.42	0.55	0.48	0.42	0.53	0.48	0.42	0.53	0.50	0.43	0.57
										0.47	0.38	0.54
										0.45	0.36	0.53
										0.50	0.43	0.59
										0.41	0.34	0.49
										0.52	0.44	0.60
β_3				-0.01	-0.02	0.01						
K	4			5			9			19		
LL	-5653			-5653			-5653			-5654		
AIC	11312			11314			11322			11344		
BIC	11331			11340			11373			11459		
DIC	11806			11805			11810			11820		

Table 5-6 Parameter Estimation and Model Diagnostics (PLN models)
(year 1993-year 2002)

	PLN_10yrs			PLN_T_10yrs			PLN_M_10yrs			PLN_VC_10yrs		
	mean	95% BCI		mean	95% BCI		mean	95% BCI		mean	95% BCI	
β_0	-8.15	-7.50	-8.88	-8.37	-7.86	-9.08	-7.27	-5.39	-8.24	-7.21	-6.03	-8.22
							-7.31	-5.42	-8.28	-8.08	-6.94	-9.26
							-7.27	-5.37	-8.24	-7.97	-6.55	-9.67
							-7.34	-5.45	-8.31	-8.19	-7.12	-9.36
							-7.28	-5.39	-8.26	-7.66	-6.35	-8.82
							-7.32	-5.41	-8.30	-8.34	-6.79	-9.66
							-7.31	-5.42	-8.29	-8.39	-7.03	-9.59
							-7.28	-5.38	-8.26	-8.08	-7.12	-9.39
							-7.19	-5.28	-8.17	-8.77	-7.87	-9.86
							-7.22	-5.30	-8.21	-10.15	-8.98	-11.18
β_1	0.58	0.50	0.66	0.59	0.54	0.66	0.50	0.31	0.61	0.46	0.35	0.56
										0.58	0.46	0.71
										0.59	0.46	0.76
										0.56	0.45	0.69
										0.58	0.44	0.71
										0.57	0.42	0.72
										0.61	0.47	0.73
										0.57	0.46	0.69
										0.63	0.52	0.73
										0.80	0.69	0.90
β_2	0.48	0.42	0.54	0.50	0.46	0.54	0.46	0.37	0.52	0.50	0.44	0.58
										0.47	0.38	0.55
										0.45	0.38	0.53
										0.51	0.44	0.58
										0.41	0.34	0.48
										0.52	0.44	0.59
										0.47	0.40	0.54
										0.49	0.42	0.55
										0.53	0.46	0.60
										0.48	0.42	0.55
β_3				0.01	0.00	0.01						
K	4			5			13			31		
LL	-9544			-9543			-9537			-9536		
AIC	19094			19094			19098			19132		
BIC	19115			19122			19181			19339		
DIC	19661			19658			19663			19669		

Table 5-7 Parameter Estimation and Model Diagnostics (Poisson AR(1) models)

(year 1993-year 1998)

	P_AR(1)_6yrs			**P_AR(1)+T_6yrs**			**P_NAR(1)_6yrs**			**EB_6yrs**		
	mean	95% BCI		mean	95% BCI		mean	95% BCI		mean	95% BCI	
β_0	-8.45	-9.10	-7.59	-8.16	-9.06	-7.38	-8.25	-9.24	-7.42	-8.81	-9.96	-7.66
β_1	0.62	0.52	0.70	0.59	0.51	0.68	0.60	0.50	0.71	0.69	0.56	0.88
β_2	0.48	0.42	0.53	0.47	0.41	0.54	0.47	0.42	0.52	0.48	0.42	0.53
β_3				-0.01	-0.03	0.01						
r	0.95	0.93	0.97	0.95	0.93	0.97	0.97	0.93	1.00			
K	5			6			5			4		
LL	-5526			-5521			-5535					
AIC	11060			11052			11078					
BIC	11086			11084			11104					
DIC	**11751**			11750			11756					

Table 5-8 Parameter Estimation and Model Diagnostics (Poisson AR(1) models)

(year 1993-year 2002)

	P_AR(1)_10yrs			**P_AR(1)+T_10yrs**			**P_NAR(1)_10yrs**			**EB_10yrs**		
	mean	95% BCI		mean	95% BCI		mean	95% BCI		mean	95% BCI	
β_0	-9.04	-8.38	-9.73	-8.77	-9.53	-7.88	-8.77	-8.08	-9.57	-9.25	-10.35	-8.14
β_1	0.67	0.59	0.74	0.63	0.50	0.72	0.64	0.57	0.73	0.73	0.61	0.85
β_2	0.50	0.43	0.55	0.50	0.46	0.55	0.50	0.44	0.56	0.48	0.43	0.54
β_3				0.00	-0.01	0.01						
r	0.97	0.96	0.98	0.97	0.96	0.98	0.99	0.97	1.00			
K	5			6			5			4		
LL	**-9269**			-9269			-9283					
AIC	**18546**			18548			18574					
BIC	**18574**			18582			18602					
DIC	**19489**			19491			19490					

For Poisson AR (1) models (Tables 5-7 and 5-8), P_AR (1) and time trend model P_AR (1)+T are comparable and better than the non-stationary model P_NAR (1) for these two data sets. This confirms that it should be a stationary process for the logarithm of crash data. P_AR (1) is more preferable because of a simpler form.

Poisson AR (1) has much lower values of AIC, BIC and a higher value of LL compared with PG and PLN models for the dataset of 1993 – 1998, and comparable values of the DIC of PG_T_6yrs, P_AR (1)_6yrs is deemed to be the best model for this data group and is thus used for further exploration and for comparison with the EB results.

For the 10 year dataset (1993 - 2002), all the model selection criteria including LL, AIC, BIC and DIC favor the P_AR (1) _10 yrs model. This model is used to estimate the true mean of crashes for 1999 – 2002 to evaluate the performance of the ranking method and criteria.

5.6.3 Evaluation of FB and EB for Hot Spot Identification

Time frames 3-4-10 and 6-4-10 were used to conduct a comparison of the FB method with the EB method for hotspot identification. According to the model selection result from 6 years of ranking data, P_AR (1)_3yrs was deemed to be the best model for the three years of ranking dataset and was used for the comparison study. The data for 1996 to 1998 is summarized in Table 5-9. Compared with the data in Table 5-2, it can be found that traffic volume on major roads continues to grow while traffic volume on minor roads stays almost the same. Crashes in 1996 to 1998 stay almost the same and crashes in the evaluation period increases by 0.07 crashes/site.year.

Table 5-9 Summary data for 726 California Unsignalized Intersections

	year 1996 - year 1998			
	Mean	Standard Deviation	Maximum	Minimum
Total Crashes / site. year	1.33	1.79	13	0
$AADT_{Major}$	8358	4347	28604	2917
$AADT_{Minor}$	655	867	7800	100

The posterior Poisson mean or expected crashes and the PSI in ranking periods, which are available for the EB method, were used for the comparison study. Crash counts in the ranking periods were used only for reference purposes.

As noted earlier, there are five evaluation criteria applied, including sensitivity and specificity, sum of crash count, sum of Poisson mean, sum of the PSI and sum of the PPSI in next period (1999 - 2002).

It should be noted there are two ranking periods: one is 1993 to 1998 and the other is 1996 to 1998. The evaluation period is 1999 to 2002. Also note that the estimate of true mean of the evaluation period was derived from the model by using 10 years of data (from 1993 to 2002). In other words, models P_AR(1)_3yrs and P_AR(1)_6yrs were used for ranking, respectively, and model P_AR(1)_10yrs was used for evaluation purpose.

Before comparing the results of FB with EB, it is necessary to investigate the estimates from these two methods since the Poisson mean, PSI and PPSI might be different due to the different developed models and error structures. It was found that the sum of the PM in the evaluation period for the 726 sites from EB is 4297 crashes while there are 4279 crashes from FB. This confirms that the estimation of expected crashes from both methods is indeed comparable, but the EB method provides a slightly higher value for the evaluation period (1999 –2002). Nevertheless, this suggests that it is appropriate to use PM as an evaluation criterion for comparison of both methods. The sum of the PM from EB was adjusted by multiplying a ratio $^{4279}/_{4297} = 0.9956$ for a better comparison. However, PSI and PPSI are quite different due to the different structures of random effects and cannot be employed as evaluation criteria for the comparison; however, they can be used to evaluate the performance of ranking criteria by each method. For example, the ranking criteria, PM and the PSI, can be evaluated based on the evaluation criteria, sum of the PSI and the sum of the PPSI for the FB and EB methods, respectively.

The evaluation results from the FB P_AR(1) models and the EB method are presented in Figures 5-1 to 5-4 for Time Frame 6-4-10, and Figures 5-5 to 5-8 for Time Frame 3-4-10. The P_AR(1) model provides better results than those from the EB method ranked by PM or PSI in terms of higher values of sensitivity, specificity, and sum of Poisson mean for both time frames. It is easy to see that the P_AR(1) model, which is ranked by PM, provides the best results since it has the highest values of sensitivity, specificity, and sum of Poisson mean, while the EB method ranked by PSI provides the poorest results as shown in Figures 5-1 to 5-8. If the crash count in 1999 – 2002 is used as an evaluation criterion, then both methods ranked by PM provide almost similar results, and better ones than by PSI.

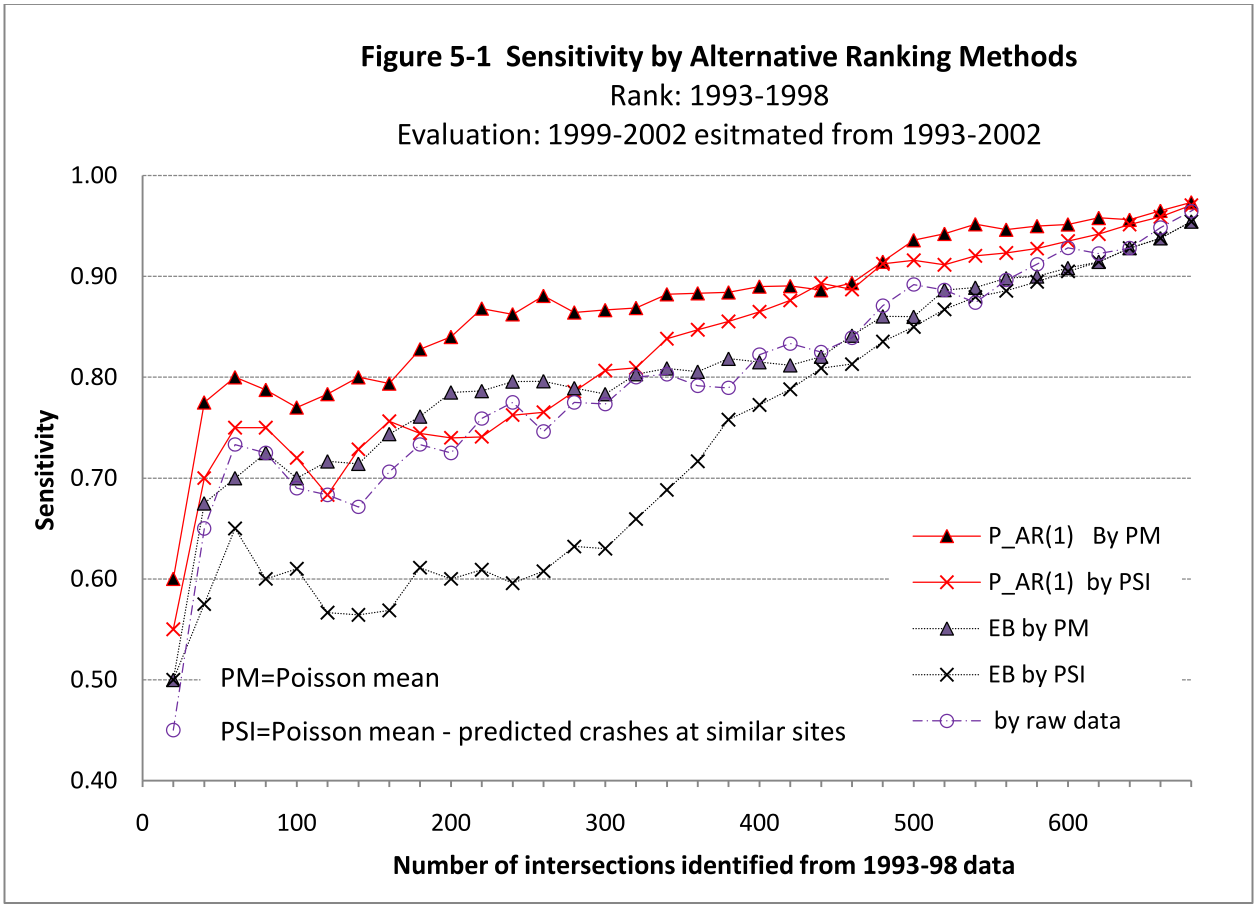


Figure 5-2 Specificity by Alternative Ranking Methods
Rank: 1993-1998
Evaluation: 1999-2002 esitmated from 1993-2002

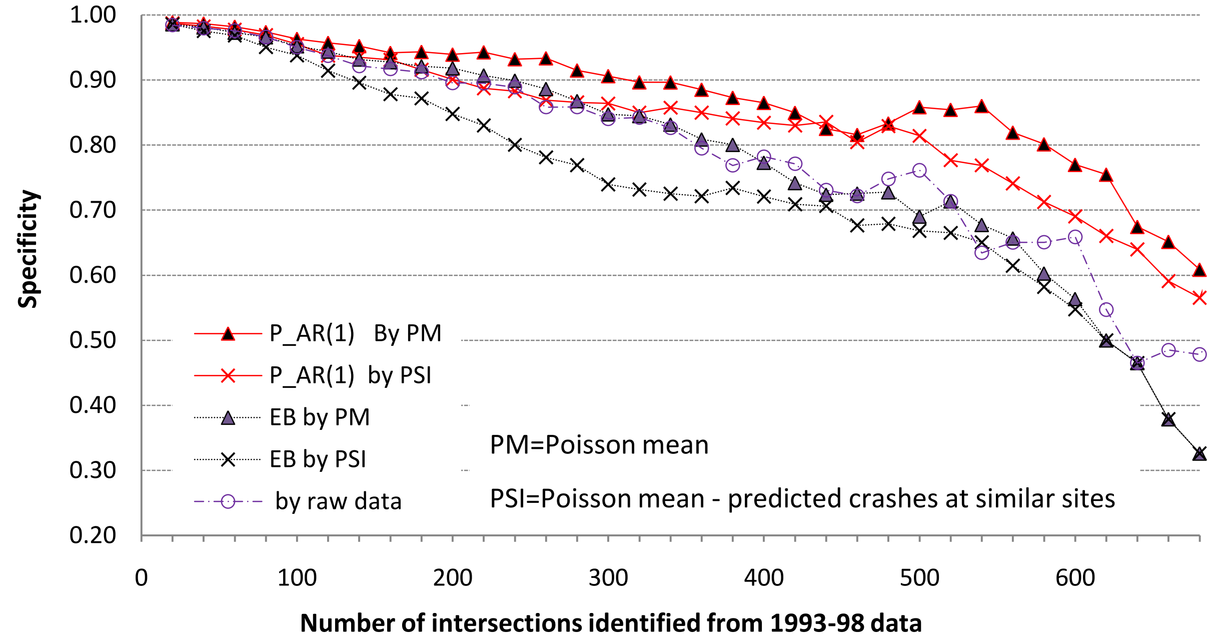
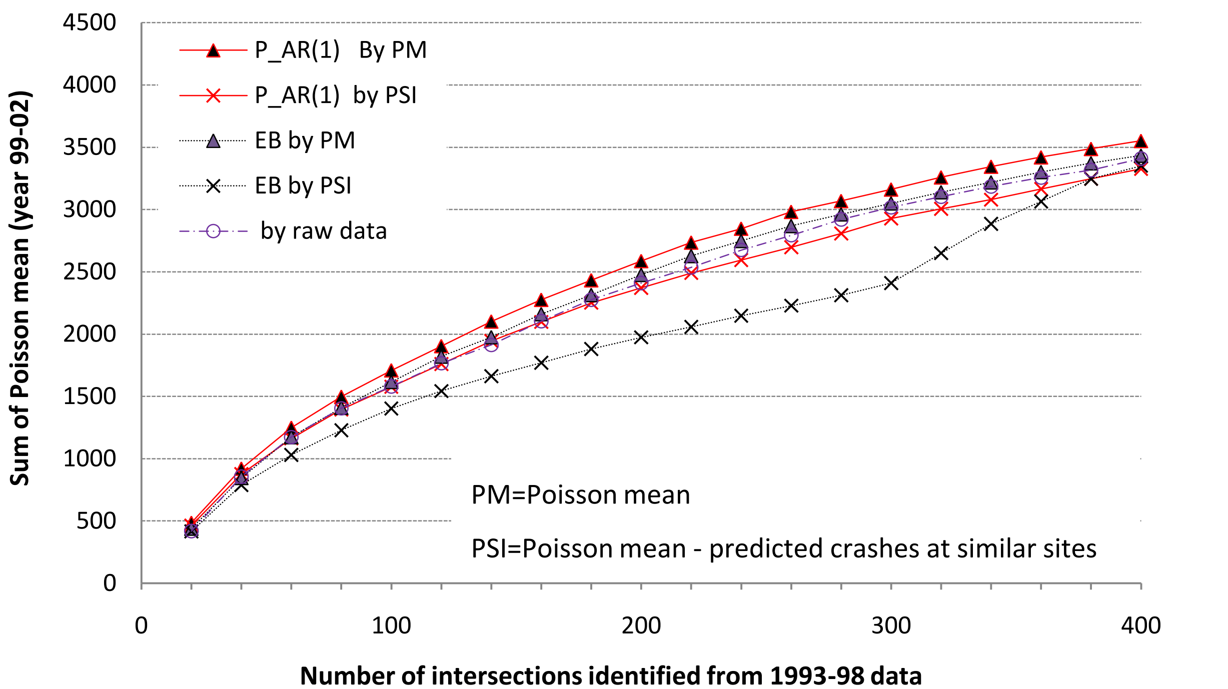


Figure 5-3 Sum of PM (1999-2002) by Alternative Ranking Methods
Rank: 1993-1998 Evaluation: 1999-2002 esitmated from 1993-2002



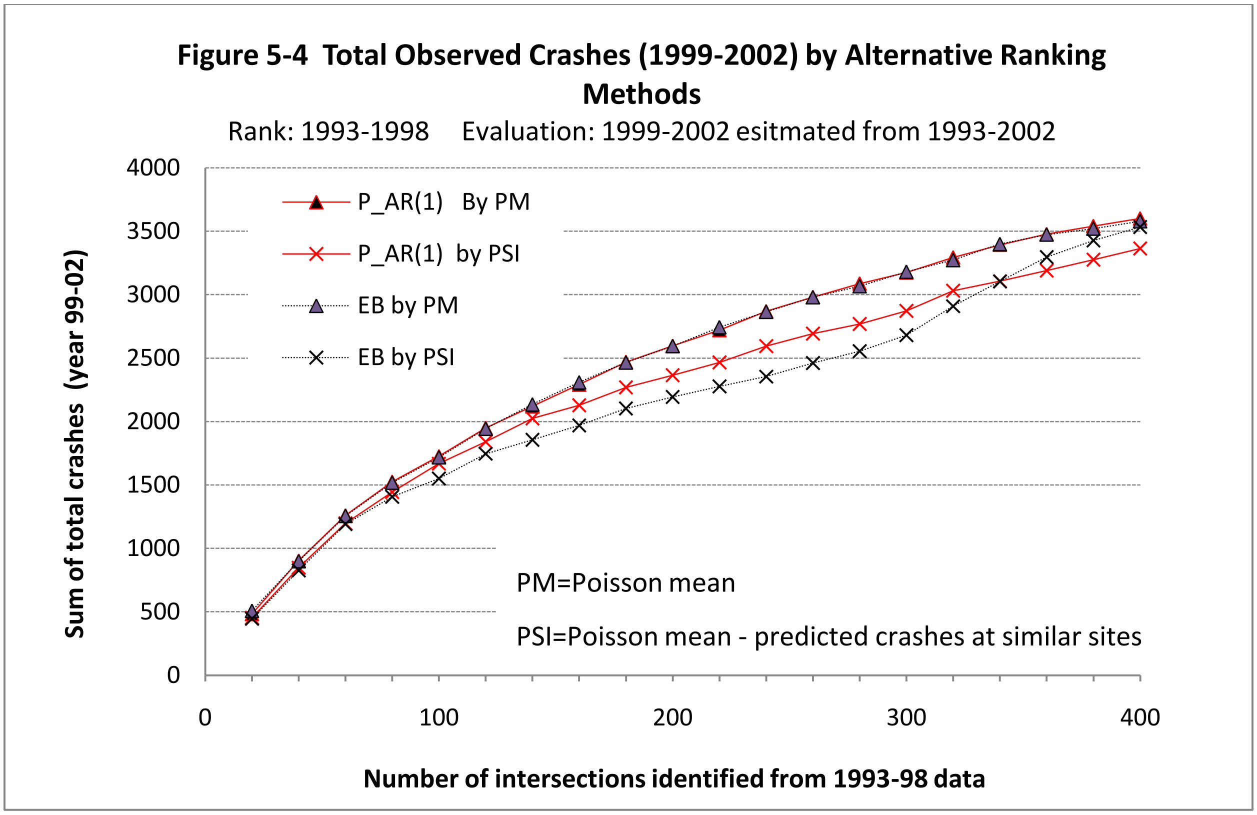


Figure 5-4 Total Observed Crashes (1999-2002) by Alternative Ranking Methods
Rank: 1993-1998 Evaluation: 1999-2002 esitmated from 1993-2002
Sum of total crashes (year 99-02)
P_AR(1) By PM
P_AR(1) by PSI
EB by PM
EB by PSI
PM=Poisson mean
PSI=Poisson mean - predicted crashes at similar sites
Number of intersections identified from 1993-98 data

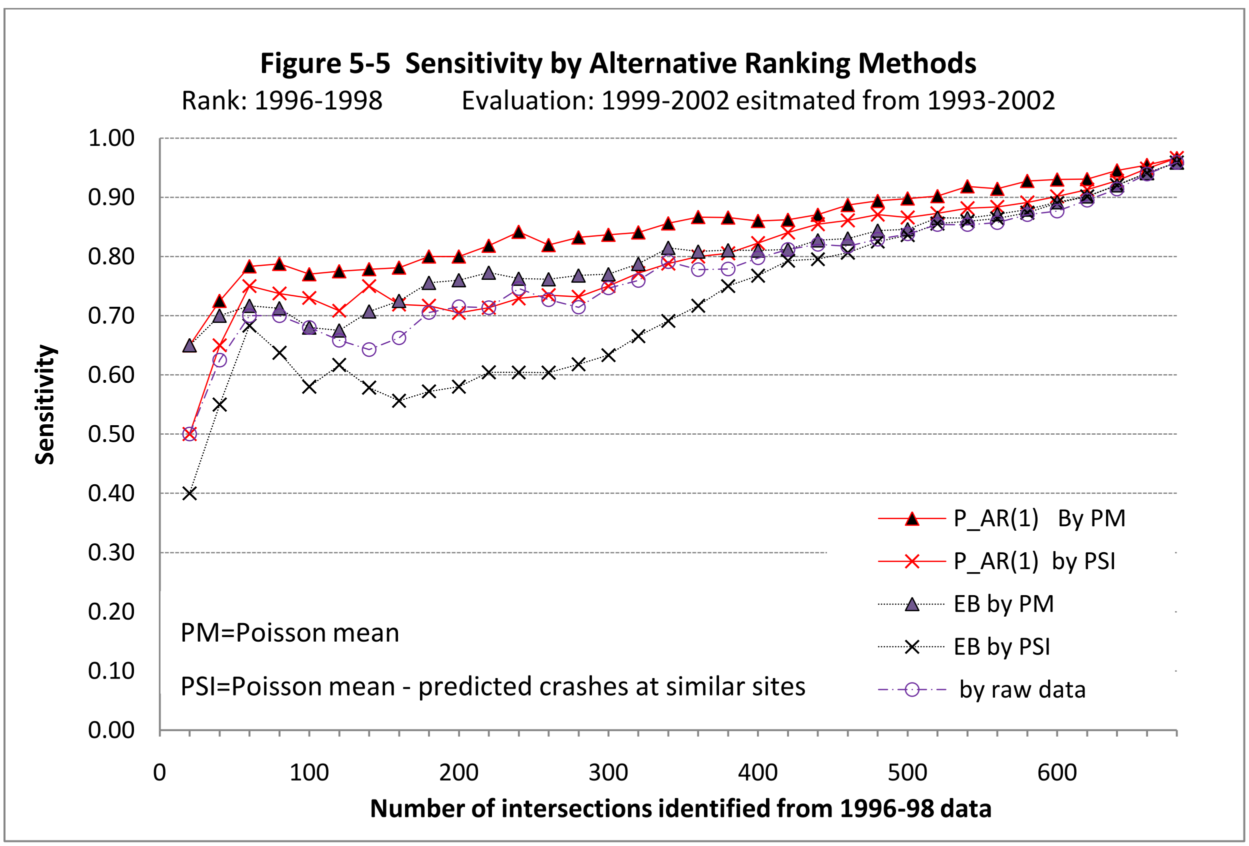


Figure 5-5 Sensitivity by Alternative Ranking Methods
Rank: 1996-1998 Evaluation: 1999-2002 esitmated from 1993-2002
Sensitivity
P_AR(1) By PM
P_AR(1) by PSI
EB by PM
EB by PSI
by raw data
PM=Poisson mean
PSI=Poisson mean - predicted crashes at similar sites
Number of intersections identified from 1996-98 data

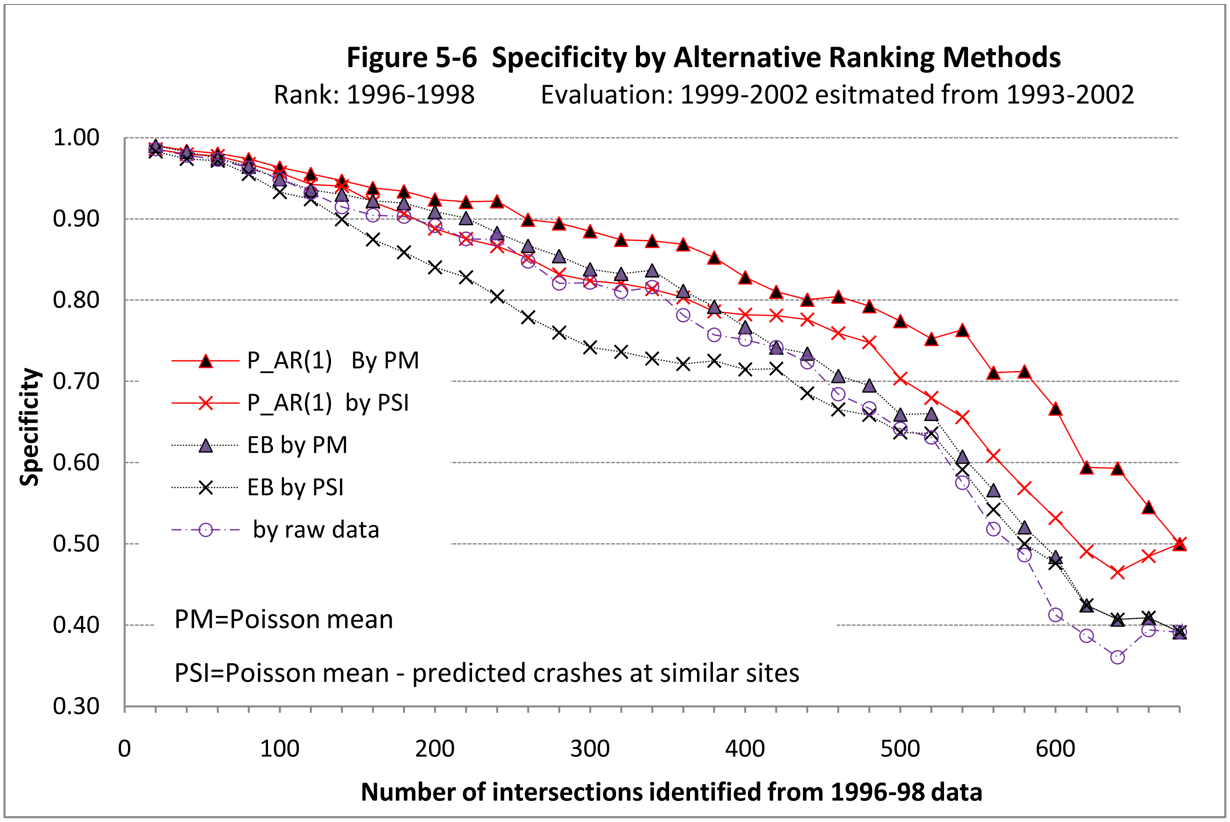


Figure 5-6 Specificity by Alternative Ranking Methods
Rank: 1996-1998 Evaluation: 1999-2002 esitmated from 1993-2002
Specificity
1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
P_AR(1) By PM
P_AR(1) by PSI
EB by PM
EB by PSI
by raw data
PM=Poisson mean
PSI=Poisson mean - predicted crashes at similar sites
0
100
200
300
400
500
600
Number of intersections identified from 1996-98 data

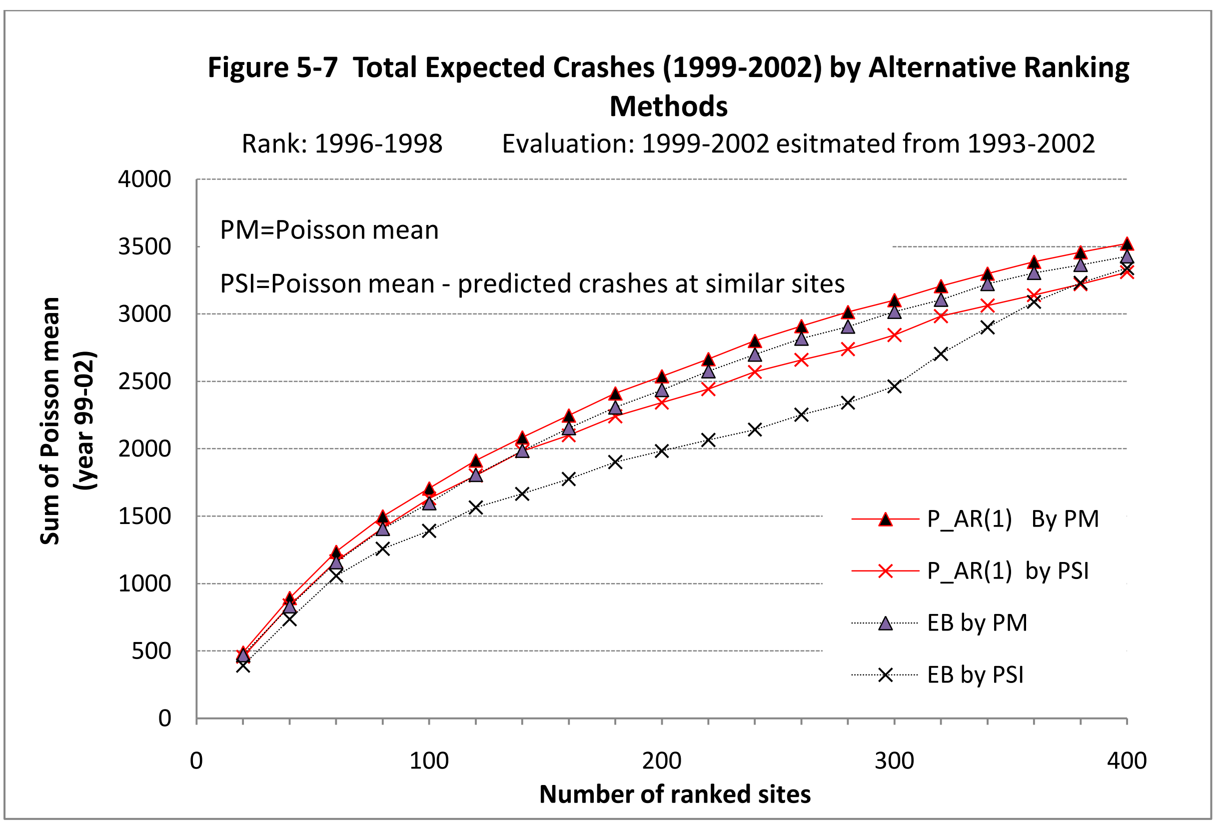
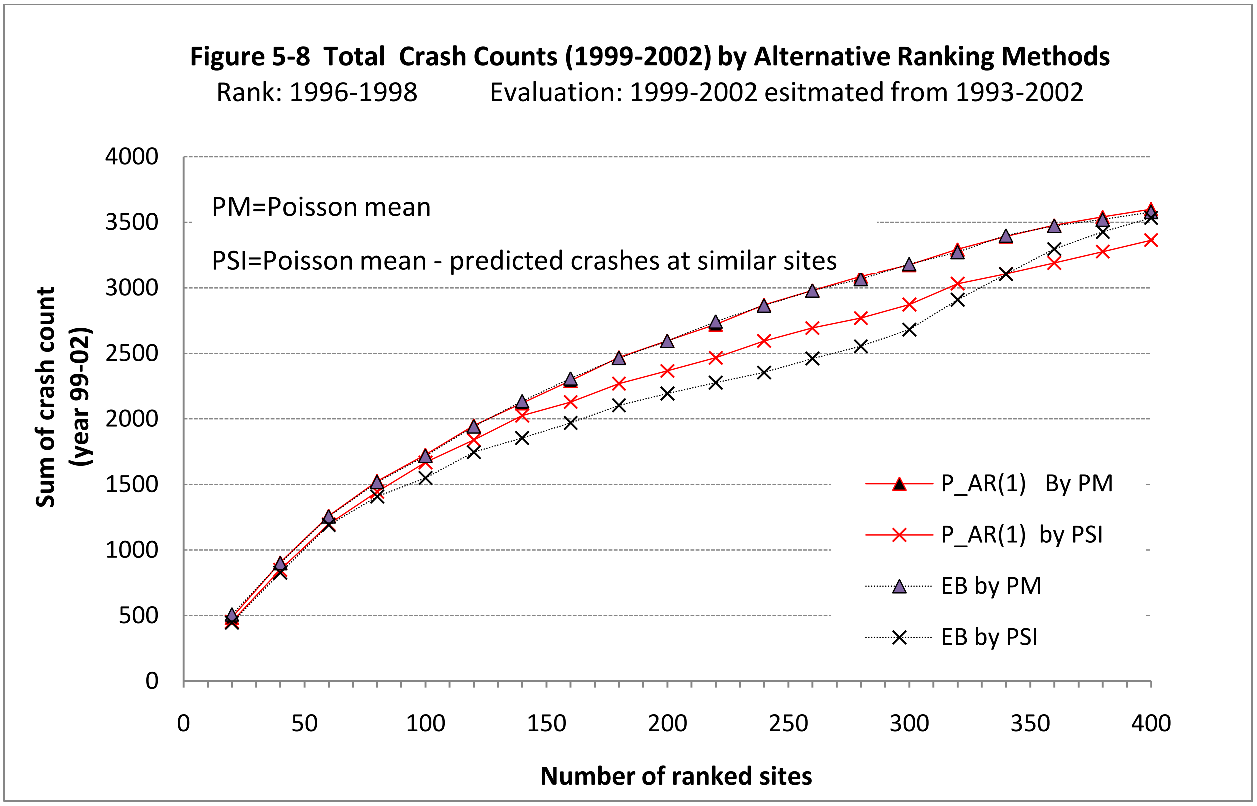


Figure 5-7 Total Expected Crashes (1999-2002) by Alternative Ranking Methods
Rank: 1996-1998 Evaluation: 1999-2002 esitmated from 1993-2002
Sum of Poisson mean (year 99-02)
4000
3500
3000
2500
2000
1500
1000
500
0
PM=Poisson mean
PSI=Poisson mean - predicted crashes at similar sites
P_AR(1) By PM
P_AR(1) by PSI
EB by PM
EB by PSI
0
100
200
300
400
Number of ranked sites



The evaluation results from the top ranked 10, 20, 30, 40 and 50 sites by the P_AR(1) model and the EB method for both time frames are listed in Table 5-10. For the top ranked 10 sites (about 1.38 % of all sites), both the FB and EB methods have the same value of sensitivity and specificity, regardless whether the ranking is by PM or PSI for the data of the two time frames. However, it cannot be concluded that the EB method obtains the same top ten ranked sites and provides comparably promising results to the FB method. The evaluation criterion, sum of the PM in the second period, might provide more information for comparison of the two methods. From Table 5-10, it can be seen that the FB method indeed provides much better results than the EB method in terms of a higher value of the sum of the PM. This might be caused by the tiny difference in the PM of the sites in the ranking period and might be due to lack of homogeneity between the ranking period and the evaluation period. In fact, even if the data do meet the homogeneity assumption for the very few top ranked sites, it is still problematic to evaluate the ranking results based only on sensitivity in that it cannot differentiate the values of the PM for false positive identification in the few top ranked sites. For example, for the top 10 most hazardous ranked sites, false identification of one site causes a 10% difference in sensitivity while the Poisson mean may just have a minor difference (say, 0.01 crashes/year). For this case, it is strongly recommended that the sum of the PM in the second period should be taken as a major evaluation criterion. In fact, from Table 5-10, the sum of the PM ranked by PM from FB method is 263 crashes in 1999 -2002 (per 6 years of ranking data) for the top 10 ranked sites, while the EB method gives just 242 crashes (multiply the ratio of total expected crashes from the FB over that from the EB which is $^{4279}/_{4297} = 0.9956$). The sum of the PM ranked by PSI by the FB method also has a much higher value than that from the EB method. Similarly, the three years of ranking data (1996- 1998) have the same pattern. This proves that the FB method can provide much better ranking results than the EB method for the top ranked 10 sites in terms of much higher values in the sum of the PM, even though they both have the same values of sensitivity and specificity.

With top ranked 20, 30, 40 and 50 sites, the FB method still provides better ranking results ranked by PM or PSI in contrast to the EB method as identified by two datasets (1993-1998 and 1996-1998) as shown in Table 5-10 and Figures 5-1 to 5-3, and Figures 5-5 to 5-7 based on the evaluation criteria sensitivity and specificity, and sum of the PM. In most cases, the sum of the crash counts from the FB method is also greater than that from the EB as is evident from Table 5-10. In fact, the FB method ranked by PM has the best ranking results. Thus, it can be concluded that the FB method is superior to the EB method for network ranking based on the evaluation results obtained.

5.6.4 Sensitivity Analysis of Ranking Criteria

A total of eight ranking criteria; the PM, the PSI, the PPSI, crash counts, expected rank of the posterior distribution of PM (denoted as expected rank), model rank and median rank of the posterior distribution of PM (defined as mode rank and median rank, respectively) and the probability of being the worst in terms of the highest value of PM, were explored using 3 years of data (1996-1998) and 6 years of data (1993-1998), respectively, for ranking by the previously identified best Poisson AR (1) model. A total of five evaluation criteria including sensitivity and specificity, sum of crash counts, sum of the PM, sum of the PSI and sum of the PPSI in the second period (1999 – 2002) were employed to conduct the evaluation study.

Table 5-10 Comparison of Evaluation Results of EB and FB

Evaluation period: year 1999 - year 2002 estimated by the models developed using 10 years' data

Hotspots identified from 1996-1998 data					**Hotspots identified from 1993-1998 data**			
FB: Poisson AR(1) model					By Poisson mean			
ranked sites	Sensitivity	Specificity	$\sum tot_{1999-2002}$	$\sum PM_{1999-2002}$	Sensitivity	Specificity	$\sum tot_{1999-2002}$	$\sum PM_{1999-2002}$
10	0.20	0.99	241	248	0.30	0.99	263	263
20	0.65	0.99	484	488	0.60	0.99	485	485
30	0.70	0.99	717	713	0.70	0.99	702	698
40	0.73	0.98	904	893	0.78	0.99	932	920
50	0.74	0.98	1055	1047	0.80	0.99	1116	1097
FB: Poisson AR(1) model					By PSI=Poisson mean-predicted crashes at similar sites			
10	0.30	0.99	235	243	0.40	0.99	243	244
20	0.50	0.99	453	457	0.55	0.99	454	463
30	0.47	0.98	641	643	0.57	0.98	707	698
40	0.65	0.98	848	839	0.70	0.98	899	876
50	0.74	0.98	1034	1014	0.74	0.98	1047	1022
EB method					By Expected Crashes			
ranked sites	Sensitivity	Specificity	$\sum tot_{1999-2002}$	$\sum PM_{1999-2002}$	Sensitivity	Specificity	$\sum tot_{1999-2002}$	$\sum PM_{1999-2002}$
10	0.20	0.99	241	224	0.30	0.99	260	243
20	0.65	0.99	508	469	0.50	0.99	476	439
30	0.63	0.98	715	664	0.67	0.99	702	651
40	0.70	0.98	899	831	0.68	0.98	917	846
50	0.70	0.98	1087	1005	0.72	0.98	1116	1028
EB method					By PSI			
10	0.30	0.99	235	209	0.40	0.99	243	209
20	0.40	0.98	446	391	0.50	0.99	468	416
30	0.40	0.97	641	569	0.50	0.98	700	625
40	0.55	0.97	827	736	0.58	0.98	892	788
50	0.64	0.97	1034	921	0.64	0.97	1016	901

The objectives of this aspect of the research were to evaluate the ranking criteria given the best FB model, using the above evaluation criteria, and to identify the relationship between the ranking criteria. To achieve the first objective, the evaluation results from alternative ranking criteria were compared to identify the criterion that provides the most promising results. To reach the second goal, the identified sites by the most popular criterion, PM, and most recommended criterion, expected rank, were treated as the base condition, respectively, then ranked sites by other criteria were compared with those by PM and expected rank. The percentage of the same sites were calculated, which may or may not take into consideration the exact order within that top ranked group, e.g., for the top 10 ranked sites identified by PM, the same sites are located in the top 10 sites ranked by median rank, but only 4 sites have the exact same order when ranked by PM. In this example, the two criteria are deemed to be exactly the same in terms of top ranked sites without considering the order, but are only 40% in similarity when considering the order of the top 10 ranked sites. In this way, insights into the ranking criteria were obtained in terms of how much of the identified sites by different criteria were the same in terms of ranked sites, and in terms of the exact order of the ranked sites.

Since there could have been several sites with the same median rank, or same mode rank or the same probability of being the worst, a second level of ranking criterion, the expected rank, was added to resolve ties. The evaluation results from the eight ranking criteria are presented in Figures 5-9-5-14 for ranking data in 1993 – 1998 and in Figures 5-15 – 5-18 for ranking data in 1996 – 1998. The information associated with the eight ranking criteria in the ranking and evaluation periods in terms of the sum of the PM, sum of the PSI, sum of the PPSI and sum of total crashes, and sensitivity and specificity among these two periods are tabulated in Tables 5-11 – 5-14 for both time frames.

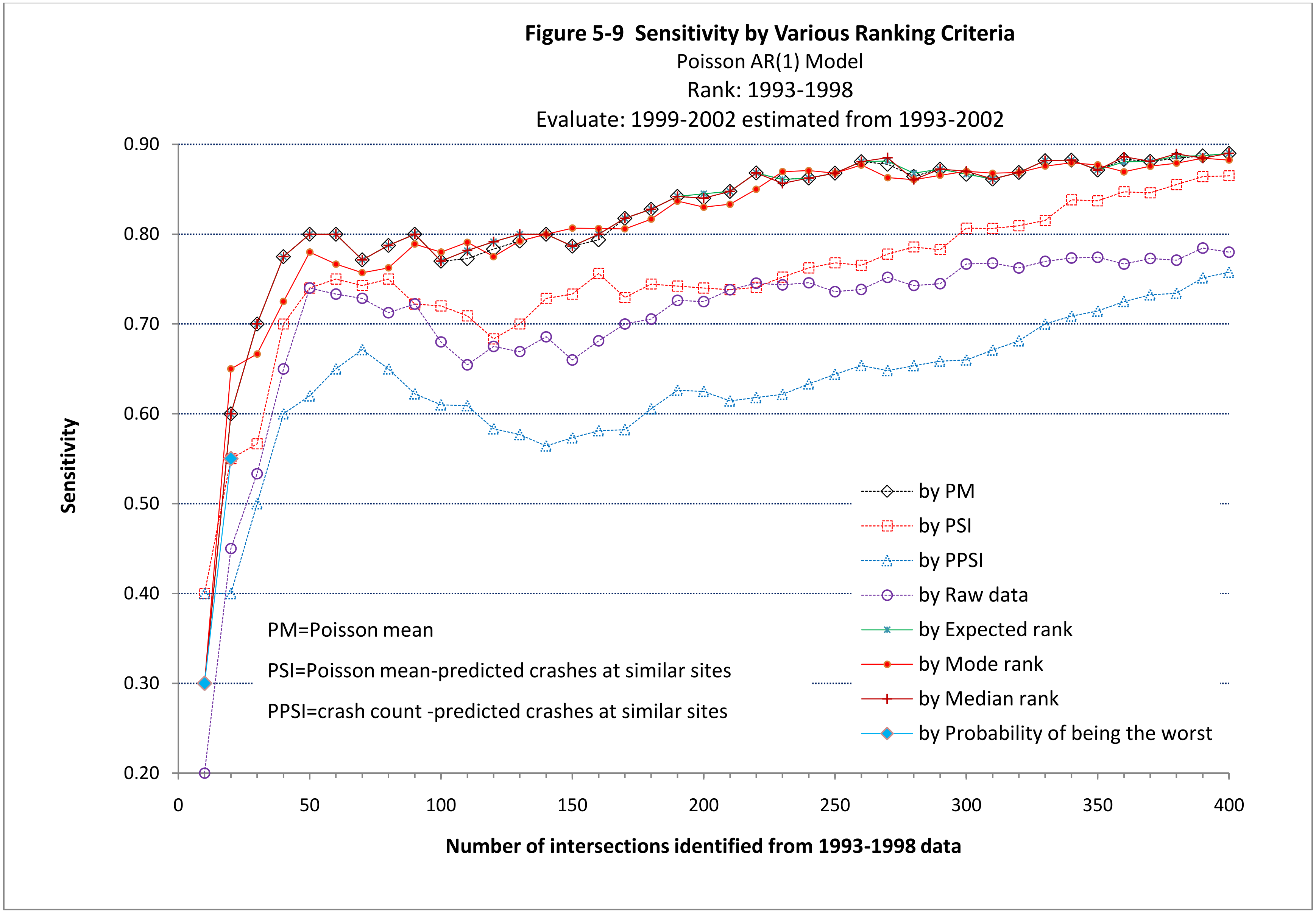


Figure 5-9 Sensitivity by Various Ranking Criteria
Poisson AR(1) Model
Rank: 1993-1998
Evaluate: 1999-2002 estimated from 1993-2002
Sensitivity
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0
50
100
150
200
250
300
350
400
Number of intersections identified from 1993-1998 data
by PM
by PSI
by PPSI
by Raw data
by Expected rank
by Mode rank
by Median rank
by Probability of being the worst
PM=Poisson mean
PSI=Poisson mean-predicted crashes at similar sites
PPSI=crash count -predicted crashes at similar sites

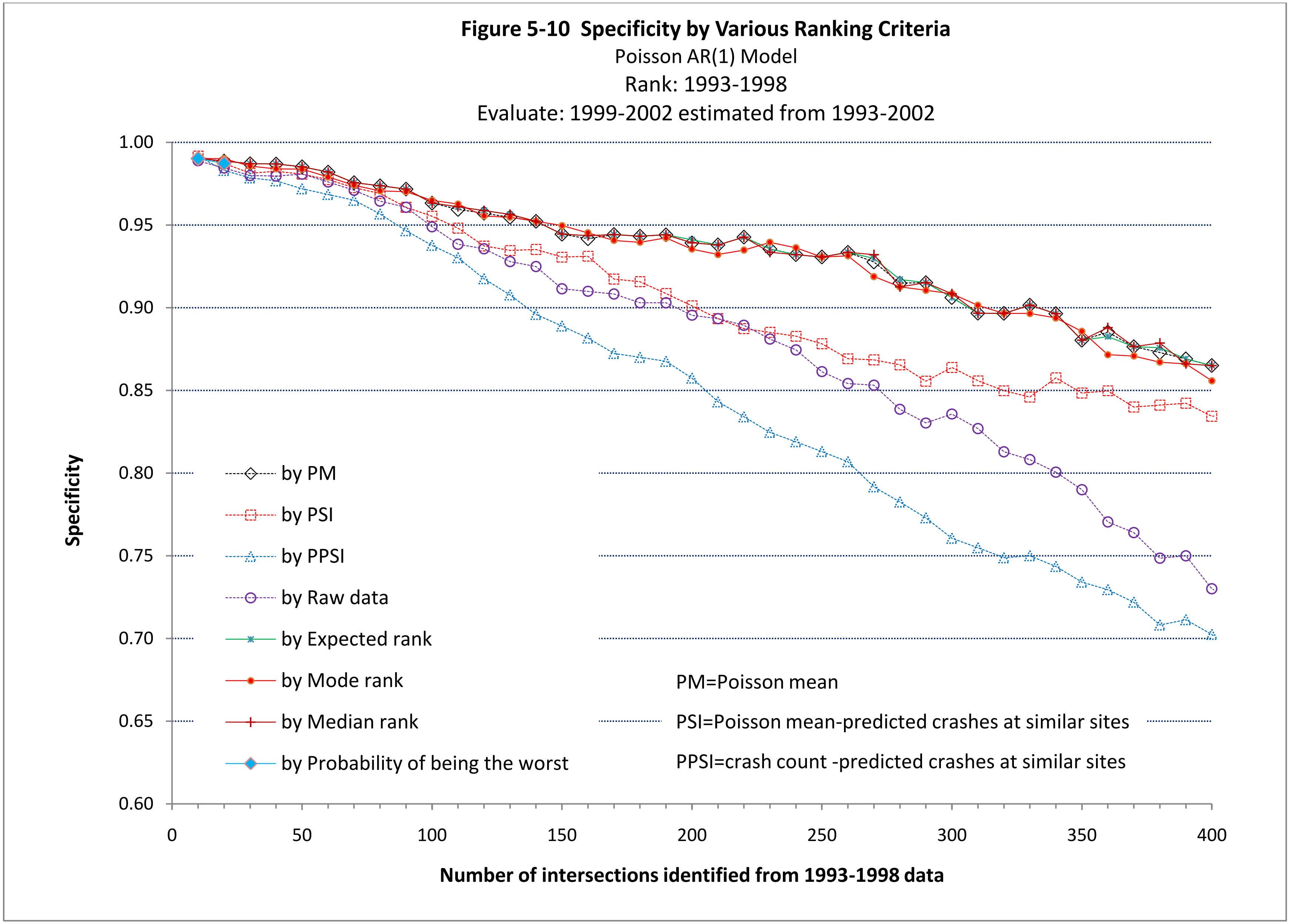


Figure 5-10 Specificity by Various Ranking Criteria
Poisson AR(1) Model
Rank: 1993-1998
Evaluate: 1999-2002 estimated from 1993-2002
Specificity
1.00
0.95
0.90
0.85
0.80
0.75
0.70
0.65
0.60
0
50
100
150
200
250
300
350
400
Number of intersections identified from 1993-1998 data
by PM
by PSI
by PPSI
by Raw data
by Expected rank
by Mode rank
by Median rank
by Probability of being the worst
PM=Poisson mean
PSI=Poisson mean-predicted crashes at similar sites
PPSI=crash count -predicted crashes at similar sites

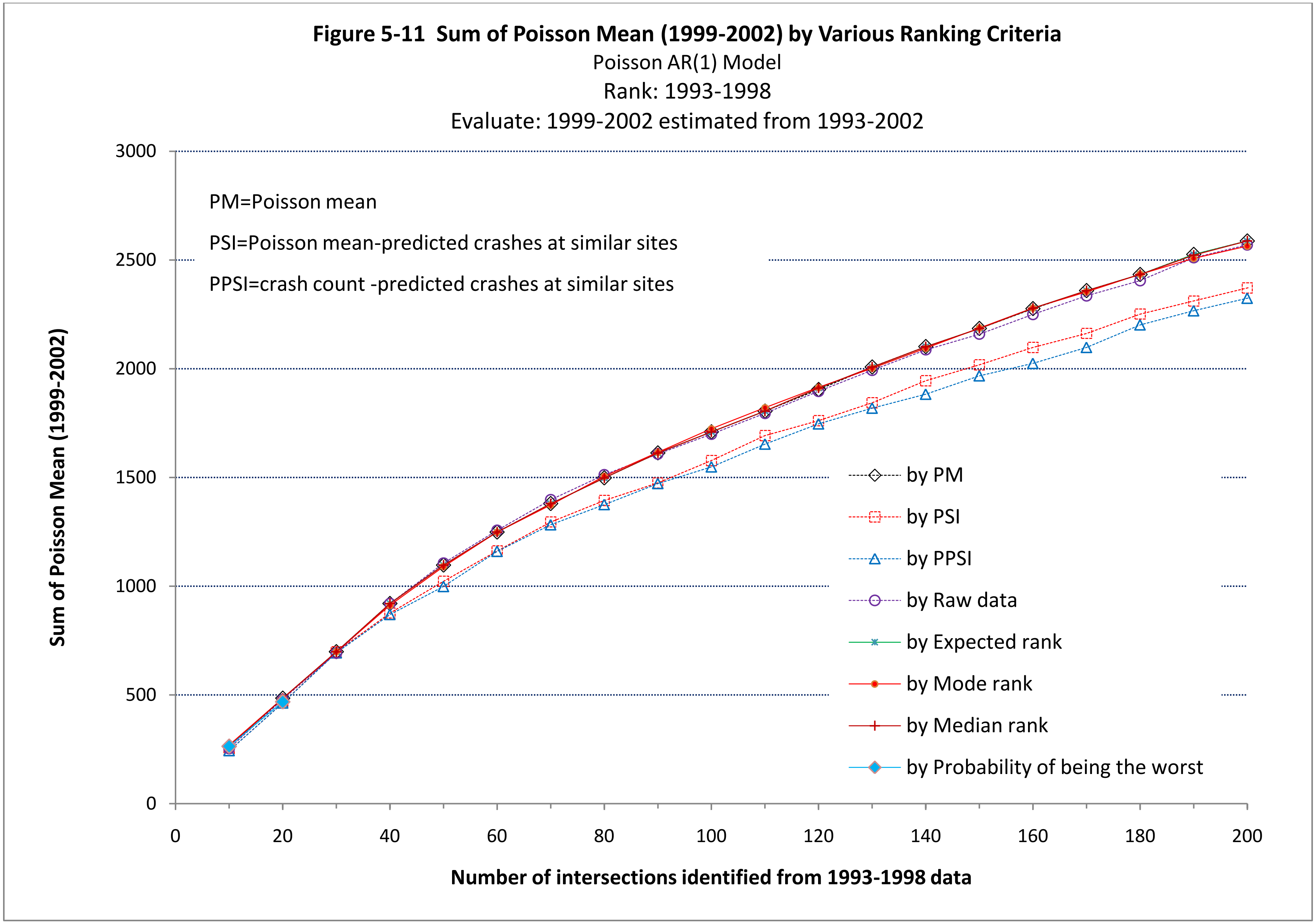


Figure 5-11 Sum of Poisson Mean (1999-2002) by Various Ranking Criteria
Poisson AR(1) Model
Rank: 1993-1998
Evaluate: 1999-2002 estimated from 1993-2002
PM=Poisson mean
PSI=Poisson mean-predicted crashes at similar sites
PPSI=crash count -predicted crashes at similar sites
by PM
by PSI
by PPSI
by Raw data
by Expected rank
by Mode rank
by Median rank
by Probability of being the worst
Sum of Poisson Mean (1999-2002)
Number of intersections identified from 1993-1998 data
0
500
1000
1500
2000
2500
3000
0
20
40
60
80
100
120
140
160
180
200

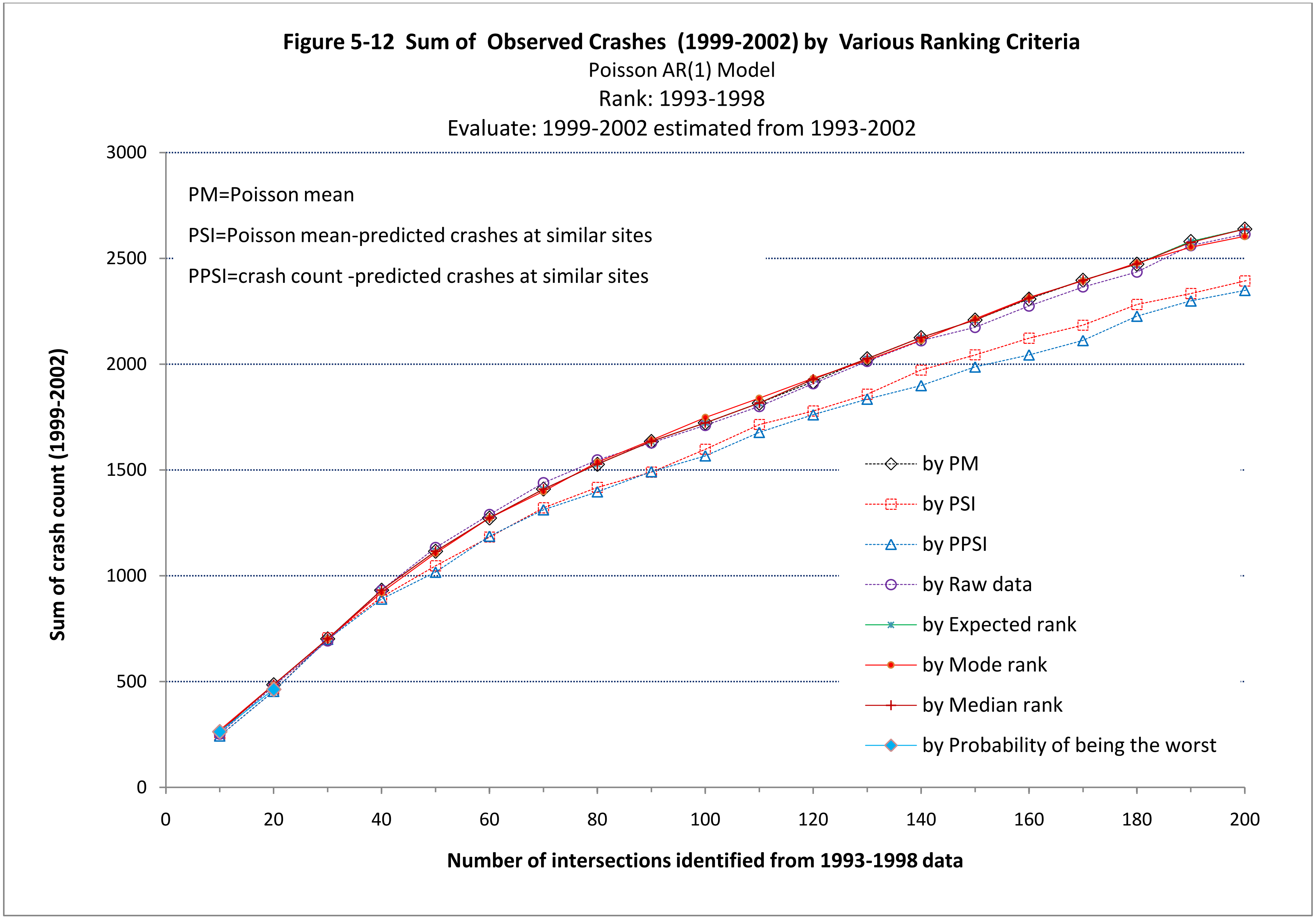


Figure 5-12 Sum of Observed Crashes (1999-2002) by Various Ranking Criteria
Poisson AR(1) Model
Rank: 1993-1998
Evaluate: 1999-2002 estimated from 1993-2002
PM=Poisson mean
PSI=Poisson mean-predicted crashes at similar sites
PPSI=crash count -predicted crashes at similar sites
Sum of crash count (1999-2002)
Number of intersections identified from 1993-1998 data
by PM
by PSI
by PPSI
by Raw data
by Expected rank
by Mode rank
by Median rank
by Probability of being the worst
0
500
1000
1500
2000
2500
3000
20
40
60
80
100
120
140
160
180
200

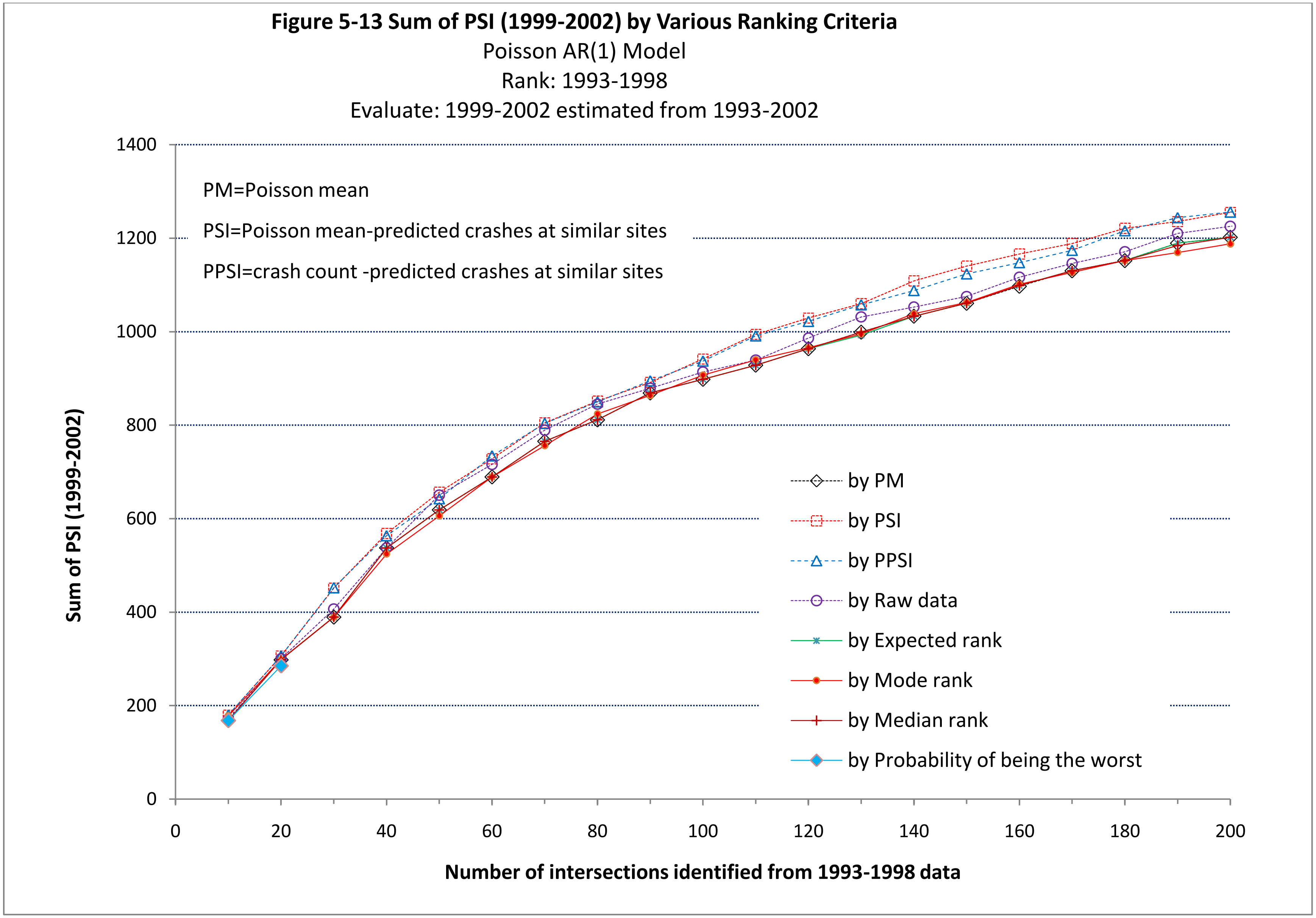


Figure 5-13 Sum of PSI (1999-2002) by Various Ranking Criteria
Poisson AR(1) Model
Rank: 1993-1998
Evaluate: 1999-2002 estimated from 1993-2002
PM=Poisson mean
PSI=Poisson mean-predicted crashes at similar sites
PPSI=crash count -predicted crashes at similar sites
by PM
by PSI
by PPSI
by Raw data
by Expected rank
by Mode rank
by Median rank
by Probability of being the worst
Sum of PSI (1999-2002)
Number of intersections identified from 1993-1998 data
0
200
400
600
800
1000
1200
1400
0
20
40
60
80
100
120
140
160
180
200

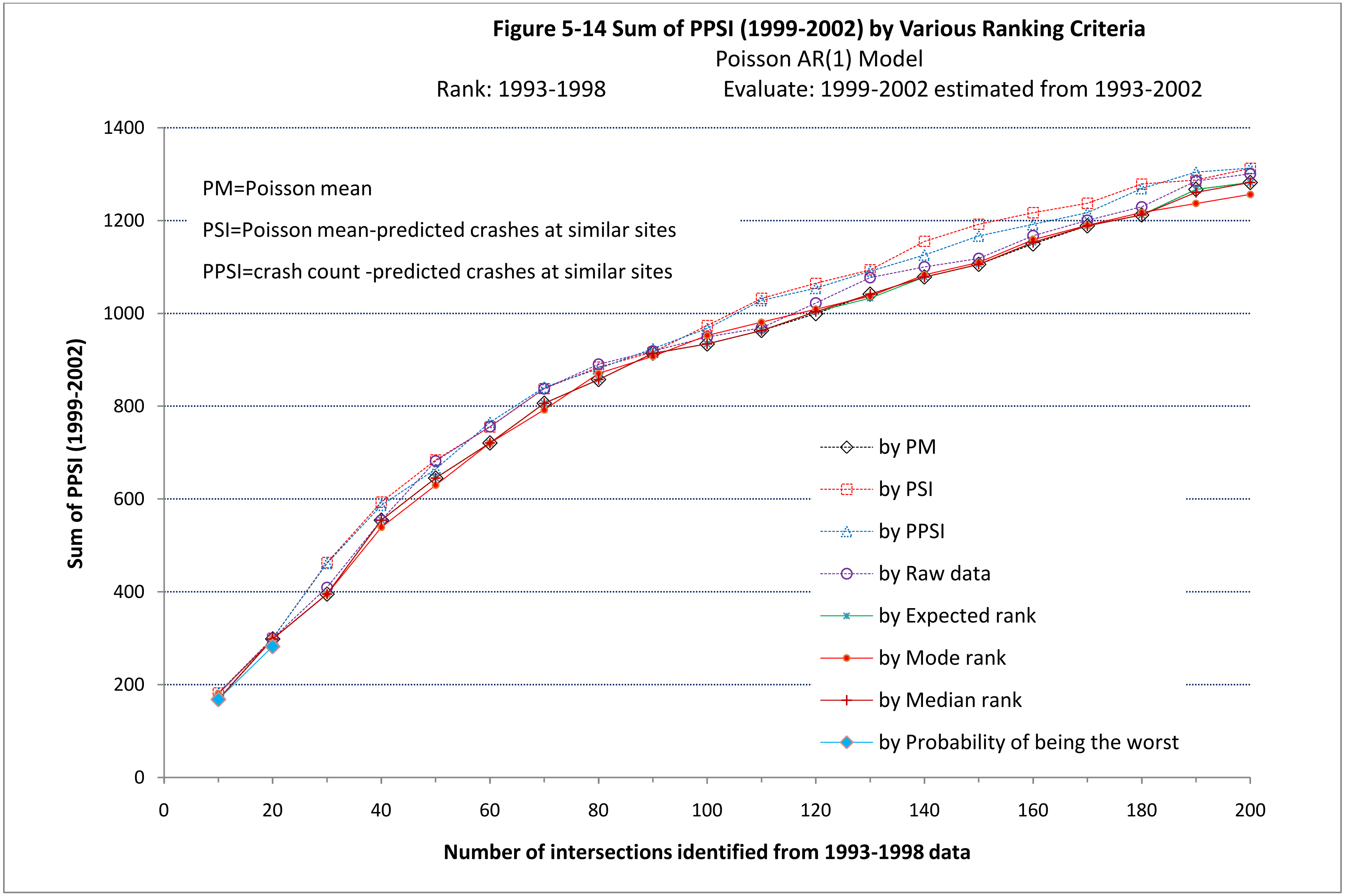


Figure 5-14 Sum of PPSI (1999-2002) by Various Ranking Criteria
Poisson AR(1) Model
Rank: 1993-1998
Evaluate: 1999-2002 estimated from 1993-2002
PM=Poisson mean
PSI=Poisson mean-predicted crashes at similar sites
PPSI=crash count -predicted crashes at similar sites
Sum of PPSI (1999-2002)
Number of intersections identified from 1993-1998 data
by PM
by PSI
by PPSI
by Raw data
by Expected rank
by Mode rank
by Median rank
by Probability of being the worst
0
200
400
600
800
1000
1200
1400
0
20
40
60
80
100
120
140
160
180
200

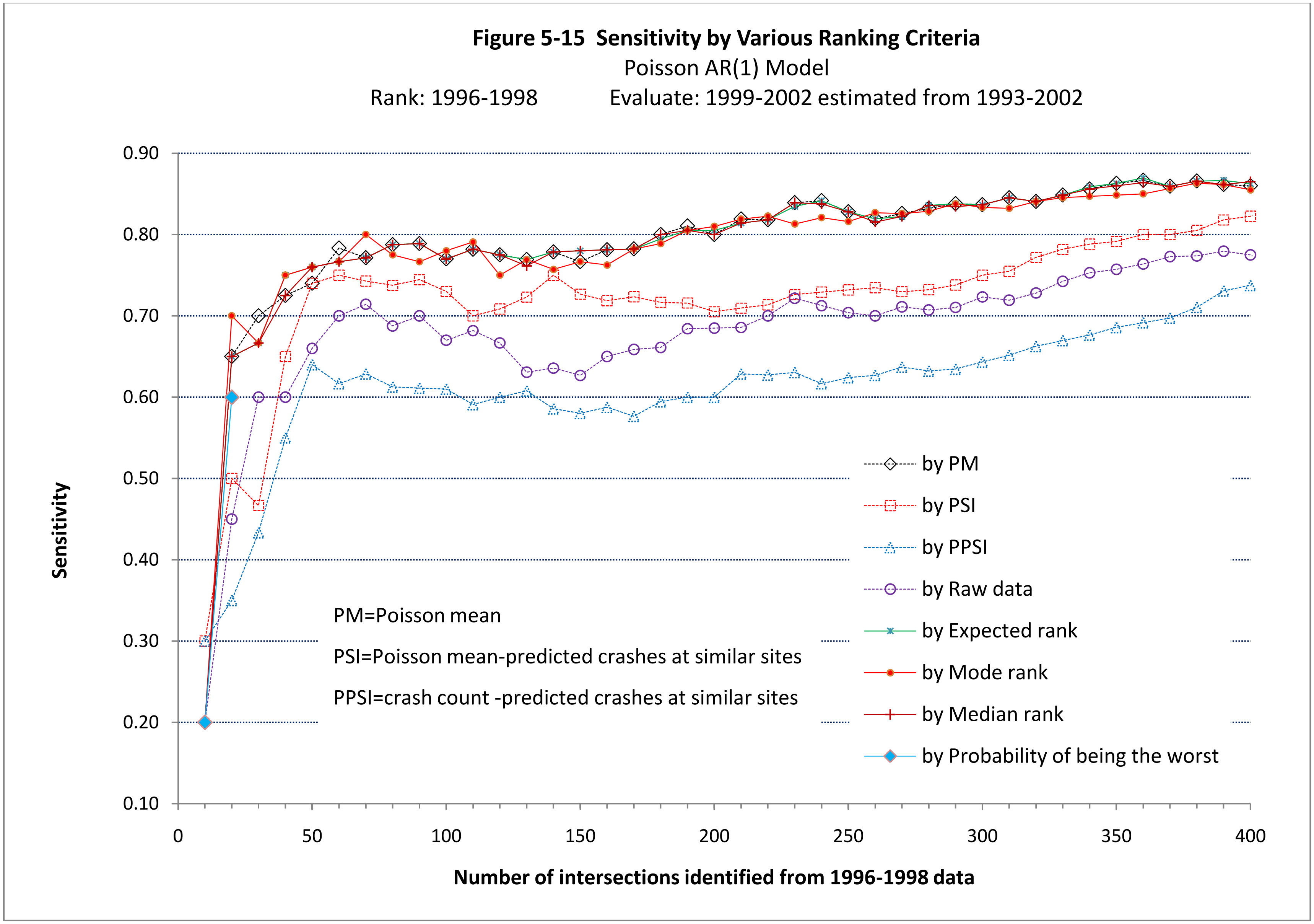


Figure 5-15 Sensitivity by Various Ranking Criteria
Poisson AR(1) Model
Rank: 1996-1998
Evaluate: 1999-2002 estimated from 1993-2002
Sensitivity
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0
50
100
150
200
250
300
350
400
Number of intersections identified from 1996-1998 data
PM=Poisson mean
PSI=Poisson mean-predicted crashes at similar sites
PPSI=crash count -predicted crashes at similar sites
by PM
by PSI
by PPSI
by Raw data
by Expected rank
by Mode rank
by Median rank
by Probability of being the worst

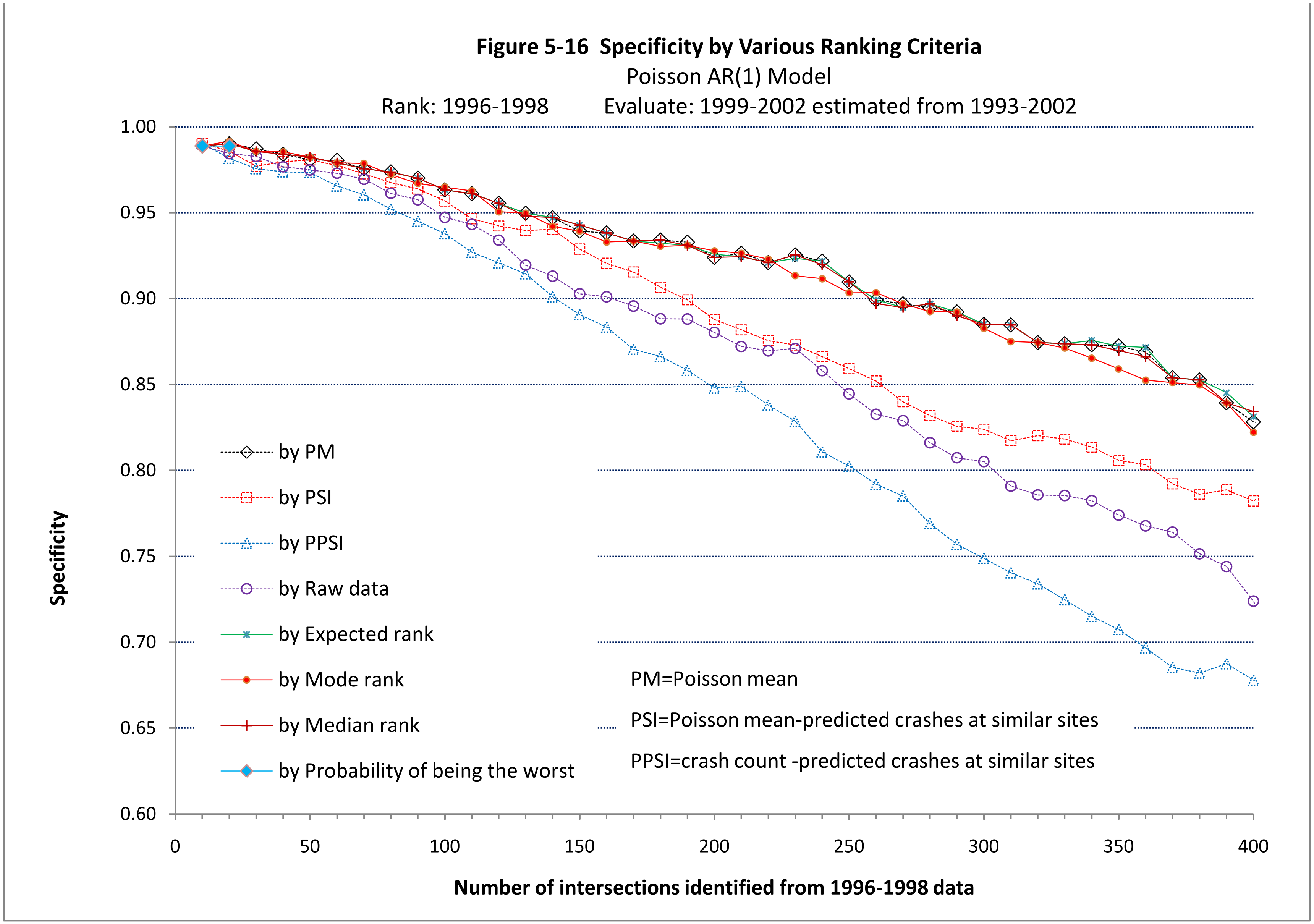


Figure 5-16 Specificity by Various Ranking Criteria
Poisson AR(1) Model
Rank: 1996-1998
Evaluate: 1999-2002 estimated from 1993-2002
Specificity
Number of intersections identified from 1996-1998 data
1.00
0.95
0.90
0.85
0.80
0.75
0.70
0.65
0.60
0
50
100
150
200
250
300
350
400
by PM
by PSI
by PPSI
by Raw data
by Expected rank
by Mode rank
by Median rank
by Probability of being the worst
PM=Poisson mean
PSI=Poisson mean-predicted crashes at similar sites
PPSI=crash count -predicted crashes at similar sites

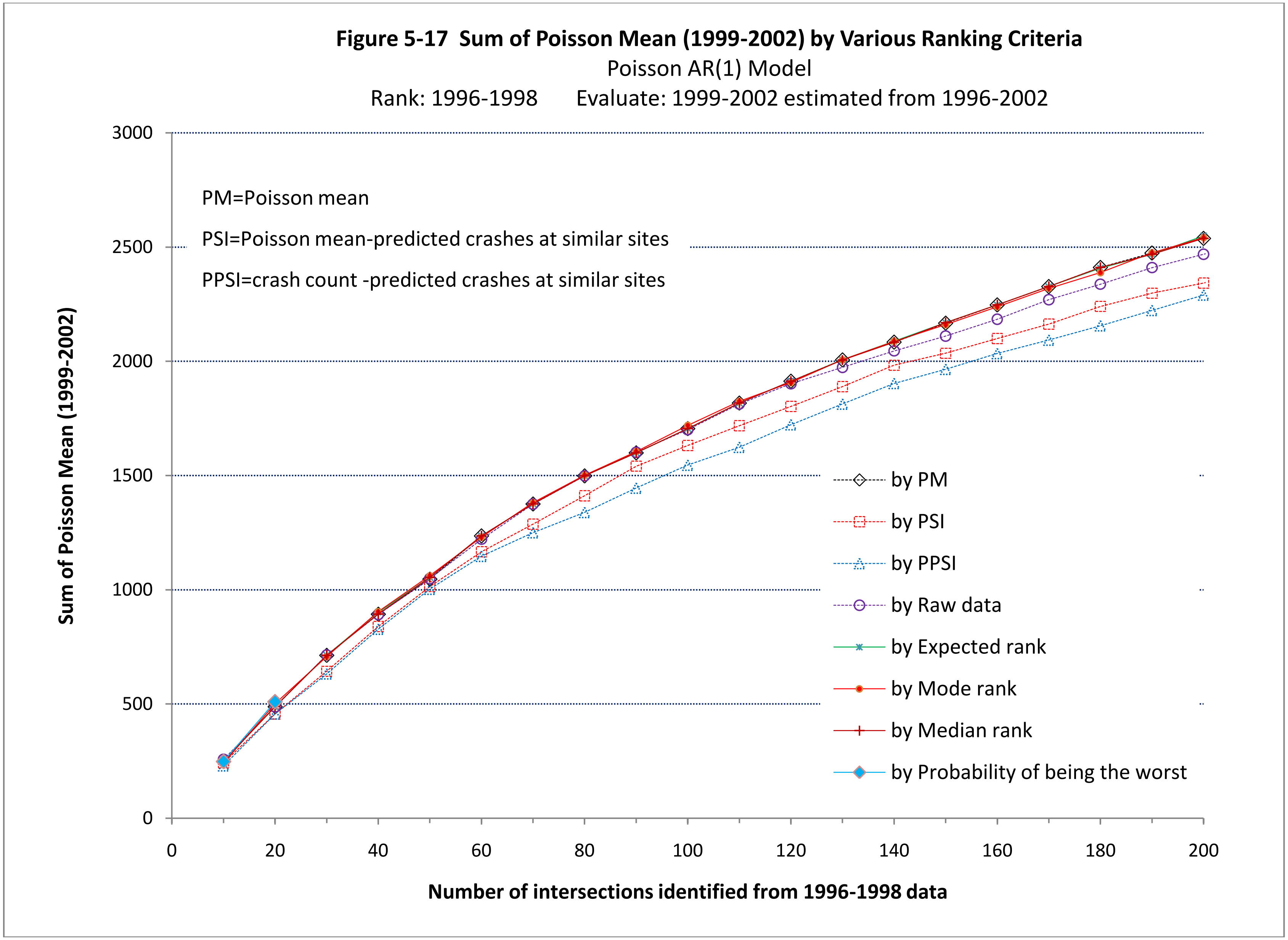


Figure 5-17 Sum of Poisson Mean (1999-2002) by Various Ranking Criteria
Poisson AR(1) Model
Rank: 1996-1998 Evaluate: 1999-2002 estimated from 1996-2002
PM=Poisson mean
PSI=Poisson mean-predicted crashes at similar sites
PPSI=crash count -predicted crashes at similar sites
Sum of Poisson Mean (1999-2002)
Number of intersections identified from 1996-1998 data
by PM
by PSI
by PPSI
by Raw data
by Expected rank
by Mode rank
by Median rank
by Probability of being the worst
0
500
1000
1500
2000
2500
3000
0
20
40
60
80
100
120
140
160
180
200

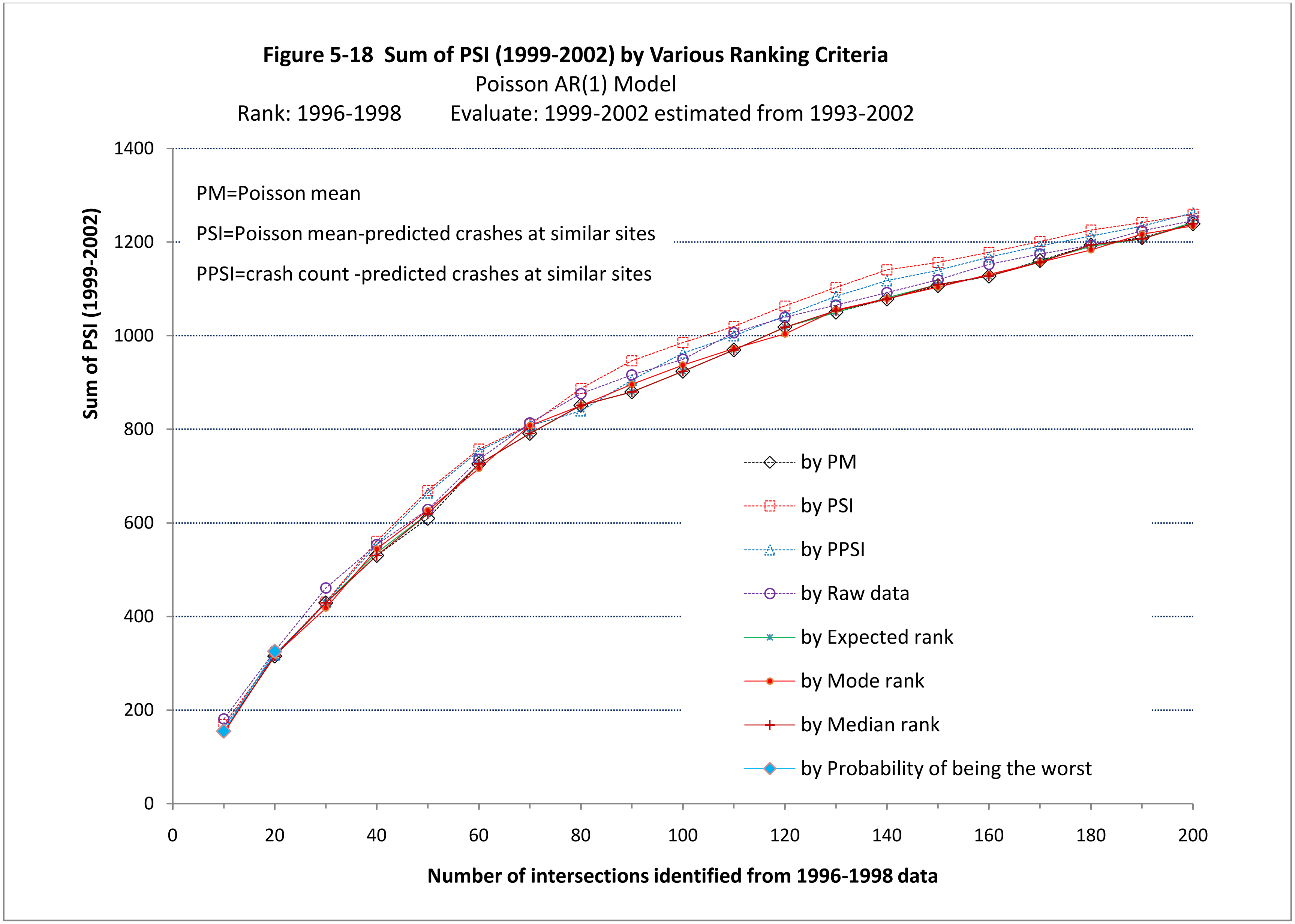


Figure 5-18 Sum of PSI (1999-2002) by Various Ranking Criteria
Poisson AR(1) Model
Rank: 1996-1998 Evaluate: 1999-2002 estimated from 1993-2002
PM=Poisson mean
PSI=Poisson mean-predicted crashes at similar sites
PPSI=crash count -predicted crashes at similar sites
Sum of PSI (1999-2002)
Number of intersections identified from 1996-1998 data
by PM
by PSI
by PPSI
by Raw data
by Expected rank
by Mode rank
by Median rank
by Probability of being the worst
0
200
400
600
800
1000
1200
1400
0
20
40
60
80
100
120
140
160
180
200

A. Analysis of ranking results from various ranking criteria

a. The probability of being the worst

It is easy to see that very few sites were identified as black spots based on the probability of being the worst. As a matter of fact, 40 sites in 1993-1998, 49 sites in 1996-1998, and only 20 sites in the evaluation period, have a probability that is greater than 0 of being the worst. The majority of the sample has a probability of 0. This ranking criterion, however, provides comparably good results for the identified top ranked sites from the graphs and tables. Sometimes it provides exactly the same results as those identified by PM in terms of the evaluation results. i.e., for the top 10 ranked sites, the criterion provides identical results as those ranked by PM and expected rank in terms of the sum of crash counts, sum of the PM, sum of the PSI, sum of the PPSI in the ranking period and evaluation criteria in the second period for the two time frames from the tables. For the top 20 ranked sites, this ranking criterion has less desirable results in comparison to the PM and expected rank from Tables 5-12 and 5-14 for the 6 year of ranking data. However, for the ranking period of 3 years, this criterion provides much better results in terms of higher values of PM, PSI, crash counts, and PPSI in the evaluation period. Indeed, this is the best ranking result for the top 20 ranked hazardous sites.

b. Posterior Poisson mean

Unlike other criteria, such as expected rank, mode rank, and the probability of being the worst, which involve extensive data process procedures, PM is easy to obtain during the modeling process. From the tables and graphs, PM can provide solid ranking results based on the evaluation criteria except for the sum of the PSI and sum of the PPSI. These evaluation criteria show that the PM gives the best or near best ranking results from the graphs and tables of the two time frames. This conclusion indicates that reliable ranking results might be obtained by the easily available ranking criterion, the PM of each site.

c. Posterior expected rank

The posterior expected rank provides almost the same ranking results as the PM from the graphs. Information from the top ranked 10 to 50 sites in the ranking and evaluation periods are listed in Tables 5-11 to 5-14. It can be seen that this criterion seems to provide exactly the same results as the PM in terms of the information in the ranking and evaluation periods from the graphs listed in Tables 5-11 and 5-12 for 1993-1998. For the second set of data (1996-1998), the expected rank has the same results as the PM for the top 10, 20 and 30 ranked sites, and for the top 40 and 50 ranked sites, the expected rank has higher values for the evaluation criteria, which indicates that it can provide a better ranked list than the PM. Note that this small difference may be a result of the procedure used to obtain the expected rank, as mentioned previously. Further examination of these two ranking criteria will be presented later.

d. Posterior median rank

Unlike other rank based criteria, the posterior median rank is available in WinBUGs output. It is used as a comparison of the expected rank and to see if it can provide comparably promising results. From Figures 5-9 to 5-18, the line which represents the sites ranked by median rank is totally overlapped by the lines ranked by PM and expected rank for the two ranking data. From Table 5-12, it can be found that for the top 10, 20 and 30 ranked sites, median rank has the same results as expected rank from the information in the ranking and evaluation periods for both time frames. For the top 40 and 50 ranked sites, the information in the ranking period for these two criteria is slightly different for 1993-1998 as shown in Table 5-12, but the evaluation results are completely the same, meaning that the sites that are not identified by expected ranking, but have very similar properties as those ranked by it, are identified by median ranking.

For the 1996-1998 ranking data, median rank has a slightly worse result than the expected ranking for the top 40 ranked sites, but provides the same results as PM from all the evaluation criteria. For the top 50 ranked sites, both median and expected ranks have the same evaluation results.

From above evaluation results, it can be found that generally, median rank can provide the same promising results as expected rank and slightly better results than PM. This indicates that median rank might be used as a substitute for expected rank in hotspot identification.

Table 5-11 Summary of Evaluation Results by Various Ranking Criteria (P_AR(1))

Hotspots identified from 1993-1998 data

	By PM= Poisson mean									
ranked sites	$\sum tot_{93\text{-}98}$	$\sum PM_{93\text{-}98}$	$\sum PSI_{93\text{-}98}$	$\sum PPSI_{93\text{-}98}$	Sensitivity	Specificity	$\sum tot_{99\text{-}02}$	$\sum PM_{99\text{-}02}$	$\sum PSI_{99\text{-}02}$	$\sum PPSI_{99\text{-}02}$
10	391	368	239	262	0.30	0.99	263	263	168	168
20	732	688	438	483	0.60	0.99	485	485	298	298
30	1033	975	560	615	0.70	0.99	702	698	389	395
40	1297	1221	704	777	0.78	0.99	932	920	537	554
50	1522	1436	784	865	0.80	0.99	1116	1097	619	645
By PSI=Poisson mean-predicted crashes at similar sites										
ranked sites	$\sum tot_{93\text{-}98}$	$\sum PM_{93\text{-}98}$	$\sum PSI_{93\text{-}98}$	$\sum PPSI_{93\text{-}98}$	Sensitivity	Specificity	$\sum tot_{99\text{-}02}$	$\sum PM_{99\text{-}02}$	$\sum PSI_{99\text{-}02}$	$\sum PPSI_{99\text{-}02}$
10	386	355	266	297	0.40	0.99	243	244	180	182
20	722	671	458	509	0.55	0.99	454	463	306	300
30	1015	947	614	682	0.57	0.98	707	698	451	463
40	1250	1162	742	830	0.70	0.98	899	876	568	593
50	1460	1353	854	962	0.74	0.98	1047	1022	657	684
By PPSI=crash counts-predicted at similar sites										
ranked sites	$\sum tot_{93\text{-}98}$	$\sum PM_{93\text{-}98}$	$\sum PSI_{93\text{-}98}$	$\sum PPSI_{93\text{-}98}$	Sensitivity	Specificity	$\sum tot_{99\text{-}02}$	$\sum PM_{99\text{-}02}$	$\sum PSI_{99\text{-}02}$	$\sum PPSI_{99\text{-}02}$
10	386	355	266	297	0.40	0.99	243	244	180	182
20	722	671	458	509	0.40	0.98	454	463	306	300
30	1009	940	614	683	0.50	0.98	700	693	452	461
40	1248	1159	742	831	0.60	0.98	890	870	563	587
50	1449	1339	853	963	0.62	0.97	1017	998	643	664
By Crash Frequency										
ranked sites	$\sum tot_{93\text{-}98}$	$\sum PM_{93\text{-}98}$	$\sum PSI_{93\text{-}98}$	$\sum PPSI_{93\text{-}98}$	Sensitivity	Specificity	$\sum tot_{99\text{-}02}$	$\sum PM_{99\text{-}02}$	$\sum PSI_{99\text{-}02}$	$\sum PPSI_{99\text{-}02}$
10	393	367	253	279	0.20	0.99	252	253	171	169
20	733	686	448	494	0.45	0.98	478	479	300	299
30	1036	973	588	650	0.53	0.98	693	693	407	409
40	1297	1221	704	777	0.65	0.98	932	920	537	554
50	1525	1434	812	901	0.74	0.98	1133	1106	650	682

Table 5-12 Summary of Evaluation Results by Various Ranking Criteria (P_AR(1))

Hotspots identified from 1993-1998 data

	By Posterior Expected Rank									
ranked sites	$\sum tot_{93-98}$	$\sum PM_{93-98}$	$\sum PSI_{93-98}$	$\sum PPSI_{93-98}$	Sensitivity	Specificity	$\sum tot_{99-02}$	$\sum PM_{99-02}$	$\sum PSI_{99-02}$	$\sum PPSI_{99-02}$
10	391	368	239	262	0.30	0.99	263	263	168	168
20	732	688	438	483	0.60	0.99	485	485	298	298
30	1033	975	560	615	0.70	0.99	702	698	389	395
40	1297	1221	704	777	0.78	0.99	932	920	537	554
50	1522	1436	784	865	0.80	0.99	1116	1097	619	645
	By Posterior Median Rank									
ranked sites	$\sum tot_{93-98}$	$\sum PM_{93-98}$	$\sum PSI_{93-98}$	$\sum PPSI_{93-98}$	Sensitivity	Specificity	$\sum tot_{99-02}$	$\sum PM_{99-02}$	$\sum PSI_{99-02}$	$\sum PPSI_{99-02}$
10	392	367	245	269	0.30	0.99	263	263	168	168
20	732	688	438	483	0.60	0.99	485	485	298	298
30	1033	975	560	615	0.70	0.99	702	698	389	395
40	1294	1220	692	762	0.78	0.99	932	920	537	554
50	1517	1434	773	852	0.80	0.99	1116	1097	619	645
	By Posterior Mode Rank									
ranked sites	$\sum tot_{93-98}$	$\sum PM_{93-98}$	$\sum PSI_{93-98}$	$\sum PPSI_{93-98}$	Sensitivity	Specificity	$\sum tot_{99-02}$	$\sum PM_{99-02}$	$\sum PSI_{99-02}$	$\sum PPSI_{99-02}$
10	391	368	239	262	0.30	0.99	271	269	178	180
20	732	688	438	483	0.65	0.99	485	485	298	298
30	1033	975	560	615	0.67	0.99	702	698	389	395
40	1297	1221	704	777	0.73	0.98	921	912	524	539
50	1522	1436	784	865	0.78	0.98	1107	1090	606	630
	By Probability of being the worst									
ranked sites	$\sum tot_{93-98}$	$\sum PM_{93-98}$	$\sum PSI_{93-98}$	$\sum PPSI_{93-98}$	Sensitivity	Specificity	$\sum tot_{99-02}$	$\sum PM_{99-02}$	$\sum PSI_{99-02}$	$\sum PPSI_{99-02}$
10	391	368	239	262	0.30	0.99	263	263	168	168
20	732	687	440	485	0.55	0.99	463	469	285	282

Table 5-13 Summary of Evaluation Results by Various Ranking Criteria (P_AR(1))

Hotspots identified from 1996-1998 data

	by PM= Poisson mean									
ranked sites	$\sum tot_{96-98}$	$\sum PM_{96-98}$	$\sum PSI_{96-98}$	$\sum PPSI_{96-98}$	Sensitivity	Specificity	$\sum tot_{99-02}$	$\sum PM_{99-02}$	$\sum PSI_{99-02}$	$\sum PPSI_{99-02}$
10	226	204	140	162	0.20	0.99	241	248	155	147
20	408	364	241	284	0.65	0.99	484	488	315	313
30	557	502	299	352	0.70	0.99	717	713	429	436
40	696	626	365	432	0.73	0.98	904	893	530	547
50	826	741	426	507	0.74	0.98	1055	1047	609	624
by PSI=Poisson mean-predicted crashes at similar sites										
ranked sites	$\sum tot_{96-98}$	$\sum PM_{96-98}$	$\sum PSI_{96-98}$	$\sum PPSI_{96-98}$	Sensitivity	Specificity	$\sum tot_{99-02}$	$\sum PM_{99-02}$	$\sum PSI_{99-02}$	$\sum PPSI_{99-02}$
10	228	201	150	177	0.30	0.99	235	243	169	164
20	400	349	253	304	0.50	0.99	453	457	318	317
30	550	481	331	401	0.47	0.98	641	643	431	431
40	684	597	401	488	0.65	0.98	848	839	561	573
50	809	706	462	565	0.74	0.98	1034	1014	669	692
by PPSI=crash counts-predicted at similar sites										
ranked sites	$\sum tot_{96-98}$	$\sum PM_{96-98}$	$\sum PSI_{96-98}$	$\sum PPSI_{96-98}$	Sensitivity	Specificity	$\sum tot_{99-02}$	$\sum PM_{99-02}$	$\sum PSI_{99-02}$	$\sum PPSI_{99-02}$
10	225	197	150	178	0.30	0.99	219	230	162	154
20	400	349	253	304	0.35	0.98	453	457	318	317
30	544	472	330	402	0.43	0.98	626	632	430	427
40	681	593	400	488	0.55	0.97	835	827	553	564
50	805	700	461	566	0.64	0.97	1023	1002	664	687
by Crash Frequency										
ranked sites	$\sum tot_{96-98}$	$\sum PM_{96-98}$	$\sum PSI_{96-98}$	$\sum PPSI_{96-98}$	Sensitivity	Specificity	$\sum tot_{A}$	$\sum PM_{A}$	$\sum PSI_{A}$	$\sum PPSI_{A}$
10	228	202	149	175	0.20	0.99	255	258	181	178
20	408	361	248	295	0.45	0.98	484	487	326	326
30	559	497	317	377	0.60	0.98	722	715	461	471
40	699	623	381	456	0.60	0.98	903	892	554	567
50	829	739	437	524	0.66	0.97	1062	1049	628	647

Table 5-14 Summary of Evaluation Results by Various Ranking Criteria (P_AR(1))

Hotspots identified from 1996-1998 data

	By Posterior Expected Rank									
ranked sites	$\sum tot_{96\text{-}98}$	$\sum PM_{96\text{-}98}$	$\sum PSI_{96\text{-}98}$	$\sum PPSI_{96\text{-}98}$	Sensitivity	Specificity	$\sum tot_{99\text{-}02}$	$\sum PM_{99\text{-}02}$	$\sum PSI_{99\text{-}02}$	$\sum PPSI_{99\text{-}02}$
10	226	204	140	162	0.20	0.99	241	248	155	147
20	408	364	241	284	0.65	0.99	484	488	315	313
30	557	502	299	352	0.67	0.99	717	713	429	436
40	696	626	365	432	0.75	0.99	916	901	536	554
50	827	741	428	510	0.76	0.98	1066	1054	619	638
	By Posterior Median Rank									
ranked sites	$\sum tot_{96\text{-}98}$	$\sum PM_{96\text{-}98}$	$\sum PSI_{96\text{-}98}$	$\sum PPSI_{96\text{-}98}$	Sensitivity	Specificity	$\sum tot_{99\text{-}02}$	$\sum PM_{99\text{-}02}$	$\sum PSI_{99\text{-}02}$	$\sum PPSI_{99\text{-}02}$
10	226	204	140	162	0.20	0.99	241	248	155	147
20	408	364	241	284	0.65	0.99	484	488	315	313
30	557	502	299	352	0.67	0.99	717	713	429	436
40	696	626	365	432	0.73	0.98	904	893	530	547
50	827	741	428	510	0.76	0.98	1066	1054	619	638
	By Posterior Mode Rank									
ranked sites	$\sum tot_{96\text{-}98}$	$\sum PM_{96\text{-}98}$	$\sum PSI_{96\text{-}98}$	$\sum PPSI_{96\text{-}98}$	Sensitivity	Specificity	$\sum tot_{99\text{-}02}$	$\sum PM_{99\text{-}02}$	$\sum PSI_{99\text{-}02}$	$\sum PPSI_{99\text{-}02}$
10	226	203	141	164	0.20	0.99	234	242	151	143
20	405	364	232	273	0.70	0.99	501	501	317	319
30	555	501	295	346	0.67	0.99	707	707	418	421
40	695	625	367	434	0.75	0.99	918	904	544	564
50	823	738	425	506	0.76	0.98	1078	1061	627	651
	By Probability of being the worst									
ranked sites	$\sum tot_{96\text{-}98}$	$\sum PM_{96\text{-}98}$	$\sum PSI_{96\text{-}98}$	$\sum PPSI_{96\text{-}98}$	Sensitivity	Specificity	$\sum tot_{99\text{-}02}$	$\sum PM_{99\text{-}02}$	$\sum PSI_{99\text{-}02}$	$\sum PPSI_{99\text{-}02}$
10	226	204	140	162	0.20	0.99	241	248	155	147
20	405	363	233	274	0.60	0.99	511	510	325	329

e. Posterior mode rank

Mode rank generally provides similar results as expected rank and even better results at times. For example, for the three years of ranking data (Table 5-14), it provides slightly better results for the top 20, 40 and 50 ranked sites based on the evaluation criteria in the second time period, while expected rank has somewhat better results for the top 10 and 30 ranked sites. For ranking data in 1993-1998, mode rank has the same or somewhat better results for the top 10, 20 and 30 ranked sites while expected rank is better for the top 40 and 50 ranked sites.

f. PSI and PPSI

The evaluation results from the PSI are not as good as had been expected and this, by the way, confirms the results by Elvik (2008a) who used the EB method, and sensitivity and specificity as evaluation criteria. Only when based on the criteria, sum of the PSI and sum of the PPSI, will the PSI criterion give the best results. For the other evaluation criteria, the PSI provides very poor results. The PPSI provides even worse ranking results than the PSI. As a matter of fact, it is the worst ranking criterion based on other evaluation criteria other than the sum of the PSI and sum of the PPSI in the second period for the ranked sites. This indicates that PSI and PPSI cannot be used alone as ranking criteria. This is theoretically correct since sites that do not have safety issues should not be selected for further investigation even if they can be greatly improved with treatment.

g. Raw data (crash count)

It is not surprising to see that the performance of raw data is poor, if not the worst, in comparison to all of the evaluation criteria. This is mainly due to the RTM problem.

In all, PM, expected rank, mode rank, median rank and the probability of being the worst from the posterior seem to be reliable ranking criteria. Raw data cannot be used as a ranking criterion alone but might be combined with other criteria to conduct a ranking analysis. PSI provides the worse results out of all five reliable ranking criteria, but is better than PPSI, except for the sum of the PSI and sum of the PPSI. It might be used as a ranking criterion if and only if the objective of the network ranking is to find an ordered list which has the greatest potential for improvement regardless of current safety site conditions, which sounds unreasonable. Thus we can conclude that PSI and PPSI actually cannot be used as ranking criteria. However, PSI might be used as a second level criterion for ranking, for example, safety issues within sites with a small difference of the PM or expected rank, PM or the expected rank could be regarded similarly. Then, priority is placed on sites with high values of PSI. In addition, the sum of the PSI could be used as an evaluation criterion for hotspot identification.

B. Best ranked results by ranking criteria

From the above figures and tables, the following best ranking results can be obtained based on all of the evaluation criteria, except for the sum of the PSI and sum of the PPSI, where PSI has the best ranking results.

a. Ranking data in 1993-1998

The top 10 best ranked sites were obtained by mode rank (Table 5-12).

The top 20 best ranked sites were identified by mode rank (Table 5-12). The Poisson mean, expected rank, and median rank provide the same evaluation results, but with less sensitivity value (false identification of 1 site).

The top 30 best ranked sites were ranked by Poisson mean, median rank and expected rank while mode rank has the same evaluation results except that there is a somewhat lower sensitivity (false identification of 1 site).

The top 40 and 50 best ranked sites were ranked by Poisson mean, median rank and expected rank (Tables 5-11 and 5-12).

For the number of ranked sites greater than 50, generally, Poisson mean, median rank and expected rank provide somewhat better results than mode rank in terms of sensitivity and specificity as is evident from Figures 5-9 and 5-10. There are no obvious differences from Figure 5-11, and it seems that the previously mentioned 3 ranking criteria and mode rank have the same promising results in terms of the sum of the Poisson mean.

b. Ranking data in 1996-1998

The top 10 best ranked sites were obtained by Poisson mean, expected rank, median rank and the probability of being the worst (Tables 5-13 and 5-14).

The top 20 best ranked sites were obtained by the probability of being the worst (Table 5-14).

The top 30 best ranked sites were obtained by Poisson mean, expected rank, and median rank (Tables 5-13 and 5-14).

The top 40 and 50 best ranked sites were identified by mode rank (Table 5-14).

For the number of ranked sites greater than 50, generally, Poisson mean, median rank, expected rank and mode rank provide the same promising results.

From the above analysis, the proposed ranking criterion (mode rank) provides the best results, and at times, this is the case for the very top ranked sites; thus it is of interest to explore this ranking criterion for network ranking.

C. Comparison of ranked sites by other criteria through posterior Poisson mean and posterior expected rank

Based on the above analysis, it is good to know the similarities among the ranking criteria in terms of identified sites. To this end, the ranked sites by PM and expected rank were used as a baseline, and sites identified by other criteria are compared with those ranked by PM or expected rank using $\mathrm{P}_{\text{without order}}$ and $\mathrm{P}_{\text{with order}}$, where $\mathrm{P}_{\text{without order}}$ and $\mathrm{P}_{\text{with order}}$ are

$$\mathrm{P}_{\text{without order}} = \frac{\mathrm{NS}_{\text{without order}}}{\mathrm{N1}} \qquad (5-12)$$

$$\mathrm{P}_{\text{with order}} = \frac{\mathrm{NS}_{\text{with order}}}{\mathrm{N1}} \qquad (5-13)$$

$\mathrm{NS}_{\text{without order}} =$ number of identical sites that occur in the top ranked group ranked by both types of ranking criteria without considering the order of the ranked sites in that group,

$\mathrm{NS}_{\text{with order}} =$ number of the same sites which have the exact order in the top ranked group by both criteria, in consideration of the order of the ranked sites in the top ranked group, and

N1= cutoff number of top ranked sites.

The higher value of $\mathrm{P}_{\text{without order}}$ and $\mathrm{P}_{\text{with order}}$ gives better and consistent ranking results provided by both criteria, and vice versa. If the ranked list identified by PM is used as a baseline, and $\mathrm{P}_{\text{without order}} = 1$ for the top 20 ranked sites, this indicates the top 20 ranked sites are the same as ranked by PM, but the order may be different. However, if $P_{\text{with order}} = 1$, this means that both types of ranking criteria provide an identical ordered ranking list. The comparison results in terms of $P_{\text{without order}}$ and $P_{\text{with order}}$ to the posterior Poisson mean and posterior expected rank for various ranking criteria including EB are shown in Figures 5-19 to 5-24 for 1993-1998 and in Figures 5-25 and 5-28 for 1996-1998.

a. The order of ranked sites is not considered

The comparison results with ranked by Poisson mean or by expected rank are shown in Figures 5-19 to 5-22 for 1993-1998 and in Figures 5-25 to 5-26 for 1996-1998.

It can be seen that expected rank and median rank provide exactly the same ranked sites as PM for the top 10 to 100 ranked sites for 1993-1998 as shown in Figure 5-19 to 5-22 if the order of each site is not considered in the top ranked group. This is slightly different for the second ranking period data (1996-1998) for the top 40 and 50 ranked sites. Generally, the ranked sites are almost the same as identified by expected rank, median rank and PM. As a result, the shape of Figures 5-19, 5-20 (relation to Poisson mean) and 5-21, 5-22 (relation to expected rank) for 1993-1998, 5-25 (relation to Poisson mean) and 5-26 (relation to expected rank) for 1996-1998, are almost the same. The posterior mode rank provides at least 95% of the same ranked sites as PM or by posterior expected rank except for the top 10 ranked sites, where mode rank has 90% of the same sites as those by expected rank or PM. PPSI and PSI provide very different sites in comparison to those ranked by PM or expected rank. The expected crashes for EB provides more than 90% of the same sites as those from PM, except for the top ten sites which are ranked with data from 1993-1998, for which 80% of the expected crashes are identified from the same sites as those by PM. PSI ranked by the EB method provides the most different sites from those ranked by PM or expected rank.

It should be noted, however, this does not necessarily mean that PM, expected rank and median rank provide the best results. Rather, they provide a basic idea on the amount of similarity in the ranked results identified by these various ranking criteria. In fact, sometimes mode rank can provide even better ranking results as analyzed above.

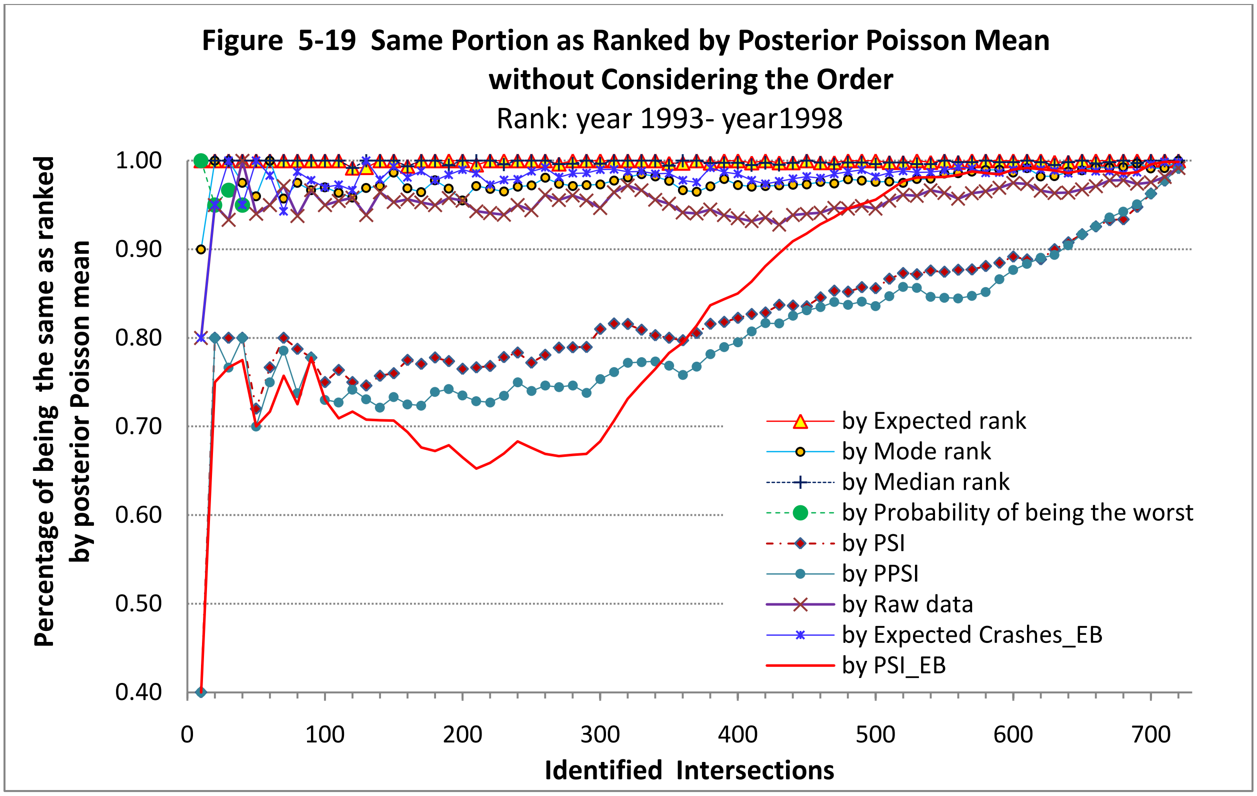


Figure 5-19 Same Portion as Ranked by Posterior Poisson Mean without Considering the Order

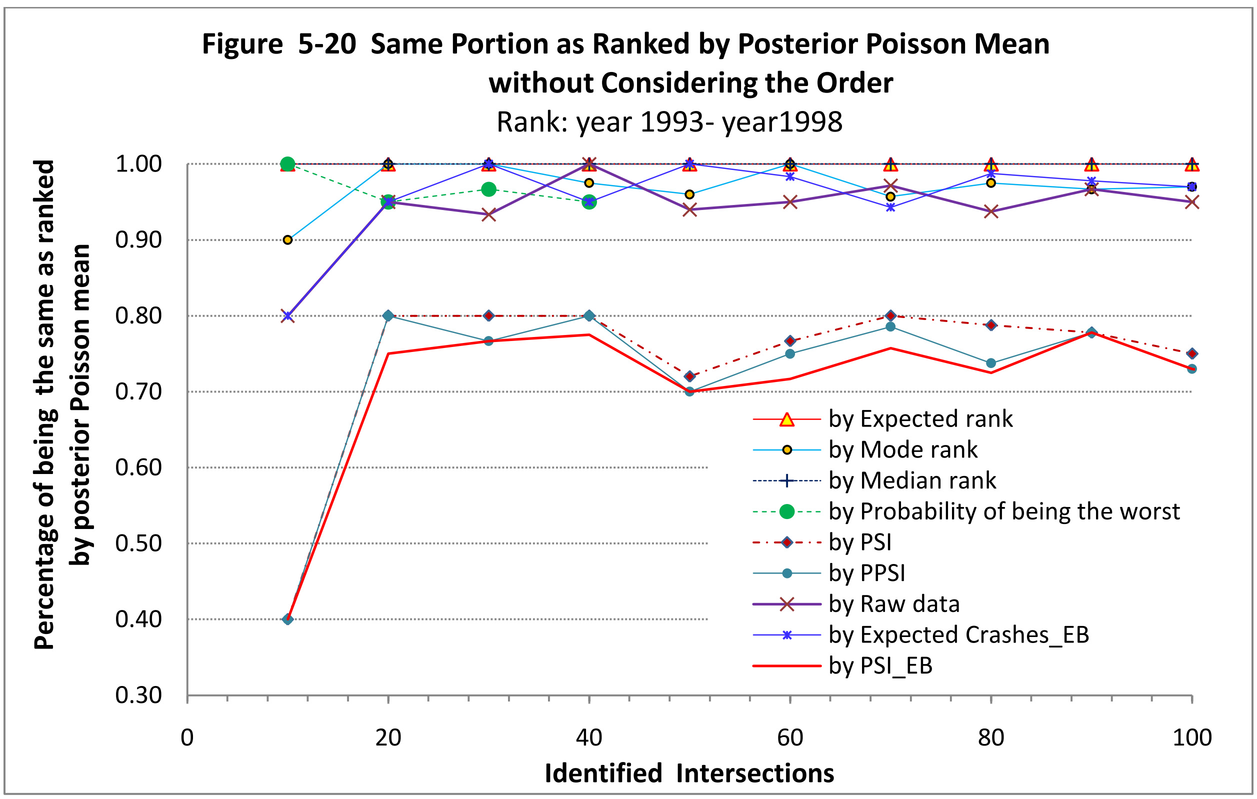


Figure 5-20 Same Portion as Ranked by Posterior Poisson Mean without Considering the Order

Figure 5-21 Same Portion as Ranked by Posterior Expected Rank without Considering the Order

Rank: year 1993- year1998

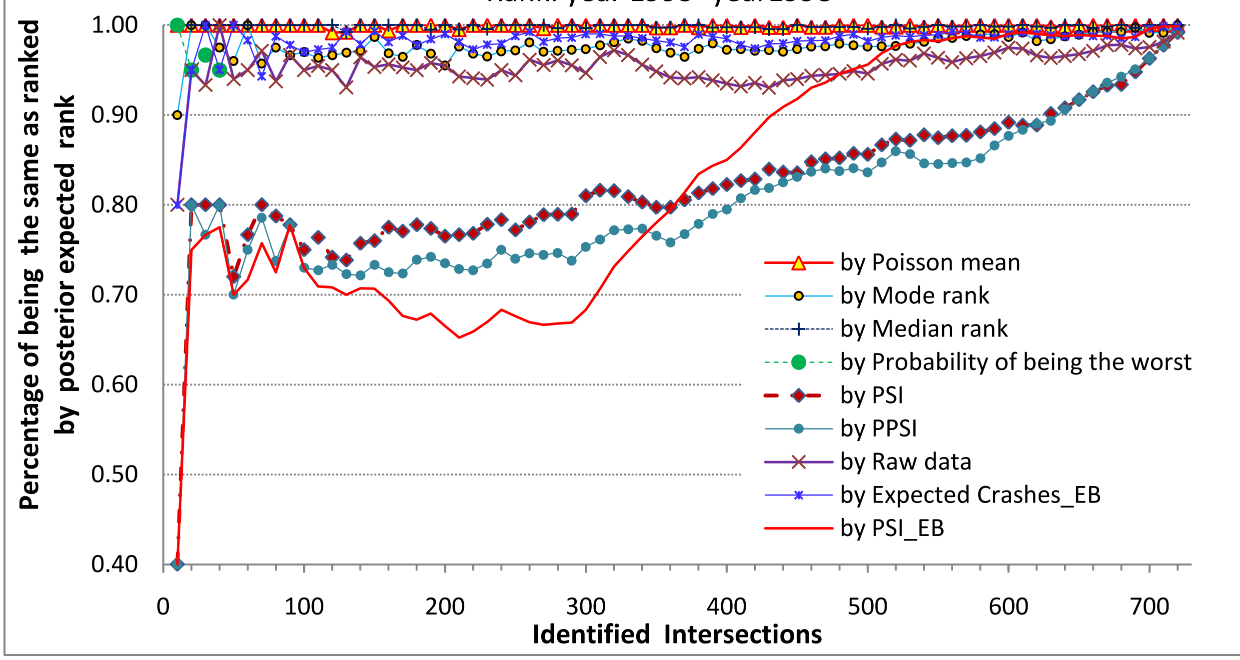
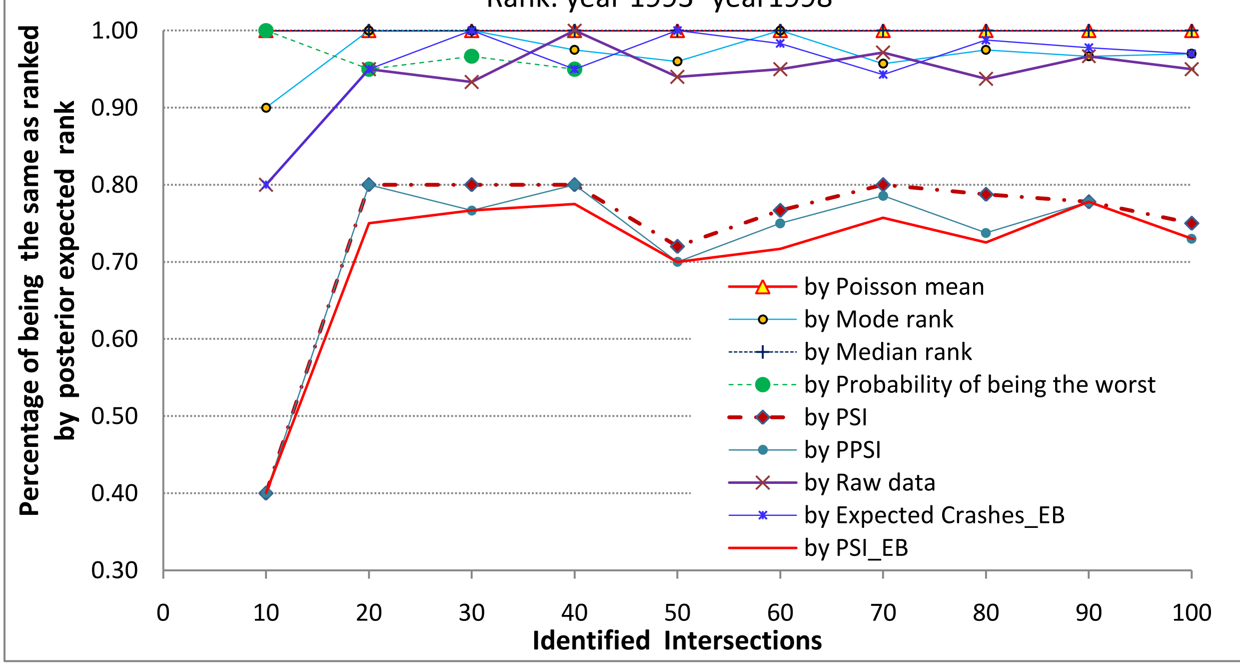


Figure 5-22 Same Portion as Ranked by Posterior Expected Rank without Considering the Order

Rank: year 1993- year1998



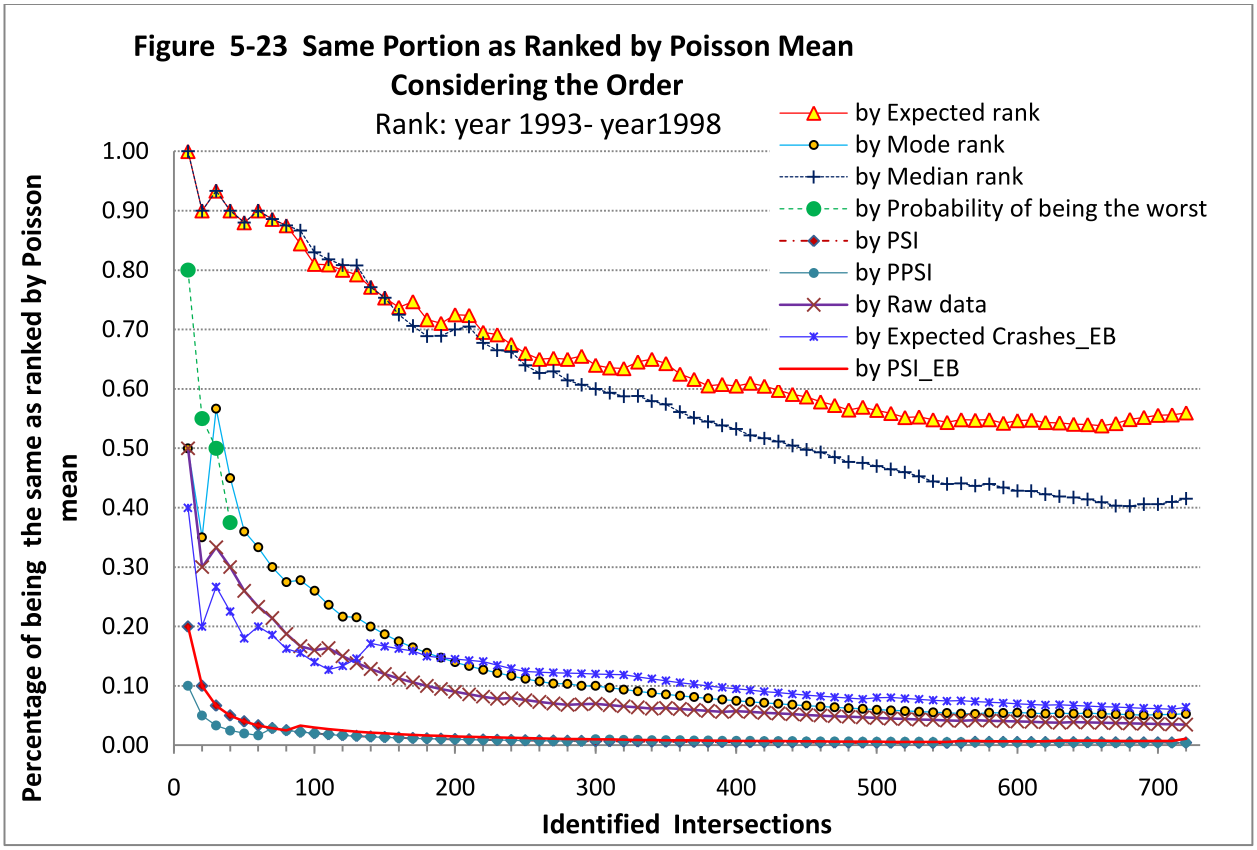


Figure 5-23 Same Portion as Ranked by Poisson Mean
Considering the Order
Rank: year 1993- year1998
by Expected rank
by Mode rank
by Median rank
by Probability of being the worst
by PSI
by PPSI
by Raw data
by Expected Crashes_EB
by PSI_EB
Percentage of being the same as ranked by Poisson mean
Identified Intersections
1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
0
100
200
300
400
500
600
700

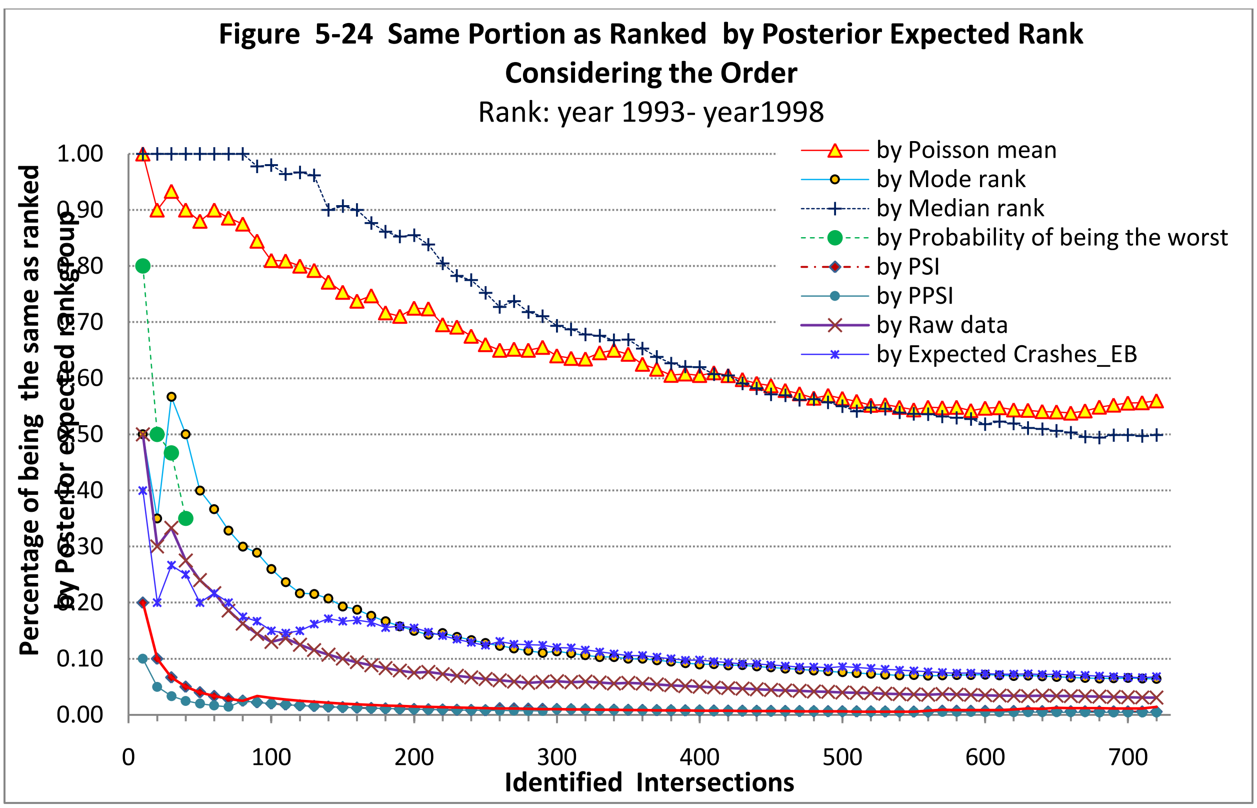
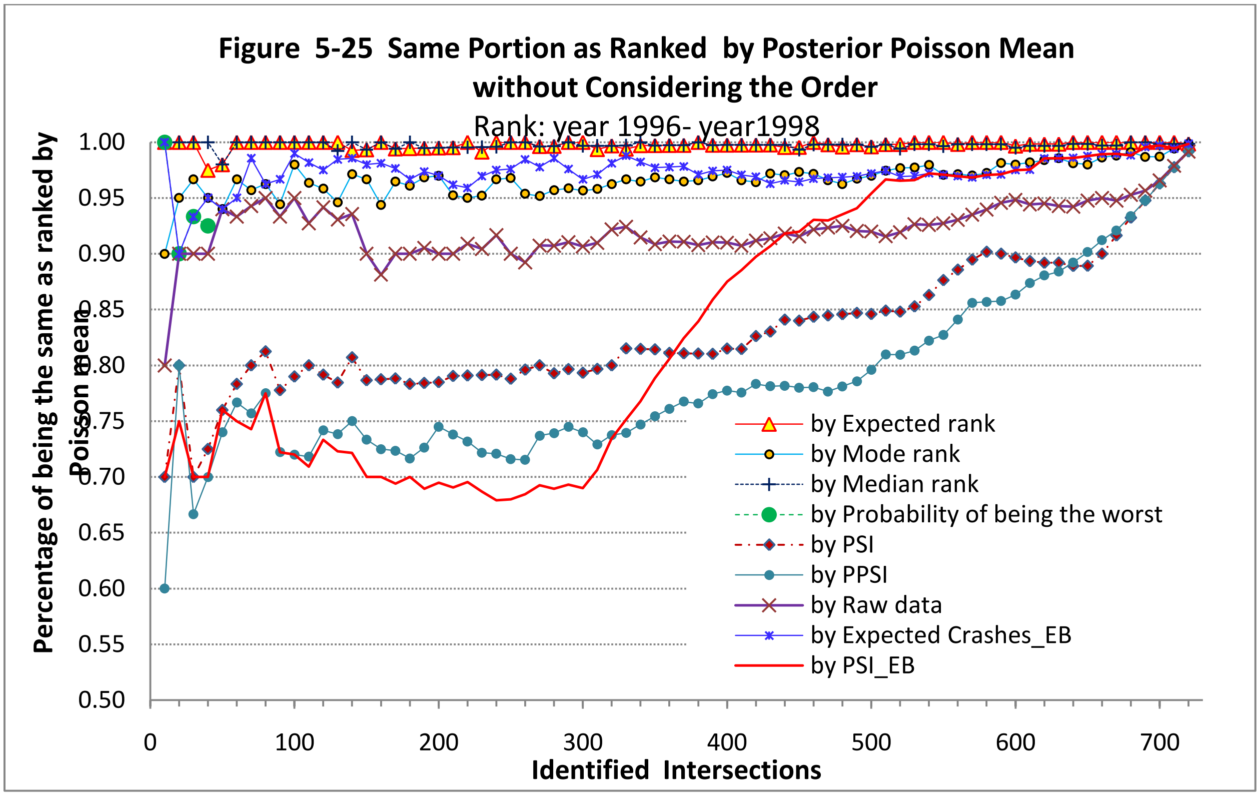


Figure 5-24 Same Portion as Ranked by Posterior Expected Rank
Considering the Order
Rank: year 1993- year1998
by Poisson mean
by Mode rank
by Median rank
by Probability of being the worst
by PSI
by PPSI
by Raw data
by Expected Crashes_EB
Percentage of being the same as ranked by Posterior expected rank group
Identified Intersections
1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
0
100
200
300
400
500
600
700

b. In consideration of the order of ranked sites

The results are presented in Figures 5-23 to 5-24 for 1993-1998 and Figures 5-27 to 5-28 for 1996-1998.



If the order of the ranked sites is considered, it can be seen that the results by PM is actually quite different from those by expected and median ranks even if they provide almost the same results without consideration of the order. Median rank provides the most similar list to expected rank. Median rank has the same ordered top 80 ranked sites as by the expected rank using data in 1993-1998, while it provides the same ordered top 30 sites as by the expected rank using three years of ranking data. This confirms that median rank might be a reliable substitute for expected rank in hot spot identification. If the ranked order of each site is a concern, EB by expected crashes provides a far different list in comparison to those by Poisson mean, expected rank, median rank, mode rank and the probability of being the worst, especially for the top ranked limited sites. This might be a hint that the EB method provides much poorer ranking results if the ranked order of each site is a concern.

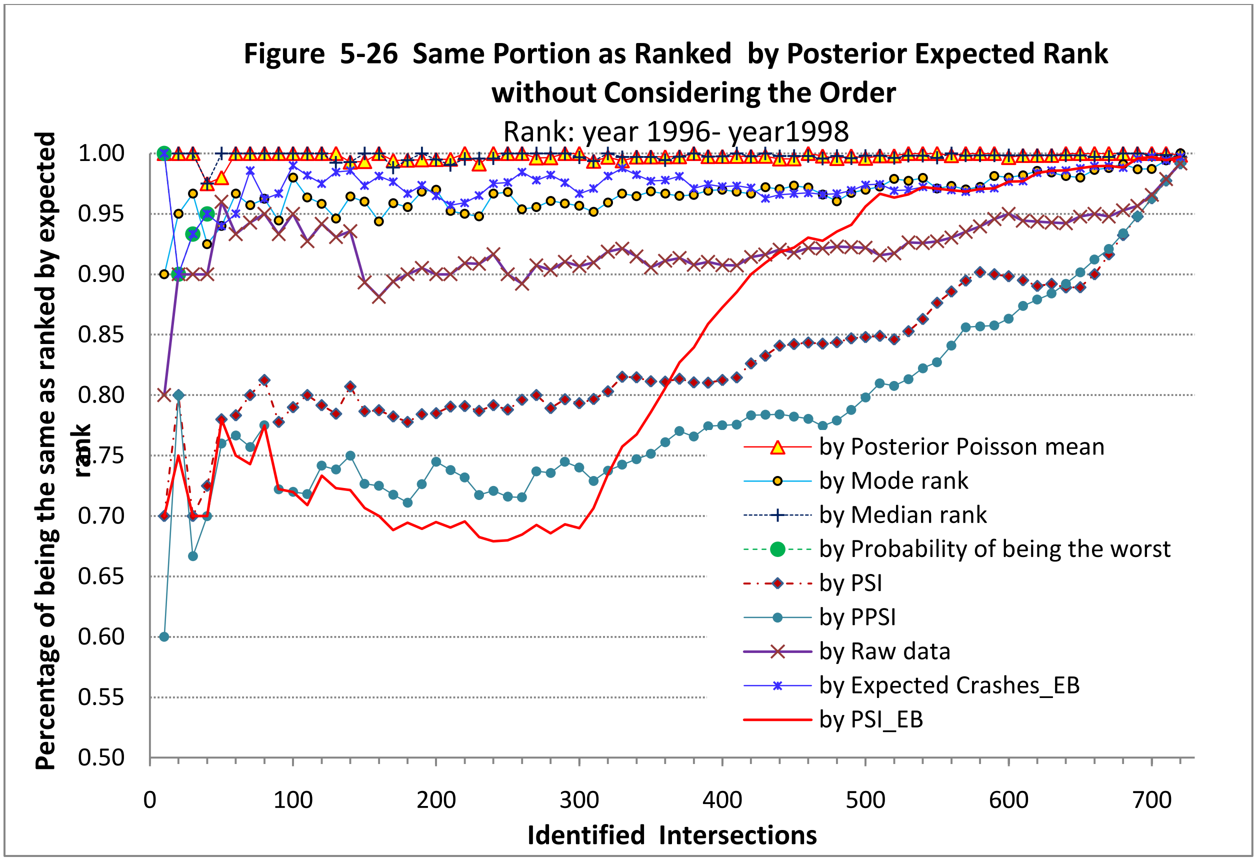


Figure 5-26 Same Portion as Ranked by Posterior Expected Rank
without Considering the Order
Rank: year 1996- year1998
Percentage of being the same as ranked by expected rank
Identified Intersections
by Posterior Poisson mean
by Mode rank
by Median rank
by Probability of being the worst
by PSI
by PPSI
by Raw data
by Expected Crashes_EB
by PSI_EB

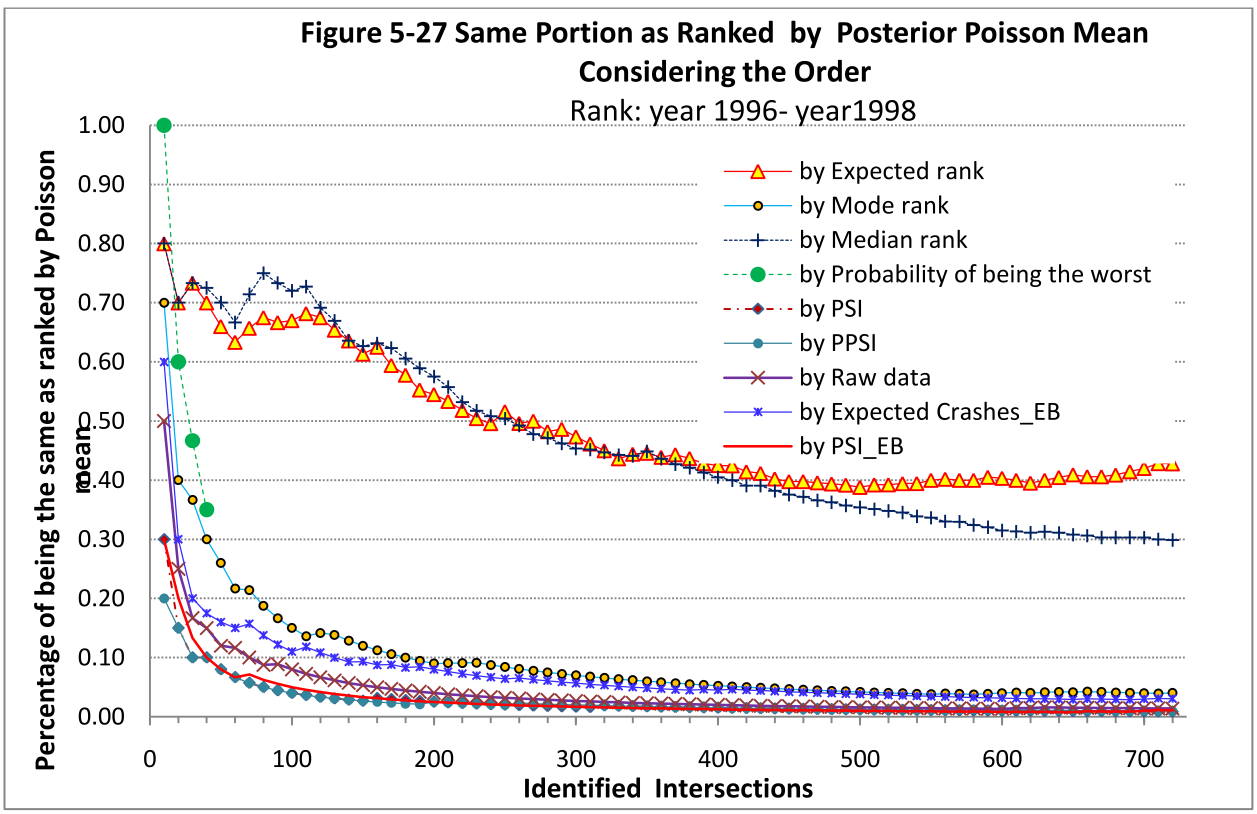
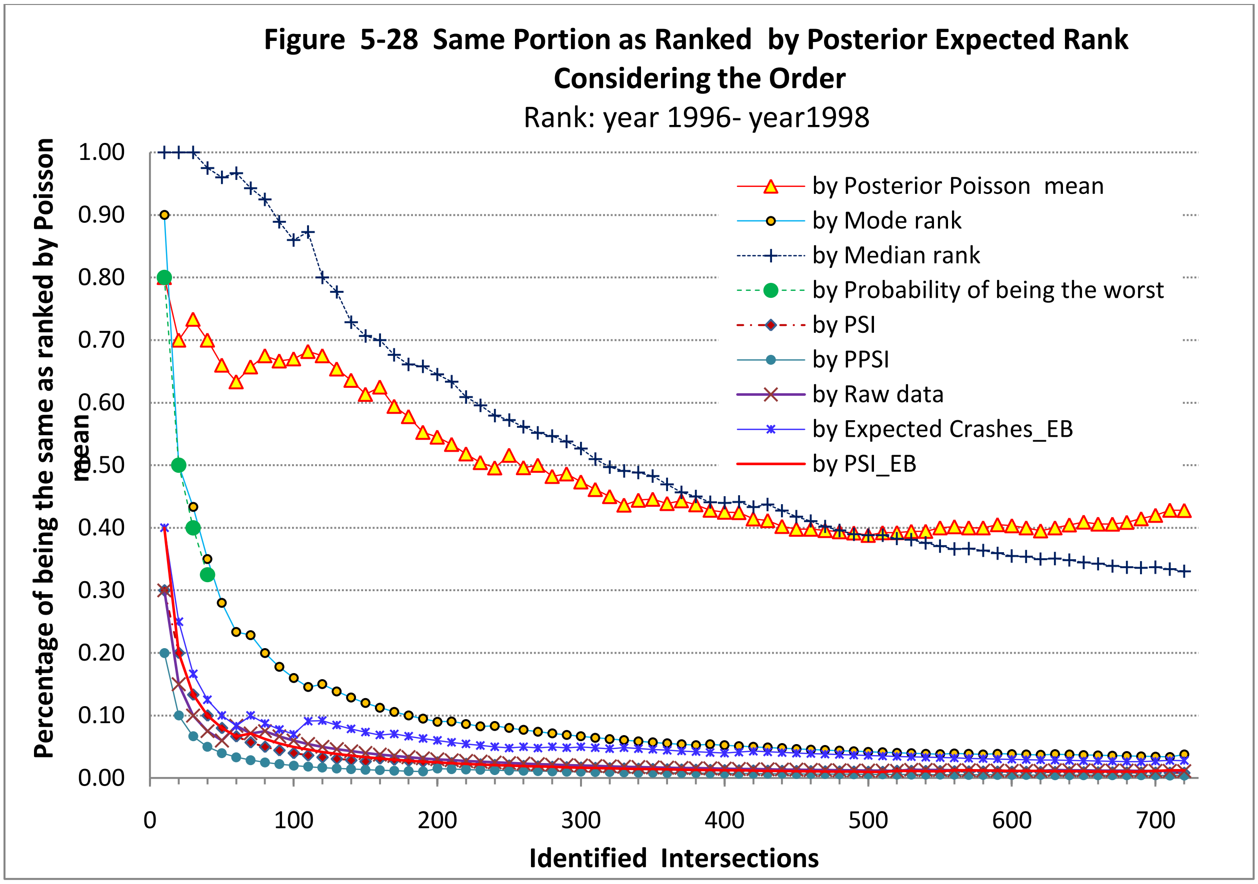


Figure 5-27 Same Portion as Ranked by Posterior Poisson Mean
Considering the Order
Rank: year 1996- year1998
Percentage of being the same as ranked by Poisson mean
Identified Intersections
by Expected rank
by Mode rank
by Median rank
by Probability of being the worst
by PSI
by PPSI
by Raw data
by Expected Crashes_EB
by PSI_EB



5.6.5 Sensitivity Analysis of Crash Data History

A. Sensitivity analysis of crash data history for the FB method

Two time periods (1993-1998 and 1996-1998) were used to conduct a ranking analysis as described above. The question of interest in this analysis is: are the ranking results the same from these two data sets with different crash data history, and if not, which one provides the better results?

The evaluation results for the top 50 ranked sites from these two periods by various ranking criteria are presented in Tables 5-15 and 5-16. It can be seen that the period with six years of ranking does not always provide better results than the shorter ranking period of three years, as expected. For the top 20 and 30 ranked sites, the results from the three years of data by reliable criteria, such as PM, median rank, expected rank, mode rank and the probability of being the worst, are better than for the 6 years of data. Specifically, in the top 20 ranked sites, 3 years of ranking data with the probability of being the worst provides the best ranking results in that it has much higher values for the sum of the PM, sum of the PSI, even the sum of the PPSI and sum of crash counts of the ranked sites in the second period as shown in Table 16. For the top 10, 40 and 50 ranked sites, generally, the six years of ranking data provide better results than the three years of ranking data. This indicates that the three years of data and six years of data indeed provide comparable results for the top ranked sites.

B. Effects of crash data history for the EB and FB methods

The information on the top 50 ranked sites in the second period, by PM and PSI from EB and FB methods which uses three years of data and six years of data, respectively, is listed in Table 5-10. The evaluation results of the ranked sites by PM are presented in Figures 5-29 to 5-31. For the top 20, 40 and 60 ranked sites, the EB method prefers short data history, more so than the FB method as seen from Figure 5-29. From Table 5-10, the top 20 and 30 ranked sites from EB or FB ranked by PM both have better results with three years of data rather than with six years. With PSI, generally, a longer data history gives better ranking results. Note that PSI cannot be used as a good ranking criterion, base on previous studies.

As the number of ranked sites increases, using 6 years of ranking data provides better results than 3 years for both EB and FB methods as shown in Figures 5-29 to 5-31.

For the top ranked limited sites, the short data history might provide better or comparable results. However, this conclusion needs further study which uses different sample sizes and data history.

Table 5-15 Summary of Evaluation Results by Various Ranking Criteria from Different Historical Data (P_AR(1))

	Hotspots identified from 1996-1998 data						**Hotspots identified from 1993-1998 data**					
	Poisson AR(1) model					By PM= Poisson mean						
ranked sites	Sensitivity	Specificity	$\sum tot_A$	$\sum PM_A$	$\sum PSI_A$	$\sum PPSI_A$	Sensitivity	Specificity	$\sum tot_A$	$\sum PM_A$	$\sum PSI_A$	$\sum PPSI_A$
10	0.20	0.99	241	248	155	147	0.30	0.99	263	263	168	168
20	0.65	0.99	484	488	315	313	0.60	0.99	485	485	298	298
30	0.70	0.99	717	713	429	436	0.70	0.99	702	698	389	395
40	0.73	0.98	904	893	530	547	0.78	0.99	932	920	537	554
50	0.74	0.98	1055	1047	609	624	0.80	0.99	1116	1097	619	645
	Poisson AR(1) model					By PSI=Poisson mean-predicted crashes at similar sites						
ranked sites	Sensitivity	Specificity	$\sum tot_A$	$\sum PM_A$	$\sum PSI_A$	$\sum PPSI_A$	Sensitivity	Specificity	$\sum tot_A$	$\sum PM_A$	$\sum PSI_A$	$\sum PPSI_A$
10	0.30	0.99	235	243	169	164	0.40	0.99	243	244	180	182
20	0.50	0.99	453	457	318	317	0.55	0.99	454	463	306	300
30	0.47	0.98	641	643	431	431	0.57	0.98	707	698	451	463
40	0.65	0.98	848	839	561	573	0.70	0.98	899	876	568	593
50	0.74	0.98	1034	1014	669	692	0.74	0.98	1047	1022	657	684
	Poisson AR(1) model					By PPSI=crash counts-predicted at similar sites						
ranked sites	Sensitivity	Specificity	$\sum tot_A$	$\sum PM_A$	$\sum PSI_A$	$\sum PPSI_A$	Sensitivity	Specificity	$\sum tot_A$	$\sum PM_A$	$\sum PSI_A$	$\sum PPSI_A$
10	0.30	0.99	219	230	162	154	0.40	0.99	243	244	180	182
20	0.35	0.98	453	457	318	317	0.40	0.98	454	463	306	300
30	0.43	0.98	626	632	430	427	0.50	0.98	700	693	452	461
40	0.55	0.97	835	827	553	564	0.60	0.98	890	870	563	587
50	0.64	0.97	1023	1002	664	687	0.62	0.97	1017	998	643	664
	Poisson AR(1) model					By Crash Frequency						
ranked sites	Sensitivity	Specificity	$\sum tot_A$	$\sum PM_A$	$\sum PSI_A$	$\sum PPSI_A$	Sensitivity	Specificity	$\sum tot_A$	$\sum PM_A$	$\sum PSI_A$	$\sum PPSI_A$
10	0.20	0.99	255	258	181	178	0.20	0.99	252	253	171	169
20	0.45	0.98	484	487	326	326	0.45	0.98	478	479	300	299
30	0.60	0.98	722	715	461	471	0.53	0.98	693	693	407	409
40	0.60	0.98	903	892	554	567	0.65	0.98	932	920	537	554
50	0.66	0.97	1062	1049	628	647	0.74	0.98	1133	1106	650	682

Notes: 1. Subscript A means year 1999-2002

Table 5-16 Summary of Evaluation Results by Various Ranking Criteria from Different Historical Data (P_AR(1))

	Hotspots identified from 1996-1998 data						**Hotspots identified from 1993-1998 data**					
	By Posterior Expected Rank											
ranked sites	Sensitivity	Specificity	$\sum tot_A$	$\sum PM_A$	$\sum PSI_A$	$\sum PPSI_A$	Sensitivity	Specificity	$\sum tot_A$	$\sum PM_A$	$\sum PSI_A$	$\sum PPSI_A$
10	0.20	0.99	241	248	155	147	0.30	0.99	263	263	168	168
20	0.65	0.99	484	488	315	313	0.60	0.99	485	485	298	298
30	0.67	0.99	717	713	429	436	0.70	0.99	702	698	389	395
40	0.75	0.99	916	901	536	554	0.78	0.99	932	920	537	554
50	0.76	0.98	1066	1054	619	638	0.80	0.99	1116	1097	619	645
By Posterior Median Rank												
ranked sites	Sensitivity	Specificity	$\sum tot_A$	$\sum PM_A$	$\sum PSI_A$	$\sum PPSI_A$	Sensitivity	Specificity	$\sum tot_A$	$\sum PM_A$	$\sum PSI_A$	$\sum PPSI_A$
10	0.20	0.99	241	248	155	147	0.30	0.99	263	263	168	168
20	0.65	0.99	484	488	315	313	0.60	0.99	485	485	298	298
30	0.67	0.99	717	713	429	436	0.70	0.99	702	698	389	395
40	0.73	0.98	904	893	530	547	0.78	0.99	932	920	537	554
50	0.76	0.98	1066	1054	619	638	0.80	0.99	1116	1097	619	645
By Posterior Mode Rank												
ranked sites	Sensitivity	Specificity	$\sum tot_A$	$\sum PM_A$	$\sum PSI_A$	$\sum PPSI_A$	Sensitivity	Specificity	$\sum tot_A$	$\sum PM_A$	$\sum PSI_A$	$\sum PPSI_A$
10	0.20	0.99	234	242	151	143	0.30	0.99	271	269	178	180
20	0.70	0.99	501	501	317	319	0.65	0.99	485	485	298	298
30	0.67	0.99	707	707	418	421	0.67	0.99	702	698	389	395
40	0.75	0.99	918	904	544	564	0.73	0.98	921	912	524	539
50	0.76	0.98	1078	1061	627	651	0.78	0.98	1107	1090	606	630
By Probability of being the worst												
ranked sites	Sensitivity	Specificity	$\sum tot_A$	$\sum PM_A$	$\sum PSI_A$	$\sum PPSI_A$	Sensitivity	Specificity	$\sum tot_A$	$\sum PM_A$	$\sum PSI_A$	$\sum PPSI_A$
10	0.20	0.99	241	248	155	147	0.30	0.99	263	263	168	168
20	0.60	0.99	511	510	325	329	0.55	0.99	463	469	285	282

Notes: 1. Subscript A means year 1999-2002

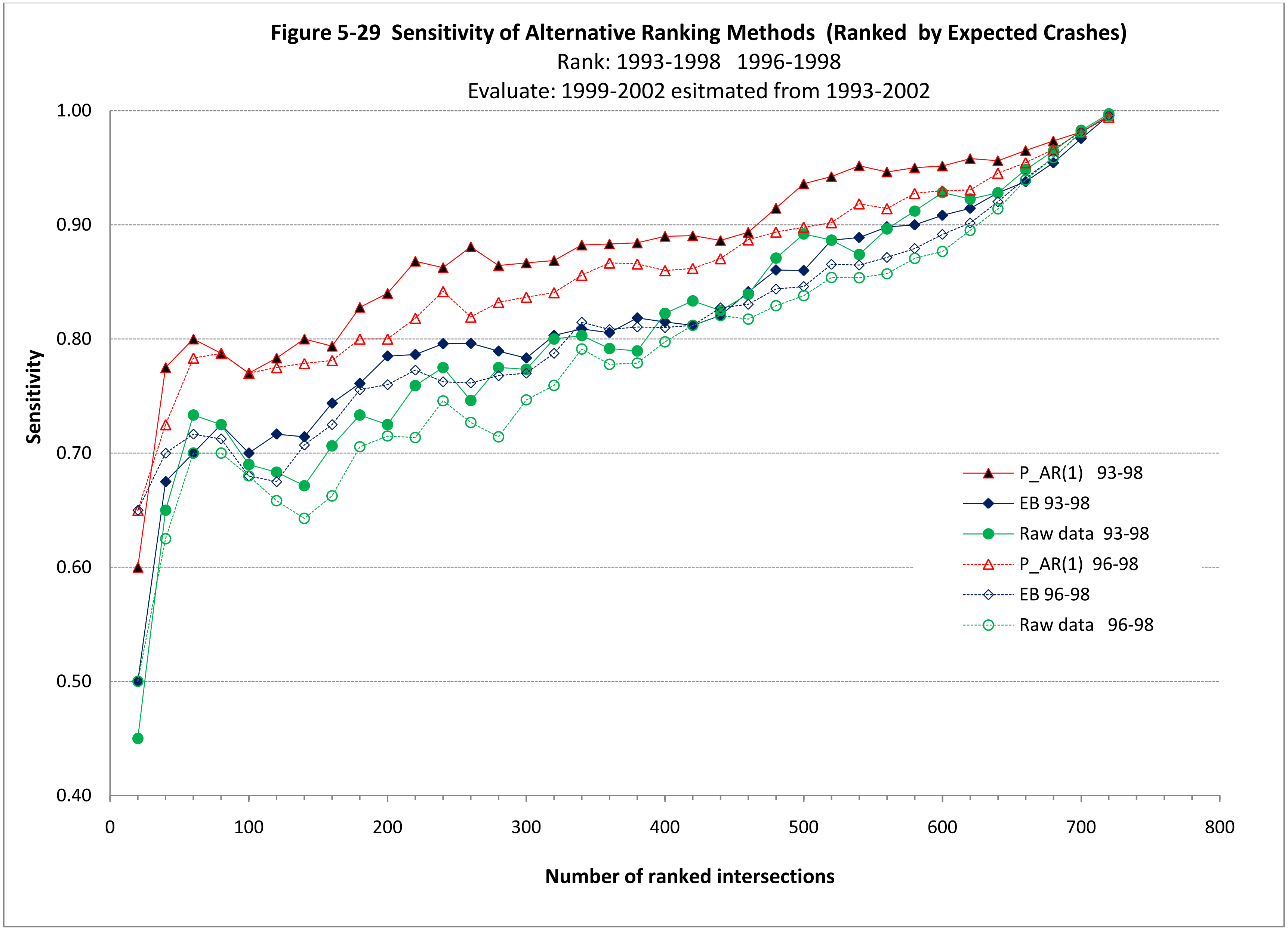


Figure 5-29 Sensitivity of Alternative Ranking Methods (Ranked by Expected Crashes)
Rank: 1993-1998 1996-1998
Evaluate: 1999-2002 esitmated from 1993-2002
Sensitivity
1.00
0.90
0.80
0.70
0.60
0.50
0.40
0
100
200
300
400
500
600
700
800
Number of ranked intersections
P_AR(1) 93-98
EB 93-98
Raw data 93-98
P_AR(1) 96-98
EB 96-98
Raw data 96-98

Figure 5-30 Specificity of Alternative Ranking Methods (Ranked by Expected Crashes)

Rank: 1993-1998 1996-1998

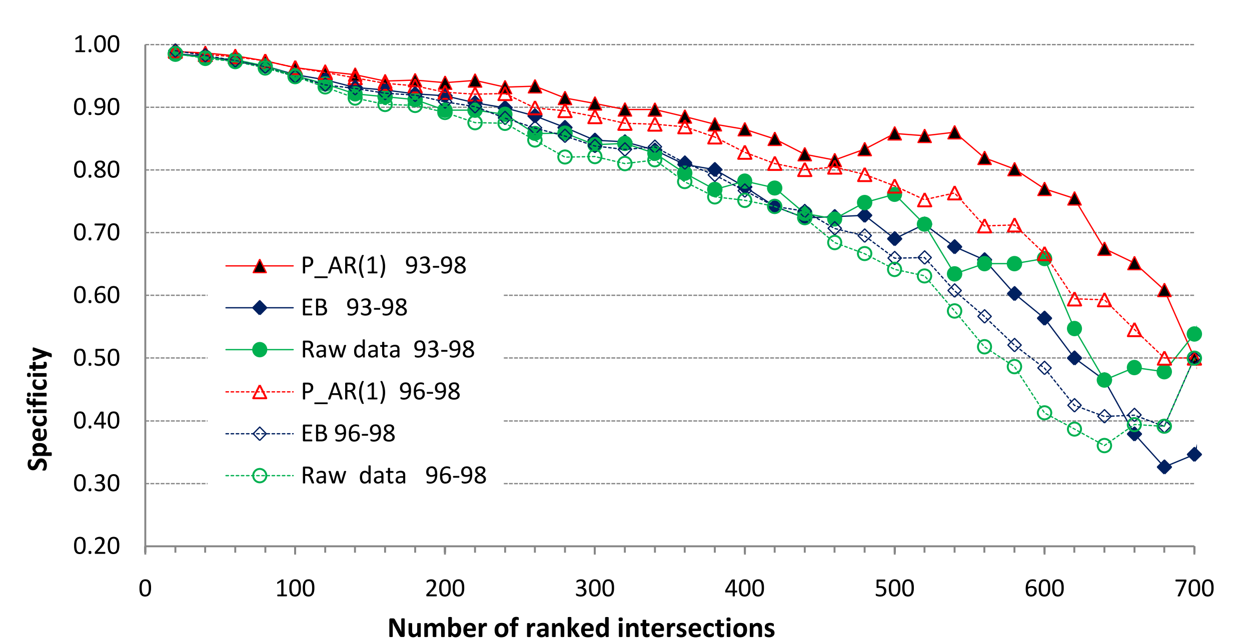
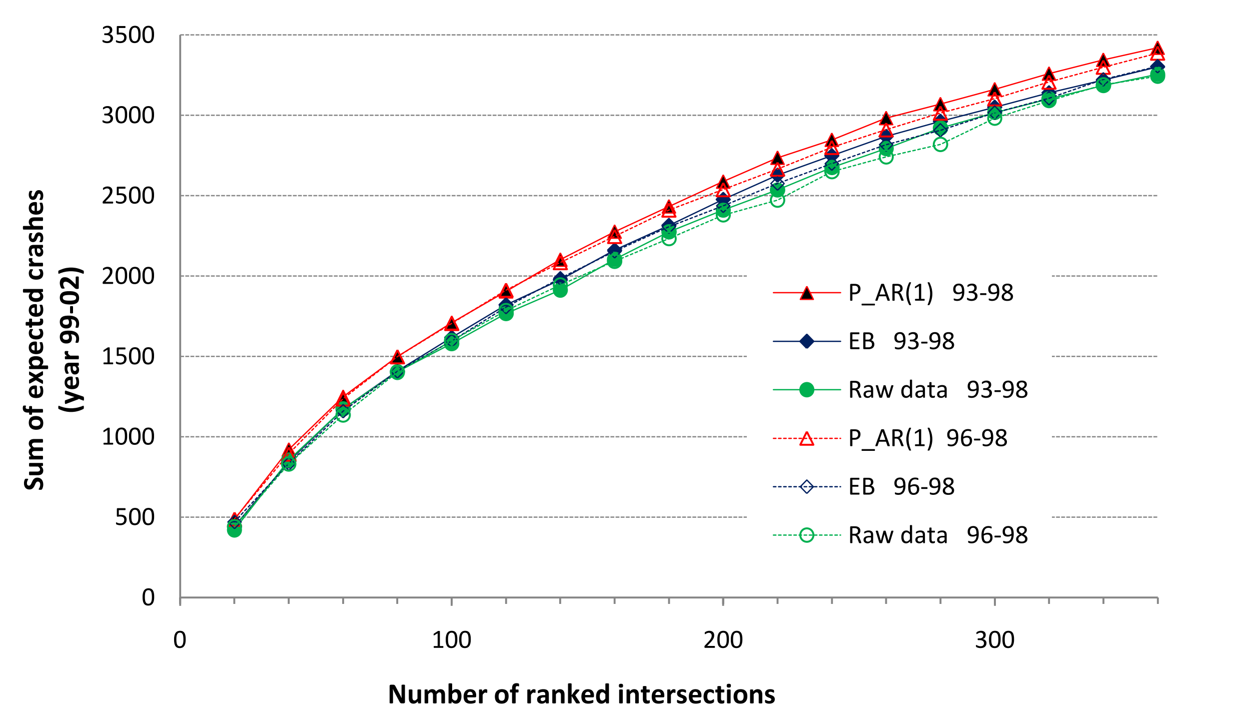


Figure 5-31 Sum of Expected Crashes (1999-2002) (Ranked by PM)

Rank: 1993-1998 1996-1998

Evaluate: 1999-2002 esitmated from 1993-2002



5.7 FB WITH MULTILEVEL SEVERITY DATA

5.7.1 Advantages of the FB Method over the EB Method

In hotspot identification, it is not always that only one level of crash severity is used to identify hazardous sites. Most often, multiple severity levels of crashes might be used for network ranking. For such data, there is only one way to conduct network ranking using the EB method. That is, SPFs need to be individually developed and the estimation of expected crashes is calculated for each level of severity. Then, the expected crashes of each level of severity can be combined by assigning different weights to rank the sites. This procedure completely neglects possible correlation among these multiple severity data and is time consuming. Another drawback with the EB method concerns the ranking criteria: only PM and PSI can be used to identify black spots and as previously found, the PSI cannot provide promising results. The ranking results of the EB were also shown to be not as good as for the FB.

For the FB method, there may be two ways to deal with such multi-severity data: a univariate approach and a multivariate approach. The models with all severity crash data can be developed simultaneously using both the univariate approach and multivariate approach. The expected crashes for each level of severity or combined severity for each site can be done within that model development procedure, and posterior Poisson mean or other decision parameters and median rank of the posterior distribution of decision parameters (Poisson mean in this study) can also be obtained during the procedure using WinBUGs software. All the procedures are integrated. Moreover, the outputs of the MCMC procedure of FB methods enable more solid ranking criteria to be explored, such as posterior expected rank and mode rank, and the probability of being the worst. Furthermore, it is more flexible to explore different distributions with FB such as PG, PLN, etc., while only Poisson or NB is available for EB. More importantly the possible correlation among these severity data can be properly addressed with FB. In all, the FB provides many advantages over the EB method for network ranking, especially using multilevel severity crash data.

5.7.2 Data Summary

There are a total of five levels of crash severity in this study. These are: Sev1: fatal (K), Sev2: incapacitating-injury (A), Sev3: non-incapacitating injury (B), Sev4: minor injury (C), Sev5: PDO. Different weights are assigned to these crashes: 5 to severity 1, 4 to severity 2, 3 to severity 3, 2 to severity 4 and 1 to severity 1 (PDO crashes). A total of 436 top ranked sites based on combined crashes calculated by Equation 5-5 were selected for this study. Similarly, data from 1993 to 1998 were used to identify hazardous sites while data from 1993 to 2002 were employed to estimate the true mean of the second period (1999 – 2002) for the evaluation of the ranked results. The summary information of the data is shown in Table 5-17. It can be seen that the traffic volume on major roads increases during the evaluation period, but on a limited scale. Since it is a systematic pattern, and the amount is not large, sensitivity and specificity can still be used as evaluation criteria, but the sum of the Poisson mean might be a better evaluation criterion.

Table 5-17 Summary for the 436 California Unsignalized Intersections

year 1993- year 1998				
Crashes /site.year	Mean	Standard Deviation	Maximum	Minumum
Sev 1	0.05	0.22	2	0
Sev 2	0.10	0.33	4	0
Sev 3	0.42	0.70	5	0
Sev 4	0.42	0.72	6	0
Sev 5	1.02	1.25	8	0
$AADT_{Major}$	9102	4520	29732	2950
$AADT_{Minor}$	839	1010	7800	100
year 1993 - year 2002				
Crashes /site. year	Mean	Standard Deviation	Maximum	Minumum
Sev 1	0.04	0.21	2	0
Sev 2	0.09	0.31	4	0
Sev 3	0.40	0.71	8	0
Sev 4	0.43	0.76	8	0
Sev 5	1.09	1.36	9	0
$AADT_{Major}$	9456	4727	29732	2900
$AADT_{Minor}$	842	1015	7800	100

5.7.3 Multivariate FB Model vs. Univariate FB Model

Previously, the univariate FB method was examined for hot spot identification and it was concluded that the FB provides much better ranking results than the EB method, and that ranking criteria which include Poisson mean, posterior expected rank, median rank, mode rank and the probability of being the worst are reliable ranking criteria and can provide promising results. If the order of each ranked site is not a concern (i.e., it is only based on the cutoff number of ranked sites), then the Poisson mean, median rank and expected rank have almost the same results. Otherwise, the ranked list is quite different in terms of order for each site by these three ranking criteria. Expected rank was proven to be a better ranking criterion than the other two while median ranking can provide very similar results as expected rank, especially for the top ranked limited sites. This aspect of the research will further evaluate the FB method with a special case: multiple crash data severity levels where univariate and multivariate FB approaches are applied. All previously applied ranking and evaluation criteria will be used here for this evaluation.

Based on the model selection results for a single level severity case, Poisson AR (1) and multivariate Poisson log normal AR (1) (denoted as MVPLN AR (1) hereafter) are deemed as the best models for univariate and multivariate approaches, respectively, in this study. The model framework is presented below.

A. Multivariate Poisson Log Normal AR(1) FB Model

Crash counts $\mathbf{Y_{it}} = (Y_{it}^1, Y_{it}^2, \ldots Y_{it}^L)$ can be described as L severities of multivariate crash records at location i (where i=1,2…N) in year t (t=1,2…J). Each severity crash is assumed to be independently Poisson distributed. That is:

$$y_{i,t}^{k} | \boldsymbol{\beta^k}, \boldsymbol{X}, \omega_{i,t}^{k}, \varepsilon_i^k \sim \text{Pois}\left(\lambda_{i,t}{}^{k}\right) \qquad k = 1,2, \ldots 5 \qquad (5-14)$$

$$\lambda_{i,t}{}^{k} = \mathrm{e}^{\omega_{i,t}{}^{k}} \mathrm{e}^{\varepsilon_i{}^{k}} \mu_{i,t}{}^{k} \qquad (5-15)$$

$$\mu_{i,t}{}^{k} = \exp^{\beta_0{}^{k}} \mathrm{ML}_{i,t}^{\beta_1{}^{k}} \mathrm{XST}_{i,t}^{\beta_2{}^{k}} \qquad (5-16)$$

where

$\mu_{i,t}{}^{k}$ = Expected crashes of type k at sites similar to site i in year t, which is the same as Equation 5-8,

$\boldsymbol{\beta^k}$ = Coefficient vector for type k crashes including $\beta_0{}^k, \beta_1{}^k, \beta_2{}^k$,

$ML_{i,t}$ = AADT on major roads at intersection i in year t,

$Xst_{i,t}$ = AADT on minor roads at intersection i in year t,

$\varepsilon_i{}^k$ =Random effects at site i for type k crashes, following a multiple normal distribution, and

$\omega_{i,t}{}^k$ = Random time effects at site i for type k crashes in year t and similar to Equation3-30,

$$\omega_{i,t}{}^k = r^k \omega_{i,t-1}{}^k + u_{i,t}{}^k \qquad \mathrm{t} \geq 2 \qquad (5-22)$$

$u_{i,t}{}^k$ = unstructured white noise, $u_{i,t}{}^k \sim \mathrm{N}(0, \sigma^{2^k})$, $\mathrm{t} \geq 1$

$$\omega_{i,1}{}^k \sim \mathrm{N}\left(0, \sigma^{2^k}/(1 - \mathrm{r}^{2^k})\right) \qquad (5-23)$$

where,

σ^{2^k} = Variance of white noise for type k crashes

r^k = Time series correlation coefficients of type k crashes

The vector $\boldsymbol{\varepsilon_i} = (\varepsilon_i^1, \varepsilon_i^2, \dots, \varepsilon_i^5)'$ is assumed to be multivariate normal distributed to account for the correlations among crashes of different severities, that is: $(\varepsilon_i^1, \varepsilon_i^2, \dots, \varepsilon_i^5)' \sim N_5(0, \boldsymbol{\Sigma})$.

$\boldsymbol{\Sigma}$ = an unrestricted 5×5 covariance matrix between different severity/type of crashes.

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{15} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{25} \\ \cdots & \cdots & \cdots & \cdots \\ \sigma_{51} & \sigma_{52} & \cdots & \sigma_{55} \end{pmatrix}$$

The MVPLN can be seen to have an additive form (logarithm) of random effects that accounts for the extra-variation between sites with correlated random errors among crash severities within a site. The covariance between the counts, $y_{i,t}^k$ and $y_{i,t}^m$, can be positive or negative depending on the sign of the (k,*m*)th element of Σ. Thus, the correlation structure of the crash counts is unrestricted.

C. Univariate Poisson Log Normal AR(1) FB Model

A univariate P-AR (1) model has the same form of ${\mu_{i,t}}^{k}$ as MVPLN. The only difference is that the vector $\varepsilon_{\mathrm{i}} = \varepsilon_{\mathrm{i}}^{1}, \varepsilon_{\mathrm{i}}^{2}, \dots, \varepsilon_{\mathrm{i}}^{5}$ is not included since time correlated random effects are already introduced.

5.7.4 Model Comparison and Parameter Estimation

The parameter estimation and model selection criteria such as LL, AIC, BIC and DIC are tabulated in Table 5-18 for the ranking data in 1993-1998 and evaluation data in 1993-2002. Due to the difference of the random effect structure and different data for model development, the parameter estimations are not the same, but generally follow the same pattern. These model selection criteria, however, provide conflicting results: LL, AIC, BIC strongly favor a univariate approach in that it has a higher value of LL and lower value of AIC and BIC while DIC strongly prefers the MVPLN AR(1) model for data in both periods. Thus, the ranking results from both approaches are evaluated as a byproduct to identify which model selection criteria is the best.

5.7.5 Evaluation of Alternative FB Approaches for Hot Spot Identification

It is worthwhile to mention that measures such as crash counts, Poisson mean, PSI and PPSI, regardless as ranking or evaluation criteria, provide the combined results of five levels of severity data. For example, as a ranking criterion, PM_i is the sum of the weighted Poisson means for five severity levels of crashes during the ranking period at site i and the same procedure is applied to obtain other criteria.

Similar to the univariate FB study, the PM might not be the same due to the structure of the random effects. In order to use the sum of the PM in the second period as an evaluation criterion for comparison of the FB methods, the PM of the MVPLN_AR (1) in the evaluation period is multiplied by the ratio of the sum of the PM for the P_AR (1) in all 436 sites, which is 6693, over the sum of the PM from the MVPLN_AR (1), which is 6667. The sum of the PM estimated from the two models is indeed quite comparable.

Evaluation criteria which include sensitivity and specificity, the sum of the PM and sum of crash counts in the second period (1999-2002) were used to conduct the evaluation analysis. The sites were ranked by PM, PSI, expected rank, median rank, mode rank and the probability of being the worst by using data in 1993-1998. Crash counts at each site were also used as a rank criterion for reference. The results are presented in Figures 5-32 to 5-41. The evaluation results for the top 100 ranked sites are also given for clarity of the comparison in that these top ranked sites are the major concern.

It can be seen that the MVPLN_AR (1) model has an obvious advantage over the univariate P_AR (1) model from the evaluation criterion -- the sum of the PM over all the corresponding ranking criteria. From sensitivity and specificity, at least for the top 100 ranked sites which count for about 23% of the whole group, MVPLN_AR (1) provides better results in terms of higher values. For ranked lists with more than 100 sites, sometimes P_AR (1) provides better results, e.g., the top 110-150 ranked sites by all the ranking criteria except PSI, in which P_AR (1) is better than MVPLN_AR (1). MVPLN_AR (1) provides better results than P_AR (1) at least for the top 23% ranked sites based on the sum of the PM, and sensitivity and specificity. For hot spot identification, the top 20% or fewer ranked sites are typically of most concern. Thus MVPLN_AR (1) is superior to univariate P_AR (1) for multilevel severities of crash data for typical network ranking applications. If changes in traffic volumes are taken into consideration, the sum of the PM might be used as a better evaluation criterion than sensitivity and specificity. Hence, based on that, MVPLN_AR (1) is systematically better than the univariate PLN AR (1) model as seen in Figures 5-39 to 5-41. This might indicate that DIC is a better model selection criterion than the others, such as LL, AIC and BIC in that it favors the MVPLN_AR (1) model. In fact, the posterior mean of the random effects of each crash severity level is strongly correlated for data from 1993-1998 and 1993-2002; the covariance and correlation matrix are shown in Tables 5-19 and 5-20. For this reason, DIC might be used as a key model selection criterion later in the research.

Table 5-18 Parameter Estimation from Alternative Approaches

		year 1993 - year 1998						year 1993 - year 2002					
		P_AR(1)_6yrs			MVPLN_AR(1)_6yrs			P_AR(1)_10yrs			MVPLN_AR(1)_10yrs		
		mean	95% BCI		mean	95% BCI		mean	95% BCI		mean	95% BCI	
Sev 1	β_0	-8.51	-12.71	-3.64	-8.82	-12.56	-4.58	-7.11	-10.61	-4.32	-6.28	-9.92	-2.71
	β_1	0.51	0.03	0.94	0.55	0.13	0.92	0.33	-0.01	0.70	0.25	-0.13	0.60
	β_2	0.08	-0.13	0.28	0.07	-0.14	0.27	0.11	-0.05	0.26	0.09	-0.07	0.26
	r	0.87	0.11	1.00	-0.53	-0.95	0.47	0.89	0.50	0.99	-0.04	-0.73	0.75
Sev 2	β_0	-6.33	-8.59	-3.27	-5.79	-8.08	-3.39	-5.72	-8.50	-3.45	-3.87	-6.03	-1.44
	β_1	0.27	-0.05	0.51	0.22	-0.03	0.46	0.19	-0.08	0.47	0.02	-0.26	0.26
	β_2	0.20	0.06	0.35	0.19	0.05	0.33	0.20	0.08	0.31	0.16	0.03	0.30
	r	0.72	0.23	1.00	0.28	-0.55	0.81	0.86	0.69	0.94	0.05	-0.70	0.60
Sev 3	β_0	-3.89	-5.29	-2.58	-3.78	-5.06	-2.54	-4.43	-5.74	-3.36	-4.10	-5.35	-2.92
	β_1	0.15	0.00	0.30	0.14	0.00	0.28	0.18	0.03	0.32	0.16	0.04	0.30
	β_2	0.24	0.17	0.32	0.23	0.16	0.31	0.26	0.19	0.33	0.24	0.16	0.31
	r	0.88	0.74	0.98	-0.06	-0.78	0.85	0.93	0.87	0.98	-0.13	-0.86	0.74
Sev 4	β_0	-7.61	-9.27	-5.86	-6.93	-8.64	-5.05	-8.56	-10.11	-7.29	-7.89	-9.20	-6.56
	β_1	0.48	0.28	0.66	0.41	0.20	0.58	0.56	0.40	0.73	0.50	0.37	0.64
	β_2	0.35	0.27	0.44	0.34	0.26	0.42	0.38	0.30	0.45	0.36	0.28	0.43
	r	0.90	0.76	0.99	0.16	-0.65	0.80	0.92	0.85	0.96	0.29	-0.15	0.68
Sev 5	β_0	-6.03	-7.57	-5.06	-5.56	-6.79	-4.21	-7.01	-8.56	-6.20	-6.51	-7.35	-5.49
	β_1	0.41	0.31	0.57	0.36	0.22	0.50	0.50	0.41	0.64	0.45	0.36	0.54
	β_2	0.34	0.28	0.41	0.33	0.28	0.39	0.37	0.32	0.44	0.36	0.31	0.43
	r	0.95	0.90	0.99	0.10	-0.40	0.56	0.96	0.93	0.98	0.38	-0.04	0.80
No. of Parameters: K		25			40			25			40		
Log likelihood: LL		-8481			-8504			-13937			-13941		
AIC		17002			17078			27914			27952		
BIC		17119			17283			28042			28175		
DIC		17840			17698			29256			29052		

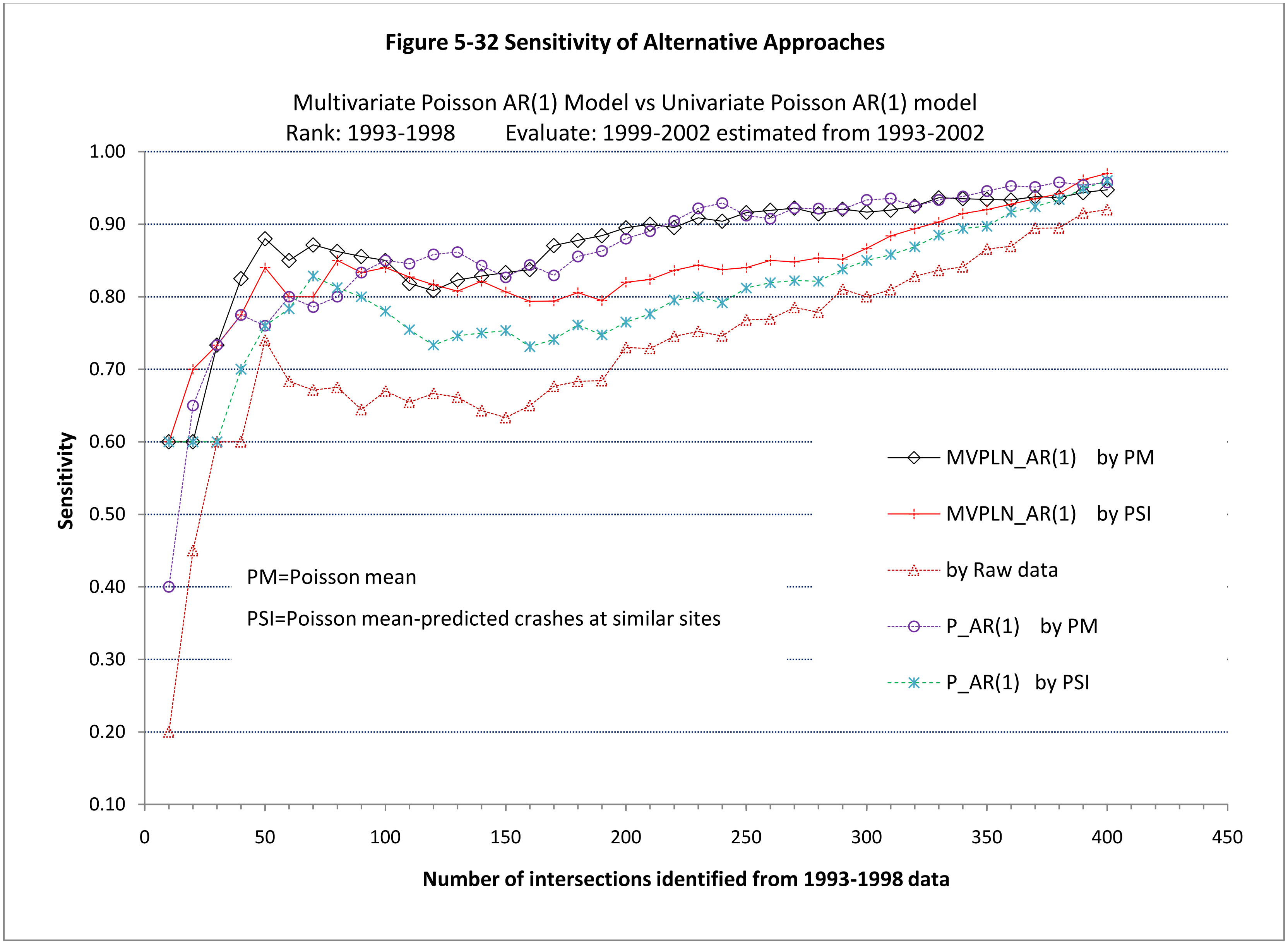


Figure 5-32 Sensitivity of Alternative Approaches
Multivariate Poisson AR(1) Model vs Univariate Poisson AR(1) model
Rank: 1993-1998 Evaluate: 1999-2002 estimated from 1993-2002
Sensitivity
1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0
50
100
150
200
250
300
350
400
450
Number of intersections identified from 1993-1998 data
MVPLN_AR(1) by PM
MVPLN_AR(1) by PSI
by Raw data
P_AR(1) by PM
P_AR(1) by PSI
PM=Poisson mean
PSI=Poisson mean-predicted crashes at similar sites

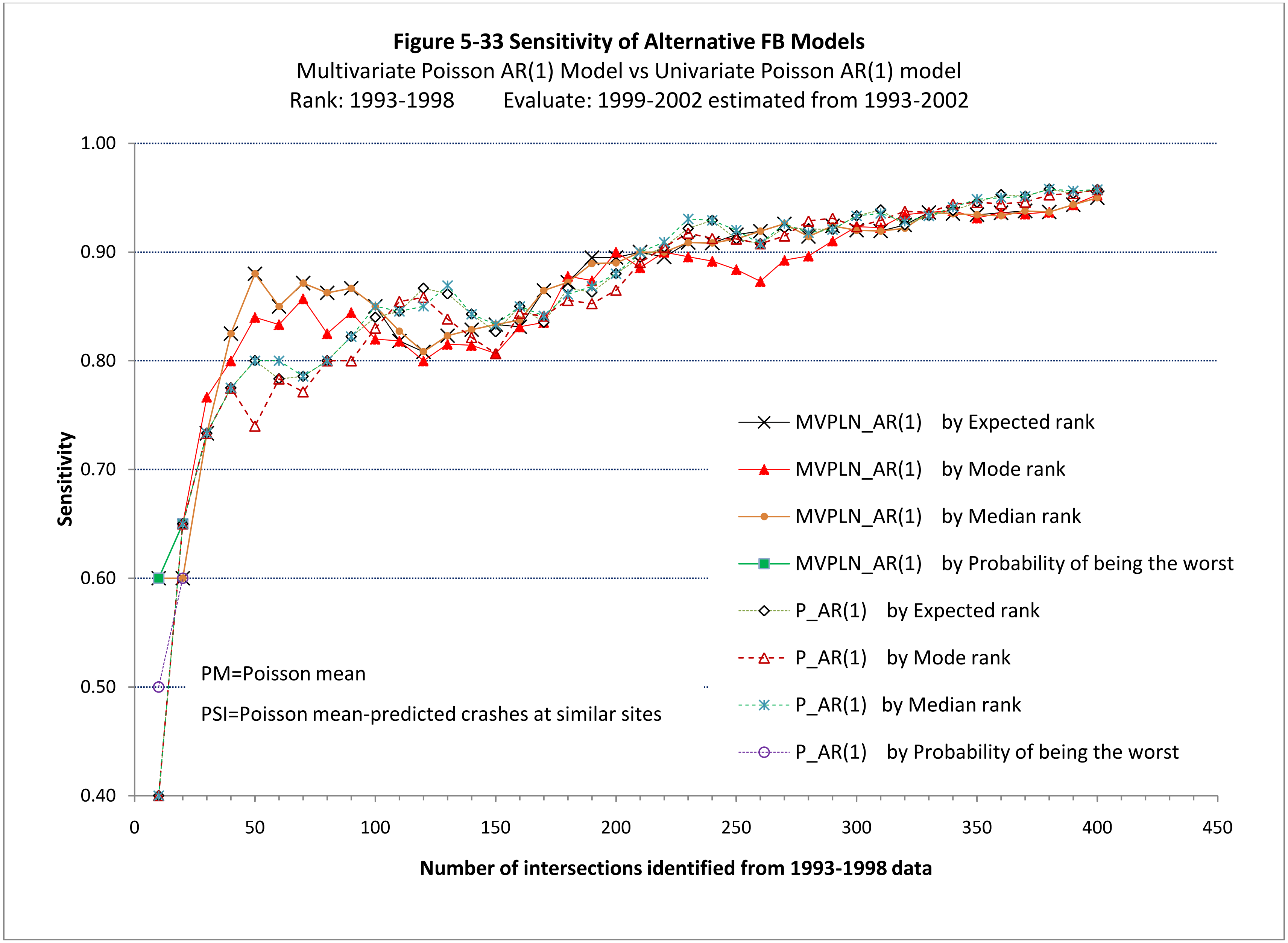


Figure 5-33 Sensitivity of Alternative FB Models
Multivariate Poisson AR(1) Model vs Univariate Poisson AR(1) model
Rank: 1993-1998 Evaluate: 1999-2002 estimated from 1993-2002
Sensitivity
1.00
0.90
0.80
0.70
0.60
0.50
0.40
0
50
100
150
200
250
300
350
400
450
Number of intersections identified from 1993-1998 data
MVPLN_AR(1) by Expected rank
MVPLN_AR(1) by Mode rank
MVPLN_AR(1) by Median rank
MVPLN_AR(1) by Probability of being the worst
P_AR(1) by Expected rank
P_AR(1) by Mode rank
P_AR(1) by Median rank
P_AR(1) by Probability of being the worst
PM=Poisson mean
PSI=Poisson mean-predicted crashes at similar sites

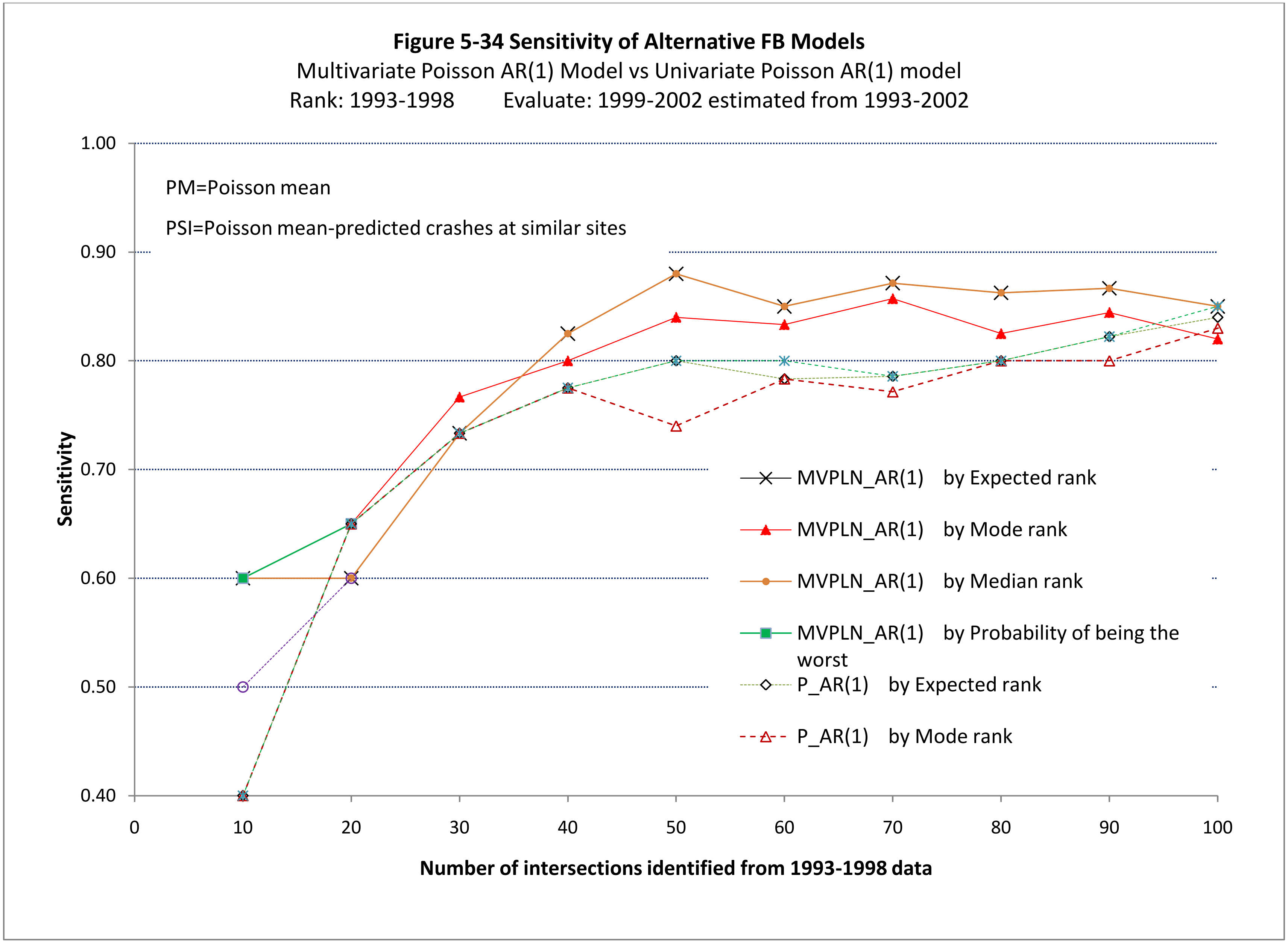


Figure 5-34 Sensitivity of Alternative FB Models
Multivariate Poisson AR(1) Model vs Univariate Poisson AR(1) model
Rank: 1993-1998
Evaluate: 1999-2002 estimated from 1993-2002
PM=Poisson mean
PSI=Poisson mean-predicted crashes at similar sites
Sensitivity
1.00
0.90
0.80
0.70
0.60
0.50
0.40
0
10
20
30
40
50
60
70
80
90
100
MVPLN_AR(1) by Expected rank
MVPLN_AR(1) by Mode rank
MVPLN_AR(1) by Median rank
MVPLN_AR(1) by Probability of being the worst
P_AR(1) by Expected rank
P_AR(1) by Mode rank
Number of intersections identified from 1993-1998 data

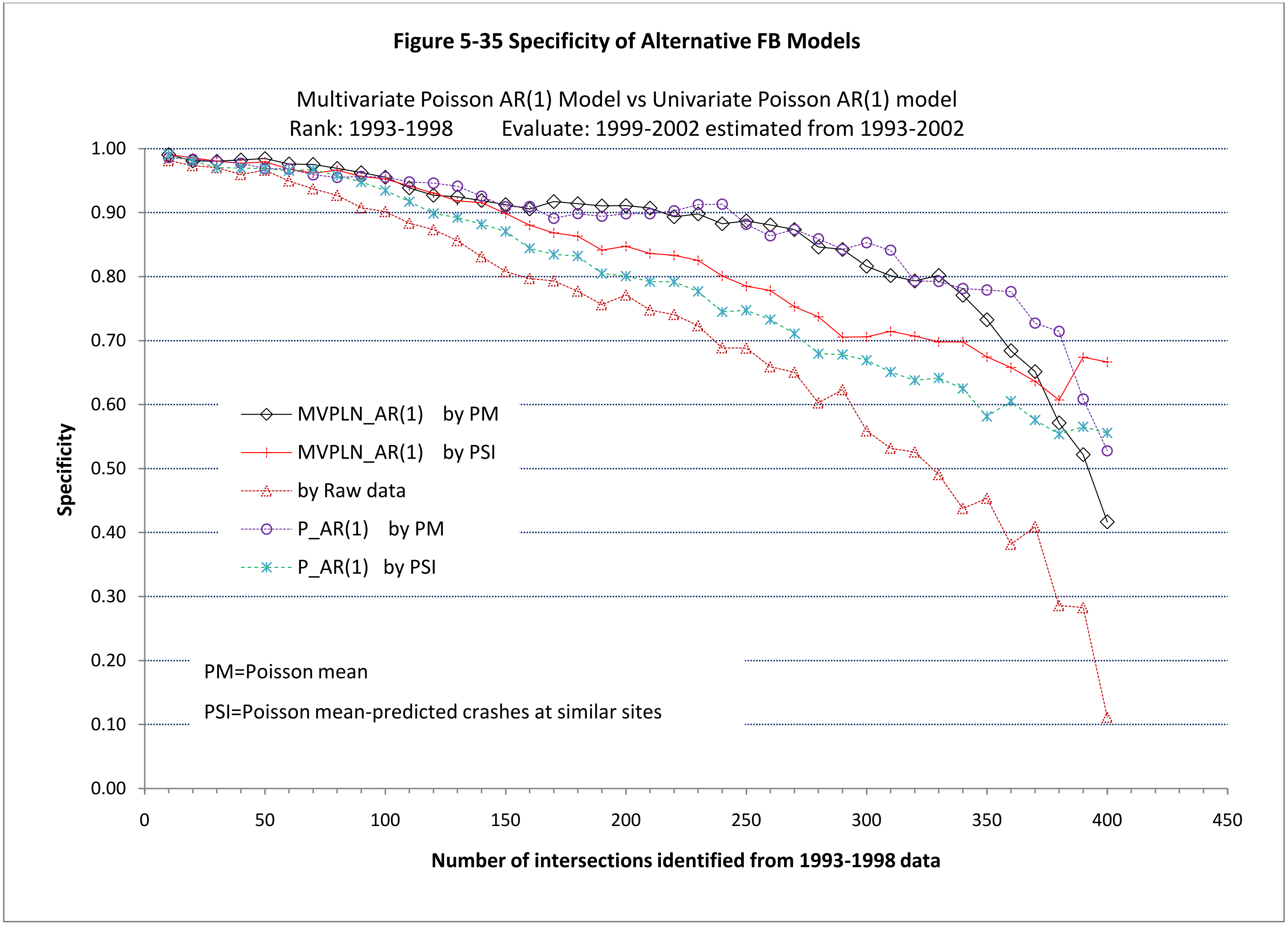
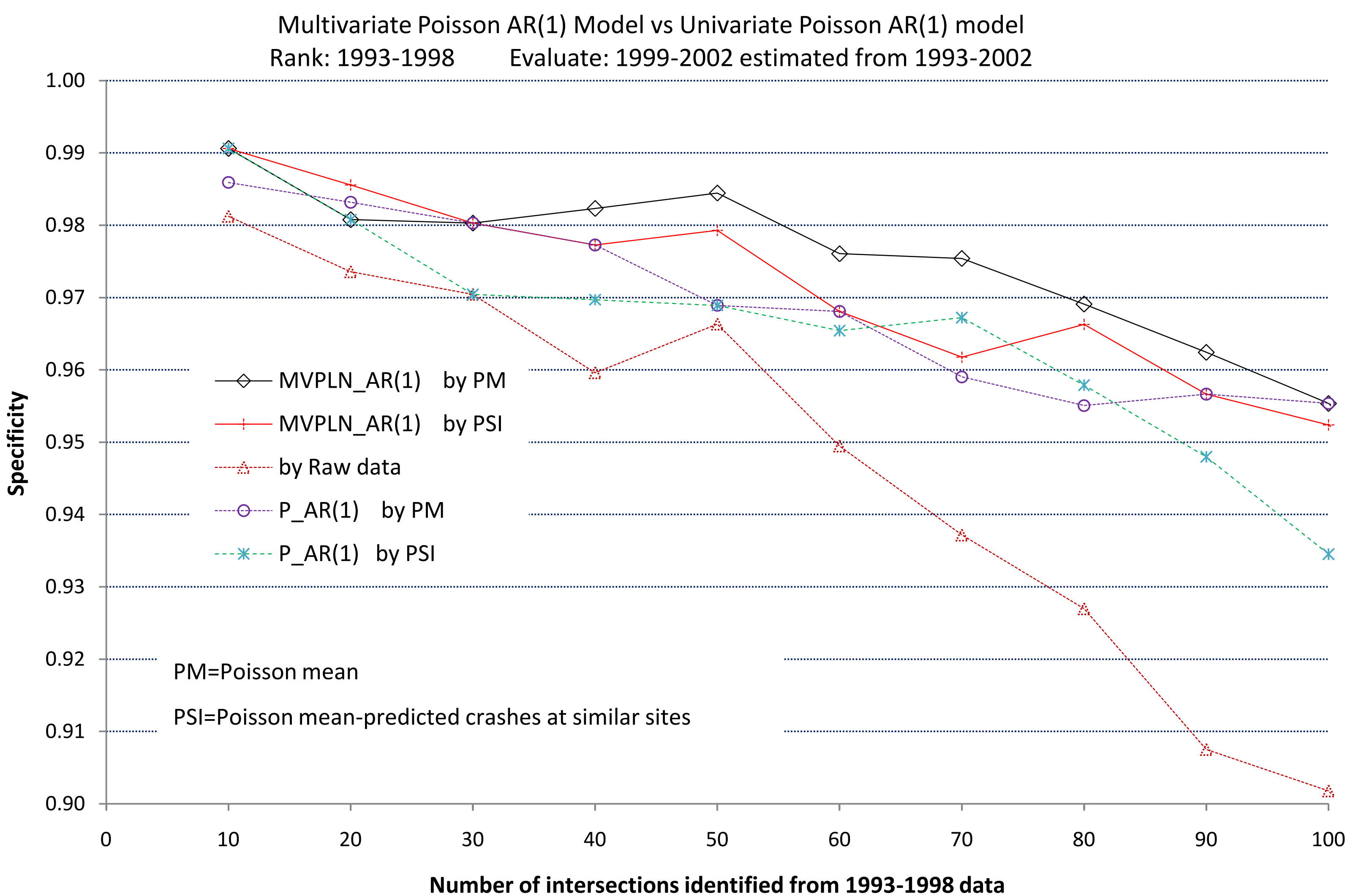


Figure 5-35 Specificity of Alternative FB Models
Multivariate Poisson AR(1) Model vs Univariate Poisson AR(1) model
Rank: 1993-1998 Evaluate: 1999-2002 estimated from 1993-2002
Specificity
Number of intersections identified from 1993-1998 data
MVPLN_AR(1) by PM
MVPLN_AR(1) by PSI
by Raw data
P_AR(1) by PM
P_AR(1) by PSI
PM=Poisson mean
PSI=Poisson mean-predicted crashes at similar sites
1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
0
50
100
150
200
250
300
350
400
450

Figure 5-36 Specificity of Alternative FB Models



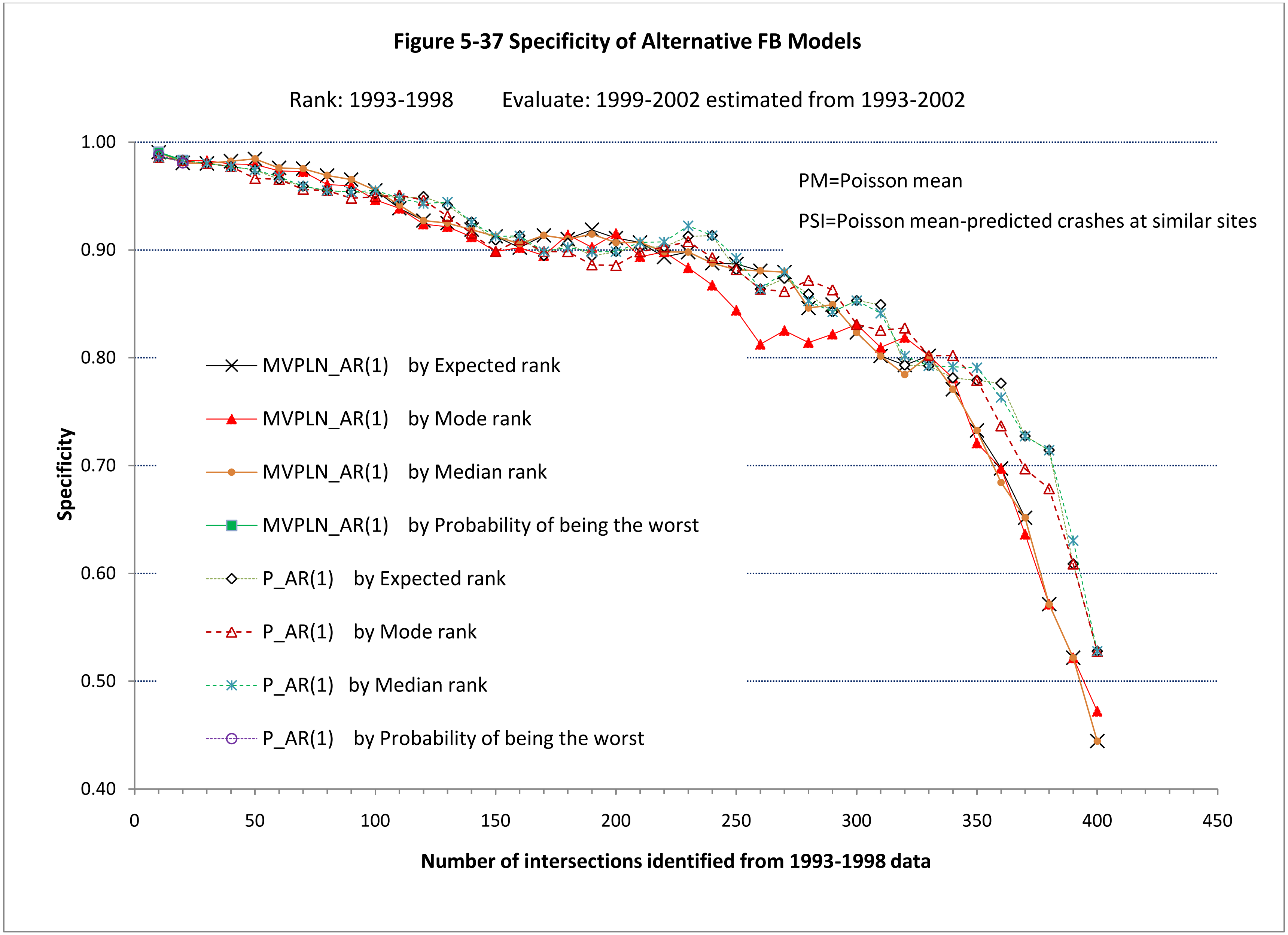


Figure 5-37 Specificity of Alternative FB Models
Rank: 1993-1998
Evaluate: 1999-2002 estimated from 1993-2002
PM=Poisson mean
PSI=Poisson mean-predicted crashes at similar sites
MVPLN_AR(1) by Expected rank
MVPLN_AR(1) by Mode rank
MVPLN_AR(1) by Median rank
MVPLN_AR(1) by Probability of being the worst
P_AR(1) by Expected rank
P_AR(1) by Mode rank
P_AR(1) by Median rank
P_AR(1) by Probability of being the worst
Specificity
1.00
0.90
0.80
0.70
0.60
0.50
0.40
0
50
100
150
200
250
300
350
400
450
Number of intersections identified from 1993-1998 data

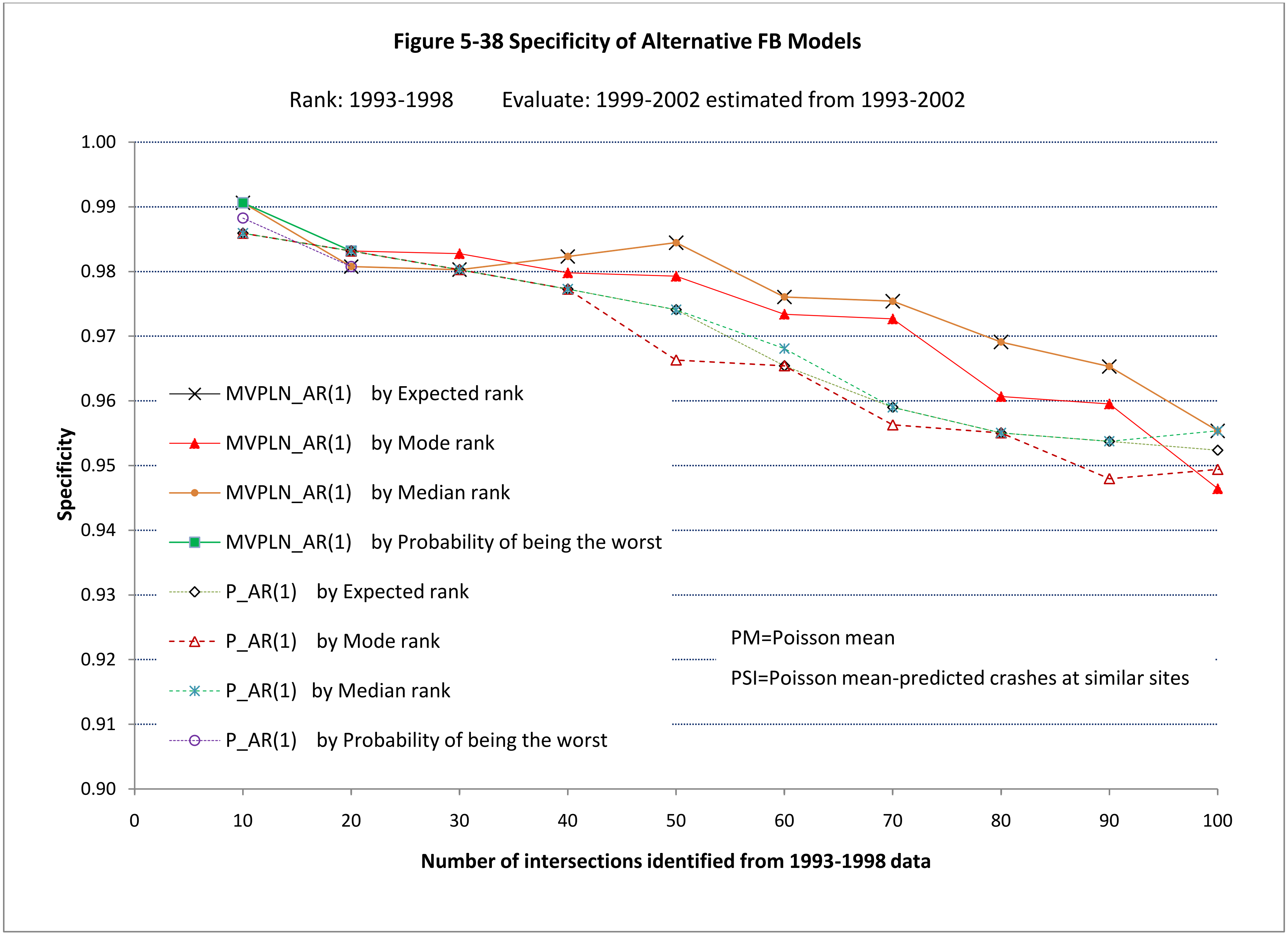


Figure 5-38 Specificity of Alternative FB Models
Rank: 1993-1998
Evaluate: 1999-2002 estimated from 1993-2002
Specificity
1.00
0.99
0.98
0.97
0.96
0.95
0.94
0.93
0.92
0.91
0.90
0
10
20
30
40
50
60
70
80
90
100
Number of intersections identified from 1993-1998 data
MVPLN_AR(1) by Expected rank
MVPLN_AR(1) by Mode rank
MVPLN_AR(1) by Median rank
MVPLN_AR(1) by Probability of being the worst
P_AR(1) by Expected rank
P_AR(1) by Mode rank
P_AR(1) by Median rank
P_AR(1) by Probability of being the worst
PM=Poisson mean
PSI=Poisson mean-predicted crashes at similar sites

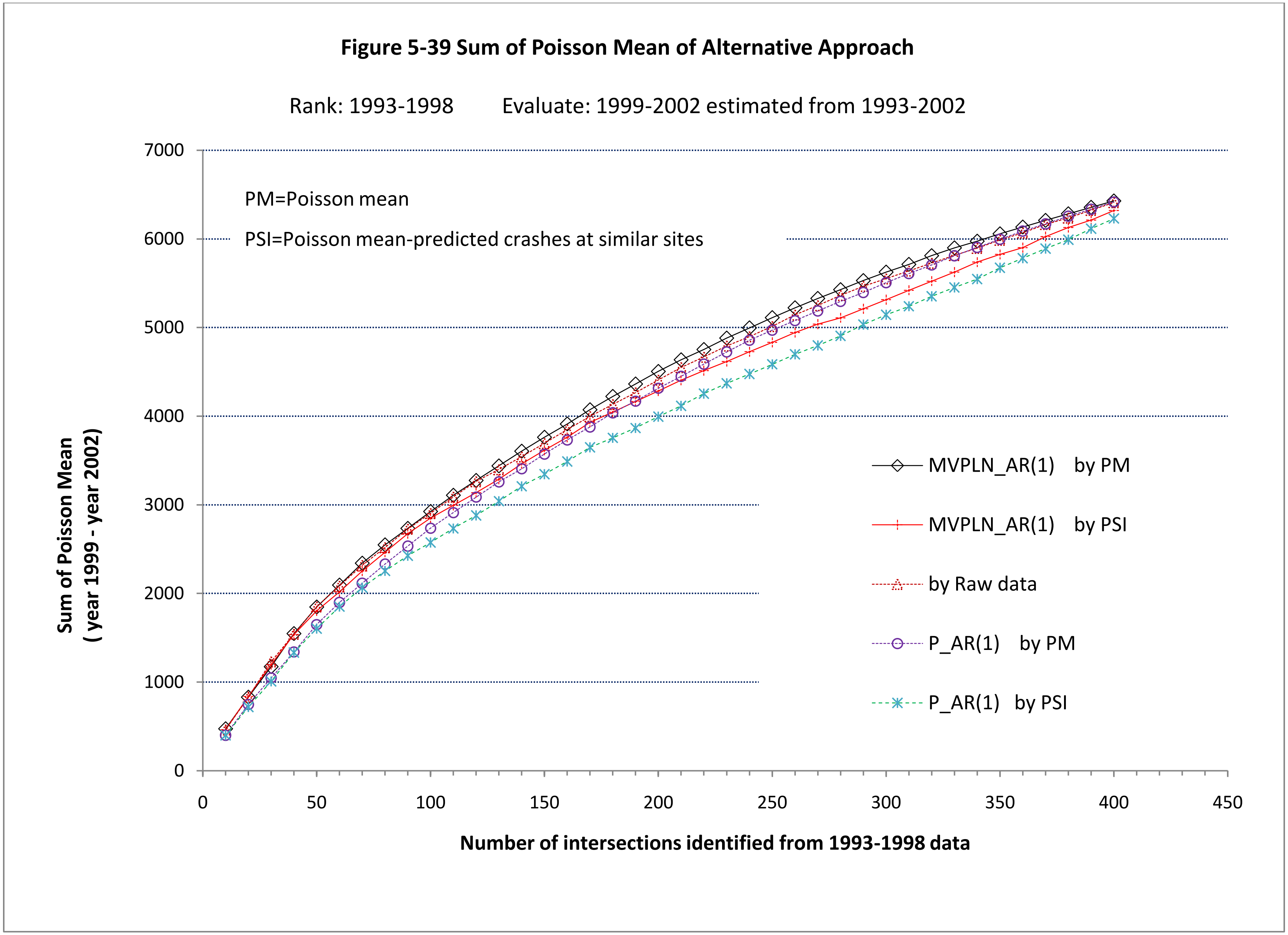


Figure 5-39 Sum of Poisson Mean of Alternative Approach
Rank: 1993-1998
Evaluate: 1999-2002 estimated from 1993-2002
PM=Poisson mean
PSI=Poisson mean-predicted crashes at similar sites
Sum of Poisson Mean
(year 1999 - year 2002)
Number of intersections identified from 1993-1998 data
MVPLN_AR(1) by PM
MVPLN_AR(1) by PSI
by Raw data
P_AR(1) by PM
P_AR(1) by PSI
0
1000
2000
3000
4000
5000
6000
7000
0
50
100
150
200
250
300
350
400
450

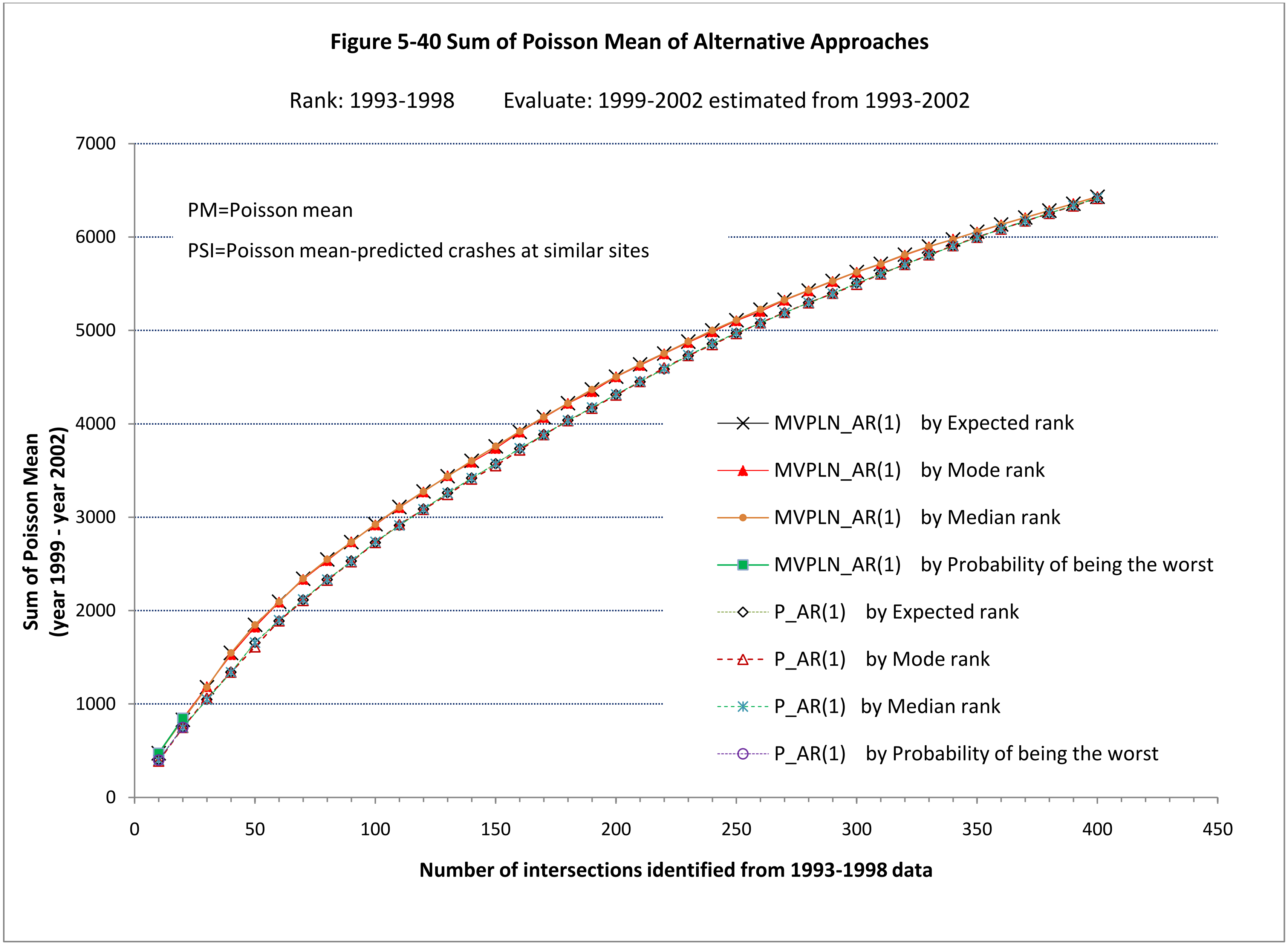


Figure 5-40 Sum of Poisson Mean of Alternative Approaches
Rank: 1993-1998 Evaluate: 1999-2002 estimated from 1993-2002
PM=Poisson mean
PSI=Poisson mean-predicted crashes at similar sites
Sum of Poisson Mean (year 1999 - year 2002)
Number of intersections identified from 1993-1998 data
MVPLN_AR(1) by Expected rank
MVPLN_AR(1) by Mode rank
MVPLN_AR(1) by Median rank
MVPLN_AR(1) by Probability of being the worst
P_AR(1) by Expected rank
P_AR(1) by Mode rank
P_AR(1) by Median rank
P_AR(1) by Probability of being the worst
0
1000
2000
3000
4000
5000
6000
7000
0
50
100
150
200
250
300
350
400
450

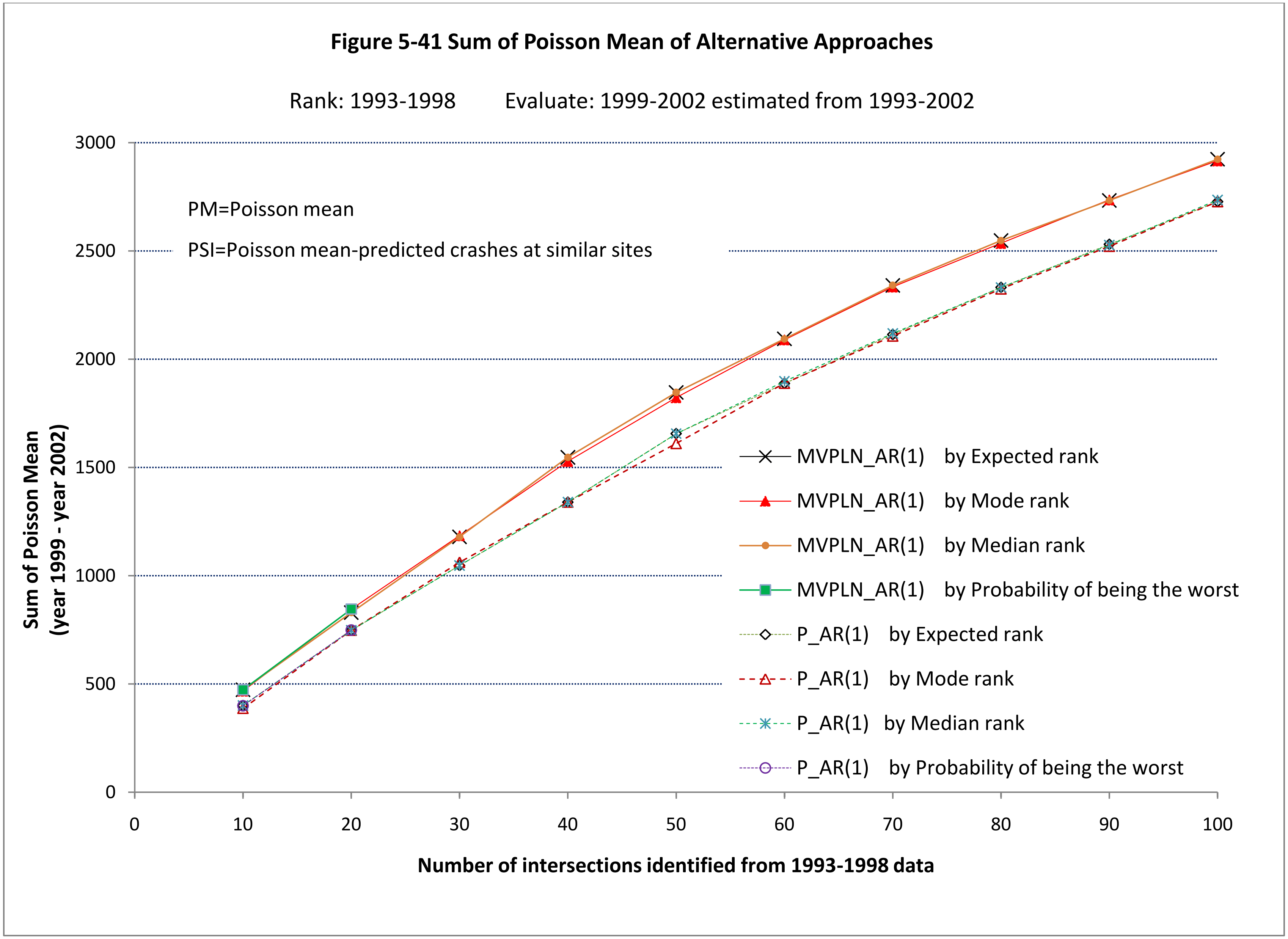


Figure 5-41 Sum of Poisson Mean of Alternative Approaches
Rank: 1993-1998 Evaluate: 1999-2002 estimated from 1993-2002
PM=Poisson mean
PSI=Poisson mean-predicted crashes at similar sites
Sum of Poisson Mean (year 1999 - year 2002)
Number of intersections identified from 1993-1998 data
3000
2500
2000
1500
1000
500
0
0
10
20
30
40
50
60
70
80
90
100
MVPLN_AR(1) by Expected rank
MVPLN_AR(1) by Mode rank
MVPLN_AR(1) by Median rank
MVPLN_AR(1) by Probability of being the worst
P_AR(1) by Expected rank
P_AR(1) by Mode rank
P_AR(1) by Median rank
P_AR(1) by Probability of being the worst

Table 5-19 Posterior Means of Covariance Matrix (Σ) of random effects

	Year 1993 - Year 1998				
	Severity 1	*Severity 2*	*Severity 3*	*Severity 4*	*Severity 5*
Severity 1	0.17				
Severity 2	0.14	0.14			
Severity 3	0.11	0.11	0.11		
Severity 4	0.09	0.11	0.11	0.14	
Severity 5	0.07	0.09	0.10	0.12	0.14
	Year 1993 - Year 2002				
	Severity 1	*Severity 2*	*Severity 3*	*Severity 4*	*Severity 5*
Severity 1	0.21				
Severity 2	0.20	0.24			
Severity 3	0.16	0.20	0.20		
Severity 4	0.12	0.17	0.18	0.18	
Severity 5	0.08	0.14	0.15	0.16	0.18

Table 5-20 Posterior Means of Correlation Matrix of random effects

	Year 1993 - Year 1998				
	Severity 1	*Severity 2*	*Severity 3*	*Severity 4*	*Severity 5*
Severity 1	1				
Severity 2	0.92	1			
Severity 3	0.77	0.88	1		
Severity 4	0.62	0.78	0.91	1	
Severity 5	0.43	0.63	0.79	0.90	1
	Year 1993 - Year 2002				
	Severity 1	*Severity 2*	*Severity 3*	*Severity 4*	*Severity 5*
Severity 1	1				
Severity 2	0.91	1			
Severity 3	0.76	0.92	1		
Severity 4	0.61	0.83	0.93	1	
Severity 5	0.41	0.67	0.81	0.92	1

5.7.6 Sensitivity Analysis of Ranking Criteria

This part of the study conducted a sensitivity analysis of the ranking criteria for both multivariate and univariate approaches. Evaluation criteria, which include sensitivity and specificity, the sum of the PM and sum of the PSI in the evaluation period, were used to identify the most appropriate criterion from all of the ranking criteria. Hence, this can be considered as a confirmation of the previous conclusion from the single severity study. From that study, it was concluded that PM, expected rank, and median rank are generally deemed to be comparable to the best ranking criteria if the order of the individually ranked sites is not a concern within the group; otherwise expected rank provides better results than PM. Mode rank provides comparable results, but sometimes may provide better results, especially for the top ranked sites. The probability of being the worst is also a comparably good ranking criterion, but is only available for top ranking a few sites in that the majority in the group has a probability of zero for being the worst.

A. MVPLN_AR (1) model

The evaluation with all ranking criteria is presented in Figures 5-42 to 5-47. Based on the evaluation criteria, sensitivity, specificity and sum of the PM, similar results are obtained as for a single level crash analysis in that PM, expected rank and median rank provide similarly promising ranking results. Mode rank and the probability of being the worst have the best results for the top 10, 20 and 30 ranked sites (the probability of being the worst only has 20 ranked sites in the evaluation period) in terms of all evaluation criteria except for the sum of the PSI. For the evaluation criteria, sensitivity, specificity and the sum of the PSI of the top 10, 20 and 30 ranked sites, it seems that PSI has good or better results. However, based on the sum of the PM, it is the worst ranking criteria. In general, PSI does not have good results as other explored ranking criteria, except for raw count data which always provides poor ranking results. The reason that PSI performs well in terms of sensitivity for the top 10 or 20 ranked sites, but does not perform well in terms of the sum of the PM, is probably the error caused by a tiny difference of the PSI in the ranking period for a small sample size. When the number of ranked sites increases, this error diminishes. To eliminate such errors, it may be better by introducing multilevel of criteria as proposed earlier. It can be concluded that there is no difference in the ranked sites if the difference in the decision parameters (such as expected rank, mode rank, etc) is located within a similar range, say, the difference of PM is less than 0.05 crashes/(site-year). Then, the second level of ranking criteria, such as PSI, can be introduced and so on. Ranking results might be improved in this way.

.

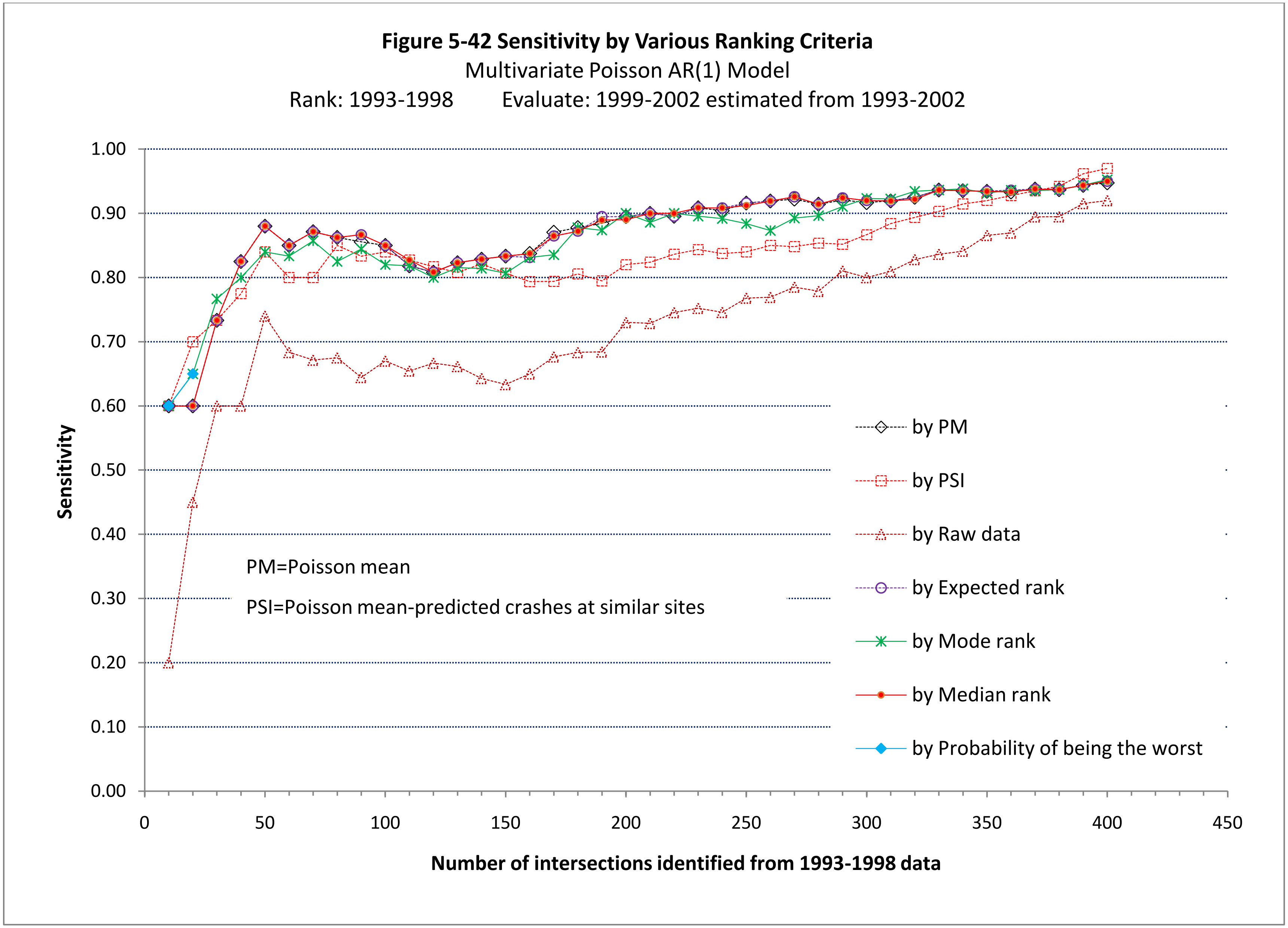


Figure 5-42 Sensitivity by Various Ranking Criteria
Multivariate Poisson AR(1) Model
Rank: 1993-1998 Evaluate: 1999-2002 estimated from 1993-2002
Sensitivity
1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
0
50
100
150
200
250
300
350
400
450
Number of intersections identified from 1993-1998 data
PM=Poisson mean
PSI=Poisson mean-predicted crashes at similar sites
by PM
by PSI
by Raw data
by Expected rank
by Mode rank
by Median rank
by Probability of being the worst

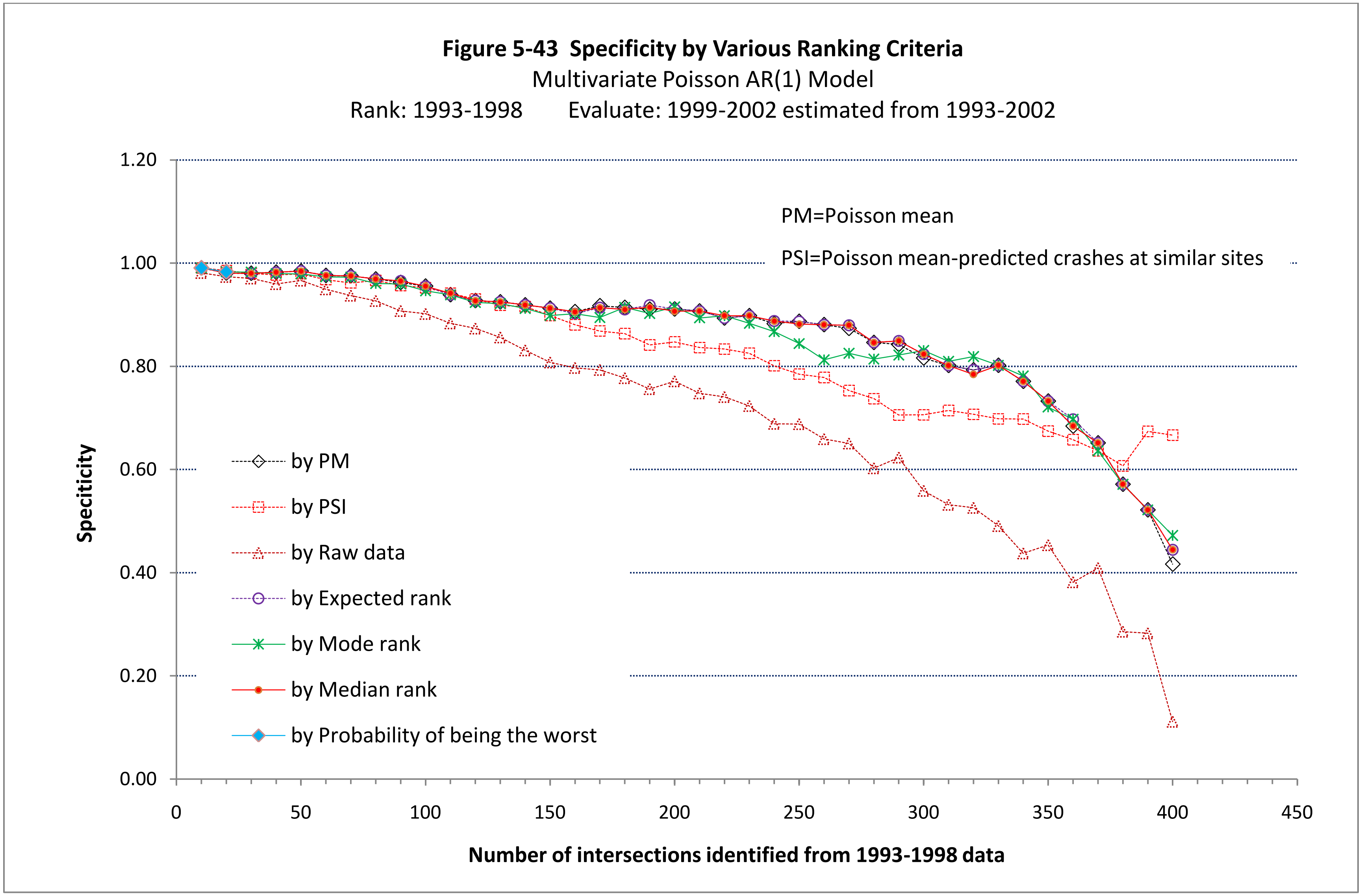


Figure 5-43 Specificity by Various Ranking Criteria
Multivariate Poisson AR(1) Model
Rank: 1993-1998 Evaluate: 1999-2002 estimated from 1993-2002
PM=Poisson mean
PSI=Poisson mean-predicted crashes at similar sites
Specificity
1.20
1.00
0.80
0.60
0.40
0.20
0.00
0
50
100
150
200
250
300
350
400
450
by PM
by PSI
by Raw data
by Expected rank
by Mode rank
by Median rank
by Probability of being the worst
Number of intersections identified from 1993-1998 data

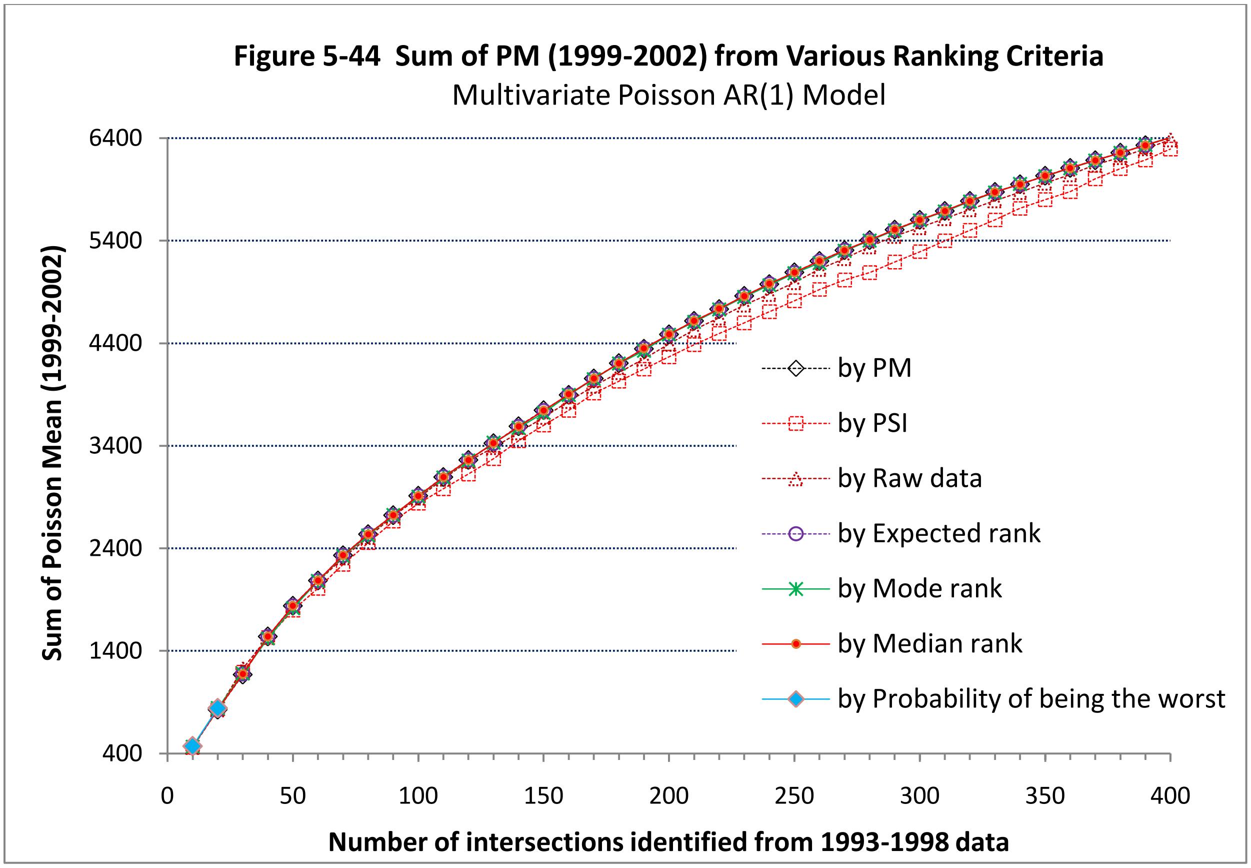


Figure 5-44 Sum of PM (1999-2002) from Various Ranking Criteria
Multivariate Poisson AR(1) Model
Sum of Poisson Mean (1999-2002)
6400
5400
4400
3400
2400
1400
400
0
50
100
150
200
250
300
350
400
Number of intersections identified from 1993-1998 data
by PM
by PSI
by Raw data
by Expected rank
by Mode rank
by Median rank
by Probability of being the worst

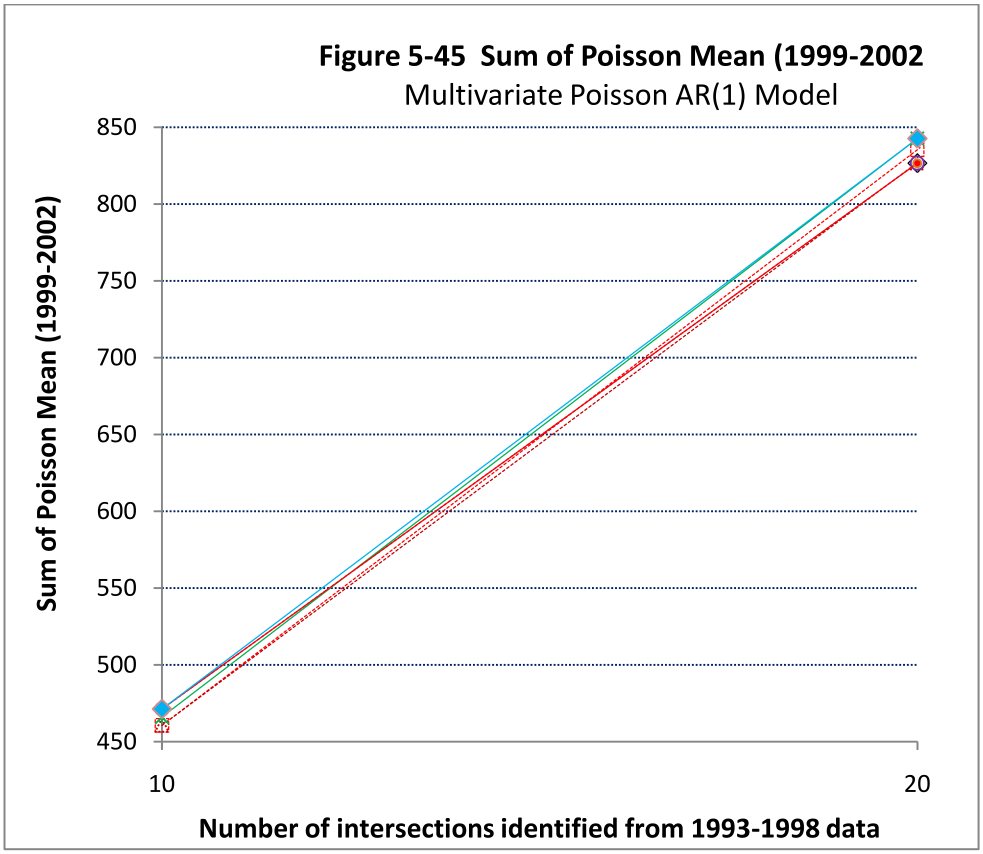


Figure 5-45 Sum of Poisson Mean (1999-2002
Multivariate Poisson AR(1) Model
Sum of Poisson Mean (1999-2002)
850
800
750
700
650
600
550
500
450
10
20
Number of intersections identified from 1993-1998 data

Figure 5-46 Sum of PSI (1999-2002) from Various Ranking Criteria
Multivariate Poisson AR(1) Model

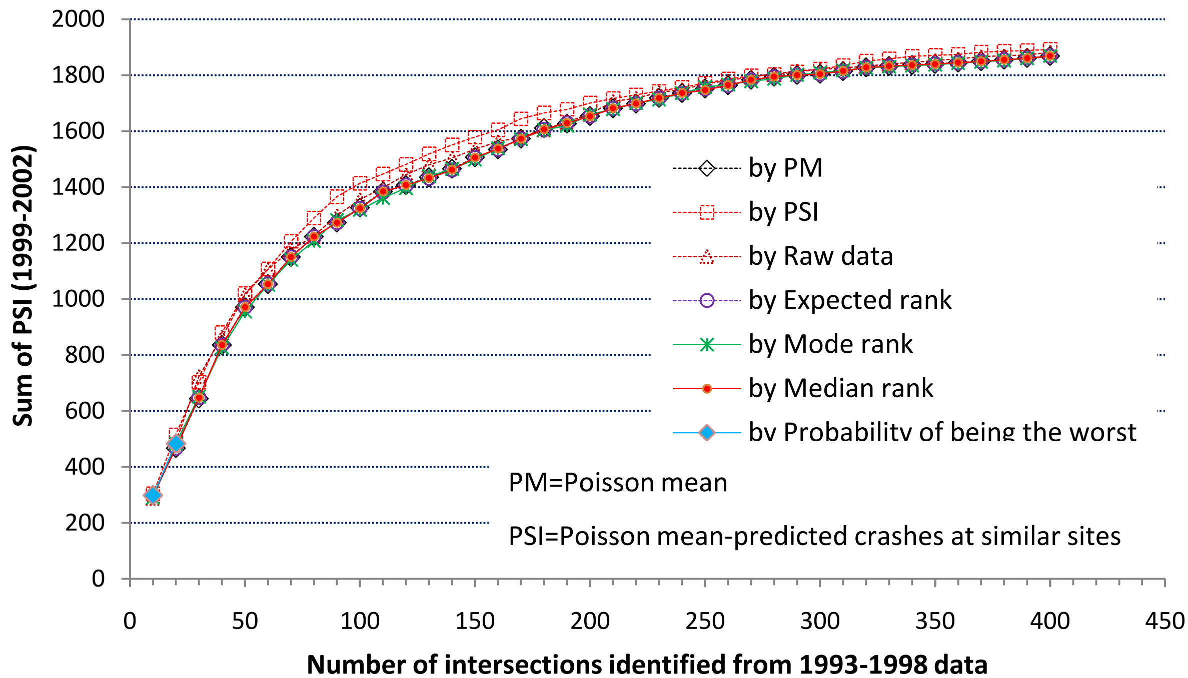
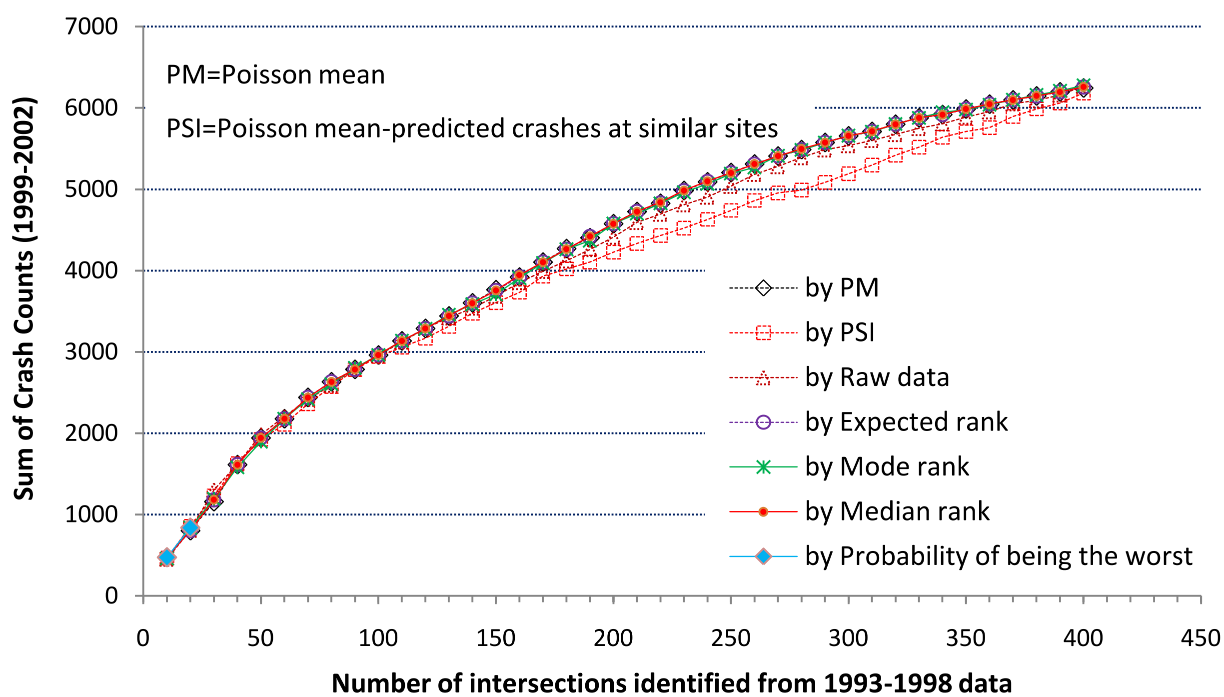
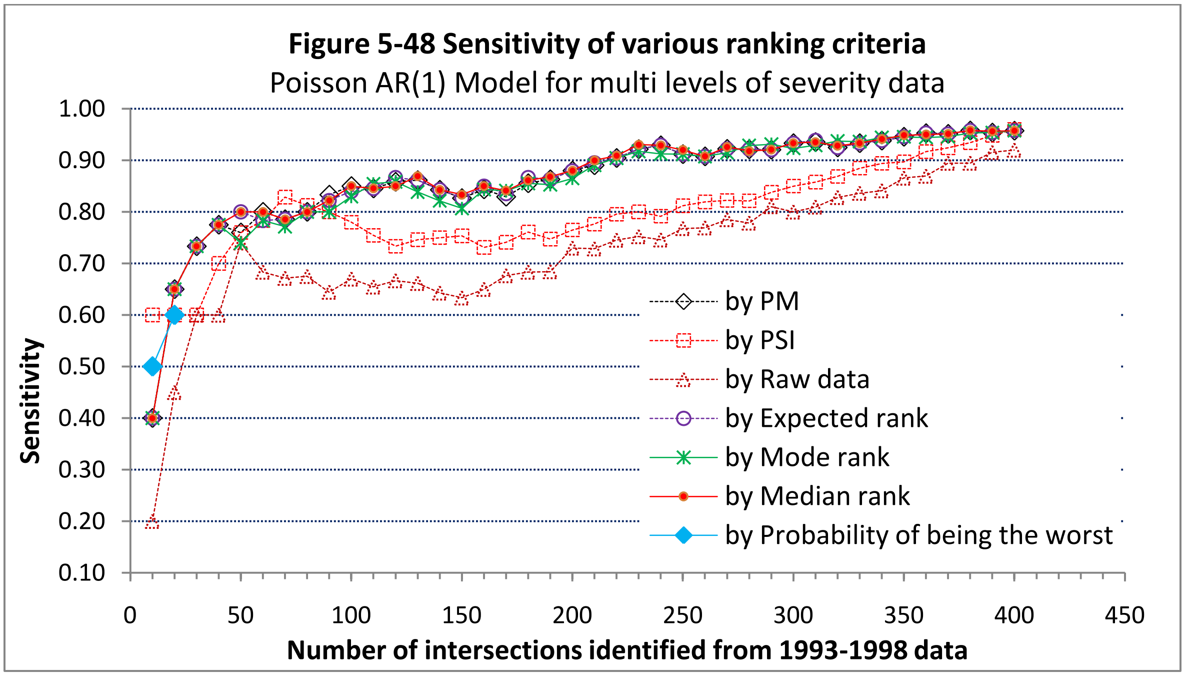


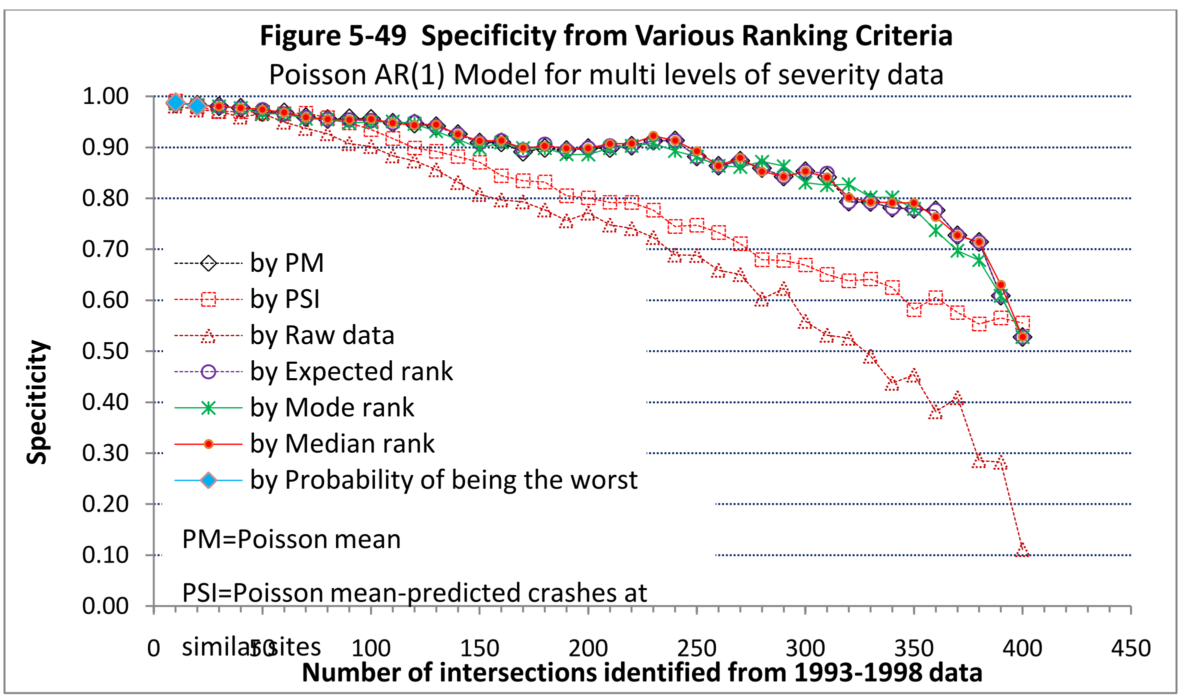
Figure 5-47 Sum of Observed Crashes (1999-2002) from Various Ranking Criteria
Multivariate Poisson AR(1) Model



B. Univariate P_AR (1) model

The evaluation results of the ranking criteria from the univariate P_AR(1) model are presented in Figures 5-48 to 5-51. The similar pattern as the multivariate approach can be observed in these figures.





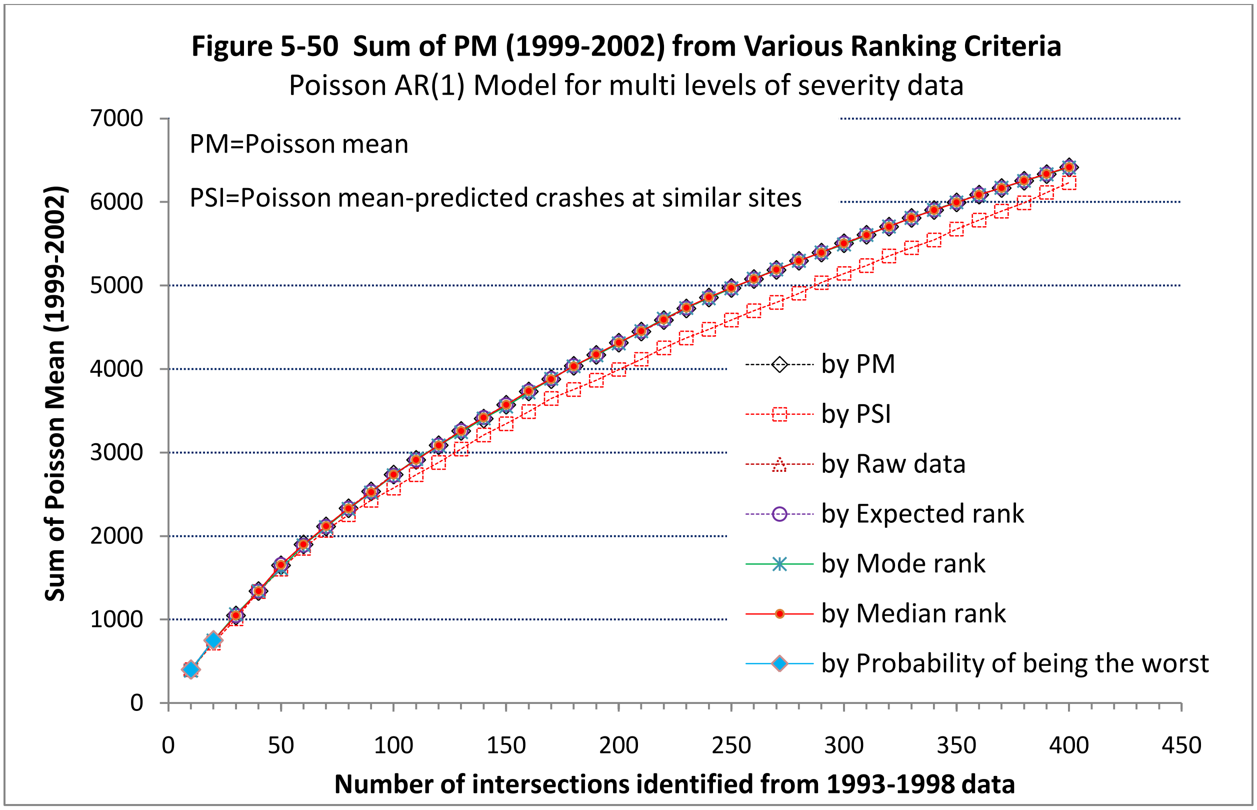


Figure 5-50 Sum of PM (1999-2002) from Various Ranking Criteria
Poisson AR(1) Model for multi levels of severity data
PM=Poisson mean
PSI=Poisson mean-predicted crashes at similar sites
Sum of Poisson Mean (1999-2002)
Number of intersections identified from 1993-1998 data
by PM
by PSI
by Raw data
by Expected rank
by Mode rank
by Median rank
by Probability of being the worst
0
1000
2000
3000
4000
5000
6000
7000
0
50
100
150
200
250
300
350
400
450

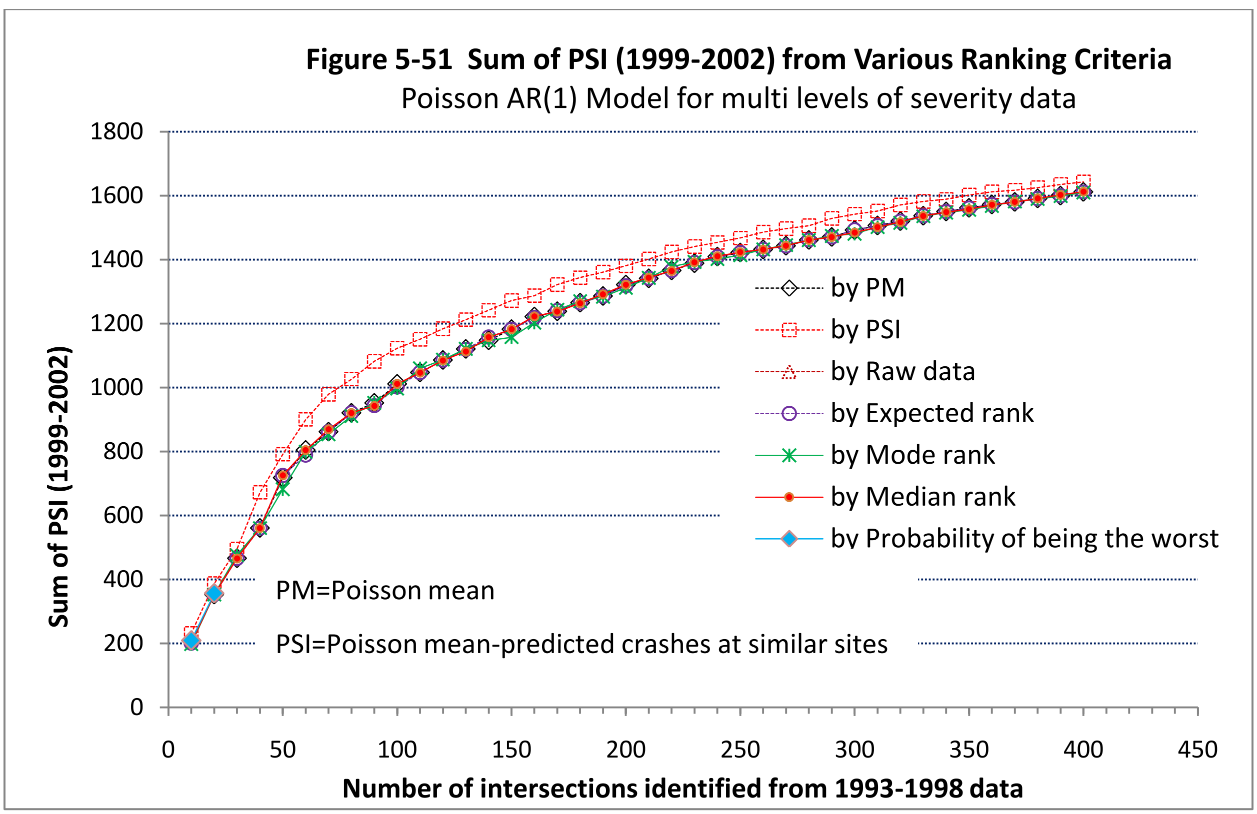


Figure 5-51 Sum of PSI (1999-2002) from Various Ranking Criteria
Poisson AR(1) Model for multi levels of severity data
Sum of PSI (1999-2002)
Number of intersections identified from 1993-1998 data
by PM
by PSI
by Raw data
by Expected rank
by Mode rank
by Median rank
by Probability of being the worst
PM=Poisson mean
PSI=Poisson mean-predicted crashes at similar sites
0
200
400
600
800
1000
1200
1400
1600
1800
0
50
100
150
200
250
300
350
400
450

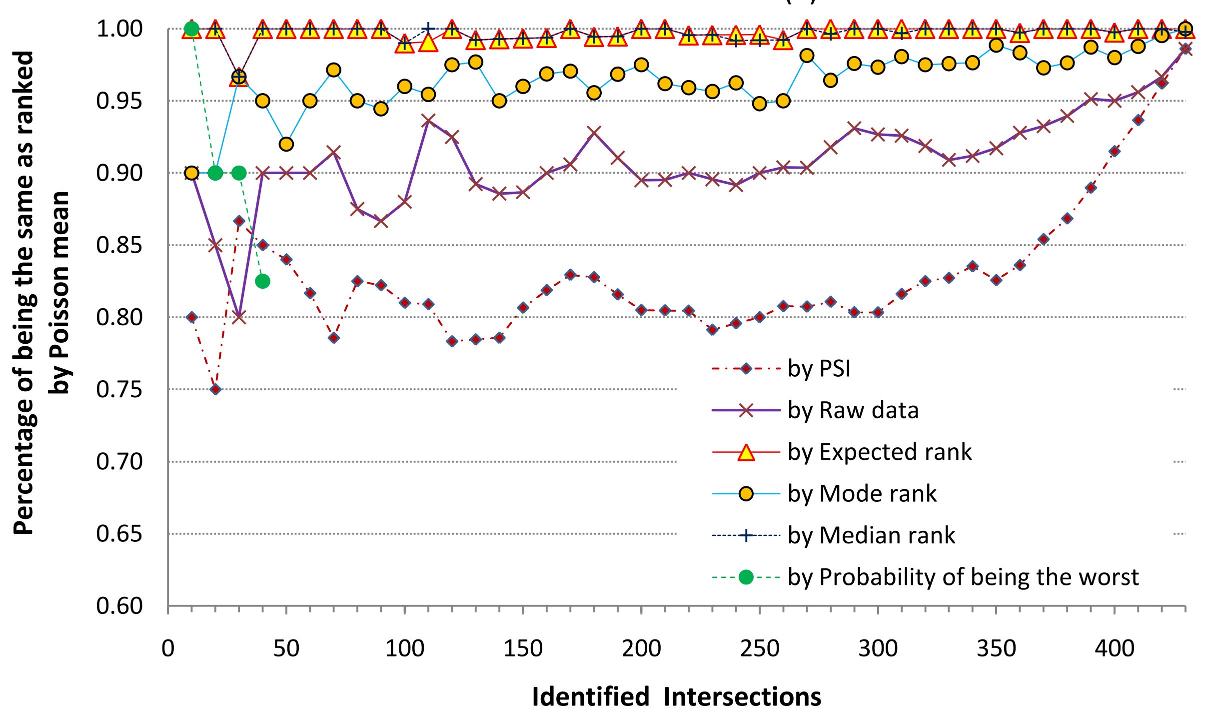
A. Comparison of ranked sites by other criteria with by PM or expected rank

Following the same procedure as before, the set of ranked sites by PM or expected rank was used as a baseline, and sites identified by other criteria were compared with those by PM or by expected rank. The percentages of identical sites occurring in both by PM or expected rank and by the criterion of interest was calculated by Equations 5-17 and 5-18 and are shown in Figures 5-52 to 5-59 for multivariate and univariate approaches.

The median and expected ranks closely identify the same sites as PM, while mode ranking results in at least 90% of the same sites if the order of the ranked individual site is not a concern. PSI provides the greatest differences in ranking in comparison to PM or expected rank, while sites ranked by raw data are also systematically different from those ranked by PM. Thus, it is further confirmed that PSI cannot be used as a major ranking criterion, but might, nevertheless, be employed as a secondary criterion.

Figure 5-52 Same Portion as Ranked by Poisson Mean without Considering the Order

Multivariate Poisson AR(1) Model



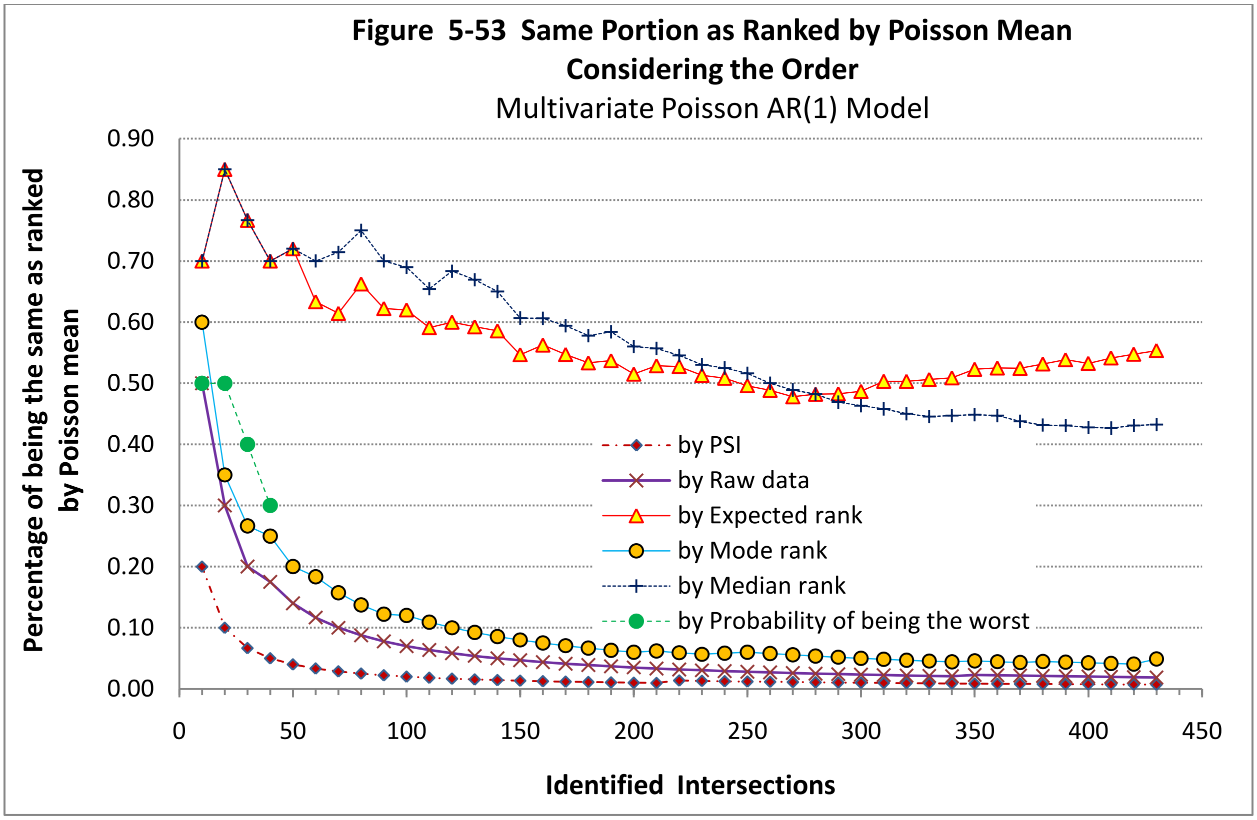


Figure 5-53 Same Portion as Ranked by Poisson Mean
Considering the Order
Multivariate Poisson AR(1) Model
Percentage of being the same as ranked by Poisson mean
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
0
50
100
150
200
250
300
350
400
450
Identified Intersections
by PSI
by Raw data
by Expected rank
by Mode rank
by Median rank
by Probability of being the worst

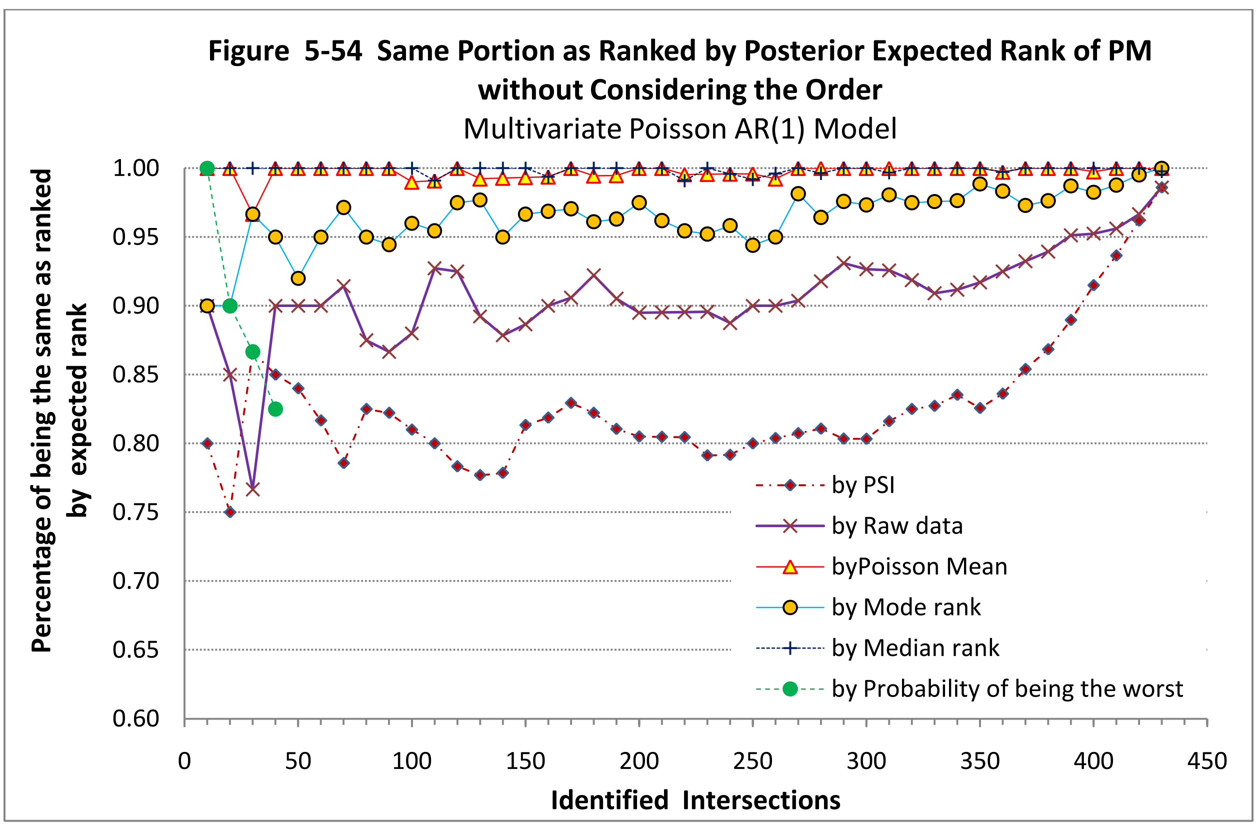


Figure 5-54 Same Portion as Ranked by Posterior Expected Rank of PM
without Considering the Order
Multivariate Poisson AR(1) Model
Percentage of being the same as ranked by expected rank
1.00
0.95
0.90
0.85
0.80
0.75
0.70
0.65
0.60
0
50
100
150
200
250
300
350
400
450
Identified Intersections
by PSI
by Raw data
byPoisson Mean
by Mode rank
by Median rank
by Probability of being the worst

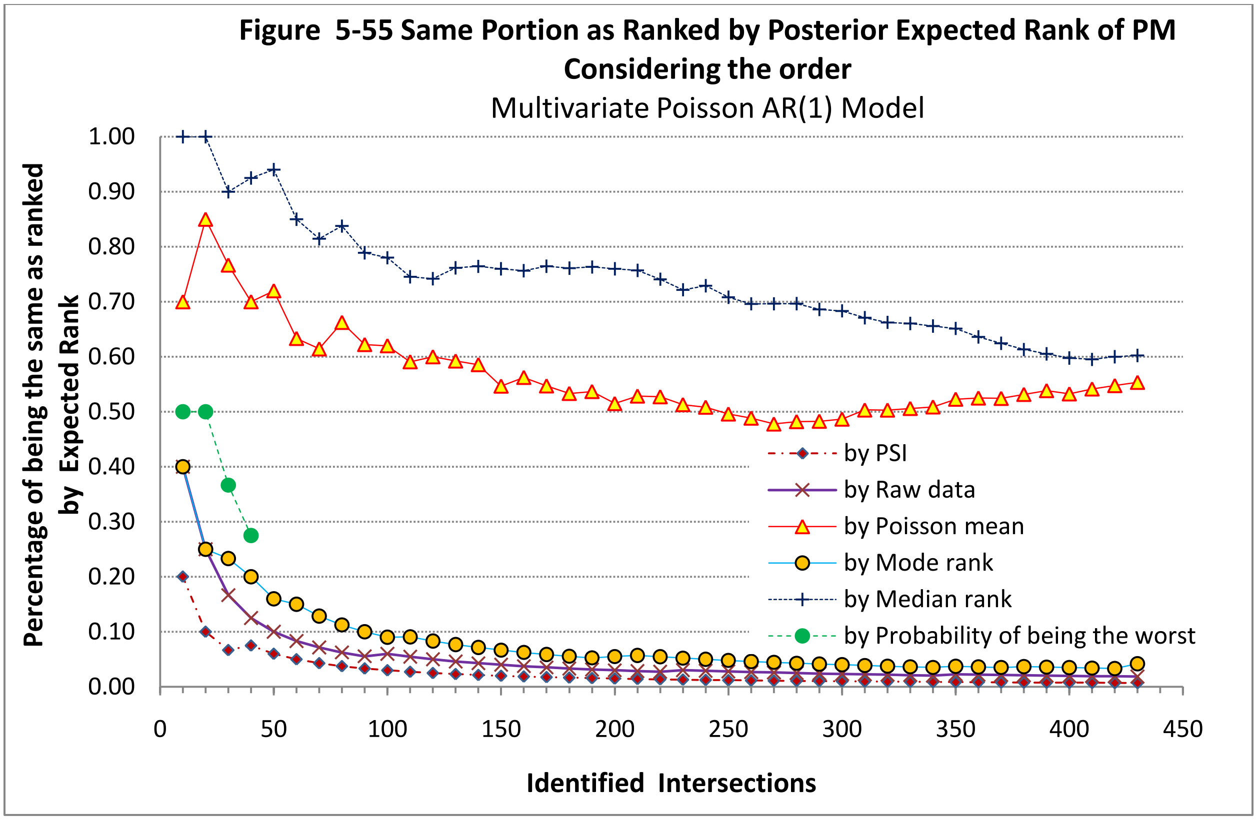


Figure 5-55 Same Portion as Ranked by Posterior Expected Rank of PM
Considering the order
Multivariate Poisson AR(1) Model
Percentage of being the same as ranked by Expected Rank
Identified Intersections
by PSI
by Raw data
by Poisson mean
by Mode rank
by Median rank
by Probability of being the worst
1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
0
50
100
150
200
250
300
350
400
450

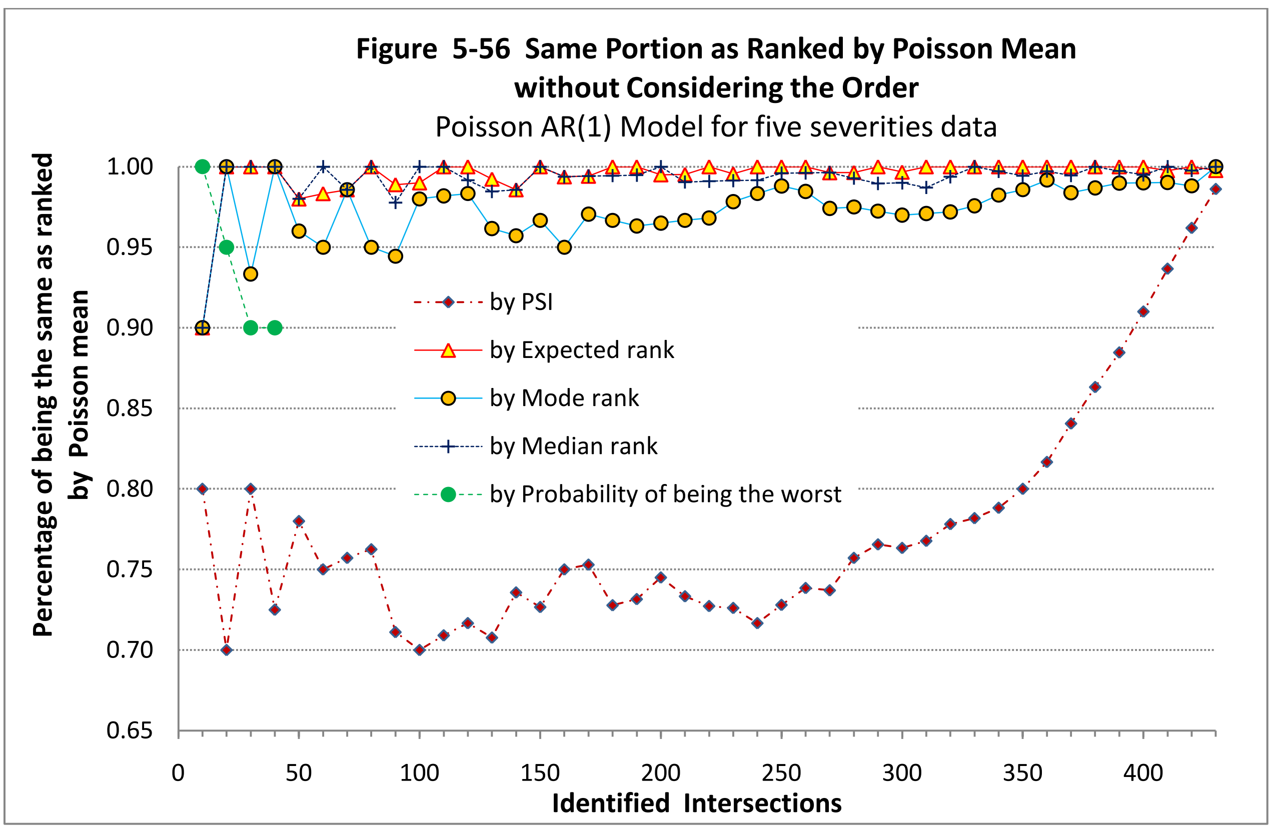


Figure 5-56 Same Portion as Ranked by Poisson Mean
without Considering the Order
Poisson AR(1) Model for five severities data
Percentage of being the same as ranked by Poisson mean
Identified Intersections
by PSI
by Expected rank
by Mode rank
by Median rank
by Probability of being the worst
1.00
0.95
0.90
0.85
0.80
0.75
0.70
0.65
0
50
100
150
200
250
300
350
400

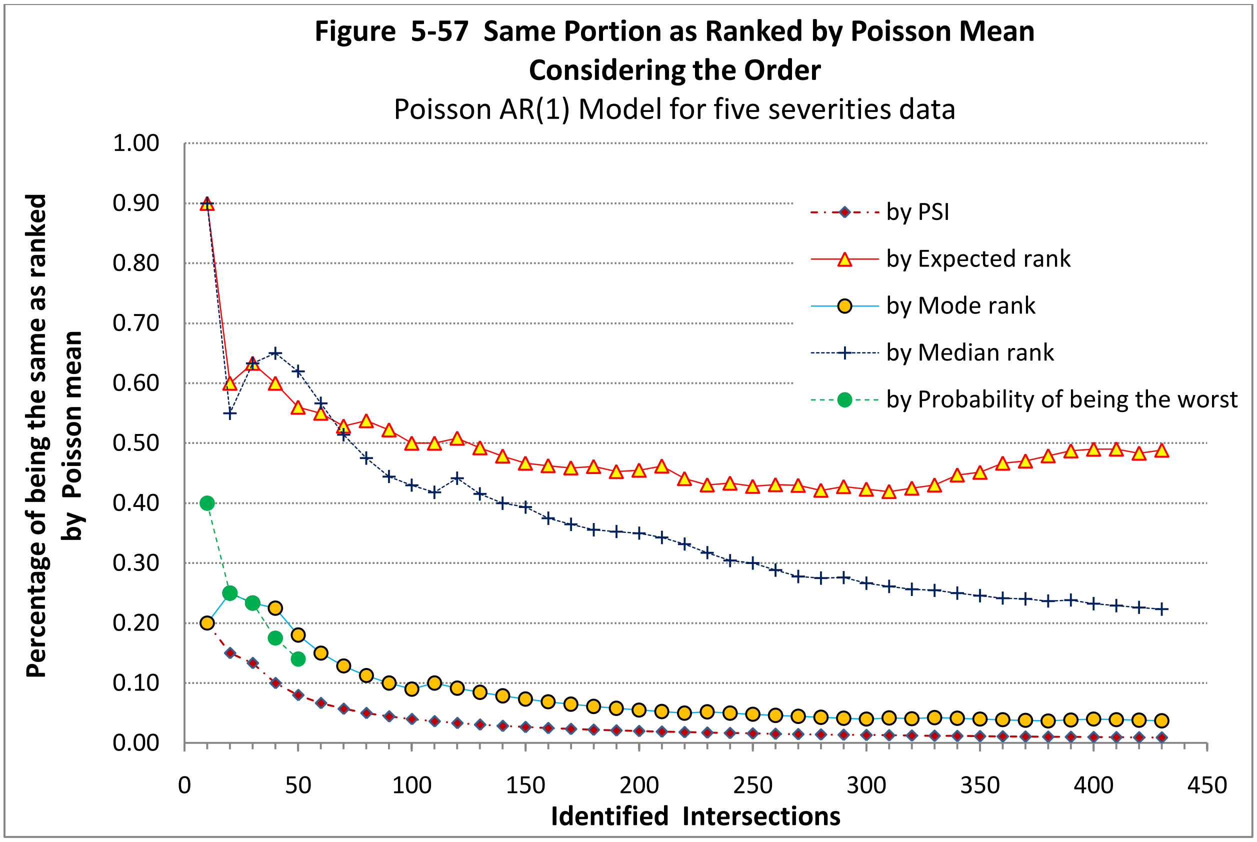


Figure 5-57 Same Portion as Ranked by Poisson Mean
Considering the Order
Poisson AR(1) Model for five severities data
Percentage of being the same as ranked by Poisson mean
Identified Intersections
by PSI
by Expected rank
by Mode rank
by Median rank
by Probability of being the worst
1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
0
50
100
150
200
250
300
350
400
450

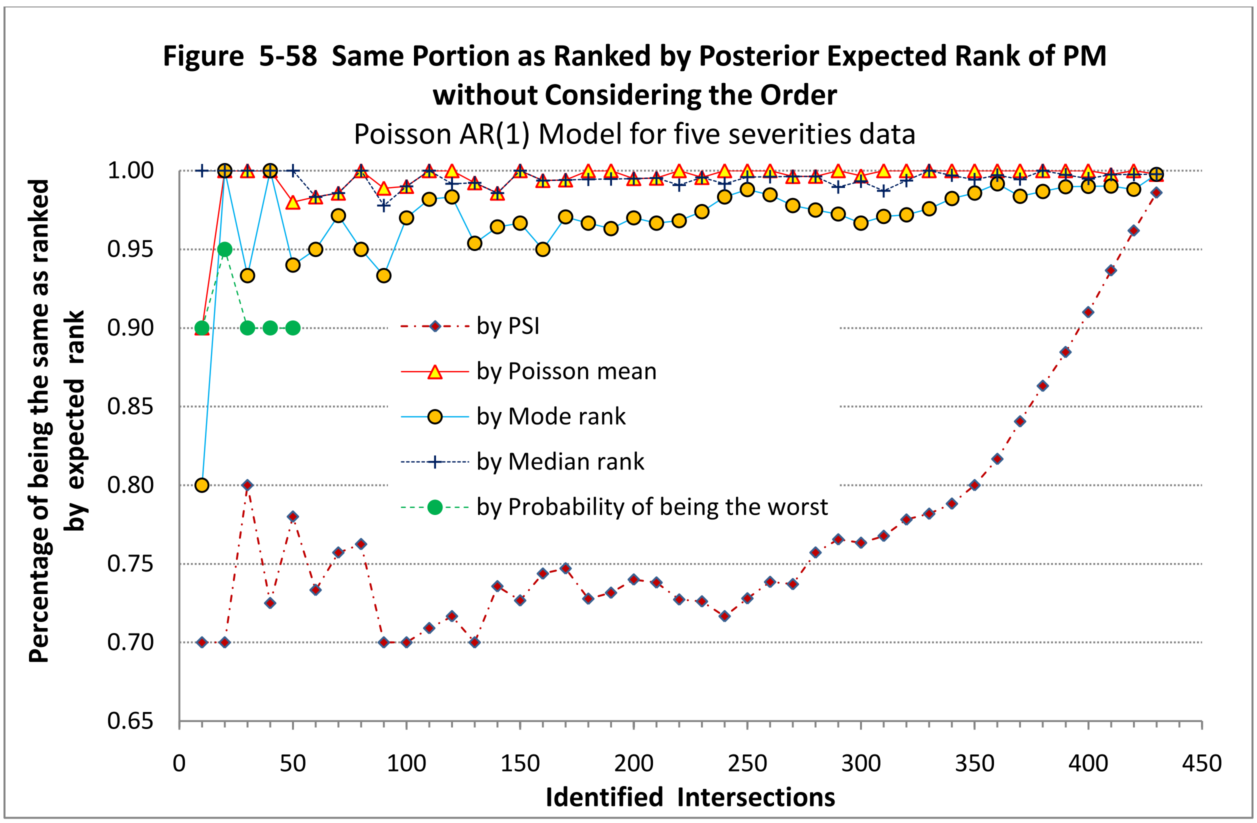
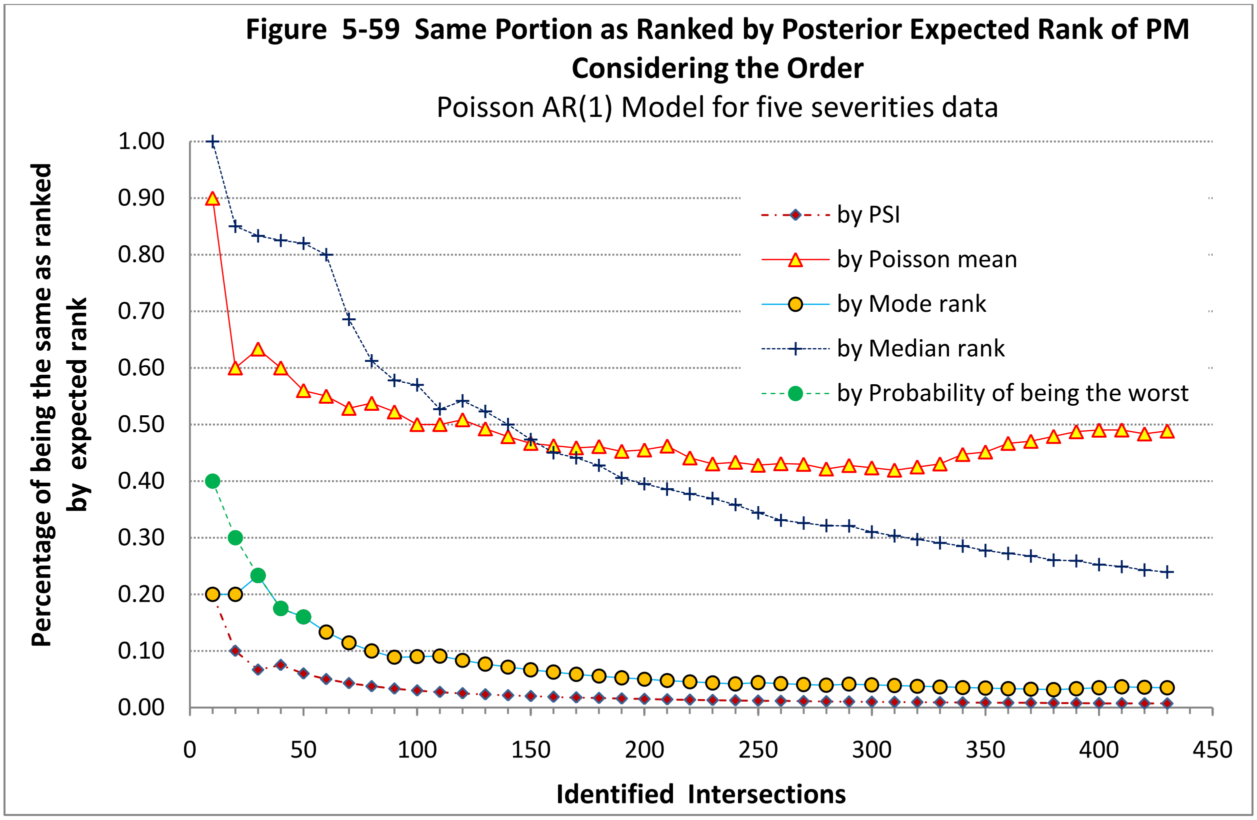


Figure 5-58 Same Portion as Ranked by Posterior Expected Rank of PM
without Considering the Order
Poisson AR(1) Model for five severities data
Percentage of being the same as ranked by expected rank
Identified Intersections
by PSI
by Poisson mean
by Mode rank
by Median rank
by Probability of being the worst
1.00
0.95
0.90
0.85
0.80
0.75
0.70
0.65
0
50
100
150
200
250
300
350
400
450



If the order of the ranked individual site is taken into consideration, the ranked pattern is quite different from those that do not consider the order, as is evident from Figures 5-53, 5-55, 5-57 and 5-59. The lists ranked in order by median, PM and expected rank are different while median rank provides the closest ranking to expected rank, but the similarity of results by median rank and by expected rank diminishes in comparison to the previous results for a single level of severity.

5.8 SUMMARY

Ten years of data from 1993-2002 that detail 726 unsignalized four legged intersections in California were used to evaluate the FB method for hot spot identification in comparison with the EB method, while 436 top ranked sites with five levels of severity data based on combined crash counts were selected for an evaluation study of the multivariate FB method. A thorough evaluation of the univariate FB versus EB method for single level severity data and multivariate FB versus univariate FB for multilevel of severity data, as well as the performance of various ranking and evaluation criteria, was presented in this chapter.

For the univariate FB study with 726 sites, a total of 11 FB models were developed. Poisson AR (1) was identified to be the best model for comparison with the EB method and for further study and the AR (1) model was applied to multilevel severity crashes. Two time frames (1996-1998 and 1993-1998) were used to rank the sites for the evaluation study. The period 1999-2002 was selected as the evaluation period. Estimates of the true mean for the evaluation period were derived from the model developed using a total of 10 years of data. It is reasonable to be believed the 10 years of data can provide a better estimation of the true mean for the evaluation period.

A total of 8 ranking criteria, which include posterior Poisson mean, posterior expected rank, posterior mode rank, posterior median rank, posterior probability of being the worst, raw data, posterior PSI and posterior PPSI have been examined for the developed FB models. Specifically, the mode rank of the posterior distribution of the Poisson mean was proposed as a ranking criterion for the evaluation study and it proved to be promising in that it can sometimes provide the best results, especially for top ranked sites. In addition, the sum of the Poisson mean and sum of the PSI in the evaluation period are proposed as evaluation criteria. The sum of the Poisson mean was found to be a solid evaluation criterion; especially for limited numbers of top ranked sites. As well, it does not assume homogeneity, unlike sensitivity and specificity. The evaluation criteria include sensitivity and specificity, the sum of the PM, sum of the PSI, sum of crash counts and sum of the PPSI. The following conclusions can be obtained from this aspect of the study:

- It was found that FB provides better results than the EB method in terms of higher sensitivity, specificity, sum of the PM and even sum of crash counts in the evaluation period regardless of whether ranking is by PM or PSI (see Table 5-10).

- Posterior expected rank, median rank and PM provide almost the same results if the order of the individual sites in the ranked group is not considered. Expected rank has somewhat better ranking results than PM and median rank.

- If the order of the individual sites in a ranked group is a concern, expected and median ranks of the posterior PM provide different lists. Median rank provides the closest ranked ordered list to those identified by expected rank, while PM could not provide good results if the order of the ranked sites is a concern.

- Mode rank provides at least 90% of the same identified sites as PM or expected rank without taking into consideration the order of the ranked group. However, there is a substantial difference in rank order in comparison to PM or expected rank. It is shown to provide the best ranking results, especially for the top ranked group.

- The probability of being the worst has the fewest ranked sites in that the majority has zero probability as the worst site. For the top ranked sites, the probability of being the worst may provide the best ranking results.

- Short data history (3 years) can provide better ranking results than longer data history (6 years) for the where identification of only limited top ranked limited is of interest, as is common in a black spot identification program. As the number of ranked sites increases, a longer data history generally provides better results. Further study is necessary to determine the optimal data history for hot spot identification.

- For multilevel severity data, a multivariate approach is better for network ranking than the univariate approach based on the evaluation results, but with longer modeling time. From the model selection results, DIC might be the best criterion compared to others such as AIC and BIC.

- Where only a few top ranked sites are of interest, sensitivity may not be a good evaluation criterion because one false positive can cause a huge difference in sensitivity while the decision parameter may just have a minimal difference (i.e. 10.5 crashes versus 10.51 crashes). In such cases, the sum of the PM might be used as a major evaluation criterion. To eliminate the effect caused by the small differences in decision parameters, multilevel ranking criteria might be necessary. For example, within some small range, where the primary decision parameters produce are essentially the same, a second level ranking criterion could be implemented, and so on. However, this suggestion needs further study.

- It is shown that PSI cannot provide good ranking results in that it has lower values of sensitivity, specificity, sum of the PM and sum of crash counts. It is only a good ranking criterion if it is based on the sum of the PSI or sum of the PPSI. In addition, it provides the most different ranked sites when PM or expected rank is used. PSI might be used as a second level ranking criterion while other reliable criteria, such as expected rank, PM or mode rank etc., are used as first level ranking criteria.

CHAPTER 6 EVALUATION OF FB METHOD FOR TREATMENT EFFECT STUDY

6.1 INTRODUCTION

Network ranking and treatment effect analysis are two major tasks in road safety study. The FB method for network ranking was explored and evaluated in Chapter 5. Treatment effect analysis, one of the most important tasks for road safety analysts, is explored in this Chapter. This aspect of the research has recently been published (Lan et al., 2009; Lan and Persaud, 2010), and some of the documentation below is taken from those sources.

For the past two decades, the empirical Bayesian (EB) method *(*Hauer, 1997; Sayed and Rodrigez, 1999; Hauer, 2002; Turner-Fairbank Highway Research Center, 1999; Persaud and Nguyen, 1998; Persaud et al, 2002*)* has been used successfully to perform this evaluation. A recent paper (Persaud and Lyon 2007) has summarized experience to date with this approach for evaluating safety treatments.

Recently, with the availability of the software package WinBUGS *(*Spiegelhalter et al., 2003*)*, fully Bayesian (FB) approach has been suggested as a useful alternative to the empirical Bayes approach (Lan et al. 2009; Persaud et al. 2010; Carriquiry and Pawlovich 2005; Pawlovich et al. 2006) in that it is believed to require less data for untreated reference sites, it better accounts for uncertainty in data used, and it provides more detailed causal inferences and more flexibility in selecting crash count distributions.

One study of treatment effect analysis using univariate FB was conducted by Pawlovich et al. (2006). This study introduced treatment effect coefficients into the model and employed matched pairs to estimate treatment effects. Pawlovich et al. developed an accident rate model and then used the model to estimate expected crashes in the after period for both the treated sites and the matched reference sites. A 25% reduction in crash frequency per mile was found in their study, which is close to the 24% reduction obtained from the Naïve before-after method *(*Pawlovich et al., 2006). The approach employed by Pawlovich et al. is similar in principle to a conventional comparison group C-G study (Hauer, 1997).

Crash data normally are collected at different severity levels (i.e. fatal, injured, PDO etc.) and pertain to different types (e.g., total, read end, right angle and left turn). Since collisions type and severity level could be correlated, it is natural to believe that a multivariate FB approach might be better for safety analysis based on both crash type and severity. As a matter of fact, the multivariate Poisson (MVP) model was introduced by Tsionas as early as 2001 and several researchers such as Karlis and Meligkotsidou (2005) as well as Ma and Kockelman (2006) have worked on this method. For example, Ma and Kockelman (2006) applied MVP regression approach to assess the effects of covariates on collision counts at different severity levels. The drawbacks of MVP, such as the assumption of equal and nonnegative covariance terms, as well as the inability to account for overdispersion, were subsequently revealed. As a result, this method was not really implemented in treatment effect analysis.

To overcome the shortcomings of MVP, the multivariate Poisson Log normal (MVPLN) model approach was introduced and applied to road safety analysis (Park and Lord 2007; Ma et al. 2008; Aguero-Valverde and Jovanis 2009; El-Basyouny and Sayed 2006; Park et al. 2009). For example, Park and Lord (2007) applied MVPLN approach to 451 three-leg unsignalized intersections in California. They developed a MNPLN model for crash frequency by 5 injury severities using the software MATLAB (Demuth et al., 2006). For comparison, they also developed a univariate Poisson model and a univariate Negative Binomial (NB) model for each crash severity level using SAS. They analyzed the resulting regression coefficients and posterior correlation coefficients matrix of random effects of these five severities. They concluded that MVPLN can provide more accurate estimates in terms of lower standard deviations of parameters for one severity crashes and the posterior correlation matrix showed correlation of latent effects. However, it should be noted that this pattern does not occur in crashes of the other four severities.

Ma et al. (2008) employed MVPLN for Washington State rural two-lane highway crashes with five injury severity levels. Crashes collected from over 7773 homogeneous segments were used to conduct the study using the software R. Univariate EB Poisson and NB models were also investigated. Their results indicated that MVPLN provided better predictions. Aguero-Valverde et al. (2009) and El-Basyouny et al. (2009) both developed MVPLN models and univariate FB Poisson Log normal (PLN) models for crash severity modeling and site ranking using WinBUGs. They found that the MVPLN model is superior to the univariate FB model in terms of overall performance of the model in that MVPLN has a lower DIC value.

Park et al. (2009) applied the MVPLN before-after approach for different severity crash data to evaluate the effect of decreasing posted speed limit on Korean expressways. Their before-after approach is quite different from our before-after FB approach (Lan et al. 2009; Persaud et al. 2010; Lan and Persaud, 2010) in that they introduce an intervention model. 33 treatment sites and 203 reference sites are used to calibrate the intervention model to obtain the expected crash frequency with treatment and to predict the expected crashes without treatment in the after period. The final MVPLN results are compared with those from Naïve (traditional method) and EB approaches. Overall, the MVPLN results are between those from the Naïve and EB, but are quite different from the results of the EB and naïve analyses. It is possible that the limited sample of 33 treatment sites used to calibrate the intervention model and obtain the corresponding coefficients related to treatment may have been too small. If so, then this could be a major limitation of this method in that typically the sample of treatment sites available for evaluation can be quite limited.

Both the MVPLN and univariate FB approaches use Markov Chain Monte Carlo (MCMC) (Gamerman, 2006; Gilks; 1996; Brooks, 1998) methods to derive the posterior distribution of estimates. Previous researchers on this subject favoured MVPLN models based on the estimated regression coefficients (Park and Lord, 2007), predicted results *(*Ma et al., 2008*)* or DIC (Aguero-Valverde and Jovanis, 2009; El-Basyouny and Sayed, 2009). These researchers believe that the MVPLN modeling approach has the following advantages over univariate EB or FB: a) It takes into account correlations that exist among different severity levels; b) It can cope with both overdispersion and a fully general correlation structure in the data; c) It can simultaneously provide estimation results for crashes at different severity levels.

The motivation for this aspect of the research was to build on the pioneering work of earlier researchers who investigated only the univariate FB or the MVPLN approaches for crashes at different injury severity levels. Moreover, previous researchers employed only one function form for expected crashes. In that formulation, $\mathrm{Y}_{\mathrm{i,t}}^{\mathrm{k}}$ is the type k crash frequency at site i in year t and is assumed that $\mathrm{Y}_{\mathrm{i,t}}^{\mathrm{k}} \sim \mathrm{Pois}\ (\exp\ (\varepsilon_{\mathrm{i}}^{\mathrm{k}}) \times \mu_{\mathrm{i,t}}^{\mathrm{k}})$, where $\exp\ (\varepsilon_{\mathrm{i}}^{\mathrm{k}})$ is the random effects for type k crash at site i, $\mathrm{Y}_{\mathrm{i,t}}^{\mathrm{k}}$ follows the MVPLN (Park and Lord, 2007; Ma, et al., 2008; Aguero-Valverde and Jovanis, 2009; El-Basyouny and Sayed, 2009; Park et al., 2009), PG or PLN based on a different form of $\exp\ (\varepsilon_{\mathrm{i}}^{\mathrm{k}})$. The corresponding expected crash frequency $\mu_{\mathrm{i,t}}^{\mathrm{k}}$ has only one form with fixed coefficients in each of the above studies. However, with the powerful software WinBUGs, more forms of $\mu_{\mathrm{i,t}}^{\mathrm{k}}$ such as those incorporating time varying coefficients can be further investigated. Finally, the univariate FB and MVPLN approach have not been examined for before-after treatment effect analysis.

To this end, this part of the research addressed these three knowledge gaps to complement, rather than duplicate, the extensive work by others in this research area. First, a simulation study was employed to explore the univariate FB for treatment effect study. Then the MVPLN approach was explored and evaluated for two cases (a group with high crash counts and one with low crash counts). Finally, the function forms of expected crashes both for univariate and multivariate FB methods were explored while addressing temporal effects.

Two objectives of this study need to be achieved: 1. to examine the univariate FB before after method to see if it can address the regression to the mean (RTM) problem which is common in road safety analysis, and; 2. to examine the MVPLN FB method to see if it can address the RTM problem and if it is superior to the univariate FB before after method, using different type crashes. To this end, two types of data -- univariate crash data and multivariate crash data were used to conduct these before-after FB studies. Four forms of $\mu_{i,t}^{k}$ as mentioned in Chapter 5 were developed and evaluated, and the results of the evaluation study are presented. The study evaluated variations of the before-after univariate FB and MVPLN FB methods by analyzing a hypothetical treatment with no effect at the sites. Then the outcomes from the univariate FB were compared with those from the Naive method, and the results from MVPLN method were compared with those obtained from univariate PG, PLN evaluations and from the naive method. The detailed results are presented below.

6.2 BEFORE-AFTER FB METHOD METHODOLOGY

A before-after FB study, which is similar to the approach used by Aul (2006), was used to evaluate the FB method for treatment effect analysis. The before-after FB approach is similar to the EB approach in that untreated reference group data are used to make inferences and to account for possible effects unrelated to the treatment. This FB method also includes data on the treated sites in the before period to develop inferential models. On the other hand, the EB approach only uses data from reference sites for this purpose.

For convenience, the univariate FB is used here to explain the before-after FB method for treatment effect analysis. Note that the same principles can be applied to the multivariate FB approach. Crash counts are typically time series data across years and therefore it is proper and necessary to include time effects into the model structure. The following simple model structure can be used to represent time series crash data:

$$\textit{Observed series} = \textit{time effect} + \textit{regression term} + \textit{random effect} \qquad (6\text{-}1)$$

where the "regression term" $\mu_{i,t}$ is of the same form as safety performance functions (SPFs) used in EB studies *(*Sayed and Rodrigez, 1999; Persaud and Nguyen, 1998; Persaud et al, 2002*)*, and "random effects" accounts for latent variables across the sites. Normally there are four ways to address time effect, as explained in Chapter 5: Poisson autoregressive model combining time effects and random effects together by an AR model; time multiplier model; time varying coefficients model which combines time effects and regression term; and time trend model. For treatment effect analysis, however, the Poisson AR model is not applicable since the countermeasure implementation year should normally be excluded from the analysis. Therefore, only three methods can be applied to deal with time effects for the before-after FB study.

The following simple model can be used to explain the before-after FB procedures:

$$Y_{i,t} \sim \text{Pois}\,(\lambda_{i,t})$$

$$\lambda_{i,t} = \mu_{i,t} e^{\varepsilon_i} \qquad (6-2)$$

$$\mu_{i,t} = e^{\beta_0} ML_{i,t}^{\beta_1} XST_{i,t}^{\beta_2} \qquad (6-3)$$

As can be seen in Equation 6-2, the basic form of FB models is a product format, if Equation 6-2 is transformed into logarithm, then it becomes an additive format.

Given the observed crash count $Y_{i,t}$ in the "after" period at treated site $_{i}$, the major task of treatment effect analysis is to compare this count with what level of safety $\lambda_{i,t}$ in the after period would have been expected had the treatment not been implemented. The procedure for predicting the expected number of crashes $\lambda_{i,t}$ in the after period without treatment includes two steps (Aul, 2006):

Step 1: Assuming $Y_{i,t}$~Pois ($e^{\varepsilon_i} \times \mu_{i,t}$), posterior distributions of the parameters are calibrated by Markov Chain Monte Carlo (MCMC) methods using the data from reference sites and the before period of treated sites.

Step 2: The corresponding expected total crashes $\lambda_{i,t}$ without treatment can then be obtained and used as an estimate of $\lambda_{i,t}$, given the traffic volumes at each treated site in the after period. The change in safety is the difference between the predicted $\lambda_{i,t}$ in the after period without treatment and the safety $Y_{i,t}$ in the same period with the treatment in place. The treatment effects can then be calculated, either in terms of a crash frequency change or in terms of a percentage change in crashes.

6.3 EVALUATION APPROACH

6.3.1 Measurements of Treatment Effect

Two measurements can be used to quantify the treatment effect: expected crash reduction (CR) and expected crash reduction rate (CRR) combining their standard deviations. The calculation of these two measurements is presented below:

Again, assuming $\mathrm{Y}_{i,t} \sim \mathrm{Pois}\left(\lambda_{i,t}\right)$,

For treatment group: $i = 1, 2, \dots, n_T \qquad t = 1, 2, \dots, t_Y - 1$

and,

For reference group: $i = 1, 2, \dots, n_R \qquad t = 1, 2, \dots, t_Y, t_Y + 1, \dots, t_Y + t_Z$

where,

t_Y = treatment implementation year

t_Z = the number of years after treatment

n_T = number of treated sites

n_R = number of reference sites

n = total number of sites including treatment sites and reference sites, $n = n_T + n_R$.

Then CR and CRR can be obtained by

Crash Reduction:

$$CR = \sum_{i=1}^{NT} \sum_{t=t_Y+1}^{t_Y+t_Z} \lambda_{i,t} - \sum_{i=1}^{NT} \sum_{t=t_Y+1}^{t_Y+t_Z} \mathrm{Y}_{i,t} \qquad (6-4)$$

Crash Reduction Rate:

$$CRR = 1 - \frac{\sum_{i=1}^{NT} \sum_{t=t_Y+1}^{t_Y+t_Z} \mathrm{Y}_{i,t}}{\sum_{i=1}^{NT} \sum_{t=t_Y+1}^{t_Y+t_Z} \lambda_{i,t}} \qquad (6-5)$$

where, $\lambda_{i,t}$= expected crashes without treatment for intersection i in year t in the after period.

For the Naïve method, the values of $\lambda_{i,t}$are obtained directly from the before period counts while they are predicted from the developed model using the FB method. CRR is a relative crash reduction measurement and is deemed to be a better measurement for treatment effect analysis.

6.3.2 Evaluation Approach

In order to properly evaluate the FB method, including univariate and multivariate FB methods, for before-after treatment evaluation a total of five datasets were used. First, three simulated datasets were used to examine the univariate FB method. Then two groups of California unsigalized intersections with four different crash types (total crashes, rear end crashes, right angle crashes and left turn crashes) were selected to compare the MVPLN method with the univariate FB method for different types of crashes. It should be noted that each of the selected datasets showed a significant crash change from first period (before hypothetical treatment) to the second period (after hypothetical treatment). Otherwise, the study is meaningless. A hypothetical treatment was randomly assigned to each of the five groups to identify a sample of treatment sites which the treatment groups also have a significant crash change. Then, before-after univariate FB (Lan et al., 2009; Persaud et al., 2010) was performed on the three simulated datasets while MVPLN and univariate FB studies (Lan and Persaud, 2010) were performed on of California unsigalized intersections, respectively. Untreated reference group data and the data for the treated sites in the before period were used to develop inferential models. These models were then used to predict the crash frequency for treatment sites in the after period, had the treatment not been implemented. In this way, it was expected that the FB method should estimate no change in safety if it is correct. In particular, the MVPLN FB method, if it is better, would provide better results than the univariate FB method and also estimate no change in safety (i.e., expected crashes) for the hypothetical treatment sites -- since there was no actual treatment. The details of the evaluation study are presented below.

6.4 EXPLORATION OF THE UNIVARIATE FB METHOD

Variations of the before-after univariate FB method were validated using simulated data for a hypothetical treatment known to have no effect.

6.4.1 Simulated Data

In deriving the simulated data, it was assumed, as is common, that the crash count over “similar” sites follows a negative binomial distribution (NBD). The NBD may be derived by “heterogeneous Poisson sampling” which assumes that the crash count $Y_{i,t}$ at a site over time is Poisson distributed with unknown mean $\mu_{i,t}$ per unit of time at site i and that these means $\mu_{i,t}$ follow a Gamma distribution over similar sites, such that

$$E(Y) = E(\mu) \qquad (6-6)$$

and

$$Var(Y) = E(\mu) + \frac{E^2(\mu)}{\varphi} \qquad (6-7)$$

where, φ is the dispersion parameter of the NBD.

The data used to examine the FB methods were generated from a PG distribution (Lord, 2006). The simulation framework for the stop-controlled intersection dataset used is as follows (Lan et al, 2009):

Step 1: Randomly generate entering traffic volumes on the major road (5000 ~ 40,000 AADT) across 6 years with random variation (within 5%). This was such that most traffic volumes would be around the mean value 20,000 AADT, which is typical of traffic volumes entering a stop-controlled intersection from the major road.

Step 2: Similarly, randomly generate traffic volumes on the minor road (500 ~ 4000 AADT) across 6 years with random variation (within 5%).

Step 3: Input safety performance function (SPF) parameters. These were developed from California state data in a recent project (Bhim, 2005). The SPF used was:

$$\mu_{i,t} = 0.000106 ML_{i,t}^{0.7191} XST_{i,t}^{0.4813} \qquad (6-8)$$

Step 4: Calculate the expected number of crashes $\mu_{i,t}$ for intersection i across 6 years from the SPF.

Step 5: Generate a scale factor δ_i from a Gamma distribution with the mean equal to 1 and the dispersion parameter φ: $\delta_i \sim \mathrm{Gamma}(\varphi, \frac{1}{\varphi})$ using software GenStat (Payne, 2000). It is necessary to use the parameterization of the gamma distribution $\delta_i \sim \mathrm{Gamma}(a, b)$ when its mean and variance are defined as $E(\delta) = ab$ and $Var(\delta) = ab^2$, respectively. GenStat uses this parameterization for generating gamma distributed values. It can be shown that when $E(\delta) = 1$ and $Var(\delta) = 1/\varphi$ (where $a = \varphi$ and $/b = 1/\varphi$), the Poisson-gamma function gives rise to a NB distribution with $Var(Y) = \mu + \mu^2/\varphi$ (for detailed derivation, see Chapter 3).

Step 6: Calculate the modified mean $\lambda_{i,t}=\delta_i \times \mu_{i,t}$

Step 7: Generate a discrete value $Y_{i,t}$ for the observed count at intersection i in year t from a Poisson distribution with mean $\lambda_{i,t}$, and with the constraint that crash counts at each intersection is less than or equal to 10 each year to reflect typical values.
Step 8: Repeat step 1 to step 7 "*n*" times for the required number of intersections.

Step 5 was performed using the software GenStat while other steps were performed in Excel coding in Visual Basic. The simulation was done for sample sizes of 1000 and 4000, with dispersion parameters φ of 0.25, 0.5, 1.0 and 2.0, respectively, to reflect the range of typical values reported on relevant studies.

For the generated dataset, a Naïve before after study was performed by comparing the crashes at a site in the first 3 years (the "before" period) with those in the last two (the "after" period), for a hypothetical treatment at the start of year 4. Only those sites where the Naïve results showed a substantial apparent crash reduction (≥10%) due to regression to the mean were used to conduct the FB analysis. In order to obtain a significant crash change from Naïve method, numerous trials for the above four values of the dispersion parameter were conducted. The real treatment effect was in fact zero since the means used to generate the counts did not change materially over time. The identified sites were then randomly allocated to treatment and reference groups. The summary information of final samples for the univariate FB validation are detailed in Tables 6-1 to 6-3 and the corresponding naive results are tabulated in Table 6-4.

6.4.2 FB Model Development

This section discusses the model development for the before-after FB method. Several variations, including three PG models and four PLN models which are of the same or similar forms in Chapter 5, were developed and tested before settling on a preferred approach. These are summarized below:

A. Poisson – Gamma models

As discussed in Chapter five, the random effect e^{ε_i} in Equation 6-2 follows the Gamma distribution: $e^{\varepsilon_i} \sim \text{Gamma}(\varphi, 1/\varphi)$ and $\varphi \sim \text{Gamma}(0.01, 0.01)$. There are three PG models based on the different forms of $\mu_{i,t}$ in this study.

Table 6-1 Summary of Simulated Dataset 1

(Total crashes in first 3 years ≥26, φ=0.25)

Hypothetical Treatment Group: 47 sites				
Variables	mean	Std.	max	min
Years before	3	0	3	3
Years after	2	0	2	2
crashes/site.year before	8.82	0.39	9	8
crashes/site.year after	7.60	1.50	9	4
$AADT_{major}$ before	21057	9280	38574	5497
$AADT_{minor}$ before	2819	788	3985	898
$AADT_{major}$ after	21072	9270	38641	5274
$AADT_{minor}$ after	2833	800	3933	857
Hypothetical Reference Group: 42 sites				
Variables	mean	Std.	max	min
Years	6	0	6	6
crashes/site.year	8.13	1.27	9	8
$AADT_{major}$	22059	9335	40393	5107
$AADT_{minor}$	2824	863	4092	612

Table 6-2 Summary of Simulated Dataset 2

(Total crashes in first 3 years ≥19, φ=1.0)

Hypothetical Treatment Group: 105 sites				
Variables	mean	Std.	max	min
Years before	3	0	3	3
Years after	2	0	2	2
crashes/site.year before	7.48	1.57	9	3
crashes/site.year after	6.60	2.05	9	1
$AADT_{major}$ before	22603	10258	40871	5750
$AADT_{minor}$ before	2685	941	4159	522
$AADT_{major}$ after	22534	10164	41268	5682
$AADT_{minor}$ after	2671	942	4138	520
Hypothetical Reference Group: 111 sites				
Variables	mean	Std.	max	min
Years	6	0	6	6
crashes/site.year	6.99	1.96	9	0
$AADT_{major}$	22135	10342	41494	4894
$AADT_{minor}$	2797	869	4166	545

Table 6-3 Summary of Simulated Dataset 3

(Total crashes in first 3 years ≥22, φ=0.25)

Hypothetical Treatment Group: 229 sites				
Variables	mean	Std.	max	min
Years before	3	0	3	3
Years after	2	0	2	2
crashes/site.year before	7.99	1.14	9	4
crashes/site.year after	7.16	1.86	9	1
$AADT_{major}$ before	22709	10358	41727	4833
$AADT_{minor}$ before	2700	827	4149	655
$AADT_{major}$ after	22701	10342	41649	4910
$AADT_{minor}$ after	2697	825	4163	682
Hypothetical Reference Group: 263 sites				
Variables	mean	Std.	max	min
Years	6	0	6	6
crashes/site.year	7.58	1.54	9	1
$AADT_{major}$	23198	10053	41675	4997
$AADT_{minor}$	2820	809	4181	612

Table 6-4 Naïve Crash Reduction and Crash Reduction Rate of Simulated Datasets

Dataset 1	Total crashes in first 3 years ≥26, φ=0.25		
Variables	Whole group (89 sites)	Treatment Group (47 sites)	Reference Group (42 sites)
Naïve Crash Reduction	338	172	166
Naïve Crash Reduction Rate	14% (3%)	14% (4%)	14% (4%)
Dataset 2	Total crashes in first 3 years ≥19, φ=1.0		
Variables	Whole group (216 sites)	Treatment Group (105 sites)	Reference Group (111 sites)
Naïve Crash Reduction	584	279	305
Naïve Crash Reduction Rate	12% (2%)	12% (3%)	12% (3%)
Dataset 3	Total crashes in first 3 years ≥22, φ=0.25		
Variables	Whole group (492 sites)	Treatment Group (229 sites)	Reference Group (263 sites)
Naïve Crash Reduction	1242	570	672
Naïve Crash Reduction Rate	11% (1%)	10% (2%)	11% (2%)

Model 1: original PG model

This is the basic form of Poisson Gamma model and it does not account for time effects. The function form of $\mu_{i,t}$ is the same as Equation 5-8. This model was defined as PG_19, PG_22 and PG_26 for the three identified datasets which have 19, 22 and 26 total crashes in the first three years, respectively.

Model 2: Time multiplier PG model, similarly denoted as PG_M_19, PG_M_22 and PG_M_26, respectively here after. This is the same as defined in Chapter 5.

Model 3: Time trend model

A potential time trend $\beta_3 \times t$ in the observed crash series was included in this model as an alternative way to deal with temporal variation. This model is described as PG_T_19, PG_T_22 and PG_T_26, respectively for the three studied datasets.

B. Poisson Log Normal models

Three Poisson Log Normal models were developed based on the three different forms of $\mu_{i,t}$ which are same as those for Poisson Gamma models. As in Chapter 5, the random effects $\varepsilon_i \sim \mathrm{N}(0, \sigma^2)$ and $\sigma^2 \sim$ Inverse Gamma $(0.001, 0.001)$. In addition, one alternative time trend model was also explored.

The three PLN models are

Model 4: Regular Poisson log normal model (PLN_19, PLN_22 and PLN_26 respectively)

Model 5: Poisson log normal time multiplier model (PLN_M_19, PLN_M_22, and PLN_M_26)

Model 6: Poisson log normal time trend model (PLN_T_19 and PLN_T_22 and PLN_T_26 respectively)

Model 7: PLN time trend models with yearly random effects

$$Y_{i,t} \sim \mathrm{Pois}\ (\lambda_{i,t})$$

$$\lambda_{i,t} = \mu_{i,t} e^{\varepsilon_{i,t}} \qquad (6-9)$$

where $\mu_{i,t}$ is the same as Equation 5-9.

Unlike the previous models, where only one random effect was introduced at each site, this model has yearly random effects $\varepsilon_{i,t}$ at each site. This model can be seen as an alternative to Model 6, denoted as PLN_T_19 $\varepsilon_{i,t}$, PLN_T_22 $\varepsilon_{i,t}$ and PLN_T_26 $\varepsilon_{i,t}$ for the three datasets respectively. It was expected this model can provide similar results as Model 6.

6.4.3 Bayesian Model Comparison and Selection

Two parallel chains were run for both scenarios to obtain posterior distributions of the coefficients and crash reduction estimates. After convergence, the results in terms of log likelihood (LL), DIC and CRR from the above seven models were collected and the other two model selection criteria AIC, BIC were calculated. The models were compared and the results are listed in Table 6-5.

It can be seen that for datasets 1 and 2, the three PLN models. PLN time multiplier model (PLN_M_19, PLN_M_26), PLN time trend model (PLN_T_19 and PLN_T_26) and PLN time trend model with yearly random effects (PLN_T_19 $\varepsilon_{i,t}$ and PLN_T_26 $\varepsilon_{i,t}$) are comparable based on the model selection criteria while the PLN time trend model has slightly better results. Specifically, for dataset 1, model selection criteria LL, AIC, BIC are exactly the same for PLN_M_26 and PLN_T_26, but the DIC of PLN_T_26 is slightly better than that of PLN_M_26. Also it can be seen that PLN_T_26 provides a better estimate of CRR, i.e., the mean is zero rather than the albeit insignificant 3% of PLN_M_26. PLN_T_19 and PLN_T_26 are seen to be the best models for these two datasets.

For dataset 3, PLN_M_22 has the same DIC as PLN_T_22, but it has higher values of LL, lower values of AIC and BIC, strongly favouring this model. Thus PLN_M_22 is deemed to be the best model.

For all of the three datasets, as expected, PLN models for time trend with yearly randomly effects (PLN_T_19 $\varepsilon_{i,t}$, PLN_T_22 $\varepsilon_{i,t}$ and PLN_T_26 $\varepsilon_{i,t}$) provide similar results as regular PLN time trend models (PLN_T_19 and PLN_T_22 and PLN_T_26) in terms of the values of model selection criteria and treatment reduction rate CRR, while the later one has somewhat better values of model selection criteria and has a simpler form. Thus, the yearly random effects model was not further investigated.

It can be seen that PLN models seem to have much better performance than corresponding PG models based on all the model selection criteria: LL, AIC, BIC and DIC. Indeed, PLN models have a slightly better performance in terms of CRRs than corresponding PG models. However, it is interesting that the CRRs of PG models are quite comparable with corresponding PLN models. For example, the CRR of PLN_T_26 is 0 with standard deviation of 3% while that of PG_T_26 is 1% with a 4% standard deviation; thus, both models estimate insignificant treatment effects. A similar pattern is observed with other PLN and PG models. This result is consistent with conclusions obtained by Maher and Mountain (2009). In their study, several distributions of random effects such as the gamma distribution, the log normal distribution and the Weibull distribution were investigated for estimating regression to the mean (RTM) using four datasets, but only with one form of $\mu_{i,t}$; they concluded that the results in terms of estimating of RTM were comparable albeit the distributions of random effects were different.

One can find that all of the models considering time effects, such as PLN_T, PLN_M, PLN_T_$\varepsilon_{i,t}$, PG_T and PG_M models, successfully estimated no treatment for hypothetical treatment sites for all three datasets, whilst the PLN and PG models without time effects falsely estimated significant treatment effects for these hypothetical treatment sites. Model PLN provided incorrect estimates of treatment effects even though all of the model selection criteria very strongly favoured PLN over PG_T and PG_M models. This suggests that the function of $\mu_{i,t}$ is much more important than what distributions the random effects follow. In other words, it is meaningful if and only if models with different distributions of random effects are compared with each other with the same structure or function form of expected crashes $\mu_{i,t}$. Otherwise, a seriously biased result can be anticipated. This phenomenon is further examined later in this thesis.

Table 6-5 Treatment Effect Analysis and Model Diagnostics

Dataset 1 (Total crashes in first 3 years ≥26, φ=0.25)	47 Treatment sites			42 Reference sites			
	PLN_26	PLN_M_26	PLN_T_26	PLN_T_26 ε_{it}	PG_26	PG_M_26	PG_T_26
K	4	5	5	5	4	5	5
LL	-816	-810	**-810**	-811	-832	-827	-827
AIC	1638	1628	**1628**	1630	1669	1662	1662
BIC	1650	1644	**1644**	1646	1681	1678	1678
DIC	1644	1636	**1633**	1635	1717	1712	1708
CRR	9% (2%)	-2% (3%)	**0 (3%)**	0 (3%)	11% (2%)	-2% (4%)	1% (4%)
Dataset 2 (Total crashes in first 3 years ≥19, φ=1.0)	105 Treatment sites			111 Reference sites			
	PLN_19	PLN_M_19	PLN_T_19	PLN_T_19 ε_{it}	PG_19	PG_M_19	PG_T_19
K	4	5	5	5	4	5	5
LL	-2121	-2113	**-2112**	-2115	-2125	-2118	-2117
AIC	4248	4234	**4232**	4238	4256	4244	4242
BIC	4263	4254	**4252**	4258	4271	4264	4262
DIC	4265	4252	**4248**	4250	4353	4342	4338
CRR	7% (1%)	-1% (2%)	**0 (2%)**	0 (2%)	8% (2%)	-1% (3%)	1% (2%)
Dataset 3 (Total crashes in first 3 years ≥22, φ=0.25)	229 Treatment sites			263 Reference sites			
	PLN_22	PLN_M_22	PLN_T_22	PLN_T_22 ε_{it}	PG_22	PG_M_22	PG_T_22
K	4	5	5	5	4	5	5
LL	-4743	**-4722**	-4725	-4726	-4778	-4761	-4763
AIC	9492	**9452**	9458	9460	9562	9530	9534
BIC	9509	**9475**	9481	9483	9579	9553	9557
DIC	9511	**9475**	9475	9478	9725	9695	9693
CRR	7% (1%)	**0 (1%)**	1% (1%)	1% (1%)	8% (1%)	0 (2%)	2% (1%)

Notes:
1. K is the number of parameters
2. CRR means crash reduction rate
3. Negative sign indicates an increase in crashes
4. Standard errors are in parentheses

6.4.4 Evaluation Results of Univariate FB

Table 6-6 compares the treatment effects obtained from the Naïve and FB approaches. The Naïve method predicted a significant total crash reduction after a hypothetical treatment with no effect was implemented, which is incorrect since RTM is not accounted for. On the other hand, the CRR shows correctly that there are no significant treatment effects estimated by the FB method, suggesting that RTM has been properly accounted for, and therefore that this method can be used for observational before-after studies.

Table 6-6 Comparison of Treatment Effect Estimates from Naïve and FB Studies

Method		Crash Reduction Rate (CRR)	Treatment Effects Identified?
Naive	Dataset 1	14% (4%)	Yes
	Dataset 2	12% (3 %)	Yes
	Dataset 3	10% (2%)	Yes
FB	Dataset 1	0 (3%) (PLN_T_26)	No
	Dataset 2	0 (2%) (PLN_T_19)	No
	Dataset 3	0 (1%) (PLN_M_22)	No

Note: 1. Datasets are identified in Tables 6-1 to 6-3
2. Standard deviations are in parenthesis

6.5 EVALUATION OF THE MULTIVARIATE FB METHOD

This section describes the evaluation of the multivariate FB method using two datasets of hypothetical before-after data for California unsignalized intersections, with different types of crashes for high and low crash count groups. For the multiple type crash data, two ways can be used to conduct an FB analysis. One way is to develop univariate FB models for each type of the crash by assuming that the different types of crashes are independent. Another way is to develop multivariate FB models as a whole accounting for the possible correlations among these different types of crashes. The objective of this part of the work was to evaluate if MVPLN is superior to univariate FB for these two data cases. Thus both univariate FB and multivariate FB approaches were applied to these two datasets through various developed models. The approach showing lowest treatment effect is deemed to be the best model, since there was in fact no treatment.

6.5.1 Data Description

All data used were provided by the Highway Safety Information System (HSIS, http://www.hsisinfo.org/). Geometry, traffic volume and four types of crash data (total, right angle, left turn and rear end) were acquired for the State of California for the period 1993-2002. The unsignalized intersections included 1381 sites that are three legged and 726 sites that were 4-legged with two lanes on the major road. In order to investigate the difference of treatment effects estimated from the before-after MVPLN and univariate FB methods, a hypothetical treatment was assumed to happen at the start of year 1998. A Naïve before-after study was performed for the two groups of data, each sorted in descending order by the 1993-1997 crash counts for each type of crashes, by comparing the crash frequency at a site in the first 5 years (the "before" period) with that for the last four years (the "after" period, i.e., after 1998). It was essential that only those subgroups where the Naïve results showed a substantial apparent crash change (absolute value of CRR ≥5%) due to RTM be used to conduct the analysis. In addition, it was desirable to investigate MVPLN for both high and low crash frequency entities. The sites in the selected subgroups were then randomly allocated to identify treatment and reference groups. Table 6-7 shows information for the two groups of data, the 4-legged unsignalized group with 7-10 total crashes/site (case 1) in the first 5 years, and the 3-legged unsignalized group with 2-3 total crashes/site (case 2) in the same period.

Table 6-7 Naïve Crash Reduction Rate

Case 1: Four legged unsigalized intersections (Total crashes in first 5 years =7-10/site)			
Variable	Whole group (116 sites) Crash reduction rate	Treatment group (57 sites) Crash reduction rate	Reference group (59 sites) Crash reduction rate
Total crashes	-12% (5%)	-11% (7%)	-4% (6%)
Right Angle crashes	-18% (11%)	-26% (18%)	-3% (13%)
Left Turn crashes	-25% (15%)	-34% (22%)	-14% (20%)
Rear End crashes	14% (12%)	31% (14%)	3% (21%)
Case 2: Three legged unsigalized intersections (Total crashes in first 5 years =2-3/site)			
Variable	Whole group (364 sites) Crash reduction rate	Treatment group (170 sites) Crash reduction rate	Reference group (194 sites) Crash reduction rate
Total crashes	-18% (5%)	-21% (8%)	-17% (7%)
Right Angle crashes	-136% (79%)	-56% (72%)	-212% (137%)
Left Turn crashes	-16% (13%)	-17% (19%)	-18% (19%)
Rear End crashes	4% (15%)	17% (18%)	-13% (24%)

Notes: 1. Negative sign indicates an increase in crashes
2. Standard errors are in parentheses

As seen in Table 6-7, some of the changes in crashes in last four years before and after hypothetical treatment, though substantial, have large standard deviations. However, this was not seen as a major obstacle to proceeding with this dataset.

Summary information for treatment and reference sites for both cases is detailed in Tables 6-8 and 6-9. It can be seen that the reference sites in the two groups are close to the corresponding treated sites in terms of crashes/site-year, AADT on the major road and AADT on the minor road.

Table 6-8 Summary of Four Legged Unsignalized Intersections (Case 1: Total crashes in first 5 years =7-10/site)

57 Hypothetically treated Unsignalized Intersections			
Variable	Mean	Minimum	Maximum
Years before	5	5	5
Years after	4	4	4
Total crashes/site.year before	1.67	0	7
Total crashes/site.year after	1.78	0	7
Right Angle crashes /site.year before	0.31	0	6
Right Angle crashes/site.year after	0.35	0	4
Left Turn crashes /site.year before	0.21	0	2
Left Turn crashes/site.year after	0.28	0	3
Rear End crashes /site.year before	0.21	0	3
Rear End crashes/site.year after	0.14	0	1
$AADT_{major}$ before	9267	3200	22400
$AADT_{minor}$ before	833	101	7800
$AADT_{major}$ after	10274	4005	24000
$AADT_{minor}$ after	833	101	7800
59 Unsignalized Reference Intersections			
Variable	Mean	Minimum	Maximum
Years	10	10	10
Total crashes/site.year	1.72	0	8
Right Angle crashes/site.year	0.42	0	6
Left Turn crashes /site.year	0.23	0	3
Rear End crashes/site.year	0.14	0	2
$AADT_{major}$	9055	3100	27754
$AADT_{minor}$	674	120	4500

Table 6-9 Summary of Three Legged Unsignalized Intersections

(Case 2: Total crashes in first 5 years =2-3/site)

170 Hypothetically treated 3-legged Unsignalized Intersections			
Variable	Mean	Minimum	Maximum
Years before	5	5	5
Years after	4	4	4
Total crashes/site year before	0.49	0	3
Total crashes/site year after	0.60	0	6
Right Angle crashes /site year before	0.007	0	1
Right Angle crashes/site year after	0.009	0	1
Left Turn crashes /site year before	0.08	0	2
Left Turn crashes/site year after	0.1	0	3
Rear End crashes /site year before	0.06	0	2
Rear End crashes/site year after	0.05	0	2
$AADT_{major}$ before	8367	2900	30000
$AADT_{minor}$ before	379	100	1950
$AADT_{major}$ after	9107	2928	30761
$AADT_{minor}$ after	385	100	1950
194 Unsignalized 3-legged Reference Intersections			
Variable	Mean	Minimum	Maximum
Years	10	10	10
Total crashes/site year	0.52	0	8
Right Angle crashes/site year	0.009	0	1
Left Turn crashes /site year	0.08	0	3
Rear End crashes/site year	0.04	0	3
$AADT_{major}$	8160	2550	24800
$AADT_{minor}$	338	100	2200

6.5.2 Bayesian Model Framework

The Bayesian models for multivariate data are similar to Equations 5-14 to 5-16 in Chapter 5. That is, $Y_{i,t}^{k} \sim$Pois $(\lambda_{i,t}^{k})$,

$$\lambda_{i,t}^{k} = exp(\boldsymbol{\beta^k}.\boldsymbol{X} + \varepsilon_i^k) = \mu_{i,t}^{k} e^{\varepsilon_i^k} \qquad k = 1, 2, \dots, L \tag{6-10}$$

$$\mu_{i,t}^{k} = exp(\boldsymbol{\beta^k}.\boldsymbol{X}) \tag{6-11}$$

where,

$\lambda_{i,t}^{k}$= the modified expected crashes of type k or severity k at location *i* in year *t*.

$\mu_{i,t}^{k}$= the expected crashes of type *k* or severity k at location *i* in year *t*.

ε_i^k = the random effect for crash type k or severity k at location *i*.

The difference between MVPLN models and univariate FB models is in how to deal with the relationship among the random effects $\varepsilon_i^1, \varepsilon_i^2, \dots, \varepsilon_i^L$. If the possible correlation among these crash counts is neglected by assuming $\varepsilon_i^1, \varepsilon_i^2, \dots, \varepsilon_i^L$ independently follows a gamma or normal distribution, then the above model is univariate PG or PLN, respectively. On the other hand, the vector $\boldsymbol{\varepsilon_i} = \varepsilon_i^1, \varepsilon_i^2, \dots, \varepsilon_i^L$ can be assumed to be multivariate normally distributed to account for the correlations among different type crashes, that is: $(\varepsilon_i^1, \varepsilon_i^2, \dots, \varepsilon_i^L)' \sim N_L(0, \boldsymbol{\Sigma})$.

Below are the five different function forms of expected crashes $\mu_{i,t}^k$ for unsignalzied intersections, as described in Chapter 5.

- Model 1: Regular model (denoted as MVPLN, PLN or PG, respectively)

$$\mu_{i,t}^k = e^{\beta_0{}^k} ML_{i,t}^{\beta_1{}^k} XST_{i,t}^{\beta_2{}^k} \qquad k = 1, 2, \dots, L \qquad (6-12)$$

- Model 2: Time multiplier model (defined as MVPLN_M, PLN_M or PG_M)

$$\mu_{i,t}^k = e^{\beta_{0,t}{}^k} ML_{i,t}^{\beta_1{}^k} XST_{i,t}^{\beta_2{}^k} \qquad k = 1, 2, \dots, L \qquad (6-13)$$

The only difference of Models 1 and 2 is that intercept $\beta_{0,t}$ varies with the year

- Model 3: Time trend model (described as MVPLN_T, PLN_T or PG_T, respectively)

$$\mu_{i,t}^k = e^{\beta_0{}^k} ML_{i,t}^{\beta_1{}^k} XST_{i,t}^{\beta_2{}^k} e^{\beta_3{}^k \times t} \qquad k = 1, 2, \dots, L \qquad (6-14)$$

- Model 4: Time varying coefficients model (defined as MVPLN_VC, PLN_VC or PG_VC)

$$\mu_{i,t}^k = e^{\beta_{0,t}{}^k} ML_{i,t}^{\beta_{1,t}{}^k} XST_{i,t}^{\beta_{2,t}{}^k} \qquad k = 1, 2, \dots, L \qquad (6-15)$$

Similarly, prior distributions for all coefficients $(\beta_{0,t}^{k}, \beta_{1,t}^{k}, \beta_{2,t}^{k}, \beta_{3}^{k})$ are assumed non-informative [i.e., N(0,1000)] to reflect the lack of precise knowledge of the value of the coefficients.

- Model 5: MVPLN model with coefficients follow multivariate normal distribution (denoted as MVPLNp)

  This model is developed based on the result from previous four MVPLN models, where the best MVPLN model is identified. This model has the same form of $\mu_{i,t}^{k}$ as for the MVPLN model identified as best, but with all coefficients for different types of crashes also following the multivariate normal distribution.

  $\beta_0^1, \beta_0^2, \ldots, \beta_0^L \sim N_L(0, \boldsymbol{\Sigma}_{\boldsymbol{\beta_0}})$

  $\beta_1^1, \beta_1^2, \ldots, \beta_1^L \sim N_L(0, \boldsymbol{\Sigma}_{\boldsymbol{\beta_1}})$

  ...

  The prior of ${\Sigma_{\beta_j}}^{-1}$ is set to follow the Wishart distribution as explained in Chapter 3. It is expected that this model provides comparable results as the identified MVPLN model in terms of CRRs.

6.5.3 Evaluation Results

Two parallel chains were run for both scenarios to obtain posterior distributions of the coefficients and crash reduction estimates. After convergence, the results in terms of DIC and CRR from the above four models (MVPLN, univariate PG and PLN) were tabulated as shown in Tables 6-10 to 6-11 and Tables 6-13 to 6-14. It is worth mentioning that all CRRs and DICs for the four types of crashes, either from MVPLN or univariate FB, can be simultaneously obtained with WinBUGs. Also of note is that the computation time for univariate PLN or PG is much less than that for MVPLN.

From Tables 6-10 to 6-11 and Tables 6-13 to 6-14, it is seen that the CRRs are quite sensitive to the expression for $\mu_{i,t}^{k}$. For Model 1, the same form as EB, a treatment effect has been incorrectly identified for total crashes and left turn crashes. For the other 3 models for $\mu_{i,t}^{k}$, which consider temporal variation across years in different ways, no significant treatment effects have been detected as should be expected since, again, there was no real treatment.

This indicates the importance of including temporal effects in model development, consistent with the way rigorous EB applications have applied time trend multipliers in the SPFs used in the analysis.

For Case 1, as seen in Tables 6-10 and 6-11, the high crash frequency group, the DIC of the PLN is the lowest for each formulation of $\mu_{i,t}^{k}$; its value is at least 10 fewer than that of MVPLN and PG, suggesting that the PLN model is the best one. However, MVPLN and univariate FB provide consistent predictions of CRRs for each form of $\mu_{i,t}^{k}$ in spite of the different DICs.

For MVPLN models, DICs from MVPLN_VC_7-10 has the highest DIC value (5993) while DICs from the other 3 models stay the same (5979). Thus the other three models are deemed to be competitive in terms of DICs. Then the expected deviance for the jth model $Dbar_j$,which can be seen as a measurement of goodness of fit, was used as the second criterion to identify the best model from the 3 candidates; MVPLN_M_7-10, which includes temporal variations, is seen to be the best MVPLN model in that it has much lower values of *Dbar*. Similarly, PLN_M_7-10 was identified to be the best PLN model for comparison.

MVPLNp_M_7-10, which is similar to MVPLN_M_7-10 but with all the coefficients of different type crashes also following multivariate normal distribution, was also developed. Both models have comparable results in terms of DIC and CRRs as seen in Table 6-12. However, this model takes much longer to run than MVPLN_M_7-10 due to the extra multivariate distributions of the coefficients. Thus MVPLN_M_7-10 is deemed to be superior to MVPLNp_M_7-10, especially where computation time is an issue.

For Case 2, the low crash frequency group, the DIC from PG is much higher than that for the other two models (a DIC difference ≥ 10) for each expression of $\mu_{i,t}^{k}$. The difference in DICs from the MVPLN and PLN Models 1, 2, and 3 is greater than 6, while it is only 2 for Model 4, indicating that, generally, PLN has a better performance than MVPLN. Similar to Case 1, CRR estimates from MVPLN, PLN and PG are comparable for each form of $\mu_{i,t}^{k}$. DICs of Models 2 and 4 are much higher than those of Models 1 and 3, suggesting that the latter two are competitive. The time trend model (Model 3) is deemed the best because temporal effects are accounted for, in contrast to Model 1, which again produced incorrect CRR estimates.

Table 6-10 Comparison of Results from MVPLN, PLN and PG Models

(Case 1: total crash in first 5 years=7-10/site)

Dbar = post.mean of -2logL; Dhat = -2LogL at post.mean of stochastic nodes														
Multivariate FB					**Univariate FB**									
Model 1: MVPLN_7-10					Model 1: PLN_7-10					Model 1: PG_7-10				
	Dbar	Dhat	pD	DIC		Dbar	Dhat	pD	DIC		Dbar	Dhat	pD	DIC
lt	947	916	31	978	lt	953	925	28	980	lt	939	903	36	975
ra	1257	1193	64	1321	ra	1255	1187	68	1323	ra	1245	1184	61	1306
re	793	772	21	814	re	801	793	8	810	re	787	761	26	813
tot	2821	2776	45	2866	tot	2834	2818	16	2850	tot	2827	2774	53	2880
total	5818	5657	161	5979	total	5843	5724	120	5963	total	5798	5622	176	5973
node	mean	sd	2.50%	97.50%	node	mean	sd	2.50%	97.50%	node	mean	sd	2.50%	97.50%
CRR1	-0.10	0.04	-0.18	-0.03	CRR1	-0.08	0.03	-0.15	-0.02	CRR1	-0.10	0.04	-0.18	-0.02
CRR2	-0.23	0.11	-0.48	-0.03	CRR2	-0.21	0.12	-0.45	0.00	CRR2	-0.19	0.11	-0.43	0.00
CRR3	-0.29	0.12	-0.55	-0.07	CRR3	-0.30	0.12	-0.56	-0.09	CRR3	-0.31	0.13	-0.58	-0.08
CRR4	0.07	0.09	-0.13	0.24	CRR4	0.04	0.09	-0.14	0.20	CRR4	0.09	0.09	-0.11	0.26
Model 2: MVPLN_M_7-10					Model 2: PLN_M_7-10					Model 2: PG_M_7-10				
	Dbar	Dhat	pD	DIC		Dbar	Dhat	pD	DIC		Dbar	Dhat	pD	DIC
lt	947	908	39	987	lt	957	924	33	990	lt	939	895	45	984
ra	1248	1176	71	1319	ra	1247	1173	74	1321	ra	1236	1166	70	1305
re	785	757	29	814	re	795	778	17	811	re	781	746	35	816
tot	2805	2752	54	2859	tot	2819	2795	24	2843	tot	2812	2750	62	2874
total	5786	5593	193	**5979**	total	5818	5670	148	**5966**	total	5768	5556	211	**5979**
node	mean	sd	2.50%	97.50%	node	mean	sd	2.50%	97.50%	node	mean	sd	2.50%	97.50%
CRR1	**-0.03**	0.06	-0.15	0.09	CRR1	**0.00**	0.05	-0.10	0.10	CRR1	**-0.02**	0.06	-0.15	0.10
CRR2	**-0.13**	0.16	-0.48	0.16	CRR2	**-0.11**	0.16	-0.46	0.17	CRR2	**-0.10**	0.16	-0.45	0.18
CRR3	**-0.16**	0.18	-0.57	0.16	CRR3	**-0.14**	0.18	-0.54	0.16	CRR3	**-0.16**	0.19	-0.60	0.16
CRR4	**-0.07**	0.21	-0.54	0.27	CRR4	**-0.13**	0.21	-0.59	0.21	CRR4	**-0.06**	0.22	-0.55	0.29

Notes: 1. CRR1, CRR2, CRR3, CRR4 are crash reduction rates for total, right angle, left turn, rear end crashes respectively

2. tot: total crash ra: right angle crash lt: left turn crash re: rear end crash

Table 6-11 Comparison of Results from MVPLN, PLN and PG Models

(Case 1: total crash in first 5 years=7-10/site)

Dbar = post.mean of -2logL; Dhat = -2LogL at post.mean of stochastic nodes														
Multivariate FB					**Univariate FB**									
Model 3: MVPLN_T_7-10					Model 3: PLN_T_7-10					Model 3: PG_T_7-10				
	Dbar	Dhat	pD	DIC		Dbar	Dhat	pD	DIC		Dbar	Dhat	pD	DIC
lt	948	915	33	981	lt	954	926	28	982	lt	940	903	37	977
ra	1258	1193	65	1323	ra	1257	1189	68	1325	ra	1247	1185	62	1309
re	788	767	21	809	re	796	787	9	806	re	783	757	27	810
tot	2821	2776	46	2867	tot	2835	2817	17	2852	tot	2828	2774	54	2881
total	5815	5651	164	5979	total	5842	5720	123	5965	total	5798	5619	179	5977
node	mean	sd	2.50%	97.50%	node	mean	sd	2.50%	97.50%	node	mean	sd	2.50%	97.50%
CRR1	-0.07	0.06	-0.19	0.04	CRR1	-0.05	0.05	-0.15	0.04	CRR1	-0.07	0.06	-0.19	0.04
CRR2	-0.10	0.15	-0.42	0.17	CRR2	-0.10	0.15	-0.41	0.16	CRR2	-0.08	0.15	-0.40	0.18
CRR3	-0.33	0.20	-0.78	0.01	CRR3	-0.30	0.19	-0.73	0.02	CRR3	-0.35	0.21	-0.81	0.01
CRR4	-0.33	0.24	-0.87	0.08	CRR4	-0.38	0.24	-0.92	0.02	CRR4	-0.32	0.24	-0.86	0.09
Model 4: MVPLN_VC_7-10					Model 4: PLN_VC_7-10					Model 4: PG_VC_7-10				
	Dbar	Dhat	pD	DIC		Dbar	Dhat	pD	DIC		Dbar	Dhat	pD	DIC
lt	937	884	53	991	lt	944	895	48	992	lt	929	868	61	990
ra	1242	1154	87	1329	ra	1241	1151	90	1331	ra	1231	1144	87	1317
re	785	741	44	828	re	792	759	33	825	re	780	730	51	831
tot	2777	2708	68	2845	tot	2783	2744	39	2823	tot	2783	2706	77	2860
total	5740	5488	252	5993	total	5760	5550	210	5971	total	5723	5448	275	5998
node	mean	sd	2.50%	97.50%	node	mean	sd	2.50%	97.50%	node	mean	sd	2.50%	97.50%
CRR1	0.06	0.06	-0.06	0.17	CRR1	0.08	0.05	-0.02	0.17	CRR1	0.08	0.06	-0.05	0.19
CRR2	0.00	0.17	-0.36	0.28	CRR2	-0.01	0.17	-0.37	0.28	CRR2	0.02	0.17	-0.34	0.31
CRR3	0.02	0.19	-0.39	0.35	CRR3	0.02	0.18	-0.37	0.33	CRR3	0.02	0.20	-0.41	0.36
CRR4	-0.06	0.22	-0.56	0.30	CRR4	-0.13	0.21	-0.63	0.23	CRR4	-0.05	0.23	-0.56	0.32

Notes: 1. CRR1, CRR2, CRR3, CRR4 are crash reduction rates for total, right angle, left turn, rear end crashes respectively

2. tot: total crash ra: right angle crash lt: left turn crash re: rear end crash

Table 6-12 Comparison of Competing MVPLN Models

(Case 1: total crash in first 5 years=7-10/site)

Dbar = post.mean of -2logL; Dhat = -2LogL at post.mean of stochastic nodes									
MVPLNp_M_7-10					MVPLN_M_7-10				
	Dbar	Dhat	pD	DIC		Dbar	Dhat	pD	DIC
lt	945	906	39	984	lt	947	908	39	987
ra	1248	1177	70	1319	ra	1248	1176	71	1319
re	783	757	26	809	re	785	757	29	814
tot	2807	2752	55	2863	tot	2805	2752	54	2859
total	5784	5592	190	5974	total	5786	5593	193	5979
node	mean	sd	2.50%	97.50%	node	mean	sd	2.50%	97.50%
CRR1	-3%	6%	-16%	8%	CRR1	-3%	6%	-15%	9%
CRR2	-15%	16%	-49%	13%	CRR2	-13%	16%	-48%	16%
CRR3	-23%	18%	-62%	8%	CRR3	-16%	18%	-57%	16%
CRR4	-2%	17%	-39%	26%	CRR4	-7%	21%	-54%	27%

Similarly, MVPLN, with coefficients for different types of crashes following a multivariate normal distribution, were developed for the time trend model. The results of MVPLNp_T_2-3 and MVPLN_T_2-3 are tabulated in Table 6-15. Again, it is seen that comparable results are provided. This result further confirms that MVPLN models with coefficients following the multivariate normal distribution are unnecessary.

The results in terms of DIC and CRRs for both cases favour PLN, indicating that the different crash types may not be correlated. Another interesting finding is that PLN is superior to PG for both high and low crash cases. With a large sample size, as will be seen in Chapter 7, the PG was usually better than PLN, with much lower DICs, while the studies using several small samples with high crash counts favour PLN as shown earlier in the evaluation of the univariate FB method. Lord and Miranda-Moreno (2008) found that when crash data are characterized by low sample mean values and a small sample size, PLN offers a better alternative than the PG model in terms of stability of posterior mean value. However, from the DICs for extremely low crash counts from both cases in this study (right angle in Tables 6-10 and 6-11, rear end crashes in Tables 6-13 and 6-14), PG generally seems better than PLN and this result is quite different from that of Lord and Miranda-Moreno (2008). This suggests that there is no hard and fast rule to decide which model (PG or PLN) is better; rather it can be concluded that PLN is always a useful alternative to PG and needs to be considered when conducting safety analysis.

Table 6-13 Comparison of Results from MVPLN, PLN and PG Models

(Case 2: total crash in first 5 years=2-3/site)

Dbar = post.mean of -2logL; Dhat = -2LogL at post.mean of stochastic nodes														
Multivariate FB					**Univariate FB**									
Model 1: MVPLN_2-3					Model 1: PLN_2-3					Model 1: PG_2-3				
	Dbar	Dhat	pD	DIC		Dbar	Dhat	pD	DIC		Dbar	Dhat	pD	DIC
lt	1541	1495	47	1588	lt	1574	1555	19	1593	lt	1542	1494	48	1590
ra	262	249	12	274	ra	270	265	5	275	ra	260	246	15	275
re	1019	977	42	1061	re	1039	1011	28	1066	re	1010	965	44	1054
tot	5176	5109	67	5244	tot	5207	5190	17	5224	tot	5189	5116	73	5262
total	7997	7829	169	8166	total	8090	8021	69	8159	total	8001	7822	180	8181
node	mean	sd	2.50%	97.50%	node	mean	sd	2.50%	97.50%	node	mean	sd	2.50%	97.50%
CRR1	-0.12	0.04	-0.19	-0.05	CRR1	-0.11	0.03	-0.17	-0.05	CRR1	-0.12	0.04	-0.19	-0.05
CRR2	-0.23	0.29	-0.91	0.21	CRR2	-0.21	0.26	-0.82	0.19	CRR2	-0.22	0.28	-0.88	0.22
CRR3	-0.15	0.09	-0.35	0.01	CRR3	-0.14	0.08	-0.31	0.00	CRR3	-0.14	0.09	-0.33	0.03
CRR4	0.11	0.09	-0.09	0.27	CRR4	0.11	0.09	-0.07	0.27	CRR4	0.13	0.09	-0.06	0.29
Model 2: MVPLN_M_2-3					Model 2: PLN_M_2-3					Model 2: PG_M_2-3				
	Dbar	Dhat	pD	DIC		Dbar	Dhat	pD	DIC		Dbar	Dhat	pD	DIC
lt	1548	1496	52	1600	lt	1574	1543	31	1605	lt	1545	1488	57	1602
ra	262	245	17	280	ra	267	252	15	282	ra	258	236	23	281
re	1018	968	50	1068	re	1038	1003	36	1074	re	1010	957	53	1063
tot	5177	5103	74	5251	tot	5206	5180	27	5233	tot	5188	5106	82	5270
total	8005	7812	194	8199	total	8085	7977	108	8193	total	8001	7786	215	8216
node	mean	sd	2.50%	97.50%	node	mean	sd	2.50%	97.50%	node	mean	sd	2.50%	97.50%
CRR1	-0.02	0.05	-0.14	0.08	CRR1	-0.02	0.05	-0.12	0.08	CRR1	-0.02	0.06	-0.13	0.08
CRR2	0.32	0.23	-0.24	0.64	CRR2	0.33	0.22	-0.21	0.64	CRR2	0.33	0.23	-0.22	0.65
CRR3	-0.10	0.15	-0.42	0.16	CRR3	-0.10	0.14	-0.41	0.15	CRR3	-0.09	0.15	-0.41	0.17
CRR4	0.11	0.16	-0.26	0.38	CRR4	0.11	0.16	-0.25	0.37	CRR4	0.12	0.16	-0.24	0.39

Notes: 1. CRR1, CRR2, CRR3, CRR4 are crash reduction rates for total, right angle, left turn, rear end crashes respectively

2. tot: total crash ra: right angle crash lt: left turn crash re: rear end crash

Table 6-14 Comparison of Results from MVPLN, PLN and PG Models

(Case 2: total crash in first 5 years=2-3/site)

Dbar = post.mean of -2logL; Dhat = -2LogL at post.mean of stochastic nodes														
Multivariate FB					**Univariate FB**									
Model 3: MVPLN_T_2-3					Model 3: PLN_T_2-3					Model 3: PG_T_2-3				
	Dbar	Dhat	pD	DIC		Dbar	Dhat	pD	DIC		Dbar	Dhat	pD	DIC
lt	1545	1501	45	1590	lt	1572	1549	23	1595	lt	1543	1495	49	1592
ra	260	247	13	273	ra	267	261	6	274	ra	257	241	16	273
re	1020	979	41	1061	re	1045	1020	24	1069	re	1011	965	45	1056
tot	5174	5107	67	5241	tot	5204	5188	16	5221	tot	5185	5112	73	5258
total	8000	7835	165	**8165**	total	8088	8018	70	**8158**	total	7996	7813	183	**8179**
node	mean	sd	2.50%	97.50%	node	mean	sd	2.50%	97.50%	node	mean	sd	2.50%	97.50%
CRR1	**-0.03**	0.05	-0.13	0.06	CRR1	**-0.03**	0.05	-0.12	0.06	CRR1	**-0.03**	0.05	-0.14	0.07
CRR2	**0.24**	0.25	-0.37	0.60	CRR2	**0.22**	0.25	-0.37	0.57	CRR2	**0.26**	0.25	-0.33	0.61
CRR3	**-0.11**	0.14	-0.41	0.13	CRR3	**-0.10**	0.13	-0.37	0.14	CRR3	**-0.11**	0.14	-0.40	0.13
CRR4	**0.04**	0.16	-0.31	0.31	CRR4	**0.03**	0.16	-0.32	0.29	CRR4	**0.06**	0.16	-0.30	0.32
Model 3: MVPLN_VC_2-3					Model 3: PLN_VC_2-3					Model 3: PG_VC_2-3				
	Dbar	Dhat	pD	DIC		Dbar	Dhat	pD	DIC		Dbar	Dhat	pD	DIC
lt	1554	1487	67	1621	lt	1586	1544	42	1627	lt	1551	1478	72	1623
ra	257	232	25	283	ra	263	242	22	285	ra	255	223	32	287
re	1017	955	62	1079	re	1042	997	46	1088	re	1007	936	71	1078
tot	5173	5081	91	5264	tot	5201	5157	44	5245	tot	5184	5085	99	5282
total	8001	7755	246	8247	total	8092	7940	153	8245	total	7996	7723	273	8270
node	mean	sd	2.50%	97.50%	node	mean	sd	2.50%	97.50%	node	mean	sd	2.50%	97.50%
CRR1	0.02	0.05	-0.09	0.12	CRR1	0.01	0.05	-0.09	0.11	CRR1	0.01	0.06	-0.10	0.12
CRR2	0.38	0.22	-0.16	0.69	CRR2	0.35	0.22	-0.20	0.65	CRR2	0.40	0.23	-0.13	0.74
CRR3	-0.02	0.15	-0.36	0.25	CRR3	-0.04	0.14	-0.34	0.20	CRR3	-0.02	0.15	-0.35	0.23
CRR4	0.25	0.16	-0.10	0.51	CRR4	0.20	0.15	-0.14	0.46	CRR4	0.25	0.16	-0.10	0.51

Notes: 1. CRR1, CRR2, CRR3, CRR4 are crash reduction rates for total, right angle, left turn, rear end crashes respectively

2. tot: total crash ra: right angle crash lt: left turn crash re: rear end crash

Table 6-15 Comparison of Competing MVPLN Models

(Case 2: total crash in first 5 years=2-3/site)

Dbar = post.mean of -2logL; Dhat = -2LogL at post.mean of stochastic nodes									
MVPLNp_T_2-3					MVPLN_T_2-3				
	Dbar	Dhat	pD	DIC		Dbar	Dhat	pD	DIC
lt	1546	1502	44	1590	lt	1545	1501	45	1590
ra	260	250	11	271	ra	260	247	13	273
re	1020	980	41	1061	re	1020	979	41	1061
tot	5174	5107	67	5241	tot	5174	5107	67	5241
total	8001	7838	162	8163	total	8000	7835	165	8165
node	mean	sd	2.50%	97.50%	node	mean	sd	2.50%	97.50%
CRR1	-3%	5%	-13%	6%	CRR1	-3%	5%	-13%	6%
CRR2	18%	26%	-43%	55%	CRR2	24%	25%	-37%	60%
CRR3	-12%	14%	-41%	12%	CRR3	-11%	14%	-41%	13%
CRR4	5%	16%	-30%	32%	CRR4	4%	16%	-31%	31%

The final results from the favoured univariate PLN, the corresponding MVPLN and from the naïve before-after study are summarized in Table 6-16.

For case 1, which included 57 treatment sites and 59 reference sites, both MVPLN and PLN provided much lower CRR estimates than the Naïve method, which did not account for regression to the mean. However, differences of DICs between MVPLN and PLN are greater than 10 for all forms of $\mu_{i,t}^{k}$, suggesting that PLN is superior to MVPLN in terms of overall performance.

For case 2, the relatively low crash group that included 174 treatment sites and 190 reference sites, the difference is more dramatic in that the MVPLN and PLN correctly estimated no significant treatment effect for all types of crashes, while the Naïve method shows there is a significant total crash reduction. The differences of DICs between MVPLN and PLN are less than 10 for all forms of$\mu_{i,t}^{k}$, indicating that the differences might be reduced with the increase of sample sizes, a hypothesis that needs support with further study.

Tables 6-17 and 6-18 are the MCMC estimates of the posterior covariance matrix and correlation matrix of the latent effects ε of the MVPLN model for both data groups, respectively. It can be seen that covariances are very low for both groups, indicating weak correlations among different crash types, while, by contrast, the correlation coefficients for case 2 seem to suggest a strong correlation among different types of crashes. This is, however, a false correlation indicator because it is caused by very low crash counts for each type of crash (i.e., mainly zeroes at each site each year for each type of crash). In fact the correlation coefficients are reduced with the increased crash frequency for case 1. Further, the evaluated results themselves confirm that the data may not be strongly correlated. It can be concluded that correlation coefficients alone cannot be used for correlation identification for low sample mean data.

Table 6-16 Final Results from MVPLN, Univariate FB and Naïve

Case 1: total crash in first 5 years=7-10/site														
Fully Bayesian Before-After										**Naïve Before-After**				
Model 2: MVPLN_M_2-3					Model 2: PLN_M_2-3									
node	mean	Std	95% BCI		node	mean	Std	95% BCI		node	mean	Std	95% CI	
CRR1	-0.03	0.06	-0.15	0.09	CRR1	0.00	0.05	-0.10	0.10	CRR1	-0.11	0.07	-0.25	0.03
CRR2	-0.13	0.16	-0.48	0.16	CRR2	-0.11	0.16	-0.46	0.17	CRR2	-0.26	0.18	-0.60	0.09
CRR3	-0.16	0.18	-0.57	0.16	CRR3	-0.14	0.18	-0.54	0.16	CRR3	-0.34	0.23	-0.78	0.10
CRR4	-0.07	0.21	-0.54	0.27	CRR4	-0.13	0.21	-0.59	0.21	CRR4	0.31	0.14	0.04	0.58
Case 2: total crash in first 5 years=2-3/site														
Fully Bayesian Before-After										**Naïve Before-After**				
Model 3: MVPLN_T_2-3					Model 3: PLN_T_2-3									
node	mean	Std	95% BCI		node	mean	Std	95% BCI		node	mean	Std	95% CI	
CRR1	-0.03	0.05	-0.13	0.06	CRR1	-0.03	0.05	-0.12	0.06	CRR1	-0.21	0.08	-0.37	-0.06
CRR2	0.24	0.25	-0.37	0.60	CRR2	0.22	0.25	-0.37	0.57	CRR2	-0.56	0.72	-1.97	0.86
CRR3	-0.11	0.14	-0.41	0.13	CRR3	-0.10	0.13	-0.37	0.14	CRR3	-0.17	0.19	-0.54	0.20
CRR4	0.04	0.16	-0.31	0.31	CRR4	0.03	0.16	-0.32	0.29	CRR4	0.17	0.18	-0.17	0.51

notes: 1. BCI means Bayesian confidence interval
2. CI means confidence interval

Similar to the model comparison for the univariate FB approach, all of the models considering time effects have successfully estimated no treatment for those hypothetical treatment sites, whilst the PLN and PG models without time effects falsely assessed significant treatment for these hypothetical treatment sites. PLN Model 1 for both data groups has much less DIC of MVPLN Models 2, 3 and 4, indicating that the PLN Model 1 is superior to the three MVPLN models. However, it incorrectly estimates significant hypothetical treatment effects of CRR1, CRR2 and CRR3 for case 1 as well as significant treatment effects of CRR 1 and CRR3 for case 2, while the other three models provide the correct estimate of zero effect. This result further confirms the previous conclusion that various function forms of expected crashes should be compared accordingly by different random error distributions to select the best model. Otherwise, the estimated results may be seriously biased.

Table 6-17 Covariance Matrix of ε

Case 1: Total crashes in first 5 years =7-10/site				
	total	right angle	let turn	rear end
total	0.014			
right angle	0.050	0.383		
let turn	0.011	0.005	0.058	
rear end	0.000	-0.043	0.009	0.023
Case 2: Total crashes in first 5 years =2-3/site				
	total	right angle	let turn	rear end
total	0.009			
right angle	0.011	0.023		
let turn	0.012	0.017	0.032	
rear end	0.014	0.024	0.012	0.052

Table 6-18 Correlation-coefficients Matrix of ε

Case 1: Total crashes in first 5 years =7-10/site				
	total	right angle	let turn	rear end
total	1			
right angle	0.684	1		
let turn	0.383	0.031	1	
rear end	-0.013	-0.459	0.250	1
Case 2: Total crashes in first 5 years =2-3/site				
	total	right angle	let turn	rear end
total	1			
right angle	0.802	1		
let turn	0.717	0.638	1	
rear end	0.658	0.677	0.286	1

6.6 SUMMARY

The Fully Bayesian approach to road safety analysis has been available for some time, but has made very little impact on the way mainstream road safety evaluation studies are conducted.

This is perhaps because researchers and analysts were content with the empirical Bayes method and because the FB method was largely untested.

The objectives of this chapter were: 1. to examine if the FB before-after method can address the regression to the mean (RTM) problem and estimate no treatment effect athypothetical treatment sites for which there was in fact no treatment; 2. to explore if MVPLN is superior to the univariate FB method. Both univariate and multivariate before-after FB methods for treatment effect analysis were tested through three simulated datasets and two observed datasets. Sites were assigned randomly to hypothetical treatment and reference groups, such that the naïve before-after method would incorrectly show a significant treatment effect. It was confirmed that FB methods can indeed provide valid results, by correctly estimating no treatment effect at these hypothetical treatment sites.

Two FB testing frameworks were employed. First the univariate before-after fully Bayesian (FB) method was examined using three simulated datasets where there was a hypothetical treatment known to have no effect. Three forms of expected crashes $\mu_{i,t}$ were explored. It was found that the FB method can provide correct results, in that they estimate a treatment effect of zero. PLN and PG provide comparable results for these three datasets although they have big difference of model selection criteria LL, AIC, BIC and DIC. PLN is better than PG models in terms of model selection criteria and slightly better estimation results of CRRs. This might imply that the function form of $\mu_{i,t}$ is more important than the distribution of latent effects. The models accounting for time effects can give correct estimates of treatment effect while those that do not account for time effects provide incorrect estimates.

Finally, MVPLN, univariate PLN and PG models were evaluated for treatment effect analyses using two groups of California unsignalized intersections with different type's crashes (total, rear end, right angle and left turn). One group had relatively high crash frequencies (the total crashes in the first 5 years was between 7 and 10) while another group had lower crash frequencies (the total crashes in the same period was between 2 and 3). Four structural forms of expected crashes $\mu_{i,t}^{k}$ were developed and investigated for the univariate FB and MVPLN evaluation. For each form of $\mu_{i,t}^{k}$, it was found that MVPLN, PLN and PG provide comparable results for crash effect estimates while PLN was the best model in terms of the DIC measure. Similarly, it was found that crash effect estimates are very sensitive to the form of $\mu_{i,t}^{k}$ and that those models considering temporal effects of unobserved latent variables are superior to those that do not account for time variations. Both MVPLN and univariate FB models can simultaneously provide treatment effect estimates for all types of crashes using WinBUGs, but the computation time of the univariate FB is less. Thus, the univariate FB might be favoured when conducting before-after treatment effect analysis using different crash types.

Conclusions from the investigation in this chapter can be summarized as follows:

1. The univariate FB method was shown to be able to address the RTM problem and can provide promising results for treatment effect analysis.
2. It is essential to introduce time effects into the developed models for treatment effect analysis, i.e., the function form of expected crashes $\mu_{i,t}$ should account for time variations. .
3. The results in terms of crash reduction rate are more sensitive to function form of expected crashes $\mu_{i,t}$ than to the distributions of latent effects. Selection of the model distribution should be performed based on the same function form of expected crashes $\mu_{i,t}$; otherwise, results may be seriously biased.
4. MVPLN, PLN and PG model provide comparable estimates of treatment effect in terms of CRRs for different type crash data. Both univariate FB models and MVPLN models can simultaneously provide estimates of treatment effect for all crash types while univariate FB models have less computation time. MVPLN with coefficients following multivariate normal distribution do not provide better results and the compoutation time can be very large, so this model is not recommended.
5. Correlation coefficients alone cannot be used for the identification of correlations in data with low sample mean.
6. If only DIC is used as a model selection criterion, and DICs of alternative models are comparable, then the expected deviance could be used as a second criterion to select the best model.

CHAPTER 7 APPLICATIONS OF FULLY BAYESIAN METHOD FOR TREATMENT EFFECT ANALYSIS

7.1 INTRODUCTION

In Chapter 6, The FB before-after method for treatment effect analysis was evaluated and shown to provide promising results. Naturally, it is worthwhile then to compare the FB approach with the now conventional EB approach. To evaluate safety treatments with the EB approach, the before period crash experience at treated sites is used in conjunction with a negative binomial crash prediction model for untreated reference sites to estimate the expected number of crashes that would have occurred without treatment. This estimate is compared to the crashes observed after treatment to evaluate the effect of the treatment. This approach accounts for regression-to-the-mean effects that result from the natural tendency to select for treatment those sites with high observed crash frequencies.

This chapter provides a detailed comparison and discussion of the pros and cons of the two Bayesian approaches (EB and FB), based on, and illustrated with, empirical applications. These applications pertain to the evaluation of two treatments: the conversion of rural intersections from unsignalized to signalized control; and the conversion of road segments from a four-lane to a three-lane cross-section with two-way left turn lanes (also known as road diets). Part of The investigation of the conversion of rural intersections from unsignalized to signalized control has recently been published (Persaud et al., 2010; Lan et al., 2009) and some of the documentation below is taken from that source.

7.2 APPLICATION TO EVALUATION OF ROAD DIETS

The analysis undertaken examined the safety impacts of converting four lane roadways to 3 lane roadways where the middle lane is now a double left-turn lane, a treatment commonly known as road diets. The sites are located on the fringes of urbanized areas.

The FB analysis documented below also investigated the effect of size of the reference group. To this end, two reference samples were used to conduct the FB: 15 yoked reference sites and 296 reference sites, and the results compared with those from the EB.

7.2.1 Data Description

There were 15 treatment sites for this study. There were also 15 comparison sites which were used for an earlier FB study conducted by Pawlovich et al. (2005). In that study, comparison sites were chosen to match treatment sites in attributes including traffic volume, geometry and location (in terms of population size). Monthly crash records, traffic volumes and other road characteristic variables for all 30 sites were available for the period 1982-2004.

The 15 yoked reference sites were not enough for the EB analysis because a safety performance function could not be developed using this small sample size. For this reason, an expanded reference group of 296 sites was used to conduct for EB study. Yearly crashes, traffic volume and other variables are available for this reference group (296 sites) from 1982 to 2004 with a few missing values. Data for the 15 treated, 15 yoked reference and 296 reference sites are summarized in Tables 7-1, 7-2 and 7-3.

Table 7-1 Summary data for 15 treated segments

Variable	Mean	Minimum	Maximum
Years before	17.53	11.00	21.00
Years after	4.47	1.00	11.00
Crashes/mile-year before	23.74	4.91	56.15
Crashes/mile-year after	12.19	2.27	30.48
AADT before	7,987	4,854	11,846
AADT after	9,212	3,718	13,908
Length (miles)	1.02	0.24	1.72

Table 7-2 Summary of data for 296 untreated reference segments

Variable	Mean	Minimum	Maximum
Years	21.8	5	23
Crashes/mile-year	26.8	0.2	173.7
AADT	8,606	826	24,772
Length (miles)	0.99	0.27	3.38

Table 7-3 Summary of data for 15 untreated yorked segments

Variable	Mean	Minimum	Maximum
Years	22.9	21	23
Crashes/mile-year	15.8	0	55.9
AADT	7,006	778	15,374
Length (miles)	1.33	0.49	2.53

7.2.2 The FB Models

The time trend and time multiplier PLN and PG models were shown in Chapter 6 to be potentially the best models for this aspect of the research. The FB models were developed using the treatment and yoked reference group as well as the treatment group with 296 reference sites, respectively.

$$Y_{i,t} \sim \text{Pois}\ (\lambda_{i,t})$$

$$\lambda_{i,t} = \mu_{i,t} e^{\varepsilon_i} \qquad (7-1)$$

With respect to the gamma or log normal distribution that latent effect $\exp^{\varepsilon_i}$ follows, the developed models are called PG or PLN models accordingly, as explained before.

- Model 1: Time trend model

PLN models for the two reference groups are denoted as PLN_T_Yoked Pair, PLN_T_Reference296, respectively, accordingly defined as PG_T_Yoked Pair and PG_T_Reference296 for the PG models.

$$\mu_{i,t} = e^{\alpha} \times AADT_{i,t}^{\beta_1} \times exp^{\beta_2 \times t} \times Section\ Length_{\ i} \qquad (7-2)$$

where,

$AADT_{i,t}$ =AADT on road section i in year t,

α, β_1, β_2 = coefficients ,

α_t = yearly varying coefficients, and

e^{ε_i} = random effect at site i, follows gamma or log normal distribution

- Model 2: Time multiplier model

Similarly, PLN models for the two reference groups are denoted as PLN_M_Yoked Pair, PLN_M_Reference296, respectively, and defined as PG_M_Yoked Pair and PG_M_Reference296 for the PG models.

$$\mu_{i,t} = e^{\alpha_t} \times AADT_{i,t}^{\beta_1} \times Section\ Length_{\ i} \qquad (7-3)$$

7.2.3 Model Comparison

Again, two parallel chains were run for both initial cases to obtain posterior distributions of the parameters and crash reductions. The model selection criteria Log likelihood (LL), AIC, BIC and DIC were collected and calculated. The results obtained from the developed FB models are tabulated in Tables 7-4 and 7-5, respectively, using the two reference groups.

Table 7-4 Treatment Effect Analysis and Model Diagnostics

Reference Group: 296 sites

PLN_T_ Reference296					PG_T_ Reference296				
K	4				K	4			
LL	-23510				LL	-23500			
AIC	60442				AIC	60422			
BIC	56323				BIC	56303			
DIC	45599				DIC	45592			
node	mean	sd	2.50%	97.50%	node	Mean	sd	2.50%	97.50%
CRR	51%	1%	49%	52%	CRR	51%	1%	49%	52%
PLN_M_ Reference296					PG_M_ Reference296				
K	25				K	25			
LL	-22870				**LL**	**-22860**			
AIC	59162				**AIC**	**59142**			
BIC	67342				BIC	67322			
DIC	44324				**DIC**	**44314**			
node	mean	sd	2.50%	97.50%	node	Mean	sd	2.50%	97.50%
CRR	47%	1%	45%	49%	**CRR**	**47%**	1%	45%	49%

Note: K is the number of parameters

For 296 reference sites, all criteria except BIC favour the PG time multiplier model PG_M_Reference296 (bolded in Table 7-4); thus it is deemed to be the best model to estimate the treatment effects. However, the PLN time multiplier model PLN_M_Reference296 provides the same estimate of treatment (the same mean and standard values of CRR) even though they have different values of model selection criteria. Similarly PLN_T_Reference296 has the same value of CRR withPG_T_Reference296.

Table 7-5 Treatment Effect Analysis and Model Diagnostics

Reference Group: Yoked Pair (15 sites)

PLN_T_Yorked Pair					PG_T_Yorked Pair				
K	4				K	4			
LL	-1867				LL	-1868			
AIC	4944				AIC	4946			
BIC	4573				BIC	4575			
DIC	3759				DIC	3759			
node	mean	sd	2.50%	97.50%	node	mean	sd	2.50%	97.50%
CRR	53%	1%	51%	55%	CRR	53%	1%	51%	55%
PLN_M_Yorked Pair					**PG_M_Yorked Pair**				
K	25				K	25			
LL	**-1824**				LL	**-1823**			
AIC	**4858**				AIC	**4856**			
BIC	**5595**				BIC	**5593**			
DIC	**3693**				DIC	**3690**			
node	mean	sd	2.50%	97.50%	node	mean	sd	2.50%	97.50%
CRR	**49%**	1%	46%	51%	CRR	**49%**	1%	46%	51%

Note: K is the number of parameters

For the yoked 15 reference sites case, LL, AIC and DIC favour both PLN and PG models with time multiplier function form of expected crashes $\mu_{i,t}$ (bolded in Table 7-5). Both models were regarded as best for estimation and, indeed, again provide the same estimate of treatment effect, CRR. The same pattern was applied to the other two models, as shown in Table 7-5. The study with two data cases further confirms how sensitive treatment estimates can be to the function forms of expected crashes $\mu_{i,t}$ rather than to the distribution of latent effects.

The results of the FB analysis using the two reference groups are presented later for comparison with the results from the EB analysis, which is taken from the published paper (Persaud et al., 2010).

7.2.4 The FB Results with Comparison of EB

The results for different study groups using the FB and EB are shown in Table 7-6. Crash reduction rates estimated from Equation 6-5 for the FB analysis were converted to crash effects measured as a % change. First, it can be seen that the FB approach can provide similar results to the EB, without even considering the variances of the estimates. Even with the relatively small sample size (the yoked pair), the FB still can provide fairly good results. This sample of 15 reference sites was too small to estimate a safety performance function to apply the EB approach, suggesting perhaps that the FB has an advantage over the EB when the reference group size is restricted because of cost and other practical limitations.

It is natural to compare our FB results with those from the Pawlovich et al. (2006) FB analysis. In that study, which used the same treatment and yoked reference sites with monthly data, the yoked pair was used to play the role of the treatment sites, but without the intervention, and a crash rate (crashes per unit of average daily traffic volume) model was developed to conduct the treatment effect analysis using a comparison group approach. By contrast, this research used a before-after approach that developed a crash count model with traffic volume as an independent variable (Equations 7-2 and 7-3), recognizing that the relationship between crashes and traffic volume may not be linear, as is assumed in a crash rate model. In addition, Pawlovich et al. reported an average reduction in crashes per mile while a composite effect over all crash sites was estimated in this research, in effect giving more weight to the results for longer segments. Thus, regardless of these subtle differences in approaches, differences in the results of the two FB studies are not directly comparable because of the different outcome variables.

Table 7-6 Comparison of Crash Effects Estimated by the EB and FB Approaches for 15 road diet treatments

(negative sign indicates an increase in crashes)

(standard errors are in parentheses)

Number of reference sites	EB	FB
15 (yoked)	Not done	49% (1%)
296	47% (2%)	47% (1%)

7.3 APPLICATION TO EVALUATION OF TRAFFIC SIGNAL INSTALLATION

7.3.1 Data Description

The Highway Safety Information System (http://www.hsisinfo.org/) provided all data used in the study. Geometry, traffic volume and crash data from 1993 to 2002 were acquired for the State of California. In order to see the difference in treatment effects from the before-after FB and the more established before-after empirical Bayes (EB) approaches, the identical dataset was used to conduct the FB analysis as was used for the EB analysis conducted earlier (Bhim, 2005; Harkey et al., 2008). Tables 7-7 to 7-9 provides the summary information for each target crash type (total, rear-end, right angle and left turn) in the before and after periods at the converted intersections, which were all in rural areas and included 4 three-legged intersections and 24 four-legged intersections; of the latter, 14 had two lanes on the major and 10 had four lanes on the major.

Table 7-7 Converted three legged intersections with 2 lanes on major road

Number of Sites = 4			
Variable	mean	minimum	Maximum
Years before	1	1	6
Years after	6	4	9
Crashes/site-year before	6.458	1	153
Crashes/site-year after	3.525	0.57	7.778
Right-angle crashes/site-year before	0.083	0	0.33
Right-angle crashes/site-year after	0	0	0
Rear-end crashes/site-year before	0.083	0	0.167
Rear-end crashes/site-year after	0.215	0	0.5
Left-turn crashes/site-year before	3.125	0.33	10.33
Left-turn crashes/site-year after	0.146	0	0.33
Major road AADT before	12975	5750	19100
Minor road AADT before	5613	201	10300
Major road AADT after	15105	7400	26945
Minor road AADT after	5638	201	10300

Table 7-8 Converted four legged intersections with 2 lanes on major road

Number of Sites = 14			
Variable	mean	minimum	Maximum
Years before	4.286	1	8
Years after	5.714	2	9
Crashes/site-year before	3.303	0.125	8.6
Crashes/site-year after	3.280	0.667	9
Right-angle crashes/site-year before	0.964	0	2.5
Right-angle crashes/site-year after	0.379	0	1.667
Rear-end crashes/site-year before	0.198	0	0.75
Rear-end crashes/site-year after	0.173	0	0.5
Left-turn crashes/site-year before	0.886	0	3.25
Left-turn crashes/site-year after	0.727	0	3.25
Major road AADT before	10344	7400	18738
Minor road AADT before	2150	101	5280
Major road AADT after	11204	7762	21700
Minor road AADT after	2187	101	5280

Table 7-9 Converted four legged intersections with 4 lanes on major road

Number of Sites = 10			
Variable	mean	minimum	Maximum
Years before	3.4	1	6
Years after	6.6	4	9
Crashes/site-year before	5.557	2.667	10.5
Crashes/site-year after	5.229	1.44	10.75
Right-angle crashes/site-year before	2.15	0	7
Right-angle crashes/site-year after	0.568	0	1.167
Rear-end crashes/site-year before	0.2	0	1
Rear-end crashes/site-year after	0.44	0	1.25
Left-turn crashes/site-year before	1.507	0	3
Left-turn crashes/site-year after	1.018	0	2.667
Major road AADT before	15958	7018	25666
Minor road AADT before	2716	600	9700
Major road AADT after	18235	7155	29750
Minor road AADT after	2790	600	9646

The reference group of untreated intersections included 1,381 that were three legged, 726 that were 4-legged with two lanes on the major, and 181 that were 4-legged with four lanes on the major. The summary information of reference groups are tabulated in Tables 7-10 to 7-12.

Table 7-10 Stop Controlled 3legged intersections with 2 lanes on major road

Number of Sites = 1381			
Variable	Mean	Minimum	Maximum
Years	10	10	10
Crashes/site-year	0.84	0	19
Right-angle crashes/site-year	0.02	0	3
Rear-end crashes/site-year	0.06	0	3
Left-turn crashes/site-year	0.17	0	15
Major road AADT	9027	2550	33500
Minor road AADT	554	100	10001

Table 7-11 Stop Controlled 4legged intersections with 2 lanes on major road

Number of Sites = 726			
Variable	Mean	Minimum	Maximum
Years	10	10	10
Crashes/site-year	1.40	0	18
Right-angle crashes/site-year	0.33	0	14
Rear-end crashes/site-year	0.10	0	4
Left-turn crashes/site-year	0.23	0	9
Major road AADT	8526	2900	29732
Minor road AADT	653	100	7800

Table 7-12 Stop Controlled 4legged intersections with 4 lanes on major road

Number of Sites = 181			
Variable	Mean	Minimum	Maximum
Years	10	10	10
Crashes/site-year	1.24	0	12
Right-angle crashes/site-year	0.28	0	7
Rear-end crashes/site-year	0.10	0	3
Left-turn crashes/site-year	0.25	0	8
Major road AADT	12462	2952	36000
Minor road AADT	596	100	6000

7.3.2 FB Models

Since the treatment sites for each type of intersections are limited, three groups of intersections were combined together to estimate the treatment effect. PG models with or without trend, and PLN models with or without trend, were developed to conduct the before-after FB analysis. In addition, PG or PLN with time multiplier model was also developed based on the results presented above for the four models developed. Dummy variables are necessary to combine all groups together. The expected crashes can be modeled as:

- Models without trend

$$\mu_{i,t} = e^{a_1 T_i + a_2 F_{2i} + a_3 F_{4i}} ML_{i,t}^{b_1 T_i + b_2 F_{2i} + b_3 F_{4i}} XST_{i,t}^{c_1 T + c_2 F_2 + c_3 F_4} \quad (7-7)$$

- Models with trend

$$\mu_{i,t} = e^{a_1 T_i + a_2 F_{2i} + a_3 F_{4i}} ML_{i,t}^{b_1 T_i + b_2 F_{2i} + b_3 F_{4i}} XST_{i,t}^{c_1 T_i + c_2 F_{2i} + c_3 F_{4i}} e^{(d_1 T_i + d_2 F_{2i} + d_3 F_{4i})t} \quad (7-8)$$

- Model with time multiplier

$$\mu_{i,t} = exp^{a_{1,t} T_i + a_{2,t} F_{2i} + a_{3,t} F_{4i}} ML_{i,t}^{b_1 T_i + b_2 F_{2i} + b_3 F_{4i}} XST_{i,t}^{c_1 T_i + c_2 F_{2i} + c_3 F_{4i}} \quad (7-9)$$

where,

$\mu_{i,t}$ = expected number of crashes at intersection i in year t

i=1 to N, with N being the total number of intersections in the treatment database (including 3 and 4-legged intersections with either 2 or 4 lanes on the major road;

T_i = dummy variable, such that T_i=1, if intersection i is 3 legged, and T_i=0 otherwise; similarly,

F_{2i} = dummy variable, such that F_{2i}=1, if intersection i is 4 legged with 2 lanes on the major road, and F_{2i}=0 otherwise;

F_{4i} = dummy variable, such thatF_{4i}=1, if intersection i is 4 legged with 4 lanes on the major road, andF_{4i}=0 otherwise;

$a_1, a_2, a_3, b_1, b_2, b_3, c_1, c_2, c_3, d_1, d_2, d_3$ = Coefficients for dummy variables

$a_{1,t}, a_{2,t}, a_{3,t}$ = yearly varying coefficients

Prior distributions for all coefficients $a_1, a_2, a_3, b_1, b_2, b_3, c_1, c_2, c_3, d_1, d_2, d_3$ and $a_{1,t}, a_{2,t}, a_{3,t}$ are again assumed as non-informative N (0, 1000).

The DICs and CRRs for each type of crash are shown in Tables 7-13 and 7-14, respectively. It is seen that the PG models have much lower values than the PLN models. Based on this, the time multiplier PG model described in Equation 7-9 was applied, but it is not favoured due to the extra 27 parameters introduced. The difference of DICs for the PG with trend or without trend models is less than 10 for each type of crash, except for rear-end crashes, indicating that the two models are comparable. With this in mind, and considering that the time effects model should be always favoured given that model selection criteria are comparable, the PG_T was selected to conduct analysis for all types of crashes.

From Table 7-14, again, it is seen that PLN and PG, PLN_T and PG_T models provided very consistent estimates, as was found in the previous study. This further strengthens the conclusion that CRR is more sensitive to the function form of expected crashes than to the distributions of latent effects.

Table 7-13 DICs from Competitive Models

	Total	Rear-end	Right-angle	Left-turn
Poisson-Lognormal without trend	54164	12012	12281	19632
Poisson-Lognormal with trend	54158	11996	12274	19629
Poisson-Gamma without trend	53971	11876	12036	19428
Poisson-Gamma with trend	**53965**	**11850**	**12040**	**19423**
Time multiplier Poisson-Gamma	53986	11851	12060	19436

Table 7-14 Summary of Crash Effects Estimated by Alternative Models

	PLN	PLN_T	PG	PG_T	PG_M
Total	19% (6%)	22% (5%)	19% (5%)	**22% (5%)**	22% (5%)
Rear-end	-20% (21%)	-26% (23%)	-23% (22%)	**-27% (23%)**	-31% (24%)
Right-angle	81% (2%)	79% (3%)	81% (2%)	**79% (3%)**	78% (3%)
Left-turn	49% (6%)	52% (6%)	50% (6%)	**52% (6%)**	53% (6%)

Notes: 1. negative sign indicates an increase in crashes
2. standard errors are in parentheses

7.3.3 FB Results and Comparison with the EB

The FB results are presented in Table 7-14. They indicate highly significant reductions in left-turn, right-angle and total crashes following signal installation; for rear-end crashes, an increase of 27% was detected, but this was not statistically significant, likely because there were few of the crashes to begin with.

The EB results are also presented in Table 7-15 for comparison. It should be noted that the conventional statistical tests for the differences of the results from the FB and EB are not relevant since these are two estimates from the same sample. Nevertheless, visual inspection does suggest that the results from the two methods are comparable. The results are consistent across methods. Notably, the standard errors from the FB method are smaller than for the EB method, in contrast to the indication by Carriquiry et al. (2005) that the standard deviation from FB can be relatively large. This is likely because they introduced an intervention model, for which only 15 treatment sites were used to calibrate the intervention coefficients to obtain the expected crash frequency with treatment in the after period.

Table 7-15 Summary of Crash Effects Estimated by the FB and EB Methods
(negative sign indicates an increase in crashes)
(standard errors are in parentheses)

	FB Method Poisson Gamma with trend	Empirical Bayes (EB) Method
Total	22% (5%)	16% (6%)
Rear-end	-27% (23%)	-26% (27%)
Right-angle	79% (3%)	72% (5%)
Left-turn	52% (6%)	49% (7%)

7.4 SUMMARY

A detailed comparison of the two Bayesian approaches (EB and FB) was presented, based on, and illustrated with the evaluation of two treatments: the conversion of rural intersections from unsignalized to signalized control; and the conversion of road segments from a four-lane to a three-lane cross-section with two-way left turn lanes (also known as road diets).

The results suggest that the EB and FB results for treatment evaluation studies are comparable while FB method provides smaller standard deviations, which indicates a more stable estimate. This would suggest it is still appropriate to conduct treatment effect analysis using EB for univairate crash data and different type of crash data, but that it is essential in so doing to account for temporal trends in crash frequency. This conclusion is quite different from the network ranking investigation in which it was found that the FB method is superior to the EB method. This is probably because autoregressive models are not applicable to the FB before-after method in that the conversion year of treatment needs to be excluded for analysis.

It must be said, however, that the FB method is much more efficient than the EB for multivariate modeling of different types of crashes since all of the estimates of each type of crashes are obtained in one modeling procedure, while EB needs to conduct analysis for each type of crashes by developing and applying separate SPFs. Another advantage of the FB method is that it is available for situations where it is difficult to acquire a large enough reference group to calibrate safety performance functions required for the EB approach. For multilevel severity crash and/or spatial correlated data, the FB method is expected to provide better estimates in that it is able to deal with the inherent correlation among the crashes and/or segments.

CHAPTER 8 ACCOMPLISHMENTS, CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE STUDIES

8.1 ACCOMPLISHMENTS

The Fully Bayesian approach to road safety analysis has been available for some time, but has made very little impact on the way mainstream road safety evaluation studies are conducted. This is perhaps because researchers and analysts were content with the empirical Bayes method, and because the FB method was largely unevaluated and untested and may have been seen as too complicated to be worth the effort.

This study tried to address this gap by conducting a thorough evaluation of the FB method for two aspects of road safety analysis -- black spot identification and treatment effect analysis. To doing so, the following tasks were performed.

1. Explored various FB models with correlated data

Various FB models were proposed, developed and presented with correlated data, such as time series, spatial, temporal spatial, multivariate (with or without temporal correlation) and/or spatial correlations. Spatially related FB models were not further investigated due to the limitations of the data at hand.

2. Investigated the model selection criteria to identify best possible model selection criterion

Model selection criteria LL, AIC, BIC and DIC were obtained and compared in the evaluation studies. It was found DIC might be the best criterion in that the selected FB models can provide a better result.

3. Developed a proper approach to conduct a thorough evaluation of FB for black spot identification

Ten years of data from 1993-2002 for 726 unsignalized four legged intersections in California were used to evaluate the FB method for hot spot identification in comparison with the EB method, while 436 top ranked sites with five levels of severity data based on combined crash counts were selected for an evaluation study of the multivariate FB method. A thorough evaluation of the univariate FB versus EB method for single level severity data and multivariate FB versus univariate FB for multilevel of severity data, as well as on the performance of various ranking and evaluation criteria, was conducted.

For the univariate FB study with 726 sites, 11 FB models were developed and Poisson AR (1) was identified as the best model for comparison with the EB method and for further study and the AR (1) model was applied to multilevel severity crashes. Two time frames (1996-1998 and 1993-1998) were used to rank the sites for the evaluation study. The period 1999-2002 was selected as the evaluation period. Estimates of the true mean for the evaluation period were derived from the model developed, using a total of 10 years of data expected to be long enough to reasonably estimate the true mean for the evaluation period.

<u>4. Investigated ranking criteria and evaluation criteria and identified the possible best ranking criteria and evaluation criteria</u>

A total of 8 ranking criteria, which include posterior Poisson mean, posterior expected rank, posterior mode rank, posterior median rank, posterior probability of being the worst, raw data, posterior PSI and posterior PPSI have been examined. Specifically, the mode rank of the posterior distribution of the Poisson mean was proposed as a ranking criterion. In addition, the evaluation criteria, which include sensitivity and specificity, the sum of the PM, sum of the PSI, sum of crash counts and sum of the PPSI, were explored. The best ranking criteria and evaluation criteria were identified.

<u>5. Designed an approach to properly evaluate FB method for treatment effect analysis</u>

Two FB testing frameworks were employed. First the univariate before-after fully Bayesian (FB) method was examined using three simulated datasets. Then MVPLN, univariate PLN and PG models were evaluated for treatment effect analyses using two groups of California unsignalized intersections. Hypothetical treatment sites were assigned randomly to these datasets to separate treatment and reference sites such that the treatment group would have significant naive treatment effect. It was confirmed that FB methods can indeed provide promising results, in that they correctly estimate a treatment effect of zero at these hypothetical treatment sites.

6. Evaluated the performance of the EB and FB approaches for network ranking and treatment effect analysis to identify the advantages of FB method over the EB method

Both FB and EB methods were evaluated and compared based on ranking criteria of PM and PSI for black spot identification, using the single severity crash data. It was found that the FB method is superior to EB method in that it provides better results and it has more solid ranking criteria. For treatment effect analysis, EB and FB methods were applied to evaluation of two treatments: the conversion of rural intersections from unsignalized to signalized control; and the conversion of road segments from a four-lane to a three-lane cross-section with two-way left turn lanes (also known as road diets). The results indicate that both FB and EB method can provide comparable treatment effect estimates, while the estimate from the FB method has smaller standard deviations, indicating a more stable estimate. This would suggest it is still appropriate to conduct treatment effect analysis using EB for univairate crash data and multi-type crash data, but that it is essential in so doing to account for temporal trends in crash frequency. This conclusion is quite different from that for the network ranking investigation. This is probably because autoregressive models are not applicable to FB before-after method, in that the conversion year of treatment needs to be excluded for analysis.

8.2 CONCLUSIONS

Through the evaluation studies of the FB method for black spots identification and treatment effect analysis, the following conclusions were obtained:

- For single severity data, FB provides better results than the EB method in terms of higher sensitivity, specificity, sum of the PM and even sum of crash counts in the evaluation period, regardless whether ranking is by PM or PSI. For multilevel severity data, a multivariate approach has better performance than the univariate FB approach for network ranking.

- Posterior expected, median and mode ranks, as well as the probability of being the worst and posterior Poisson mean are good ranking criteria while expected rank has somewhat better ranking results than PM and median rank, both mode rank and the probability of being the worst may provide the best ranking results especially for the top ranked group. Expected rank, median rank and PM provide almost the same results if the order of the individual sites in the ranked group is not considered. Mode rank provides at least 90% of the same identified sites as PM or expected rank without taking into consideration the order of the ranked group. However, there is a substantial difference in rank order in comparison to PM or expected rank. It is shown that PSI cannot provide good ranking results but it might be used as a second level ranking criterion while other reliable criteria such as expected rank, PM or mode rank etc., are used as first level ranking criteria.

- Where only a few top ranked sites are of interest, sensitivity may not be a good evaluation criterion because one false positive can cause a huge difference in sensitivity, while the decision parameter may just have a minimal difference (i.e. 10.5 crashes versus 10.51 crashes). In such cases, the sum of the PM might be used as a major evaluation criterion.

- Short data history (3 years) can provide better ranking results than longer data history (6 years) for the where identification of only limited top ranked limited is of interest, as is common in a black spot identification program. As the number of ranked sites increases, a longer data history generally provides better results.

- The FB method was shown to be able to address regression to the mean problem and to provide promising results for treatment effect analysis. Both univariate FB and MVPLN models can simultaneously provide comparable estimates of treatment effect in terms of CRRs for different crash types. Univariate FB might be favoured due to overall performance and much shorter modeling time when using different types of crashes. In addition, correlation coefficients alone cannot be used for correlation identification for low sample mean data.

- The results in terms of crash reduction rate are more sensitive to function form of expected crashes $\mu_{i,t}$ than to the distributions of latent effects. Model selection of various distributions should be performed based on the same function form of expected crashes, $\mu_{i,t}$,; otherwise, seriously biased results can be anticipated. In addition, it is essential that the function form of expected crashes $\mu_{i,t}$ accounts for time variations. The model addressing time effect is always preferred unless it is strongly not favoured by the model selection criteria.

- MVPLN with coefficients following multivariate normal distribution is not recommended due to largely increased modeling running time and comparable results in terms of model selection criteria applied, and the CRRs estimated.

- DIC is a better model selection criterion than log likelihood, BIC and AIC. When only DIC is used as model selection criteria, and if DICs of alternative models are comparable, then the expected deviance might be used as a second criterion to select the best model.

- For cases such as small reference samples, multilevel severity crashes and/or spatial correlated data, FB method is anticipated to provide better estimates in that it is able to deal with the small sample problem and handle the coherent correlation among the crashes and/or segments.

8.3 RECOMMENDATIONS FOR FUTURE STUDIES

The research can be extended in several directions as follows:

1. Research for black spot identification can be extended to explore multilevel ranking criteria. For example, within some small range, where the primary decision parameters produced are essentially the same, a second level ranking criterion could be implemented, and so on. Potentially, expected rank can be used as first level criterion; PSI would then be used as the second level criterion. In this way, the ranked list might provide the most hazardous sites, while achieving the greatest possible safety improvement with a limited budget.

2. Three and six years' ranking data were used to study the effect of data history for black spots identification. Further study could be performed by exploring other data history years to find an optimal data history for hot spot identification.

3. Spatial correlated data could be used for the evaluation of black spot identification and treatment effect analysis and to identify the magnitude of difference from the FB and EB methods.

4. Finally, it would be useful to do a similar comparative evaluation of EB and FB methods using other datasets to examine the above conclusions.

REFERENCES

1. Abdel-Aty, M. and Wang, X. (2006). Crash Estimation at Signalized Intersections along Corridors: Analyzing Spatial Effect and Identifying Significant Factors. Transportation Research Record: Journal of the Transportation Research Board. Issue Number: 1953 Publisher: Transportation Research Board, pp.98-111.

2. Aguero-Valverde, J. and Jovanis, P. P. (2008). Analysis of Road Crash Frequency with Spatial Models. Transportation Research Record: Journal of the Transportation Research Board, No. 2061, Transportation Research Board of the National Academies, Washington, D.C., pp. 55–63.

3. Aguero-Valverde, J., Jovanis, P.P. (2009). Bayesian multivariate Poisson log-normal models for crash severity modeling and site ranking. Presented at the 88th Annual Meeting of the Transportation Research Board. Washington DC.

4. Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. Proc. 2nd International Symposium on Information Theory (B. N. Petrov and F. Csaki eds.) Akademiai Kiado, Budapest, pp. 267-281.

5. Aul, N. (2006). Using a Propensity Score Matching Method and a Hybrid Bayesian Method to Estimate Crash Modification Factors of Signal Installation. Transportation Research Board, Washington, D.C. CD-ROM paper.

6. Berg, A., Meyer, R., Yu, J. (2004). Deviance information criterion for comparing stochastic volatility models. Journal of Business and Economic Statistics, Vol.22, pp.107–120.

7. Bhim, R., (2005). Observational Before and After Safety Study of Installing Signals at Rural Intersections: Using the Empirical Bayes (EB) and Conventional Methods, M.A.Sc. Thesis, Ryerson University, Canada.

8. Bland, J M and Altman, D G (1994a). Statistic Notes: Regression towards the mean. British Medical Journal; 308:1499. http://www.bmj.com/cgi/content/full/308/6942/1499.

9. Bland, J M and Altman, D G (1994b). Statistics Notes: Some examples of regression towards the mean. British Medical Journal; 309:780. http://www.bmj.com/cgi/content/full/309/6957/780.

10. Bossche, F. V. , Wets, G., Lesaffre, E. (2003). A Bayesian Hierarchical Approach To model The Rank Of Hazardous Intersections For Bicyclists Using The Gibbs Sampler. Presented at the 82nd Annual Meeting of the Transportation Research Board. Washington DC.

11. Bozdogan, H. (2000). Akaike's Information Criterion and Recent Developments in Information Complexity. Journal of Mathematical Psychology, Vol. 44, pp. 62-91.

12. Brijs, T., Bossche, F. V., Wets, G. and Karlis, D.(2006). A model for identifying and ranking dangerous accident locations: a case study in Flanders. Statistica Neerlandica, Vol. 60, No. 4, pp. 457–476

13. Brijs, T., Karlis, D. Van den Bossche, F. Wets, G. (2007). A Bayesian Model for Ranking Hazardous Road Sites. Journal of the Royal Statistical Society Series A, Vol 170; No. 4, pp. 1001-1017.

14. Brooks, S. P. (1998) Markov Chain Monte Carlo Method and Its Application. The Statistician, Vol. 47, pp. 69-100.

15. Browne, M. (2000). Cross-validation methods. Journal of Mathematical Psychology, Vol. 44, pp.108-132.

16. Brüde, Ulf and Larson, J. (1988). The Use of Prediction Models for Eliminating Effects Due To Regression-To-the Mean in Road Accident Data. Accident Analysis and Prevention, Vol. 20, pp. 299–310.

17. Burnham, K. P., Anderson, D. R. (2002). Model selection and multi-model inference. Springer-Verlag.

18. Burnham, K. P. and Anderson, D. R. (2004). Multimodel Inference: Understanding AIC and BIC in Model Selection. Sociological Methods Research Vol. 33, pp. 261-304.

19. Busemeyer, J. R. and Wang,Y. (2000). Model Comparisons and Model Selections Based on Generalization Criterion Methodology. Journal of Mathematical Psychology, Vol. 44, pp.171-189.

20. Carriquiry, A. and Pawlovich,M.D (2005). From Empirical Bayes to Full Bayes: Methods For Analyzing Traffic Safety Data.
http://www.dot.state.ia.us/crashanalysis/pdfs/eb_fb_comparison_whitepaper_october2004.pdf. Accessed August 1st. 2005.

21. Celeux, G., Forbes, F., Robert, C.P. and Titterington, D.M. (2006). Deviance Information Criteria for Missing Data Models. Bayesian Analysis, Vol. 1, No. 4, pp. 651-674.

22. Cheng, W., Washington, S. (2005a). Experimental evaluation of hotspot identification methods. Accident Analysis and Prevention, Vol. 37, pp. 870–881.

23. Cheng, W. , Washington, S. (2008). New Criteria for Evaluating Methods of Identifying Hot Spots. Transportation Research Record: Journal of the Transportation Research Board, No. 2083, Transportation Research Board of the National Academies, Washington, D.C., pp. 76–85.

24. Chib, S. and Greenberg, E.(1994). Bayes inference in regression models with ARMA (p, q) errors. Journal of Econometrics, Vol. 64, pp. 183 -206

25. Chib, S. and Greenberg, E. (1995). Understanding the Metropolis-Hastings Algorithm. The american statistician, Vol. 49, No.4, pp. 327-335.

26. Congdon, P. (2001). Bayesian Statistical Modeling. JOHN WILEY & SONS, LTD.

27. Congdon, P. (2003). Applied Bayesian Modeling. JOHN WILEY & SONS, LTD.

28. Cowles, M. K. (2004). Review of WinBUGS 1.4. The American Statistician, Vol. 58, pp.330-336.

29. Davis, C E (1976). The effect of regression to the mean in epidemiologic and clinical studies. American Journal of Epidemiology. Vol. 104, pp. 493-498.

30. Deeks, J. J. (2001). Systematic Reviews of Evaluations of Diagnostic and Screening Tests. In Systematic Reviews in Health Care: Meta-Analysis in Context, 2nd ed., London, pp. 248–282.

31. Elvik, R. (1997). Evaluations of road accident black spot treatment: a case of the iron law of evaluation studies? Accident Analysis and Prevention, Vol. 29, Issue2, pp. 1964-1969.

32. Elvik, R. (2002). The Importance of Confounding in Observational Before-And-After Studies of Road Safety Measures. Accident Analysis & Prevention. Vol. 34, Issue 5, pp.631-635.

33. Elvik, R. (2008a). Comparative Analysis of Techniques for Identifying Locations of Hazardous Roads. Transportation Research Record: Journal of the Transportation Research Board, No. 2083, Transportation Research Board of the National Academies, Washington, D.C., pp. 72–75.

34. Elvik, R. (2008b). A Survey of Operational Definitions of Hazardous Road Locations in Some European Countries. Accident Analysis and Prevention, Vol. 40, pp. 1830–1835.

35. Elvik, R. (2008c). The predictive validity of empirical Bayes estimates of road safety. Accident Analysis and Prevention, Vol. 29, Issue 2, pp. 191-199.

36. El-Basyouny, K., Sayed, T. (2009). Collision prediction models using multivariate Poisson-lognormal regression. Accident Analysis & Prevention, Vol. 41, Issue 5. pp. 820-828.

37. Forster, M. R. (2000). Key Concepts in Model Selection: Performance and Generalizability. Journal of Mathematical Psychology, Vol. 44, pp.205-231.

38. Gelfand, A.E., Hills, S.E., Racine–Poon, A., and Smith, A.F.M., (1990). Illustration of Bayesian inference in normal data models using Gibbs sampling. Journal of the American Statistical Association, Vol. 85, pp. 972–985.

39. Gelfand, A. E. and Dey, D. K. (1994). Bayesian Model Choice: Asymptotics and Exact Calculations. Journal of the Royal Statistical Society. Series B (Methodological), Vol. 56, No. 3, pp. 501-514.

40. Gelman, A., Carlin, B. J., Stern, S. H. and Rubin, B. D.(2003). Bayesian Data Analysis, 2nd edition. CRC Press.

41. Geurts, K., Wets, G. (2003). Black Spot Analysis Methods: Literature Review. Technical report. http://www.steunpuntmowverkeersveiligheid.be/en/modules/publications/store/17.pdf

42. Geurts, K., Wets, G., Brijs, T., Karlis, D., Vanhoof, K. (2004). Ranking and Selecting Dangerous Accident Locations: Correcting for the Number of Passengers and Bayesian Ranking Plots. Technical report. http://www.steunpuntmowverkeersveiligheid.be/en/modules/publications/store/39.pdf

43. Geurts, K. Wets, G. Brijs, T. Vanhoof, K. (2004). Identification and Ranking of Black Spots: Sensitivity Analysis. Transportation Research Record. No. 1897, pp. 34-42.

44. Goodman, S. N. (1999). Toward Evidence-Based Medical Statistics. 2: The Bayes Factor. Ann Intern Med, Vol.130, pp.1005-1013.

45. Global status report on road safety: time for action. Geneva, World Health Organization, 2009. www.who.int/violence_injury_prevention/road_safety_status/2009.

46. Haque, M. M., Chin, H. C. and Huang, H (2010). Applying Bayesian hierarchical models to examine motorcycle crashes at signalized intersections. Accident Analysis and Prevention, Vol. 42, pp 203–212

47. Harkey, D., Persaud, B, Lyon, C. and others (2008). Crash Reduction Factors for Traffic Engineering and ITS Improvements. Final Report National Cooperative Highway Research Program (NCHRP) Project 17-25. (To appear as NCHRP Report 572).

48. Hauer, E. (1992) Empirical Bayes Approach to the Estimation of “Unsafety”: The Multivariate Regression Method. Accident Analysis and Prevention, Vol. 24, Issue5, pp. 457–477.

49. Hauer, E (1996). Identification of “Sites with Promise.” Transportation Research Record Journal of the Transportation Research Board, No. 1542, Transportation Research Board of the National Academies, Washington, D.C., pp. 54–60.

50. Hauer, E. (1997). Observational Before-After Studies in Road Safety: Estimating the Effect of Highway and Traffic Engineering Measures on Road Safety. Elsevier Science Ltd.

51. Hauer, E., Harwood, D. W. Council, F. M. and Griffith, M. S. (2002). Estimating Safety by the Empirical Bayes Method: A Tutorial. In Transportation Research Record: Journal of the transportation Research Board, No. 1784, Transportation Research Board of the National Academies,Washington, D.C., pp. 126–131.

52. Hauer, E., Kononov, J. and Allery, B.K., and Griffith M.S. (2004). Screening the Road Network for Sites with Promise. Transportation Research Record 1784, 27-32.

53. Hirst, W.M., Mountain, L.J. and Maher, M.J. (2004). Sources of error in road safety scheme evaluation: a quantified comparison of current methods. Accident Analysis & Prevention. Vol. 36, issue 5, pp.705-715.

54. Huang, H., Chin, H. C. and Haque, M. M. (2009). Empirical Evaluation of Alternative Approaches in Identifying Crash Hot Spots. Transportation Research Record: Journal of the Transportation Research Board, No. 2103, Transportation Research Board of the National Academies, Washington, D.C., pp. 32–41.

55. Karlis, D., Meligkotsidou, L.(2005). Multivariate Poisson regression with covariance structure. Statistics and Computing, Vol. 15, pp. 255–265.

56. Kass, R.E; Raftery, A. E. (1995). Bayes factors. Journal of the American Statistical Association. 90, 430; ABI/INFORM Global, pp. 773-795.

57. Kuha, J. (2004). AIC and BIC: Comparisons of Assumptions and Performance. Sociological Methods Research, Vol. 33, pp. 188-229.

58. Laird, N. M. and Louis, T. A. (1989). Empirical Bayes ranking methods. Journal of Educational Statistics, Vol. 14, pp. 29–46.

59. Lan, B., Persaud, N. B., Lyon, C., and Bhim, R. (2009). Validation of a Full Bayes methodology for observational before–after road safety studies and application to evaluation of rural signal conversions. Accident Analysis & Prevention. Vol. 41, Issue 3. pp. 574-580.

60. Lan, B., Persaud N. B. (2010). Evaluation of Multivariate Poisson Log Normal Fully Bayesian Methods for Before-After Treatment Effect Analysis. Transportation Research Board, Washington, D.C., CD-ROM paper.

61. Li, L., Li, Z. and Sui, D. Z. (2007). A GIS-based Bayesian approach for analyzing spatial–temporal patterns of intra-city motor vehicle crashes Journal of Transport Geography, Vol. 15, pp 274–285.

62. Lichstein, J.W., Simons, T.R., Shriner, S.A. and Franzreb K. E. (2002). Spatial Autocorrelation and Autoregressive Models in Ecology. Ecological Monographs, Vol. 72, No. 3, pp. 445–463.

63. Liang, K.-Y., and Zeger, S. L.(1986). Longitudinal Data Analysis Using Generalized Linear Models. Biometrika, Vol. 73, pp. 13–22.

64. Liu, C. C. and Aitkin, M. (2008). Bayes factors: Prior sensitivity and model generalizability. Journal of Mathematical Psychology, Vol. 52, Issue 6, pp. 362-375.

65. Lord, D. and Persaud, B.N. (2000). Accident Prediction Models With and Without Trend--Application of the Generalized Estimating Equations Procedure. Transportation Research Record 1717. Paper No. 00-0496.

66. Lord, D.(2006). Modeling motor vehicle crashes using Poisson-Gamma models: Examining the effects of low sample mean values and small samples size on the estimation of the fixed dispersion parameter. Accident Analysis and Prevention, Vol. 38, pp. 751-766.

67. Lu H., Reilly, C. S., Banerjee, S., Carlin, B.P. (2007). Bayesian areal wombling via adjacency modeling. Environ Ecol Stat, Vol. 14, pp. 433–452.

68. Lunn, D. J., Thomas, A., Best, N. and Spiegelhalter, D. (2000). WinBUGS – A Bayesian modeling framework: Concepts, structure and extensibility. Statistics and Computing 10, pp. 325–337.

69. Ma, J., Kockelman, M. K. (2006), Bayesian Multivariate Poisson Regression for Models of Injury Count, by Severity. Transportation Research Record: Journal of the Transportation Research Board, No. 1950, pp. 24-34.

70. Ma, J., Kockelman, M. K., Damien, P. (2008). A multivariate Poisson-lognormal regression model for prediction of crash counts by severity, using Bayesian methods. Accident Analysis & Prevention. Vol. 40, Issue 5, pp. 964-975.

71. Maher, M. J. & Summersgill, I. (1996). A comprehensive methodology for the fitting of predictive accident models. Accident Analysis & Prevention, Vol. 28, pp. 281-296.

72. Maher, M., Mountain, L. (2009). The Sensitivity of Estimates of Regression to the Mean. Accident Analysis & Prevention. Vol. 41, pp. 861-868.

73. McGuigan, D. R. D. (1981). The Use of Relationships between Road Accidents and Traffic Flow in “Black-Spot” Identification. Traffic Engineering and Control, Aug.–Sept. pp. 448–453.

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

75. Miaou, S.-P., Song, J. J. (2005). Bayesian ranking of sites for engineering safety improvements: Decision parameter, treatability concept, statistical criterion, and spatial dependence. Accident Analysis and Prevention, Vol. 37, pp. 699–720.

76. Miranda-Moreno F. L., Fu, L. (2007). Traffic Safety Study: Empirical Bayes or Full Bayes? Transportation Research Board, Washington, D.C. CD-ROM paper.

77. Myung, J. (2000). The Importance of Complexity in Model Selection. Journal of Mathematical Psychology, Vol.44, pp. 190-204.

78. Myung, J., Tang, Y. and Pitt, M. A. (2009). Evaluation and Comparison of Computational Models. Methods Enzymol. Vol. 454: pp. 287–304.

79. Newton, M.A. and Raftery, A.E. (1994). Approximate Bayesian inference by the weighted likelihood bootstrap (with Discussion). Journal of the Royal Statistical Society, series B, 56, pp. 3-48.

80. Park, E.S. and Lord, D., 2007.Multivariate Poisson–lognormal models for jointly modeling crash frequency by severity. Transportation Research Record 2019, pp. 1–6.

81. Park, E.S., Park, J., Lomax, T.J. (2009). A Fully Bayesian Multivariate Approach To Before-After Safety Evaluation. Presented at the 88th Annual Meeting of the Transportation Research Board. Washington DC.

82. Pawlovich, M.D., Li, W., Carriquiry, A., and Welch, T (2005). Iowa's Experience with "Road Diet" Measures: Impacts on Crash Frequencies and Crash Rates Assessed Following a Bayesian Approach. Transportation Research Record 1953, pp. 163-171

83. Payne, R.W.(2000) The Guide to Genstat. Lawes Agricultural Trust. Rothamsted Experimental Station, Oxford, UK.

84. Persaud, N. B. (1988). Do Traffic Signals Affect Safety? Some Methodological Issues" Transportation Research Record No. 1183, TRB, The National Academies, Washington, D.C., pp. 37-47.

85. Persaud, N. B. and Nguyen, T. (1998). Disaggregate Safety Performance Models for Signalized Intersections on Ontario Provincial Roads. Transportation Research Record 1635, pp.113-120.

86. Persaud, N. B., Hauer, E., Retting, R, Vallurupalli, R and Mucsi, K (1997). "Crash Reduction Related to Traffic Signal Removal in Philadelphia", Accident Analysis and Prevention, Vol. 29, No.6., pp. 803-810.

87. Persaud, N. B., C. Lyon, and T. Nguyen (1999). Empirical Bayes Procedure for Ranking Sites for Safety Investigation by Potential for Safety Improvement. In Transportation Research Record: Journal of the Transportation Research Board, No. 1665, TRB, National Research Council, Washington, D.C., pp. 7–12.

88. Persaud, N. B, Retting, R., Garder, P. and Lord,D. (2001) "Safety Effect of Roundabout Conversions in the U.S.: Empirical Bayes Observation Before-After Study," Transportation Research Record No. 1757, TRB, The National Academies, Washington, D.C., pp. 1-8.

89. Persaud, N. B., Lord, D. and Palmisano, J (2002). Calibration and Transferability of Accident Prediction Models for Urban Intersections, Transportation Research Record 1784, pp. 57-64.

90. Persaud, N. B, McGee, H., Lyon, Craig., and Lord, D., (2003). Development of a Procedure for Estimating the Expected Safety Effects of a Contemplated Traffic Signal Installation, Transportation Research Record No. 1840, Transportation Research Board, Washington, D.C..

91. Persaud, N. B, Council, F. M., Lyon, C., Eccles, K. A., Griffith, M. (2005). A Multi-jurisdictional Safety Evaluation of Red Light Cameras. Transportation Research Record 1922, pp. 29-37.

92. Persaud, B., Lyon, C. (2007). Empirical Bayes before–after safety studies: Lessons learned from two decades of experience and future directions. Accident Analysis & Prevention, Volume 39, Issue 3, pp. 546-555.

93. Persaud N. B., Lan, B., Lyon, C., and Bhim, R. (2010). Comparison of Empirical Bayes And Full Bayes Approaches For Before-After Road Safety Evaluations. Accident Analysis & Prevention, Vol. 42, Issue 1, Pages 38-43.

94. Pitt, M.A. and Myung, I. J. (2002). When a good fit can be bad. TRENDS in Cognitive Sciences. Vol.6, No.10, pp. 421-425.

95. Pitt, M. A., Kim, W. J. and Myung, I. J. (2003). Flexibility versus generalizability in model selection. Psychonomic Bulletin & Review, Vol.10, issue1, pp. 29-44.

96. Raftery, A. E. (1999). Bayes Factors and BIC. Sociological methods & research, Vol.27, No.3, pp. 411-427.

97. Rothman, K., and Greenland, S. (1998). Modern Epidemiology, 2nd ed. Lippincott Williams and Wilkins, Philadelphia, Pa.

98. SAS Institute (1998). SAS Institute Inc. SAS Users Manual, SAS Institute Inc, Cary, NC.

99. Sayed, T., and Rodrigez, F (1999). Accident Prediction Models for Urban Unsignalized Intersections in British Columbia. Transportation Research Record No. 1665, pp. 93-99.

100. Schluter, P. J. Deely, J. J. and Nicholson, A. J. (1997). Ranking and selecting motor vehicle accident sites using a hierarchical Bayesian model. The Statistician, vol. 46, No. 3, pp. 293-316.

101. Schwarz, G. (1978). Estimating the Dimension of a Model. The Annals of Statistics, Vol. 6, No. 2, pp. 461-464.

102. Shaddick, G., Choo, L. L., and Walker, S. G. (2007). Modelling Correlated Count Data with Covariates. Journal of Statistical Computation and Simulation, Vol. 77, No. 11, pp.945–954.

103. Shen, W. and Louis, T. A. (1998). Triple-Goal Estimates in Two-Stage Hierarchical Models. J. R. Statist. Soc. B, Vol.60, Part 2, pp. 455-471.

104. Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and Linde, A. v. d. (2002). Bayesian measures of model complexity and fit. J. R. Statist. Soc. B, 64, Part 4, pp. 583–639.

105. Spiegelhalter, D., Thomas, A., Best, N., and Lunn, D (2003). WinBUGS Version 1.4 User Manual. MRC Biostatistics Unit, Cambridge. http://www.mrc-cam.ac.uk/bugs.

106. SØrenthen, M. (2007). Sixth framework programme. Technical Report. Ripcord Iserest

107. Transport Canada (2005). Canadian motor vehicle traffic collision statistics: 2003 [online]. Available from http://www.tc.gc.ca/eng/roadsafety/tp-tp3322-2006-menu-586.htm

108. Tsionas, E.G. (1999). Bayesian Analysis of the Multivariate Poisson Distribution, Communications in Statistics-Theory and Methods, Vol. 28, pp. 431–451.

109. Tsionas, E. G. (2001). Bayesian Multivariate Poisson Regression. Communications in Statistics—Theory and Methods, Vol. 30, No. 2, pp. 243–255.

110. Tunaru, R. (2002). Hierarchical Bayesian models for multiple count data. Austrian Journal of Statistics 31 (3), pp. 221–229.

111. Turner-Fairbank Highway Research Center (1999). Crash Models for Rural Intersections: Four-Lane by Two-Lane Stop-Controlled and Two-Lane by Two-Lane Signalized. Publication NO. FHWA-RD-99-128.

112. US Department of Transport (2001). Safety Conscious Planning: Parts 1, 2, and appendices, Federal Highway Administration [online] Available from http://www.fhwa.dot.gov/planning/scp/ec041scp1.htm.

113. Wasserman, L. (2000). Bayesian Model Selection and Model Averaging. Journal of Mathematical Psychology, Vol. 44, pp. 92-107.

114. Weakliem, D. L. (2004). Introduction to the Special Issue on Model Selection. Sociological Methods Research, Vol. 33, pp. 167-187.

115. Yu, J., Meyer, R. (2006). Multivariate Stochastic Volatility Models: Bayesian Estimation and Model Comparison. Econometric Reviews, 25(2–3):361–384

116. Zeger, S. L., and Liang, K.-Y.(1986). Longitudinal Data Analysis for Discrete and Continuous Outcomes. Biometrics, Vol. 42, pp. 121–130.

APPENDIX

Probability Distributions for Modeling

This appendix provides a summary of the probability distributions used in this dissertation, which are Poisson distribution, Gamma distribution, log normal distribution, multivariate normal distribution and Wishart distribution. The probability density function (denoted PDF), its mean and variance of each probability distribution are provided below.

1. Poisson

The Poisson distribution is a discrete distribution. It is used to model the number of events occurring within a given time interval or a given space.

$$X \sim \mathrm{Pois}\ (\lambda) \qquad\qquad \lambda > 0 \text{ and } x = 0,1,2 \dots ,$$

then the PDF is

$$f(x; \lambda) = \frac{\lambda^x e^{-\lambda}}{x!}$$

The mean and variance:

$$E(X) = \lambda$$
$$\mathrm{Var}(X) = \lambda$$

2. Gamma

The Gamma distribution is a continuous probability distribution. The probability density function of the gamma distribution can be expressed in terms of the gamma function parameterized in terms of a shape parameter α and scale parameter β.

$$X \sim Gamma\ (\alpha, \beta) \qquad\qquad x > 0\ \ and\ \alpha, \beta > 0 ,$$

The PDF is

$$f(x;\alpha,\beta)=\frac{\beta^{\alpha}x^{\alpha-1}e^{-\beta x}}{\Gamma(\alpha)} \qquad x>0 \; and \; \alpha,\beta>0$$

The mean and variance of the Gamma distribution are:

$$\mathrm{E(X)}=\alpha/\beta$$

$$\mathrm{Var(X)}=\alpha/\beta^2$$

3. Lognormal

If X is a random variable with a normal distribution, then Y = exp(X) has a log-normal distribution; likewise, if X is log-normally distributed, then X = log(Y) is normally distributed.

A log normal distribution is a probability distribution of a random variable whose logarithm is normally distributed. The PDF of a log-normal distribution is:

$$X\sim \text{log normal } (\mu,\sigma^2)$$

The PDF is

$$f_X(x;\mu,\sigma)=\frac{1}{x\sigma\sqrt{2\pi}}e^{-\frac{(lnx-\mu)^2}{2\sigma^2}} \qquad x>0$$

Its expected value (mean) and variance are,

$$E(X)=e^{\mu+\frac{\sigma^2}{2}}$$

$$Var(X)=(e^{\sigma^2}-1)e^{2\mu+\sigma^2}$$

4. Multivariate normal Distribution

If a random vector $\boldsymbol{X} = (X_1,\dots,X_K)'$ is a multivariate normal distribution:

$$\boldsymbol{X}\sim N_K(\boldsymbol{\mu},\boldsymbol{\Sigma})$$

The PDF is

$$f_{\boldsymbol{X}}(x;\ \boldsymbol{\mu},\boldsymbol{\Sigma})=\frac{1}{(2\pi)^{K/2}|\boldsymbol{\Sigma}|^{1/2}}\mathrm{e}^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}$$

where $|\Sigma|$ is the determinant of Σ, and where $(2\pi)^{K/2}|\Sigma|^{1/2}$ could instead be written as $|2\pi\boldsymbol{\Sigma}|^{1/2}$. This expression reduces to the density of the univariate normal distribution if $\boldsymbol{\Sigma}$ is a scalar (i.e., a 1×1 matrix).

The vector $\boldsymbol{\mu}$ in these conditions is the expected value of X and the matrix $\boldsymbol{\Sigma}$ is the covariance matrix.

The mean and variance are

$$E(\boldsymbol{X}) = \boldsymbol{\mu}$$
$$Var(\boldsymbol{X}) = \boldsymbol{\Sigma}$$

5. Wishart Distribution

A Wishart (R, p) prior is defined for the covariance matrix Σ^{-1} of multivariate normal distribution in this study, denoted as (f_Σ) where R is the scale matrix and p is the degrees-of-freedom parameter respectively. The hyper-prior parameters R and $p \geq L$ are known, usually assuming $p = K$ for vague prior, where K is the number of severities or types of crashes in road safety analysis.

$$\boldsymbol{\Sigma}^{-\mathbf{1}} \sim W(R, p)$$

The parameterization of the Wishart probability density function (pdf) is

$$f(\boldsymbol{\Sigma}^{-1}; R, p) = |\mathrm{R}|^{\frac{p}{2}} |\Sigma^{-1}|^{\frac{p-L-1}{2}} \exp^{-0.5\mathrm{Tr}(\mathrm{R}\boldsymbol{\Sigma}^{-1})}$$

The mean and variance are

$$E(X) = p\Sigma^{-1}$$
$$Var(X) = p({\sigma_{ij}}^2 + \sigma_{ii}\sigma_{jj})$$