Effects of Highway Geometric Features and Pavement Quality on Traffic Safety

A Thesis Presented for the
Doctor of Philosophy
Degree
The University of Tennessee, Knoxville

Ximiao Jiang
August 2012

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my advisor, Dr. Baoshan Huang, and my co-advisor, Dr. Stephen Richards, for their guidance, support, help, and encouragement throughout my doctoral study. Without their insightful suggestions on the research topics, I could not have been where I am now.

I earnestly thank my advisor for Statistics Master's Degree, Dr. Russell Zaretzki, for his professional guidance and his enduring assistance on both my PhD's research and overall academic well-being. A special thank has to be addressed to my committee member, Dr. Lee Han, who provided extensive support and understanding on my research.

Great thanks go to Mr. Tom Every in the Department of Transportation, who has provided the required data in my research. I would also like to recognize the Highway Safety Information System for providing Illinois Highway Safety Database.

Great thanks go to Dr. Xiang Shu for his help and suggestions on my research. I would like to extend my appreciation to my colleagues and friends, Dr.Qiao Dong, Dr. Hao Wu, Dr. Jingsong Chen, Sheng Zhao, Changjun Zhou, Benjamin Bowers for their help and friendship. I would also like to recognize Tennessee Department of Transportation and Southeastern Transportation Center for the financial support.

Finally, I would like to thank my wife, Bin Zhang. She has always been supportive in my research career. Further thanks go to my parents, my sister and my brother in law. Their best wish, although from thousands of miles away, is always my best gift. Their love and encouragement deserve special recognition.

ABSTRACT

Curbs are commonly used on roadways to serve for drainage managing, access controlling and other positive functions. However, curbs may also bring about unfavorable effects on drivers' behavior and vehicles' stability when hitting curbs, especially for high-speed roadways. In addition, numerous pavements have been experiencing wear off rapidly in recent decades, which significantly affected driving quality. However, previous study of pavement management factors as related to the happening and outcome of traffic-related crashes has been rare. The objective of this dissertation is to evaluate the influences of outside shoulder curbs, pavement management factors, and other tradition traffic engineering factors on the occurrence and outcome of traffic-related crashes.

The Illinois Highway Safety Database from 2003 to 2007 and the Tennessee crash data from 2004 to 2009 was employed in this research. A few advantage statistics models were built to study the effects of curbs, pavement quality and other typical factors on both the occurrence and the outcome of crashes. These models include: the Zero-Inflated Negative Binomial models (ZINB), the Zero-Inflated Ordered Probit (ZIOP) model, the random effect Poisson and Negative Binomial model, as well as the Bayesian Ordered Probit (BOP) model.

The findings of this study suggest that the employment of curbed outside shoulders on high-speed roadways would not pose any significantly harmful effect on the occurrence

of crashes. On high-speed roadways with curbed outside shoulders in terms of the crash frequency, reducing speed limit from 55 mph to 45 mph would not achieve any safety benefit. Crashes occurring on roadways with curbed outside shoulders are more likely to be no and minor injury related as compared to crashes on roadways without curbs. The increase of speed limit from 45 to 55 has relatively small effects on single vehicle crashes occurring on roadways with curbed outside shoulders. Rough pavements were associated with higher overall crash frequency but lower level of injury severity given that a two vehicle involved rear-end, head-on and angle crash has occurred. Pavements with more severe distress were related to lower crash frequency.

Keywords: *Curb, Pavement Quality, Crash Frequency, Injury severity*

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ACRONYMS

AADT	Annual Average Daily Traffic
AIC	Akaike's Information Criteria
BOP	Bayesian Ordered Probit
DIC	Deviance Information Criteria
HSIS	Highway Safety Information System
HPMA	Highway Pavement Management Application
LTPP	Long Term Pavement Program
NB	Negative Binomial
NCHRP	National Cooperative Highway Research Program
OP	Ordered Probit
PSI	Present Serviceability Index
PDI	Pavement Distress Index
RD	Rutting Depth
SBC	Schwarz Bayesian Criteria
TRIMS	Tennessee Roadway Information Management System
TDOT	Tennessee Department of Transportation
VMR	Variance to Mean Ratio
ZIP	Zero-Inflated Poisson
ZINB	Zero-Inflated Negative Binomial
ZIOP	Zero-Inflated Ordered Probit

PART 1 RESEARCH BACKGROUND AND OBJECTIVES

1.1 Problem Statement

Traffic safety has been a major concern for the general public and various government agencies for decades. In 2002, the World Health Organization (WHO 2002) reported that traffic crashes ranked tenth among top killers in terms of years of life lost world wide. In 2004, 42,636 people lost their lives in U.S. motor-vehicle crashes (NHTSA 2005a). In the U.S., traffic crashes cause more loss of human life (as measured in human-years) than almost any other cause-falling behind only cancer and heart disease (NHTSA 2005b).

Traffic safety is complexly determined by various components of human-environment-vehicle system. Numerous traffic accident studies have attempted to determine the impacts of all kinds of factors on the occurrence and outcome of traffic accidents. However, the research results so far are still controversial, and until this point, a few factors have seldom been accounted for in traffic safety researches.

First of all, there is a gap in the safety evaluation process between pavement management and highway design groups. Traditionally, skid resistance and surface texture, roughness (e.g., IRI), pavement friction and surface conditions (ruts, faults, potholes, cracks, spalls, etc.) related to different paving materials and types are the most common pavement management indicators of safety problems. While these pavement engineering factors are important for pavement design and maintenance, they are not utilized for traffic safety considerations by highway design and planning groups. In fact, these pavement management factors are complexly correlated with highway design factors, such as highway type, profile, alignment, the number of lanes, lane width, shoulder types, median types, traffic characteristic, etc. For example, curve segments have a higher demand for surface friction to reduce lane-departure crash rates. Therefore, there is a need for integrating pavement management and road geometric factors into traffic safety management for high speed facilities.

In addition, roadway geometric design has always been a major concern in traffic safety research. As the development of suburbanization on city peripheries, the increasing traffic volume may exceed the capacity of the original two-lane rural highways, which will prompt the need to increase the number of lanes. However, the increase number of lanes will bring about a series of issues regarding drainage performance, speed limit control, as well as cross section design. Particularly, whether the installation of curbs for drainage purpose will conflict with original speed limit in terms of traffic safety is still questionable.

1.2 Research Objectives and Significance

The objective of this study is to investigate the influence of typical geometric features, pavement quality, as well as their interactions with general environmental, human and vehicle factors on the occurrence and outcome of traffic related accidents. The original contribution includes:

- Incorporate real-time pavement quality measurements into statistics models to investigate their influences on both crash frequency and traffic related crash injury on State Route roadways;

- Account for both the temporal and spatial correlations in evaluating the effects of pavement management factors on crash frequency;
- Employ large-scale dataset to investigate the safety effects of curbs and their interactions with speed limits on traffic safety;
- First introduce Zero-Inflated Ordered Probit model into the study of traffic safety issues;
- Fully evaluate the performance of the Bayesian method in comparison with Maximum Likelihood Estimate method in traffic safety prediction models.

1.3 Research Methodology

This dissertation primarily include four parts of research. In the first part, the influences of outside shoulder curbs on the occurrence of crashes were studied utilizing Zero-Inflated Negative Binomial (ZINB) model; in the second part, the influences of outside shoulder curbs on the injury severity of traffic-related crashes were evaluated employing the Zero-Inflated Ordered Probit (ZIOP) model; in the third part, the effects of pavement management factors on the occurrence of crashes were estimated utilizing random effect Bayesian models, including: Poisson, Negative Binomial (NB), One Random Effect Poisson (OREP), One Random Effect Negative Binomial (ORENB), Two Random Effect Poisson (TREP), Two Random Effect Negative Binomial (TRENB) models; in the fourth part, the effects of pavement management factors on the outcome of traffic-related crashes were evaluated employing both the traditional Ordered Probit (OP) model and the Bayesian Ordered Probit (BOP) model. The overall research approach is shown in Figure 1.1.

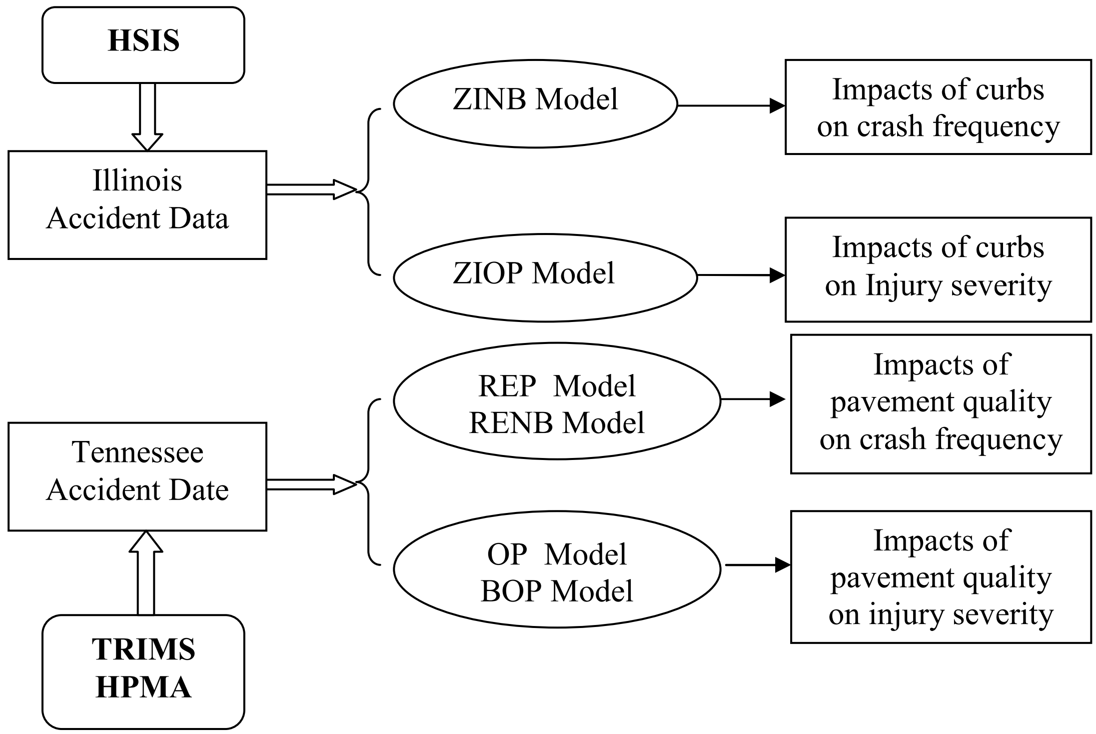


Figure 1.1 Research Flow Chart

PART 2 INFLUENCE OF CURBS ON TRAFFIC CRASH FREQUENCY IN HIGH-SPEED ROADWAYS

This part is based on a published journal paper (Jiang, et al. 2011). This paper was submitted to the journal Traffic Injury Prevention in January, 2011. The authors have revised a few issues according to the suggestions of reviewers in February and March, 2011. This paper was final accepted in April, 2011, and published in July, 2011.

The authors that have contributed to this paper include Dr. Xuedong Yan, Dr. Baoshan Huang, Dr. Stephen Richards and myself. Dr. Yan's major contribution is collecting the data for this research, and providing help in statistics modeling. Dr. Huang and Dr. Richards have contributed to providing funding and the consultant help in highway geometric design related knowledge. Myself has contributed to the data mining, statistics modeling construction, literature review, result interpretation and the composing of the text.

2.1 Abstract

Curbs are commonly used on roadways to serve for drainage managing, access controlling and other positive functions. However, curbs may also bring about unfavorable effects on drivers' behavior and vehicles' stability when hitting curbs, especially for high-speed roadways. The objective of this part of research is to investigate if the presence of curbs along outside shoulders has produced adverse effects on traffic safety on high-speed roadways, and whether increasing speed limits has created any further harmful effects.

In this study, the Illinois Highway Safety Database from 2003 to 2007 were selected to

evaluate the effects of curbs over traffic safety on 2-lane and 4-lane non-freeways with the speed limits of 45, 50 and 55 mph. The Wilcoxon/Kruskal-Wallis nonparametric tests were conducted to compare the road-segment crash rates between three types of outside shoulders (curbed shoulder, soft flush shoulder, and hard flush shoulder) and between the three types of speed limits. In addition, the Zero-Inflated Negative Binomial models were developed for all the roadway segments combined, as well as the curbed outside shoulder only segments. The models were used to estimate the influences of curbed outside shoulder, speed limit level, as well as other roadway characteristics on the crash frequency.

It was found that road segments with different types of outside shoulders were from different populations in terms of the distribution of crash rates, so did segments with different speed limits. Further, the crash frequency analysis indicates that the curbed outside shoulders are not likely to result in a higher crash frequency compared to the other two types of outside shoulders. In addition, there was no evidence that a decrease in speed limit results in reduction in crash frequencies for road segments with curbed outside shoulders.

The findings of this study suggest that the employment of curbed outside shoulders on high-speed roadways would not pose any significantly harmful effect on the occurrence of crashes, and on high-speed roadways with curbed outside shoulders, reducing speed limit from 55 mph to 45 mph would not achieve any safety benefit.

2.2 Introduction

Due to the development of suburbanization on city peripheries, the traffic volume of the original two-lane rural highways, typically with 55 mph speed limit, may exceed their capacities. In order to enhance driving quality and reduce traffic delay, the two-lane rural highways need to be widened to four-lane highways to mitigate traffic pressure. In these highway improvement projects, curbs are frequently installed along the new four-lane highways to provide drainage and address issues such as access control, difficult terrain, and limited right-of-way.

According to AASHTO, curbs are used extensively on all types of urban highways, but should be used cautiously on high-speed rural highways because they may lead to vehicles losing control or tires going airborne during impact (AASHTO, 2005). Therefore, on one hand, the use of curbs is generally discouraged in many states in the U.S. on roadways with design speeds over 45 mph, and whenever they are necessary for drainage control or other special functions, only certain types of curbs are permitted to be used with specified requirements of placement. On the other hand, the usage of curbs when updating and widening rural two-lane highways into four-lane highways, frequently results in reducing the speed limit from 55 mph to 45 mph.

However, lowering speed limits may confuse drivers and create an enforcement problem since many drivers do not perceive the danger and try to travel as fast as they had before the road was improved. As a result, although the highway conditions are significantly enhanced, the projects may be viewed negatively by the public, design professionals, and law enforcement personnel (Baek et al., 2006). Thus, whether or not to cut speed limits is a dilemma for engineers when updating and widening two-lane rural roadways with 55 mph speed limit into four-lane highways with the presence of curbs along outside shoulders.

There is lack of literatures specifically focusing on whether and to what extent the existence of curbs would add to the occurrence of traffic crashes on high-speed roadways. Hadi et al. (1995) evaluated the effects of cross-section elements on total fatality and injury crash rates for different types of rural and urban highways at various traffic levels. The results indicated that the presence of raised curbs in the medians of all urban highways except four-lane undivided highways has an adverse effect on vehicular safety. Similarly, Bligh and Mak found that curbs and other non-rigid or low-profile fixed objects are involved in many more rollover accidents compared to rigid fixed objects (Bligh et al., 1999). Moreover, the study conducted by Plaxico et al. reported that curbs pose a significant hazard to the security of motorcyclists. Besides, their analyses on the NASS-GES and 1996-1997 Michigan and Illinois HSIS data showed that the proportion of rollover involved in curb crashes increased as the assigned operating speed increased (Plaxico et al., 2005). However, none of these studies were specifically focused on the safety effects of curbed outside shoulders on high-speed roadways.

There are also previous research indicating that curbs would not increase crash rate for highspeed roadways. Baek et al. (2008) conducted a research to identify the mean crash rates difference between roadways at speed limit of 45 mph and 55 mph with the presence of curbs next to each direction of the outside lanes. The results suggest that there is no significant difference between them in both total and curb-involved crashes. This study was based on the data collected from 60 selected sites in North Carolina, which might be limited to draw any sound conclusion. The further research conducted by Baek and Hummer (2008) indicated that multilane highways with curbed outside shoulders were associated with fewer total collisions and equal injury collisions as compared to no curbs, and the segments with speed limit of 55 mph had more collisions than those of 45 mph. This study included 2,274 collisions occurring on 191.85 miles of directional multilane highway segments in North Carolina from 2001 to 2003. However, this study is lack of the exploration into the interaction impact of speed limit and curbed outside shoulders on the collision frequency. Besides, taking crash rate (the number of collisions per mile per year) as the predicted variable ignored the possible non-linear relationship between the number of crashes and segment length, which has been raised by increasing number of researchers (Milton et al., 1998; Qin et al., 2004; Sawalha et al., 2001).

It can be conclude from the literature review that the research on the effects of curbs on the occurrence of crashes in high speed limit conditions is very limited and the results are still controversial. Crash frequency prediction models taking into account the presence of curbed outside shoulders, as well as typical geometric and traffic features are also lacking. Hence a comprehensive analysis of curb-related safety influence considering speed limit change on high-speed roadways is desired.

The objective of this part of research aims at investigating the safety effect of curbs along outsider shoulders on high-speed roadways, utilizing the Illinois Highway Safety Database from the year 2003 to 2007. This research specifically focused on crash rate comparison and crash frequency modeling analysis of the non-freeway 2-lane and 4-lane road segments with posted speed limits of 45, 50 and 55 mph.

2.3 Methodology

2.3.1 Data Preparation

The Illinois Highway Safety Database (IHSD) was selected for this study. Illinois is one member of the Highway Safety Information System (HSIS), which is a multistate database that contains crash, roadway inventory, and traffic volume data for a select group of States. The HSIS dataset contains the subset of 1985-2007 accidents that occurred on the Illinois State-inventoried system. The Illinois transportation database includes Accident Data, Roadlog File, Bridge (Structures) File, Railroad(RR) Grade Crossing File (Council et al. 2009). The IHSD Accident Data and Roadlog File from 2003 to 2007 were extracted and linked to each other using three common variables: county, route, and milepost.

The combined dataset was further cleaned according to the following criteria:

- Freeway segments were excluded since the use of curbs on freeways is limited to special conditions in Illinois and specific features of curbs are required as well (IDOT, 2001);

- Intersection related segments were excluded as the traffic patterns are different from continuous segments;
- Speed limits of 45, 50 and 55 mph were selected since this study is focusing on high-speed roadways;
- The road segments with 2 lanes and 4 lanes were selected since they are quite common in sampled area;
- Crash rates (Equation 2.1) greater than 200 per million vehicle miles travel (MVMT) were excluded as outliers, because these observations have very low frequency and extraordinarily large variance (26124.89). Besides, they either had low AADT or short segment length according to the database, which may introduce AADT data collection errors or unidentified factors into the regression analysis.

  $$CR = CF / MVMT = \frac{CF \times 10^6}{AADT \times 365 \times l} \tag{2.1}$$

  where CR is crash rate, CF is crash frequency, l is segment length.
- Segment length of 2.74 mi was removed as an outlier since it is extraordinarily large in sampled segments;

The cleaned dataset applied in this part of research includes 48,831 total crashes occurring on 3,553.32 miles of 36,132 homogeneous segments during the 5 years from 2003 to 2007. The number of crashes occurring on each of these segments ranges from 0 to 128, with the average of 1.35 and the standard deviation of 3.64. Table 2.1 summarizes the descriptive statistics for the explanatory variables employed in this study. The outside

Table 2.1 Summary statistics of variables

Description of categorical variables					
Variables	label	Categories		Frequency	Percent (%)
Spd_limt	Speed limit	45	45 mph	4922	13.62
		50	50 mph	2235	6.19
		55	55 mph	28975	80.19
No_lanes	Number of lanes	2	2 lanes	31530	87.26
		4	4 lanes	4602	12.74
Outshtp	Outside shoulder type	Outshd_curb	Curbed outside shoulder	1807	5.00
		Outshd_soft	Soft flush outside shoulder	8684	24.04
		Outshd_hard	Hard flush outside shoulder	25641	70.96
Medtype	Median type	Med_raise	Raised median	2668	7.38
		Med_flush	Flush median	2772	7.67
		Med_none	No median	30692	84.95
Rururb	Rural or Urban	1	RURAL	27632	76.48
		2	URBAN	8500	23.52

Description of continuous variables						
Variables	Label	Min	Max	Mean	Median	Standard deviation
ThAADT	Thousand of AADT	0.10	49.37	6.14	3.47	7.30
Seg_lng	Segment length (mi)	0.01	1.88	0.09	0.06	0.13
Lanewid	Lane width (ft)	8.00	30.00	11.89	12.00	1.07
Outshwd	Outside shoulder width (ft)	0.00	12.00	5.22	4.00	3.28
Medwid	Median width (ft)	0.00	99.00	2.81	0.00	8.96

shoulder type specified in this part of research is restricted to the shoulders adjacent to both directions of the travel lanes. Thousand of AADT (ThAADT) is employed in the regression modeling as an exposure variable, which is calculated through dividing AADT by 1000, since the change of crash frequency with increment by one vehicle is meaningless.

The outside shoulder types and median types were re-categorized into new classes from the original records according to the following rules:

Median types:

- "No median" is categorized into the class of segments without medians, symbolized as Med_none;
- "Unprotected-sodded, treated earth, or gravel surface", "Painted median" and "Rumble strip or chatter bar" are categorized into the class of segments with flush medians, symbolized as Med_flush;
- "Curbed-raised median, any width" and "Positive barrier-fencing, guardrail, retaining walls, or other barriers" are categorized into the class of segments with raised medians, symbolized as Med_raise;

Outside shoulder types:

- "Earth" and "Sod" are categorized into soft flush outside shoulder type, symbolized as Outshd_soft;
- "Curb and gutter" is categorized into curbed outside shoulder type, symbolized as Outshd_curb;

- “Aggregate”, “Surface treated”, “Bituminous” and “concrete-tied” are categorized into hard flush outside shoulder type, symbolized as Outshd_hard.

2.3.2 Crash Rate Comparison

The crash rate distribution histogram (Figure 2.1) clearly illustrates that the distribution of crash rate is extremely non-normal. The normality test shows that the skewness is 4.45 and the kurtosis is 27.68, which strongly rejects the null hypothesis that the crash rate follows the normal distribution.

To compare crash rates between different outside shoulder types in related to various speed limit levels, Wilcoxon/ Kruskal-Wallis tests were conducted. Wilcoxon Rank Sum test is a non-parametric test based solely on the order of each observation from two samples, and the Kruskal–Wallis test is the extension of Wilcoxon Rank Sum test used for more than two groups’ comparison.

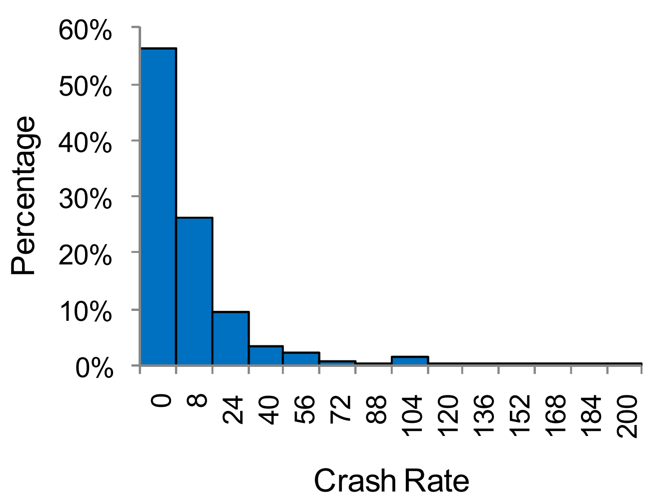


Figure 2.1 The Distribution of Crash Rate

To apply this test, independent random samples of sizes $n_1, n_2, \dots, n_k$ from k populations are obtained in the first place, and then the data from all these samples is arranged in ascending order in a single series (in the case of repeated value, assign ranks to them by averaging their rank position). Once this is completed, ranks of the different samples are separated and summed up. Assume that there are n observations in total, let R_i be the sum of the ranks of the n_i observations in the ith sample. The Kruskal-Wallis statistic can be expressed as:

$$H = \frac{12}{n(n+1)} \sum_{i=1}^{k} \frac{R_i^2}{n_i} - 3(n+1) \qquad (2.2)$$

where H is the Kruskal-Wallis Test statistic. When the sample sizes are large and all k populations have the same continuous distribution, then H has an approximate chi-square distribution with $k-1$ degrees of freedom. If the calculated value of Kruskal-Wallis Test is less than the chi-square table value, then the null hypothesis will be accepted. If the calculated value of Kruskal-Wallis Test is greater than the chi-square table value, then we will reject the null hypothesis and say that these samples come from different populations. For this research, the null hypothesis is that segments of different shoulder types or speed limits are from identical populations in terms of the distribution of crash rates, and the alternative hypothesis is that they are from different populations, only with respect to the distribution of crash rates.

In order to identify underlying associations between variables in the Kruskal-Wallis test, the log-linear model was employed. The log-linear model computes the means of cell counts in contingency tables by describing the association patterns among a set of categorical variables without specifying any variable as a response variable.

To investigate associations between categorical variables, the saturated log-linear model is first established, where the main effects and all interactions are included. Then insignificant interactions are removed in descending order, and a new model will be fitted in each step. The new model is deemed as reasonably well if the p value of the likelihood ratio test is greater than 0.05, and all insignificant interactions are excluded.

In this part of research, associations between speed limit, outside shoulder type, and rural or urban, number of lanes, median type was investigated. To our point of interest, only the main effects and two-way interaction in related to speed limit and outside shoulder type were analyzed in the model. the higher order interaction may exist but will not affect the judgement. All the variables of significant interaction ($p<0.05$) in this model can be considered as highly related.

2.3.3 Crash Frequency Regression Analysis

Poisson model is known as the simplest and most common model for count data regression analysis, which has been widely used in crash frequency modeling studies (Jovanis et al. 1990, Joshua et al. 1993, Miaou et al. 1993). It assumes that the observed count data y_i ,given the vector of covariate x_i, follows a Poisson distribution. The density function of y_i can be expressed as Equation 2.3.

$$P\left(Y_i = y_i \middle| x_i\right) = \frac{e^{-\mu_i} \mu_i^{y_i}}{y_i!}, \quad y_i = 0,1,2,... \tag{2.3}$$

where, the parameter u_i, conditional mean number of events for each covariate x_i, is given by Equation 2.4.

$$\mu_i = Exp(x'_i \, \beta) \tag{2.4}$$

where β is a $(k+1)\times 1$ parameter vector (β_0 is the coefficient for intercept, and β_1, β_1, $\beta_2 \ldots \beta_k$ are for k regressors).

In the Poisson regression, the conditional variance of the count variable is equal to the conditional mean, as shown in Equation 2.5.

$$V\left(y_i \middle| x_i\right) = E\left(y_i \middle| x_i\right) = \mu_i \tag{2.5}$$

However, this assumption is contradict to the fact that the vehicle accident data are always significantly overdispersed relative to its mean. In this situation, the common Poisson regression model is inappropriate as it can result in bias and inconsistent of parameter estimates. Therefore, the more general probability distribution like Negative Binomial (NB) regression became popular in modeling crash count data in the last few years, since it can address the issue of overdispersion (Milton et al. 1998, Shankar et al. 1995, Carson et al. 2001).

The general form of NB regression model can be expressed as Equation 2.6. There are many variants of NB, while in this part of research, we will focus on the negative binomial model with quadratic variance function (p=2).

$$u_i = Exp(x_i\beta + \varepsilon_i) = Exp(x_i\beta)Exp(\varepsilon_i)\ ,$$

$$Exp(\varepsilon_i) = Gamma(\alpha^{-1}, \alpha^{-1}) \tag{2.6}$$

where, x_i is the covariate of road segment geometric and traffic features in each record including the intercept; u_i is the conditional mean of the crash frequency y_i; β is the parameter coefficients vector to be estimated for each independent variable including intercept; e^{ε_i} is a heterogeneity component accounting for unobserved heterogeneity in the crash count data , which is independent of x_i . It is assumed that e^{ε_i} follows a $Gamma(\alpha^{-1}, \alpha^{-1})$ distribution with $E(e^{\varepsilon_i}) = 1$ and $V(e^{\varepsilon_i}) = \alpha$.

The negative binomial distribution (the density of each crash count) can be derived as:

$$f(y_i|x_i) = \frac{\Gamma(y_i + \alpha^{-1})}{y_i!\Gamma(\alpha^{-1})}\left(\frac{\alpha^{-1}}{\alpha^{-1} + \mu_i}\right)^{\alpha^{-1}}\left(\frac{\mu_i}{\alpha^{-1} + \mu_i}\right)^{y_i},\quad y_i = 0,1,2,... \tag{2.7}$$

However, there is always large density of zeros in crash count data, which can hardly be predicted by traditional NB model. For this situation, zero inflated regression models were developed in the crash frequency related research area (Lee et al. 2002, Lord et al. 2005).

Zero-inflated count models provide a way of modeling the excess zeros in addition to allowing for overdispersion. For each road segment, there are two possible data generation processes. Process 1 is chosen with probability ω_i and process 2 with probability $1-\omega_i$. Process 1 generates only zero counts, whereas process 2 generates counts from either a Poisson or a negative binomial model. In this part of research, the probability ω_i depends on the geometric and traffic features of segment i, can be obtained from the logistic function F, as shown in Equation 2.8.

$$\omega_i = F(z'_i\,\gamma) = \Lambda(z'_i\,\gamma) = \frac{\exp(z'_i\,\gamma)}{1+\exp(z'_i\,\gamma)} \tag{2.8}$$

where z'_i is the vector of independent variables specified in the logistic regression model (geometric and traffic features) and intercept; γ is the vector of zero-inflated coefficients to be estimated.

The probability of crash frequency for segment i can be expressed as Equation 2.9.

$$p(y_i|x_i, z_i) = \begin{cases} \omega_i + (1-\omega_i)g(y_i|x_i), y_i = 0 \\ (1-\omega_i)g(y_i|x_i), y_i > 0 \end{cases} \tag{2.9}$$

where $g(y_i|x_i)$ follows either Poisson distribution or NB distribution, x_i is the vector of covariates of observation i (geometric and traffic features) specified in the model.

In general, the log-likelihood function of zero-inflated models fully incorporates the logistic and Poisson or NB process, which can be expressed as Equation 2.10 (Lambert, 1992; Greene, 1994). Therefore, zero-inflated models jointly estimate the incidence of zero crash along with the positive counts.

$$L = \sum_{i=1}^{n} \ln [p(y_i|x_{i,} z_i)] \tag{2.10}$$

where n is the total number of observations.

In this study, both NB and Zero-inflated Negative Binomial (ZINB) models were used for initial regression efforts. Akaike's Information Criteria (AIC) (Akaike, 1973), Schwarz Bayesian Criteria (SBC) (Sawa, 1978) and Vuong test (Greene, 1994) were employed to identify which one was more appropriate for this crash frequency regression analysis. The model with the smaller AIC or SBC among all competing models is deemed as the better model. The Vuong's test (Vuong, 1989) is an addition to and even outperform AIC and SBC in comparing models from different regression series. Given Equation 2.11,

$$m_i = Log\left(\frac{P_1(y_i|x_i)}{P_2(y_i|x_i)}\right) \tag{2.11}$$

where $p_N(y_i|x_i)$ is the predicted probability of observed crash frequency for segment i from model N. Then, the Vuong statistic is computed to test whether the two models are significantly different in predicting true crash count or not, which can be expressed as Equation 2.12.

$$V = \frac{\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} m_i\right)}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(m_i - \overline{m}\right)^2}} \tag{2.12}$$

where V is the Vuong statistic. If V >1.96, the first model is preferred. If V <-1.96, then the second one is preferred.

2.4 Results

2.4.1 Analysis of Crash Rates

The mean crash rates of different outside shoulder types by various speed limits are shown in Figure 2.2.

It can be observed that, compared to the other two types of outside shoulders without curb, the curbed outside shoulders were associated with higher mean crash rates for the 45 and 50 mph speed limits, while for the 55 mph speed limit there was no apparent difference in the mean crash rates between the three types of shoulders. Furthermore, the

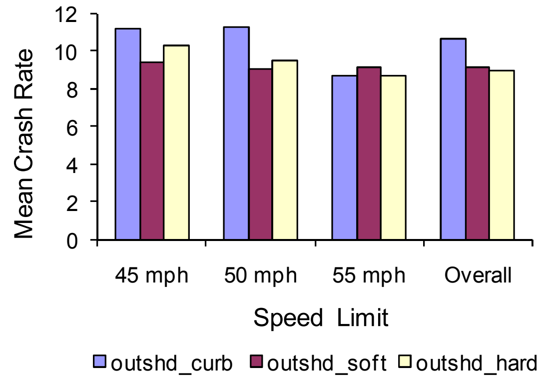


Figure 2.2 Mean crash rates for outside shoulder types by speed limit

55 mph speed limit was involved in a lower mean crash rate than the 45 and 50 mph speed limits.

In order to identify whether the crash rate distributions of road segments with different types of outside shoulders significantly varied or not, a series of Wilcoxon/Kruskal-Wallis tests with a significance level of 0.05 were executed under various speed limit conditions (45mph, 50mph, 55mph), respectively. Table 2.2 is a summary of the test results.

As is shown in Table 2.2, in each of the studied speed limit conditions, as well as the overall speed limit conditions, road segments with curbed outside shoulders always had higher mean Wilcoxon scores of crash rates than those with soft flush outside shoulders and hard flush outside shoulders. The p values of these tests are all less than 0.05, indicating that the segments with various outside shoulder types were from different populations of crash rates.

Furthermore, Wilcoxon/Kruskal–Wallis tests were also conducted on crash rates on road segments with different outside shoulder types by different speed limits. Table 2.3 summarizes the test results in detail.

It was found that for all the studied road segments combined, as well for those with different outside shoulders except the soft flush type, the mean Wilcoxon scores of crash

Table 2.2 Wilcoxon test on crash rates over outside shoulder types by speed limits

Speed limit	Outside shoulder type	N	Mean Score	DF	Chi-square	Pr>chisquare
45	Outshd_curb	1098	2699.45	2	51.32	<.0001
	Outshd_soft	560	2246.48			
	Outshd_hard	3264	2418.34			
50	Outshd_curb	291	1257.38	2	36.35	<.0001
	Outshd_soft	372	970.44			
	Outshd_hard	1572	1127.12			
55	Outshd_curb	418	15579.96	2	10.27	0.0059
	Outshd_soft	7752	14393.85			
	Outshd_hard	20805	14501.14			
Overall	Outshd_curb	1807	21083.85	2	208.44	<0.0001
	Outshd_soft	8684	17567.45			
	Outshd_hard	25641	18022.87			

Table 2.3 Wilcoxon test on crash rates over speed limits by outside shoulder types

Shoulder Type	Speed Limit	N	Mean Score	DF	Chi-square	Pr>chisquare
Outshd_curb	45	1098	934.34	2	28.90	<.0001
	50	291	957.09			
	55	418	787.34			
Outshd_soft	45	560	4488.93	2	2.66	0.2648
	50	372	4315.34			
	55	7752	4333.23			
Outshd_hard	45	3264	13876.71	2	188.25	<.0001
	50	1572	14267.45			
	55	20805	12546.08			
Overall	45	4922	19813.10	2	313.28	<0.0001
	50	2235	19896.06			
	55	28975	17628.68			

rates were always lowest for the speed limit of 55 mph, with the p value less than 0.05, indicating that the crash rates were from different populations of crash rates. This result in some ways supports the findings of Baek et al. (2006): “when there was a difference in collision rate or severity, it was usually the 55 mph segments that had the lowest rates.”

Through the Wilcoxon/Kruskal–Wallis tests over the distribution of crash rates, it can be concluded that segments with curbed outside shoulders and the other two types of outside shoulders without curb were from different populations, as were the segments with different speed limits.

However, the likelihood ratio statistics of the log-linear model (Table 2.4) show that there are significantly strong associations ($p<0.0001$) between outside shoulder types, speed limits and number of lanes, rural or urban, as well as median type. Hence, the results of the Wilcoxon/Kruskal-Wallis tests may subject to the underlying association among these geometric and roadway features, and even other factors not identified in this part of research. Bearing this concern in mind, a more sufficient analysis method is needed to precisely identify the safety effects of curbed outside shoulders and speed limits change.

2.4.2 Analysis of Crash Rates

Both NB and ZINB models were fitted with SAS *Countreg Procedure*. For the NB model and the NB regression part of the ZINB model, crash frequency (Tot_n) was specified as the dependent variable. The regressors included segment length (Seg_lng), thousand average annual day traffic (ThAADT), lane width (Lanewid), median width (Medwid),

Table 2.4 Likelihood ratio statistics for Type 3 analysis

Source	DF	Chi-Square	Pr > ChiSq
Spd_limt	2	810.31	<.0001
Outshtp	2	7990.57	<.0001
Rururb	1	54.43	<.0001
No_lanes	1	1474.58	<.0001
Spd_limt*Rururb	2	7088.94	<.0001
Spd_limt*No_lanes	2	1512.05	<.0001
Spd_limt*Medtype	4	1369.26	<.0001
Spd_limt*Outshtp	4	114.47	<.0001
Outshtp*No_lanes	2	2964.42	<.0001
Outshtp*Rururb	2	2028.09	<.0001
Outshtp*Medtype	4	1981.57	<.0001

outside shoulder width (Outshwd), speed limit (Spd_limt), number of lanes (No_lanes), rural or urban (Rururb), median type (Medtype), outside shoulder type (Outshtp). In the final version of logistic regression part of the ZINB model, only variables that are significant ($p<0.05$) were included to estimate the likelihood of zero crash frequency.

The performance of each model is shown in Table 2.5. Both models have 36132 numbers of observations, and the Newton-Raphson optimization method was employed. The result shows that both the AIC and SBC values of ZINB model are smaller than that of NB model, indicating that the ZINB model outperformed the NB model. The result of Vuong's Test is 20.50, which is much greater than 1.96, also indicating that the ZINB model is preferred.

In terms of AIC, SBC and Vuong's Test results, the ZINB model appears to bring about a clear improvement in the overall fitness of crash frequency compared to the NB model. Therefore, the ZINB model is selected and further analyzed in the remaining section of this part of research.

The ZINB model parameter estimates have two parts: NB regression and logistic regression. Table 2.6 presents the parameter estimates and the significance tests of the ZINB model. It was found that median types, number of lanes, speed limit, rural or urban, thousand AADT, segment length and lane width were significantly associated with the crash frequency, while outside shoulder type, outside shoulder width and median width were not significantly related to crash frequency. Further, the measure of _Alpha in table

Table 2.5 Model comparison between NB model and ZINB model

Model	Number of observations	Optimization method	Log-likelihood	Maximum Absolute Gradient	AIC	SBC
NegBin	36132	Newton-Raphson	-47271	4.632E-6	94582	94752
ZINB			-46196	3.656E-7	92450	92696
Vuong's Test ZINB:NB= 20.50						

Table 2.6 Parameter estimates of ZINB model

Parameter	DF	Estimate	Standard Error	t Value	P Value
Negative Binomial Regression Part					
Intercept	1	-0.307773	0.113655	-2.71	0.0068
Med_none	0	0	.	.	.
Med_flush	1	-0.145379	0.045709	-3.18	0.0015
Med_raise	1	-0.100170	0.038979	-2.57	0.0102
Nolane_2	0	0	.	.	.
Nolane_4	1	-0.185290	0.039690	-4.67	<.0001
Spdlimt_55	0	0	.	.	.
Spdlimt_50	1	0.159994	0.031707	5.05	<.0001
Spdlimt_45	1	0.151024	0.025880	5.84	<.0001
Rururb_r	0	0	.	.	.
Rururb_u	1	0.324692	0.027500	11.81	<.0001
Outshd_hard	0	0	.	.	.
Outshd_curb	1	0.013165	0.042568	0.31	0.7571
Outshd_soft	1	-0.015626	0.021488	-0.73	0.4671
Thaadt	1	0.076724	0.001640	46.78	<.0001
Seg_lng	1	3.240140	0.059981	54.02	<.0001
Outshwd	1	-0.000583	0.003139	-0.19	0.8528
Medwid	1	-0.000422	0.001380	-0.31	0.7596
Lanewid	1	-0.036865	0.009447	-3.90	<.0001
Logistic Regression Part					
Inf_Intercept	1	2.265432	0.361664	6.26	<.0001
Inf_Thaadt	1	-0.069838	0.006236	-11.20	<.0001
Inf_Seg_lng	1	-49.916395	2.070501	-24.11	<.0001
Inf_Nolane_2	0	0	.	.	.
Inf_Nolane_4	1	0.319464	0.102276	3.12	0.0018
Inf_Rururb_r	0	0	.	.	.
Inf_Rururb_u	1	-0.344662	0.085421	-4.03	<.0001
Inf_Outshwd	1	0.027374	0.008865	3.09	0.0020
_Alpha	1	0.728132	0.015750	46.23	<.0001

2.6 is 0.73, with the p-value less than 0.001, displaying a very strong overdispersion effect, indicating the superiority of the ZINB model over the Zero-inflated Poisson (ZIP) model.

The coefficient estimates of the model parameters in the NB regression part reflect how these independent variables correlated with crash frequency: when the value of independent numerical variable increases, the mean number of expected crashes increases if the coefficient is positive, or decreases if the coefficient is negative.

According to the parameter estimates in the NB regression part of this model, the p values for the variable Outshd_curb (P = 0.7571) and Outshd_soft (P = 0.4671) are much greater than the 0.05 significance level, indicating that there was no significant difference in crash frequency between segments with different types of outside shoulder. In other words, curbed outside shoulder did not show a significantly adverse effect on the occurrence of crashes compared to soft flush and hard flush outside shoulders. In addition, the large p value for variable Outshwd (P = 0.8528) suggests that the increase of outside shoulder width did not show a significant effect on the probability of crashes. However, it was found that the correlation between outside shoulder width and outside shoulder type was as high as 0.5198 in the sampled data. Therefore, the parameter estimate of outside shoulder width is plausible as it may be highly affected by the outside shoulder type.

The modeling results also show that the occurrence of crashes was significantly related to median types, number of lanes, speed limit, rural or urban, thousand AADT, segment length and lane width. The positive coefficients for ThAADT and Seg_lng indicate that the crash frequencies increased as the AADT increased by 1000 vehicles per day, and also increased with each mile increase of segment length. These findings are consistent with many previous research results (e.g. Persaud et al., 1993; Abdel-Aty et al., 2000). It is noteworthy that the effect of segment length is non-linear with the crash frequency (β=3.2401), indicating that the crash frequency is likely to increase by as large as 25.44 times with each one-mile increase of segment length. Since the segment length is limited to 1.88 mi in this study, this coefficient should only be used to address crash occurrences in short segments. This finding is consistent with the previous analysis of Milton and Mannering (1998): this exponential relationship can be explained by the fact that there are too many short segments included in the study, where crash frequency is highly sensitive to the increase of segment length.

Moreover, road segments without median have higher crash frequency than those with either flush medians or raised medians, wherein flush medians are associated with the lowest crash frequency. The p value for Medwid (0.7596) indicates that the increase of median width did not significantly affect the crash frequency. This result does not match most of the previous findings, which suggest that increases of median width may serve to the reduction of crash frequency (Abdel-Aty et al., 2000; Knuiman et al., 1993). However, median width and median types are highly correlated (correlation coefficient is as high as 0.63) in the sampled data. Hence the parameter estimate for median width is to some extent subject to the distribution of median types.

It was also observed that urban regions appeared to have a higher crash frequency than rural regions. This is to be expected since urban regions involve more complex traffic conditions, high traffic volume, congestion, bad pavement conditions and so on. 4-lane roadways were found to be associated with lower number of crashes than 2-lanes in this model. This is reasonable because this comparison was based on the assumption of same traffic exposure so that the segments with 4 lanes should have lower traffic volume per lane. In addition, the increase of lane width is predicted to reduce crash frequency in this model, which is consistent with many other research results (Hadi et al., 1995; Andrew, 1998; Garber et al., 2000)

The coefficient for the speed limit of 45 mph is 0.1510, and for 50 mph is 0.1600, where the speed limit of 55 mph is the reference level. This result indicates that crash frequencies for the speed limits of 45 and 50 mph are significantly higher than that for the speed limit of 55 mph (P values are less than 0.05 for both cases).

The logistic regression part of the model predicts the likelihood of zero crash occurrence. The modeling results reveal that the variables of segment length (Inf_Seg_lng), thousand average annual daily traffic (Inf_ThAADT), number of lanes (Inf_No_lanes), rural or urban (Inf_Rururb), and outside shoulder width (Inf_Outshwd) are significant in estimating the probability of roadways belonging to the zero crash occurrence group. In terms of the parameter coefficients estimated, the higher the traffic exposure (thousand of AADT and segment length), the lower the possibility of zero crash occurrences, and urban regions are associated with a lower probability of zero crash than rural regions. In addition, a larger outside shoulder width and higher number of lanes appear to be positively related to the likelihood of zero crash occurrences.

2.4.3 Regression Analysis for the Segments with Curbed Outside Shoulders

In order to further investigate the safety factors on the road segments with curbed outside shoulders, a ZINB crash frequency regression model was developed only for segments with curbed outside shoulders. The data applied in this analysis includes all 1,801 segments with curbed outside shoulders extracted from the previous cleaned dataset. Consequently, there were overall 8097 crashes in 113.75 miles of segments selected in this section. The number of crashes occurring on each of these segments ranges from 0 to 114 during the 5 study years. The model estimation is shown in Table 2.7.

It can be observed that thousand of AADT, segment length, number of lanes, road locations (rural or urban), speed limit of 55 mph are significant in predicting crash frequency in segments with curbed outside shoulders, whereas only intercept, thousand of AADT and segment length are significant in estimating the probability of zero crashes.

Regarding our point of interest, having a speed limit of 45 mph as the reference level, the negative coefficients of variable Spdlimt_50 (-0.1427) and Spdlimt_55 (-0.3424) indicate that the decrease of speed limit from 55 or 50 to 45 mph are not likely to have any safety benefit and may even possibly increase the crash frequency. In terms of the p values, the comparison between speed limits of 45 and 50 mph is marginally significant (p=0.0781) while the comparison between speed limits of 45 and 55 mph is essentially significant

Table 2.7 Parameter estimates of ZINB model on curbed outside shoulders

Parameter	DF	Estimate	Standard Error	t Value	P Value
Negative Binomial regression part					
Intercept	1	0.427	0.092	4.64	<.0001
Med_none	0	0	.	.	.
Med_flush	1	0.046	0.113	0.41	0.6842
Med_raise	1	0.098	0.102	0.95	0.3408
Nolane_4	0	0	.	.	.
Nolane_2	1	0.294	0.104	2.83	0.0047
Rururb_u	0	0	.	.	.
Rururb_r	1	-0.547	0.150	-3.65	0.0003
Thaadt	1	0.060	0.004	16.90	<.0001
Seg_lng	1	6.447	0.453	14.24	<.0001
Lanewid	1	-0.029	0.026	-1.12	0.2610
Medwid	1	-0.003	0.006	-0.51	0.6104
Spdlimt_45	0	0	.	.	.
Spdlimt_50	1	-0.143	0.081	-1.76	0.0781
Spdlimt_55	1	-0.342	0.085	-4.04	<.0001
Logistic regression part					
Inf_Intercept	1	2.323	0.385	6.04	<.0001
Inf_Thaadt	1	-0.079	0.015	-5.29	<.0001
Inf_seg_lng	1	-100.545	16.739	-6.01	<.0001
_Alpha	1	0.963	0.053	18.02	<.0001

($p<0.001$) . Therefore, the common practice in the U.S., reducing the speed limit from 55 mph down to 45 mph when applying curbs for updating and widening rural two-lane highways into four-lane highways, is not likely to address the curb-related safety concern.

2.5 Conclusions

There have been numerous efforts to investigate crash occurrences as related to roadway design features, environmental conditions, drivers' characteristics and traffic features. However, very few of them have specifically considered the influence of curbs on high-speed roadways. The research presented in this section was primarily motivated by the dilemma of speed selection in roadways with the presences of curbs next to each direction of road lanes.

On the basis of the Illinois Accident Data and Roadlog File from 2003 to 2007, Wilcoxon/Kruskal–Wallis tests were performed to preliminarily identify the distribution of crash rates. Different populations of crash rates were found for road segments with different outside shoulder types, as well as various speed limits. The results, however, may be confouned by underlying associations between outside shoulder type, speed limit, and other geometric or roadway features.

ZINB models were further developed in this part of research to precisely identify the effects of curbed outside shoulders combined with other road variables on the occurrence of crashes in high speed roadways. The results suggest that the presence of curbed outside shoulders in high-speed roads is not likely to result in a higher crash frequency compared to other types of outside shoulders. The decrease of speed limit from 55 to 50 or 45 mph will not have a positive effect on road safety because crash frequencies for the speed limits of 45 and 50 mph are significantly higher than that for the speed limit of 55 mph.

The ZINB models estimated here incorporate the safety effects of several roadway geometric designs and traffic features of interest to traffic and transportation engineers. Based on the ZINB models, the other conclusions that can be drawn include:

- Crash frequency increases as AADT and segment length increase;
- Road segments with 4 lanes have a lower crash frequency than those with 2 lanes;
- Roads in urban regions are associated with a higher crash frequency than those in rural regions;
- Flush median appears to be associated with the lowest crash frequency, followed by raised median. Roads without medians had the highest crash frequency;
- Outside shoulder width and median width are not significantly related to the number of crashes;
- Increases of lane width were found to be related to lower crash frequency;

However, there are other factors of interest not been considered in these models, for example, vertical and horizontal curves, intersections, traffic signs, weather, light, pavement defects and drivers' characteristics. In addition, potential interactions existing among explanatory variables (such as median type and median width, outside shoulder type and outside shoulder width) were ignored in these models, and should be examined in the future endeavors. Moreover, the crash frequency by different severity level (fatality, injury, property damage only, and so forth) across different types of crash (curb involved, run out of road, sideswipe and so forth) are also suggested as topics for the futher research.

To sum up, curbed outside shoulders are acceptable in high-speed (up to 55 mph) roadways in terms of the crash frequency, and lowering the speed limit from 55 mph to 45 mph on high-speed roadways with curbed outside shoulders is not recommended.

References

AASHTO, 2005. A policy on geometric design of highways and streets, Washington, D.C.

Abdel-Aty, M.A., Radwan, A.E., 2000. Modeling traffic accident occurrence and involvement. Accid. Anal. Prev. 32, pp. 633–642.

Akaike, H., 1973. Information theory and an extension of the maximum likelihood principle. In B.N. Petrov and F.Csaki (Eds.), Second international symposium on information theory, pp. 267-281.

Andrew, V., Bared, J., 1998. Accident Models for Two-Lane Rural Segments and Intersections, paper presented at the 77th Annual Meeting of the Transportation Research Board.

Baek, J., Hummer, J., Williams, B.M., Cunningham, C.M., 2006. Reasonable speed limits on suburban multilane highways with curbs. Transport. Res. Record 1969, pp. 10-11.

Baek, J., Hummer, J., 2008. Collision models for multilane highway segments to examine safety of curbs. Transport. Res. Record 2083, pp. 128-136.

Bligh, R.P., Mak, K.K., 1999. Crashworthiness of roadside features across vehicle platforms. Transport. Res. Record 1690, pp. 68–77.

Carson, J., Mannering. F., 2001. The effect of ice warning signs on accident frequencies and severities. Accident Anal. Prev. 33(1), pp. 99-109.

Council, F.M., Mohamedshah, Y.M., 2009. Guidebook for the Illinois state datafiles, FHWA, U.S. Department of Transportation.

Garber, N. J., Ehrhart, A. A., 2000. Effect of Speed, Flow, and Geometric Characteristics on Crash Frequency for Two-Lane Highways. Transport. Res. Record 1717, pp. 76–83.

Greene, W., 1994. Accounting for Excess Zeros and Sample Selection in Poisson and Negative Binomial Regression Models, Working Paper, Department of Economics, New York University.

Hadi, M.A., Aruldhas, J., Chow, L.F., Wattleworth, J.A., 1995. Estimating safety effects of cross-section design for various highway types using negative binomial regression. Transport. Res. Record 1550, pp. 169-176.

IDOT, 2010, Illinois Department of Transportation, Bureau of design and environment manual-cross section elements 32-2.04.

Jovanis, P.P., Chang, H.L., 1986. Modeling the relationship of accidents to miles traveled. Transport. Res. Record 1068, pp. 42-51.

Joshua, S.C., Garber, N.J., 1990. Estimating truck accident rate and involvements using linear and Poisson regression models. Transportation Planning and Technology 15(1), pp. 41-58.

Knuiman, M.W., Council, F.M., Reinfurt, D.W., 1993. Association of median width and highway accident rates. Transp. Res. Rec. 1401, pp. 70-82.

Lambert, D.,1992. Zero-inflated Poisson regression, with an application to defects in manufacturing. Technometrics. 34(1), pp.1-14.

Lee, J., Mannering, F., 2002. Impact of roadside features on the frequency and severity of run-off-roadway accidents: an empirical analysis. Accident Anal. Prev. 34(2), pp. 149-161.

Lord, D., Washington, S.P., Ivan, J.N., 2005. Poisson, Poisson-gamma and zero inflated regression models of motor vehicle crashes: balancing statistical fit and theory. Accident Anal. Prev. 37(1), pp. 35-46.

Miaou, S.P., Lum, H., 1993. Modeling vehicle accidents and highway geometric design relationships. Accident Anal. Prev. 25 (6), pp. 689–709.

Milton, J., Mannering, F., 1998. The relationship among highway geometrics, traffic-related elements and motor-vehicle accident frequencies. Transportation 25, pp. 395–413.

Persaud, B., Dzbik, L. 1993. Accident prediction models for freeways. Transp. Res. Rec. 1401, pp. 55–60.

Plaxico, C.A., Ray, M.H., Weir, J.A., Orengo, F., Tiso, P., 2005. Recommended guidelines for curb and curb-barrier installations. Transport. Res. Board, NCHRP Report 537.

Qin, X., Ivan, J.N., Ravishanker, N., 2004. Selecting exposure measures in crash rate prediction for two-lane highway segments. Accident Anal. Prev. 36, pp. 183–191.

Sawalha, Z., Sayed, T., 2001. Evaluating safety of urban arterial roadways. J. Transp. Eng. 127 (2), pp. 151–158.

Shankar, V., Mannering, F., Barfield, W., 1995. Effect of roadway geometrics and environmental factors on rural accident frequencies, Accident Anal. Prev. 27(3), pp. 371-389.

Sawa, T., 1978. Information criteria for discriminating among alternative regression models. Econometrica 46, pp. 1273-1282.

Vuong, Q.H., 1989. Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses, Econometrica 57(2), pp. 307-333.

PART 3 INVESTIGATION THE INFLUENCE OF CURBS ON SINGLE-VEHICLE CRASH INJURY SEVERITY UTILIZING ZERO-INFLATED ORDERED PROBIT MODELS

3.1 Abstract

The severity of traffic-related injuries has been studied by many researchers in recent decades. However, previous research has seldom accounted for the effects of curbed outside shoulders on traffic-related injury severity. This study introduces the Zero-inflated Ordered Probit (ZIOP) model to evaluate the influences of curbed outside shoulders, speed limit change, as well as other traditional factors on the injury severity of single-vehicle crashes. Crash data from 2003 to 2007 in the Illinois Highway Safety Database were employed in this study.

The ZIOP model assumes that injury severity comes from two distinct sources: injury propensity and injury severity when this crash falls into the *injury prone* category. The modeling results show that on one hand, single-vehicle crashes that occurring on roadways with curbed outside shoulders are more likely to be *injury prone.* On the other hand, the existence of a curb decreases the likelihood of severe injury if the crash was in the *injury prone* category. As a result, the marginal effect analysis implies that the presence of curbs is associated with a higher likelihood of no injury and minor injury involved crashes, but a lower likelihood of incapacitating injury and fatality involved crashes. In addition, in the presence of curbed outside shoulders, the increase of speed limit from 45 to 55 mph is found to add only a small impact to the injury severity of single-vehicle crashes.

Moreover, the modeling results also highlight some interesting effects caused by vehicle type, light and weather conditions, and drivers' characteristics, as well as crash type and

location. Through a comprehensive evaluation of the modeling results, the authors found that the ZIOP model significantly outperforms the tradition Ordered Probit (OP) model, and can serve as a superior alternative in future studies of crash injury severity.

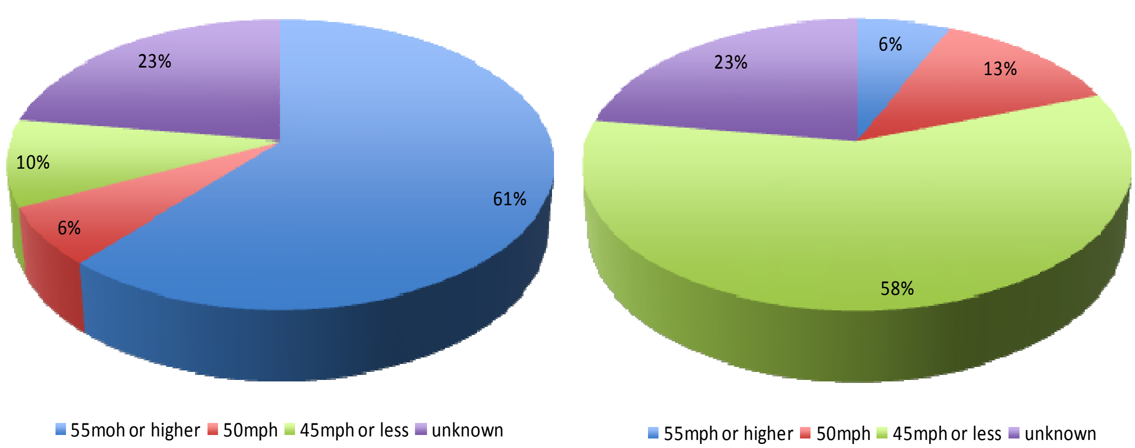
3.2 Introduction

Because of the suburbanization of city peripheries, the traffic volume of the original two-lane rural highways, typically with 55 mph speed limits, may exceed their capacities. To enhance driving quality and reduce traffic delay, many two-lane rural highways need to be widened to four-lanes to mitigate traffic pressure. In these highway improvement projects, curbs frequently are installed along the new four-lane highways to provide drainage and address issues such as access control, difficult terrain, and limited right-of-way.

According to AASHTO, curbs are used extensively on all types of urban highways but should be used cautiously on high-speed rural highways because they may cause vehicles to lose control or go airborne during impact (AASHTO, 2005).Therefore, the use of curbs is discouraged in many states in the U.S. on roadways with design speeds over 45 mph. When they are necessary for drainage control or other special functions, only certain types of curbs are permitted and must be used with specified placement requirements.

The authors conducted a survey of curb usage across the U.S. Each state's road design manuals and relevant guidelines, if any, were reviewed. The survey results show that 31of 50 states have accessible regulations pertaining to the use of curbs. These guidelines reveal the effectiveness and performance of curbs and corresponding regulations on their use in high-speed highways. The regulations vary among urban and rural highways in different states. Figure 3.1 shows the percentage of states that discourage the use of curbs on highways exceeding specific speeds. The chart on the left depicts urban areas while the right-hand chart shows rural areas.

It can be observed from Figure 3.1 that more than 60% of the US states normally use curbs on urban highways even at speed limits of 55 mph or above, while the corresponding figure is only 6% on rural highways. On the other hand, the percentage of states that discourage the use of curbs at the speed limit of 45 mph or less is 6% on urban highways, but more than 50% on rural highways. These results reveal to us that curbs are mostly discouraged on rural highways, especially at speed limits of 55 mph or above. According to the guidelines of many DOTs, the typical unfavorable effects that curbs produced are: the trend to influence drivers' behavior, which may increase the possibility



a. Curb usage in urban areas   b. Curb usage in rural areas

Figure 3.1 States usage of curbs by speed limitation on highways

to crash with the adjacent vehicles; the loss of vehicle control when impacting on curbs, even leading to roll over.

Accordingly, the usage of curbs when updating and widening rural two-lane highways to four-lane highways, frequently results in speed limit reduction from 55 mph to 45 mph. However, lowering speed limits may confuse drivers and create an enforcement problem since many drivers do not perceive the danger and try to travel as fast as they had before the road was improved. As a result, although the highway conditions are significantly enhanced, the projects may be viewed negatively by the public, design professionals, and law enforcement personnel (Baek et al., 2006). Thus, whether to cut speed limits is a dilemma for engineers when updating and widening two-lane rural roadways that have a 55 mph speed limit into four-lane highways with curbs along outside shoulders. The research presented in this part of research was motivated primarily by the inherent dilemma in speed selection for roadways with curbs next to each direction of road lanes.

3.3 Literature Review

There is a growing body of literature addressing the factors that influence the severity of traffic-related injuries. Many of these factors are consistent from single-vehicle crashes to multi-vehicle crashes. For example, drug and alcohol intake has been found to increase significantly the likelihood of severe injury (Baker et al., 2002; Zajac and Ivan, 2003; Keall et al., 2004; Smink et al., 2005);wearing restraint devices normally results in relatively less severe injuries (Valent et al., 2002; Bedard, 2002 ; Abdel-Aty, 2003);

older drivers have a higher risk of severe injury as compared to younger and middle aged drivers (Khattak et al., 1998; Lyman et al., 2002; Finison and Dubrow, 2002); females are more likely to suffer severe injury than males by most measures (Abdel-Aty and Abdelwahab, 2001; Evans, 1988&2001).

However, because single-vehicle and two-vehicle crashes often have different types of collisions, the impact of various risk factors may vary substantially between them. Therefore, a few researchers have specifically investigated factors that affect single-vehicle crash injuries. For example, Renski et al. (1999) explored the effect of speed limit increase on the severity of single-vehicle crash injuries by examining 2,729 single-vehicle crashes on Interstate Highways in North Carolina. The result suggests that increasing speed limits from 55 to 60 mph (88.5 to 96.6 km/h) and from 55 to 65 mph (88.5 to 104.6 km/h) notably increase the likelihood of sustaining minor and non-incapacitating injuries, while increasing speed limits from 65 to 70 mph (104.6 to 112.7 km/h) does not affect crash severity significantly.

Krull et al. (2000) analyzed driver injury severity from single-vehicle rollover crashes using three years of crash data from Michigan and Illinois. Results indicate that driver injury severity increases with drinking, not using seatbelts, and exceeding the posted speed limit. In addition, this research also reported that passenger cars versus pick-up trucks, daylight versus night time, rural roads versus urban roads, dry pavements versus slippery pavements, are more likely to associate with severe injuries.

Schneider et al. (2009) examined driver injury severity for single-vehicle crashes happening along horizontal curves on rural two-lane highways. Their research reveal that injuries are more likely to happen on curves with a moderate radius of between 500 and 2,800 ft. than on sharp, lower-speed curves, or on more gradual, higher-speed curves. Drivers who are older, female, uninsured, and fatigued are more likely to be injured as a result of a crash; the presence of passengers in the vehicle and not using seatbelts greatly increase the chances of injury.

Nevertheless, the effects of curbs on either the occurrence or the outcome of crashes seldom have been studied, particularly on high-speed roadways. A few researchers have investigated the effects of curbs on the occurrences of crashes. For example, Baek and Hummer (2006) conducted research to identify the mean crash rate differences between roadways with speed limits of 45 mph and those with speed limit of 55mph, both with curbs next to the outside lanes in both directions. The results suggest that there is no significant difference on crash rate in both total and curb-involved crashes when speed limit changes. This study was based on the data collected from 60 selected sites in North Carolina. The same authors (2008) studied 2,274 collisions occurring on 191.85 miles of directional multilane highway segments in North Carolina from 2001 to2003. The results indicate that multilane highways with curbed outside shoulders were associated with fewer total collisions and equal injury collisions when compared to no curbs, and segments with speed limit of 55 mph had more collisions than those with speed limits of 45 mph. Jiang et al. (2011) employed the Illinois Accident Data from 2003 to 2007 to analyze the effects of outside shoulder curbs on crash frequency. The results suggest that the presence of curbed outside shoulders on high-speed roads do not result in a significantly higher crash frequency when compared to other types of outside shoulder, and lowering speed limit from 55 mph to 45 mph in the presence of curbs is not necessary to reduce crash frequency.

To study the effects of curbs on injury severity, Plaxico et al. (2005) investigated the extent of the national safety problem related to curbs using FARS and NASS-GES data. The study summarized the statistics of curb-involved single-vehicle crashes in roadways with speed limits at 40 mph (65km/h) or higher. The results indicate that curb-related fatal crashes on roadways with speed limits of 40 mph (65 km/h) and above represent a very small percentage of total fatal crashes. Their successive research used a logistic regression model to compare the severity of all single-vehicle curb crashes and single-vehicle non-curb crashes under similar conditions. The results imply that in locations where curbs are located, single-vehicle crashes involving curbs were clearly no more severe than crashes involving other roadside objects. However, their study derived the effects of curbs only through the involvement of curbs in crashes; it did not account for the possibility that the presence of curbs may be a factor in non-curb-involved crashes. Moreover, the extent to which curbs may be involved in a crash might not be correctly identified in databases, owing to the complexity of crashes.

The objective of this study is to develop statistical models to analyze the effect of curbs along outsider shoulders on the injury levels of single-vehicle crashes. The traffic crash data were acquired from Illinois Highway Safety Database for the years 2003 to 2007.

This research specifically focused on the analysis of crashes that occurred on non-freeway, non-intersection related, two-lane and four-lane road segments with posted speed limits of 45, 50, and 55 mph.

3.4 Methodology

A large variety of statistical models have been used to evaluate the influence of various risk factors on the severity of traffic related injuries (Savolainen et al., 2011). Among them, ordered logit and probit models have been employed frequently because injury levels are inherently ordered (O'Donnell and Connor, 1996; Shimamura and Fujita, 2005). However, traditional Ordered Logit (OL) and Ordered Probit (OP) models constrain the influences of variables to be monotonic, i.e., they will either increase or decrease the likelihood of injury severity level (Washington et al. 2003).This may be inappropriate in that a variable may increase (or decrease) the likelihood of the mid-level injuries but decrease (or increase) the probability of the lowest level and highest level injuries simultaneously. For instance, Savolainen and Ghosh (2008) employed nominal Multinomial Logit model to estimate traffic safety effects. They found that airbag deployments are associated with higher probability of possible and non-incapacitating injuries, but lower chance of property damage only or incapacitating or fatal injury.

In view of this limitation, Wang and Abdel-Aty (2008) and Quddus et al. (2010) employed Generalized Ordered Logit (GOL) model in traffic injury studies. The GOL model, known as partial proportional odds model, allows some of the parameter estimates to vary across severity levels, thereby making the effects of these variables vary accordingly. However, to allow parameter estimates randomly vary across severity levels may result in over-fitting of the models, i.e., the parameter estimates based on one sampled dataset can hardly be applied to other samples. In addition, both traditional OP models and the GOL models have limited capacity in explaining a preponderance of zero observations (crashes with no injury) in traffic crash databases.

To improve on this, the authors introduce the ZIOP model into this study. The ZIOP model assumes that zeros come from two distinct sources. In the first case, Zeros may be recorded for *injury free*, which may depend on a few sound variables that ensure relatively safe crashes such as driving speed, driver's health condition, type of crash and so forth. Alternatively, there may be zeros where the results of the crashes are subject to many other conditions, for example, shoulder type, median type, lane width, etc. Here, the latter source of zeros may turn into injury-involved crashes when injury prone factors were triggered. Thus, it is likely that these two types of zeros are driven by different systems of crash mechanisms. By using the ZIOP model, the influence of each factor can vary across injury levels.

3.4.1 ZIOP Model

The central idea for the traditional OP model is that there is a latent continuous metric underlying the ordinal responses. In this particular research, let $x_i = \{ x_{i1}, \cdots, x_{iq}, \cdots, x_{iQ}\}^T$ represent variables that may affect the injury severity, where i indicates the ith crash, q indicates the qth variable and Q indicates the total number of explanatory variables. The latent variable z_i is assumed to be expressed as:

$$z_i = x_i^T \beta + \varepsilon_i \qquad\qquad i = 1, \cdots, n \tag{3.1}$$

where $\beta = \{\beta_1, \cdots, \beta_q, \cdots, \beta_Q\}$, is a vector of parameters to be estimated and ε_i is a random error term (assumed to follow an independent and identically distributed standard normal distribution). n is the total number of crashes.

Let y_i be a categorical random variable with J+1 categories that represents the injury severity. The observed variable $Y = \{y_1, \cdots y_i \cdots y_n\}$ can be connected to the latent variable $Z = \{z_1, \cdots z_i \cdots z_n\}$ through a function $g(z)$:

$$y_i = g(z_i) = \begin{cases} 0 & if -\infty = \gamma_0 < z_i \leq \gamma_1 \\ 1 & if \;\; \gamma_1 < z_i \leq \gamma_2 \\ \vdots & \vdots \\ J & if \;\; \gamma_J < z_i \leq \gamma_{J+1} = +\infty \end{cases} \tag{3.2}$$

where $\gamma = \{\gamma_0, \gamma_1, \cdots, \gamma_J, \gamma_{J+1}\}$, is the threshold value for all categories, wherein $\gamma_0 = -\infty, \gamma_{J+1} = +\infty$, and the remaining threshold values are subjected to the constraint $\gamma_1 \leq r_2 \leq \cdots \leq \gamma_J$. The function $g(.)$ is taken to be non-decreasing, so that small and large values of z_i can be interpreted as corresponding to small and large values of y_i.

Given the value of x_i, the probability that the injury severity of individual i belongs to each category is

$$\begin{aligned} &P(y_i = 0) = \Phi(\gamma_1 - x_i^T \beta) \\ &P(y_i = 1) = \Phi(\gamma_2 - x_i^T \beta) - \Phi(\gamma_1 - x_i^T \beta) \\ &P(y_i = J) = 1 - \Phi(\gamma_J - x_i^T \beta) \end{aligned} \tag{3.3}$$

where $\Phi(\cdot)$ stands for the cumulative probability function of the standard normal distribution.

To calculate β and γ, one restriction is applied to the threshold values to make γ_1=0 (Mckelvey and Zavoina, 1975). In the classical OP model, the values of β and $\gamma^* = \{\gamma_2, \cdots \gamma_J\}$ can be determined by the ML estimation method. Based on the knowledge in the above text, the likelihood function can be presented as Equation 3.4.

$$L(\beta, \gamma^*|y) = \prod_{i=1}^{n} \prod_{j=0}^{J} [\Phi(\gamma_{j+1} - x_i^T \beta) - \Phi(\gamma_j - x_i^T \beta)]^{I(y_i=j)} \quad (3.4)$$

where $I(y_i = j)$ is an indicator function: if $y_i = j$, then $I(y_i = j)$ is 1, otherwise 0. The parameters β and γcan be determined by maximizing $L(\beta, \gamma^*|y)$.

However, traditional Ordered Probit (OP) models are not able to account for across category effects of variables, and have limit capacity in explaining preponderance of zeros. To solve this drawback, the ZIOP model was first developed by Zhao and Harris (2004) to estimate different categories of tobacco consumers.

In contrast to the traditional OP model, the ZIOP model involves two latent equations: a binary probit selection equation and an OP equation. In other words, for each crash observation, there are two possible data prediction processes: *Process 1* predicts only zero as opposed to non-zero categories, denoted by *injury free* and *injury prone*, whereas *Process 2* predicts all categories of severity from the standard OP model.

Let s denote a binary variable indicating *injury free* ($s = 0$) and *injury prone* ($s = 1$). s is related to a latent variable γ^* through the criteria: $s = 0$ for $\gamma^*<0$ and $s = 1$ for $\gamma^*>0$. Similar to Equation 3.1, the latent variable γ^* represents the propensity of injury involvement and is given by

$$r_i^* = x_i^T \beta + \varepsilon_i \qquad\qquad i = 1, \cdots, n \tag{3.5}$$

where $x_i = \{ x_{i1}, \cdots, x_{ic}, \cdots, x_{iC}\}^{\mathrm{T}}$ represents covariates in identifying injury propensity, and $\beta = \{\beta_1, \cdots, \beta_c, \cdots, \beta_C\}$ is the corresponding vector of parameters to be estimated. Accordingly, the probability of a crash being in the *injury prone* category is given by (Maddala, 1983)

$$P(s = 1|x) = P(r^* > 0|x) = \Phi(x_i^T \beta) \tag{3.6}$$

Conditional on $s = 1$, the observed injury level $\tilde{y} = \{\tilde{y}_1, \cdots \tilde{y}_i \cdots \tilde{y}_n\}$ can be connected to a latent variable y^* through a function $g(w)$:

$$y_i^* = g(w_i) = w_i^T h + \mu_i \tag{3.7}$$

where $w_i = \{ w_{i1}, \cdots, w_{id}, \cdots, w_{iD}\}^{\mathrm{T}}$ represents explanatory variables in the second process of the ZIOP model; $h = \{h_1, \cdots, h_d, \cdots, h_D\}$ is associated vector of coefficients to be estimated; μ_i is a random error term (assumed to follow an independent and identically distributed standard normal distribution). The mapping between $\tilde{y}_i$ and y_i^* is obtained by

$$\tilde{y}_i = \begin{cases} 0 & if -\infty = \gamma_0 < y_i^* \le \gamma_1 \\ 1 & if \;\; \gamma_1 < y_i^* \le \gamma_2 \\ \vdots & \vdots \\ J & if \;\; \gamma_J < y_i^* \le \gamma_{J+1} = +\infty \end{cases} \tag{3.8}$$

The meaning of the parameter γ_j and the function $g(.)$ is the same to that in Equation 3.2.

In the current work, the response $\tilde{y}$ has four categories: no injury, minor injury, incapacitating injury, and fatality. In addition, the authors assume throughout this part of research that $\gamma_1 = 0$. Note that, importantly, *Process 2* also allows for zero injury level, i.e., *no injury.* Also, there is no requirement that w_i=x_i so that separate explanatory factors can be used in both equations. Under the assumption that μ is standard Gaussian, *Process 2* creates the probability of each injury level as follows, conditional on $s = 1$:

$$\begin{gathered} P(\tilde{y}_i = 0) = \Phi(-w_i^T h) \\ P(\tilde{y}_i = 1) = \Phi(\gamma_2 - w_i^T h) - \Phi(-w_i^T h) \\ P(\tilde{y}_i = J) = 1 - \Phi(\gamma_J - w_i^T h) \end{gathered} \tag{3.9}$$

While s and $\tilde{y}$ are not individually observable in terms of the zeros, they are observable via the criterion

$$y = s \cdot \tilde{y} \tag{3.10}$$

That is, to observe a $y = 0$ outcome, we require either that $s = 0$ (*injury free*) or jointly that $s = 1$ and $\tilde{y}$ =0 (*injury prone* but no injury involved). To observe a positive y, we require jointly that the crash is *injury prone* ($s = 1$) and that the crash happened to be injury involved, $\tilde{y} > 0$. Under the assumption that ε and μ identically and independently follow standard Gaussian distributions, the full probabilities for observed y are given by

$$\mathrm{P}(y) = \begin{cases} P(y=0|W,X) = P(s=0|X) + P(s=1|X)\,P\,(\tilde{y}=0|W,s=1) \\ P(y=j|W,X) = P(s=1|X)\,P(\tilde{y}=j|W,s=1) \qquad (j=1,\ldots,J) \end{cases}$$

$$= \begin{cases} \mathrm{P(y=0|W,X) = [1-\Phi(X'\beta)] + \Phi(X'\beta)\Phi(-W'h)} \\ \mathrm{P(y=j|W,X) = \Phi(X'\beta)\big[\Phi\big(u_{j+1}-W'h\big)\big] - \Phi(u_j - W'h)} \quad (j=1,\ldots,J-1) \\ \mathrm{P(y=J|W,X) = \Phi(X'\beta)[1-\Phi\big(u_J - W'h\big)} \end{cases} \tag{3.11}$$

It can be observed from this equation that the probability of a zero observation has been "inflated". The probability of overall no injury is a combination of the probability of *injury free* crash from the binary Probit process plus the probability of no injury observation conditional on *injury prone* category from the OP process.

The parameters of the full model $\theta = (\beta', h', \gamma')'$ can be consistently and efficiently estimated using maximum likelihood (ML) criteria. The log-likelihood function is given by

$$l(\theta) = \sum_{i=1}^{n} \sum_{j=0}^{J} l_{ij} \ln\,[\mathrm{P}(y_i = j|x_i, w_i, \theta)] \tag{3.12}$$

Where the indicator function l_{ij} is:

$$l_{ij} = \begin{cases} 1 & \text{if individual i chooses outcome j} \\ 0 & \text{otherwise} \end{cases} \qquad (i=1,\ldots,n; j=0,1,\ldots,J)$$

The overall structure of the ZIOP is sketched in Figure 3.2. The unconditional probability of positive injury is also a combination of the probability of being *injury prone* and the conditional probability of each injury level.

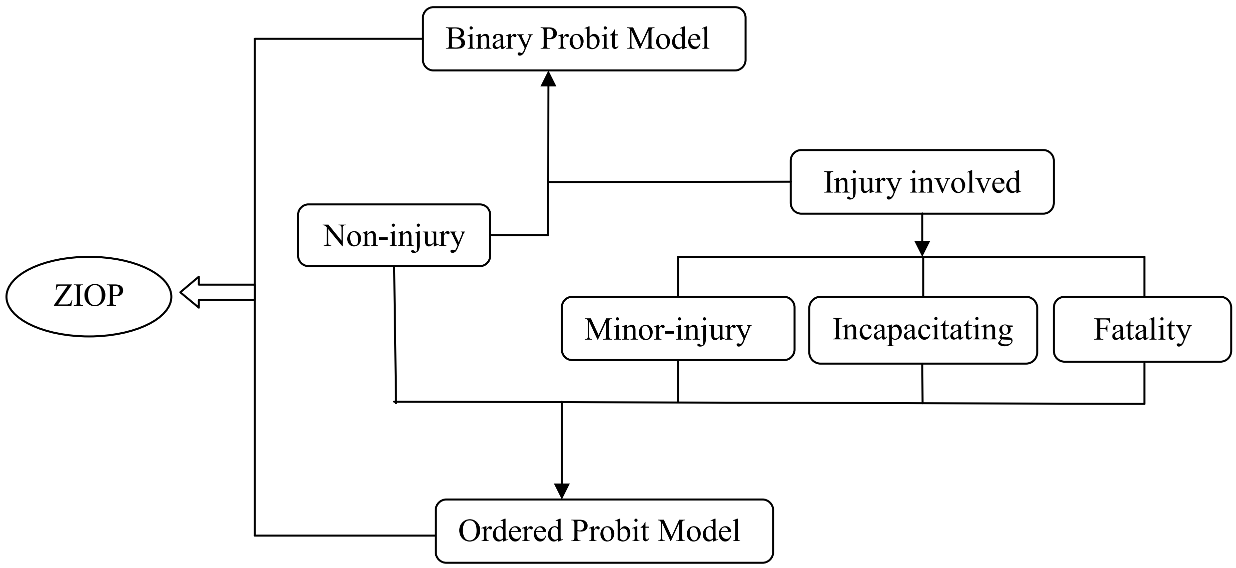


Figure 3.2 Sketch of the ZIOP model

3.4.2 Marginal Effects

In order to evaluate the amount of change in the probability of certain injury severity with the change of each factor, marginal effects were computed. In the case under study, people may be interested in the marginal effects of each variable on the probability of *injury free*, $P(s = 0)$, or on the probabilities of certain injury level given the crash is *injury prone*, $P(\tilde{y} = j|s = 1)$, or on the overall probabilities of each injury level, $P(y = j)$.

The marginal effect of a dummy variable is defined as the difference in the predicted probability of each injury level with this dummy variable assigned to 0 and 1 in the predicted models, and holding all other covariates fixed. For example, the marginal effect of curbed outside shoulder type on minor injury, can be calculated from Equation 3.13. The marginal effect of a continuous variable g can be calculated by $\partial P(.)/\partial g$.

$$\mathrm{ME}[\mathrm{P}(\mathrm{y} = 1)] = \frac{1}{n}\sum_{i=1}^{n}[P(y_i = 1|x_{i1,} \cdots, curb = 1, \cdots, x_{iC,} w_{i1}, \cdots, curb = 1, \cdots, w_{iD}) - P(y_i = 1|x_{i1,} \cdots, curb = 0, \cdots, x_{iC,} w_{i1}, \cdots, curb = 0, \cdots, w_{iD})] \tag{3.13}$$

Following the work of Harris and Zhao (2007), to derive the marginal effects on the overall probabilities for the general ZIOP model, the explanatory variables and the associated coefficients can be reorganized as

$$X = \binom{m}{\tilde{x}},\ \beta = \binom{\beta_m}{\tilde{\beta}},\ W = \binom{m}{\tilde{w}},\ h = \binom{h_m}{\tilde{h}} \tag{3.14}$$

where m represents the mutual explanatory variables that appear in both X and W; β_m and h_m denote the corresponding coefficients of m obtained from the first process and the second process, respectively; $\tilde{x}$ and $\tilde{w}$ represent distinct explanatory variables that only appear in one of the two processes; $\tilde{\beta}$ and $\tilde{h}$ are corresponding coefficients for $\tilde{x}$ and $\tilde{w}$, obtained from these two processes.

Let $X^* = (m', \tilde{x}', \tilde{w}')'$ represents the vector of overall explanatory variables, with associated coefficient vectors $\beta^* = \left(\beta'_m, \tilde{\beta}', 0'\right)'$ and $h^* = \left(h'_m, 0', \tilde{h}'\right)'$. The marginal effects of the explanatory variable vector X^* on the probability of *injury free* and *injury prone* can be expressed as

$$
\begin{aligned}
ME\{P(s=0|X)\} &= \tfrac{\partial P(s=0)}{\partial x_c} = \varphi(X'\beta)\beta_c \\
ME\{P(s=1|X)\} &= \tfrac{\partial P(s=1)}{\partial x_c} = -\varphi(X'\beta)\beta_c
\end{aligned} \tag{3.15}
$$

where $\varphi(.)$is the probability density function (*p.d.f.)* of the standard univariate normal distribution.

The marginal effects of variables on $P(y=j)$ given that the crashes are *injury prone* can be obtained following the same algorithm as illustrated above. Therefore, the overall marginal effects of variables on each injury severity level can be expressed as:

$$
\begin{aligned}
\mathrm{ME}[\mathrm{P}(\mathrm{y}=0)] &= [-\Phi(\mathrm{w}'\mathrm{h})-1]\varphi(\mathrm{X}'\beta)\beta^{*}-\Phi(\mathrm{X}'\beta)\varphi(\mathrm{w}'\mathrm{h})\mathrm{h}^{*},\\
\mathrm{ME}[\mathrm{P}(\mathrm{y}=1)] &= [-\Phi(\mathrm{w}'\mathrm{h})-1]\varphi(\mathrm{X}'\beta)\beta^{*}-\Phi(\mathrm{X}'\beta)\varphi(\mathrm{w}'\mathrm{h})\mathrm{h}^{*}\\
&\quad+[\varphi(\mathrm{w}'\mathrm{h})\Phi(\mathrm{X}'\beta)-\varphi(\gamma_2-\mathrm{w}'\mathrm{h})\Phi(\mathrm{X}'\beta)]\mathrm{h}^{*},\\
\mathrm{ME}[\mathrm{P}(\mathrm{y}=2)] &= [\Phi(\gamma_3-\mathrm{w}'\mathrm{h})-\Phi(\gamma_2-\mathrm{w}'\mathrm{h})]\varphi(\mathrm{X}'\beta)\beta^{*}\\
&\quad+[\varphi(\gamma_2-\mathrm{w}'\mathrm{h})\Phi(\mathrm{X}'\beta)-\varphi(\gamma_3-\mathrm{w}'\mathrm{h})\Phi(\mathrm{X}'\beta)]\mathrm{h}^{*},\\
&\ \vdots\\
\mathrm{ME}[\mathrm{P}(\mathrm{y}=\mathrm{J})] &= \Phi(\mathrm{w}'\mathrm{h}-\gamma_J)\varphi(\mathrm{X}'\beta)\beta^{*}+\varphi(\mathrm{w}'\mathrm{h}-\gamma_J)\times\Phi(\mathrm{X}'\beta)\mathrm{h}^{*}
\end{aligned}
\tag{3.16}
$$

3.4.3 Model Evaluation

In this study, the authors also are interested in comparing the fit of the ZIOP model against traditional OP model. Therefore, the traditional OP model also was built with all factors included. Akaike's Information Criteria (AIC) (Akaike, 1973), Bayesian Information Criteria (BIC) (Sawa, 1978), and the Vuong test (Greene, 1994) were employed to compare these two models. The model with the smaller AIC and BIC among all competing models is deemed the better model.

Vuong's test (Vuong, 1989) can outperform AIC in comparing models from different regression series. Given Equation 3.17, define

$$
m_i = Log\left(\frac{P_1(y_i|x_i, z_i)}{P_2(y_i|x_i, z_i)}\right) \tag{3.17}
$$

where $p_N(y_i|x_i, z_i)$ is the predicted probability of observed crash frequency for segment i from model N. Then, the Vuong statistic, V , is computed to test whether the two models are significantly different in predicting true crash count or not, which can be expressed as Equation 3.18.

$$V = \frac{\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} m_i\right)}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(m_i - \overline{m}\right)^2}} \tag{3.18}$$

If V>1.96, the first model is preferred with the second preferred if V < 1.96.

3.5 Data Preparation

The Illinois Highway Safety Database (IHSD) was selected for this study. Illinois is one member of the Highway Safety Information System (HSIS), which is a multistate database that contains crash, roadway inventory, and traffic volume data for a select group of states. The HSIS dataset contains the subset of 1985-2007 crashes that occurred on the Illinois State-inventoried system. The Illinois transportation database includes Accident Data, Roadlog File, Bridge (Structures) File, Railroad (RR) Grade Crossing File (Council et al. 2009). The IHSD Accident Data from 2003 to 2007 and corresponding Roadlog File were extracted and linked to each other using three common variables: county, route, and milepost. There are overall 891,682 crashes extracted from the original database.

Explanatory factors considered in the present study include: outside shoulder type, median type, number of lanes, speed limit, rural or urban, light conditions, weather conditions, peak hour, vehicle type, driver's age, driver's gender, crash type, surface year, first involve location, Annual Average Daily Traffic (AADT), lane width, and outside shoulder width. Here, AADT was calculated through dividing the original AADT by 1000, since the change of injury severity with increment by one vehicle would be too trivial. The outside shoulder type specified in this part of research is restricted to the shoulders adjacent to both directions of the travel lanes. All discrete factors were categorized into subclasses based on the original records as shown in Table 3.1.

The combined dataset was further cleaned according to the following criteria:

- Only single-vehicle crashes are studied in this part of research;
- Freeway segments were excluded since the use of curbs on freeways is limited to special conditions in Illinois and specific features of curbs are required as well (IDOT, 2001);
- Intersection related crashes were excluded as the traffic and collision patterns are different from continuous segments;
- Speed limits of 45, 50 and 55 mph (72, 80 and 88 mph) were selected since this study is focusing on high-speed roadways;
- The road segments with two lanes and four lanes were selected since they are the most common in sampled area;
- Pedestrian, pedal cyclist, motorcycle, large truck, and non-traditional vehicles involved accidents were screened out because they are minor and have different injury mechanisms;
- Crashes with evidence of not using a seatbelt were excluded to minimize the effects of seatbelts on injury severity, hence seatbelt usage will not be taken as an explanatory variable;
- Crashes happening on “potentially standard” horizontal curves (D>=2.5) were removed to eliminate the inherent effect of horizontal curves on injury severity. In

Table 3.1 Variables Classification Criteria

Variables	Description	Categories & Classification Criteria	
outshtp	Type of outside shoulder	Curb	curbed shoulder
		Non-curb	none curb shoulder
medtype	Type of median	Flush	flush median
		None	no median
		Raised	raised median
no_lanes	Number of lanes	2	2 lane
		4	4 lane
spd_limt	Speed limit	45	45 mph (72 km/h)
		50	50 mph (80 km/h)
		55	55 mph (88 km/h)
rururb	rural or urban	Rural	rural area
		Urban	urban area
light	light condition	Daylight	daylight
		Nighttime	dark lighted and unlighted, dawn and dusk
weather	weather condition	Clear	Clear
		Inclement	rainy, foggy, snowy, etc.
phour	peak hour	Yes	6:00-10:00 & 16:00-19:00
		No	10:00-16:00 & 19:00-6:00
vehtype	vehicle type	Pass_car	convertible, hatchback, sedan, etc.
		Light_trk	pickup, SUV, minivan, small bus
drvage	driver's age single-vehicle	Youngers	driver's age in between 16 and 24
		Middle_aged	driver's age in between 25 and 64
		Olders	driver's age ≥ 65
drvsex	driver's gender single-vehicle	Male	driver is male
		Female	driver is female
crashtype	type of crash	Animal	vehicle collide with animals
		Overturn	vehicle overturned in crashes
		Object	vehicle impact with objects
surfyear	year of surface construction	Pre_2000	road surface constructed before 2000
		Post_2000	road surface constructed after 2000
frst_loc	first involvement location	On_pave	crash occurred on the roadway
		Off_pave	crash occurred off the roadway

other words, all the crashes studied in this part of research can be deemed to have happened on standard ($D<2.5$) horizontal curves;

- All the crashes considered have drivers under normal conditions, i.e., the effects of alcohol, drug, fatigue and illness were eliminated.

The treated dataset applied in this part of research includes 46,159 single-vehicle crashes that occurred on state routes and highways within the five years from 2003 to 2007. These crashes occurred on road segments having AADT (in thousands) ranges from 0.025 to 79.70, with an average of 9.01; lane width ranges from 8 to 27 feet with an average of 11.85 feet; outside shoulder width ranges from 0 feet to 16 feet, with an average of 5.25 feet. In this study, the severity level of each crash was determined by the injury level of the worst-injured occupant in all the vehicles involved. Four levels of injuries were identified from the original database: No injury, Minor Injury, Incapacitating injury, and Fatality. The average percentage of each injury level in ascending order is 89.38%, 8.25%, 2.18% and 0.19%, separately. Table 3.2 provides the distribution of injury severity by each explanatory variable.

It can be observed from Table 3.2 that curb was associated with lower percentages of no injury, incapacitating injury and fatality, but a higher percentage of minor injuries in the overall sampled crashes, in comparison with non-curb shoulder type. Among other variables, the authors would like to emphasize a few variables that have shown noticeable difference between categories. For example, night time was associated with a remarkably higher percentage of no injury crashes but with a lower percentage of minor injury

Table 3.2 Injury severity distribution on key variables

Distribution of injury severity by categorical variables									
Variables	Categories	Injury Severity							
		0 (No injury)		1 (Minor injury)		2 (Incapacitating)		3 (Fatality)	
outshtp	Curb	2764	88.50%	296	9.48%	59	1.89%	4	0.13%
	Non_curb	38491	89.44%	3514	8.17%	949	2.21%	82	0.19%
medtype	Flush	4173	89.34%	389	8.33%	104	2.23%	5	0.11%
	None	33286	89.55%	2980	8.02%	829	2.23%	76	0.20%
	Raised	3796	87.93%	441	10.22%	75	1.74%	5	0.12%
no_lanes	2	32741	89.59%	2923	8.00%	814	2.23%	69	0.19%
	4	8514	88.58%	887	9.23%	194	2.02%	17	0.18%
spd_limt	45	5205	87.74%	589	9.93%	125	2.11%	13	0.22%
	50	3366	88.79%	341	8.99%	81	2.14%	3	0.08%
	55	32684	89.70%	2880	7.90%	802	2.20%	70	0.19%
rururb	Rural	27641	90.15%	2280	7.44%	678	2.21%	62	0.20%
	Urban	13614	87.84%	1530	9.87%	330	2.13%	24	0.15%
light	Daylight	10720	81.78%	1827	13.94%	519	3.96%	42	0.32%
	Nighttime	30535	92.39%	1983	6.00%	489	1.48%	44	0.13%
weather	Clear	33385	90.45%	2694	7.30%	770	2.09%	61	0.17%
	Inclement	7870	85.09%	1116	12.07%	238	2.57%	25	0.27%
phour	Yes	18629	89.74%	1640	7.90%	455	2.19%	34	0.16%
	No	22626	89.08%	2170	8.54%	553	2.18%	52	0.20%
vehtype	Pass_car	24976	88.42%	2553	9.04%	665	2.35%	54	0.19%
	Light_trk	16279	90.89%	1257	7.02%	343	1.92%	32	0.18%
drvage	Youngers	9115	83.31%	1457	13.32%	342	3.13%	27	0.25%
	Middle_aged	28941	91.45%	2101	6.64%	565	1.79%	40	0.13%
	Olders	3199	89.58%	252	7.06%	101	2.83%	19	0.53%
drvsex	Male	24117	91.01%	1863	7.03%	465	1.75%	54	0.20%
	Female	17138	87.17%	1947	9.90%	543	2.76%	32	0.16%
crashtype	Animal	30023	96.73%	817	2.63%	194	0.63%	3	0.01%
	Overturn	1078	49.13%	847	38.61%	265	12.08%	4	0.18%
	Object	10154	78.54%	2146	16.60%	549	4.25%	79	0.61%
surfyear	Pre_2000	30089	89.17%	2849	8.44%	746	2.21%	59	0.17%
	Post_2000	11166	89.93%	961	7.74%	262	2.11%	27	0.22%
frst_loc	On_pave	32393	95.73%	1179	3.48%	258	0.76%	8	0.02%
	Off_pave	8862	71.93%	2631	21.35%	750	6.09%	78	0.63%

crashes compared with day light condition; inclement weather conditions were more likely to result in more severe injuries in comparison with clear weather conditions; both younger drivers and older drivers were associated with a relatively high percentage of injury involved crashes; more than half of "overturn" involved crashes resulted in injuries, and the crashes that involved impacts with objects also resulted in a high likelihood of injury (around 23%), while collisions with animals had a much lower injury probability compared with the average proportion; last but not the least, crashes happening off pavement were more likely to result in injuries in comparison with on pavement crashes.

To quantitatively estimate the influences of these variables on the injury severity of single-vehicle crashes, the ZIOP model was built and compared with the traditional OP model. In addition, marginal effect of each variable was calculated, making the modeling results more accessible.

3.6 Results

3.6.1 Parameter Estimates

In this section we review and interpret the parameters resulting from the fit of the ZIOP model. In order to select variables for injury proneness of the accident described by Equation 3.6, a traditional Probit model for binary injury response (no injury and injury involved) was employed. Through the traditional binary Probit model, the authors identified variables with significant effects ($p<0.05$) as the covariates for the first process of the ZIOP model. These variables include: AADT, vehicle type, light conditions, weather conditions, driver's gender, driver's age, outside shoulder type, peak hour, first involved location, and crash type. The second process of the ZIOP model employed all the factors considered in this part of research. Estimates from the ZIOP model are shown in Table 3.3.

Parameter estimates provide the change in crash severity level in comparison to a reference attribute after holding fixed all of the other explanatory variables. Parameter estimates greater than 0 indicate that a particular cluster of a variable is related to a higher severity level, and vice versa.

To the authors' point of interest, the ZIOP model shows significant effects of curbed outside shoulders on traffic-related injury severity in single-vehicle crashes. In the first process, the coefficient for curb is significantly positive (p=0.0073), which means crashes that occurred on roadways with curbed outside shoulders are more likely to be *injury prone.* In the second process, curb is remarkably negative (p=0.0003), which instead indicates that conditional on being *injury prone*, a curb is more likely to be associated with a lower level of injury in comparison to no curb. This makes strong sense: On one hand, curbs are more likely to be associated with "collision with objects" crash types (59.37%) as compared to crashes associated with non-curb roads (25.73%). Collision with objects appears to have higher probability of injury as compared with animal involved crashes as shown in Table 3.3. On the other hand, crashes on roadways with curbs are less likely to be "overturn" crash type (2.08%) as compared to crashes on non-curb roadways (4.95%). Overturn has been proved to be the No.1 killer in both previous studies (Plaxico et al. (2005) and here in Table 3.3. It is worthwhile to clarify that crashes

Table 3.3 Parameter estimates from the ZIOP model

Parameter		DF	Estimate	P Value	Estimate	P Value
			Binary Probit process		Ordered Probit process	
continuous variables	AADT	0	0.0071	0.0933	-0.0053	0.0020
	Lanewid	0	-	-	-0.0828	<0.0001
	Outshwd	0	-	-	-0.0029	0.4489
vehtype	Pass_car					
	Light_trk	1	-0.2075	0.0012	0.0161	0.6044
rururb	Urban					
	Rural	1	-	-	0.0125	0.6761
no_lanes	2					
	4	1	-	-	-0.0013	0.9772
spd_limt	45					
	50	1	-	-	-0.0570	0.1589
	55	1	-	-	0.0197	0.5434
light	Daylight					
	Nighttime	1	-0.3877	<0.0001	0.0293	0.3980
weather	Clear					
	Inclement	1	0.4431	0.0001	-0.3517	<0.0001
drvsex	Female					
	Male	1	-0.0580	0.3598	-0.2114	<0.0001
drvage	Youngers					
	Middle_aged	1	-0.4605	<0.0001	0.0778	0.0292
	Olders	1	-0.8402	<0.0001	0.4421	<0.0001
median	None					
	Flush	1	-	-	0.0119	0.7936
	Raised	1	-	-	-0.0041	0.9275
outshtp	non_curb					
	curb	1	0.5228	0.0073	-0.2344	0.0003
phour	Yes					
	No	1	-0.0645	0.2735	0.0586	0.0384
surfyear	Pre_2000					
	Post_2000	1	-	-	0.0079	0.7329
frst_loc	On_pave					
	Off_pave	1	1.0696	<0.0001	0.1290	0.0341
crashtype	Animal					
	Object	1	0.5457	0.0003	0.4834	<0.0001
	Overturn	1	2.2782	0.0005	0.9675	<0.0001
γ_1	1|2	1	-	-	0.9882	<0.0001
γ_2	2|3	1	-	-	2.0784	<0.0001

on roadways with curbs installed do not necessarily mean they are curb involved. The high proportion of "collision with objects", and low proportion of "overturn" on curbed highways is quite possible. The presences of curbs produces a certain number of vehicle-curb collisions, and increases the chance of other curb-object collisions because drivers have less space to keep away from objects in emergencies. However, drivers may operate at relatively lower speeds on highways with curbs installed, which may compensate for the high risk of "collision with objects" crashes and result in less overturn.

Among other factors, light trucks were associated with a significantly lower likelihood of *injury prone* crashes as compared to passenger cars, but did not show any significant difference on the conditional injury severity. The finding on injury propensity is to some extent consistent with the result reported by Krull et al. (2000), which indicates that injury severity increases with passenger cars as opposed to pick-up trucks. The lower propensity of injury might be attributed to the relatively higher rigidity of a light truck as opposed to a passenger car, or perhaps a difference in the driver population. Meanwhile, light truck vehicles are found to have a relatively higher percentage of "overturn" crashes (5.83%) in the sampled dataset in comparison with passenger cars (4.07%), which may cancel out the positive effect of light trucks and result in insignificant effect on the probability of conditional injury severity. The unstable characteristics of light truck vehicles have been discussed by many scholars. For example, NHTSA (1997) reported that light truck vehicles have a higher rate of rollover crashes as opposed to passenger cars.

First involving location of crashes is significant in both the injury propensity estimating and injury severity predicting processes. Off-pavement crashes were more likely to be *injury prone* and associated with more severe injury conditional on being in the *injury prone* category, as opposed to on-pavement crashes. Nighttime was associated with a significantly lower probability of *injury prone* crashes, but made no difference on the conditional injury severity as compared to day time. The less *injury prone* finding is consistent with many previous studies (Krull et al., 2000; Savolainen and Ghosh, 2008), since drivers are prone to drive more carefully in night time. However, the traffic volume in the night time is lower, which may encourage drivers to maintain higher speeds, and night time visibility is poorer. These may cancel out the positive effect of driver's attention and lead to an insignificant effect on the conditional injury severity. Inclement weather conditions are more likely to be associated with *injury prone* crashes as opposed to clear weather, but also with notably lower chances of conditional severe injury. This is reasonable since inclement weather often results in slippery roads, spray and splashing and other unfavorable driving conditions that will interrupt driving maneuvers and vehicles trajectory in an emergency. On the other hand, in inclement weather, drivers paid more attention to driveways and other vehicles, and drove at lower speeds than they did in clear weather.

Driver's sex is not significant in injury propensity prediction, but is extremely significant ($p<0.0001$) in injury level estimation conditional on being in *injury prone* category. The result shows that crashes by male drivers are less likely to produce severe injuries than crashes by females. This may caused by physiological and behavioral differences, and is supported by quite a few other scholars (Abdel-Aty and Abdelwahab, 2001; Evans, 1988&2001). Middle-aged and older drivers were less likely to be involved in *injury prone* crashes. This might be because they have more experience and drive less aggressively than younger drivers. On the contrary, old drivers were associated with a significantly higher probability of severe injury conditional on being *injury prone* than were younger drivers. This makes sense considering that older drivers may have poor health and are likely to react slowly in emergencies.

The increase of AADT is not significant in injury propensity prediction, but is associated with significantly (p=0.0020) lower injury severity conditional on being in the *injury prone* group. Peak hour is not significant in affecting injury propensity, but is positively related to the injury severity conditional on being *injury prone* by having the lower probability of severe injury in comparison to non-peak hour. This is reasonable because peak hours always mean relatively low driving speeds.

As for factors considered only in the second process of the ZIOP model, the wider lanes are associated with the lower severity of injuries. Number of lanes, rural urban location, speed limit, median type, and surface year turned out to be insignificant in estimating injury severity.

3.6.2 Marginal Effects

Because of the nonlinear nature of the ordered probit model, it is difficult to directly interpret the parameters in a physically meaningful way. Therefore, the authors introduced marginal effects in order to understand the impact of various factors on crash severity. Marginal effects are simply derivatives, thus they must be evaluated at a particular point of the covariate space. There are two approaches to obtain the marginal effects: one is to evaluate the marginal effects at the joint mean of the covariates while the other option is to evaluate them separately for each observation and report the average across all observations. The authors choose to apply the second approach in this part of research.

The marginal effects of each factor on the injury level probabilities in Tables 3.4 and 3.5. For comparison purposes, the authors also include the marginal effects computed from traditional OP model. Note that for variables appearing in both X and Z, the authors have combined the two parts of the marginal effects following Equation 3.15. In Table 3.4, the marginal effects on $\mathrm{P}(\mathrm{y}=0)$ using both ZIOP model and OP model are presented. For the ZIOP model, the authors also decompose the overall marginal into two parts: the effect on *injury free*, $\mathrm{P}(\mathrm{s}=0)$, and the effect on no injury conditional on being *injury prone*, $\mathrm{P}(\mathrm{y}=0|\mathrm{s}=1)$. In Table 3.5, the marginal effects on the unconditional probabilities of all three positive injury levels are presented.

The marginal effects in Tables 3.4 and 3.5 highlight interesting differences from alternative models for some of the explanatory factors. A key example is the effect of outside shoulder. The ZIOP model shows that curbed outside shoulders were associated with a 13.10% decrease in the probability of *injury free* crashes, but a 13.63% rise in the probability of no injury conditional on being *injury prone*. The former marginal effect

Table 3.4 Marginal effects on no injury

Parameter		OP	ZIOP		
		P(y=0)	P(s=0)	P(y=0|s=1)	P(y=0)
continuous variables	AADT	0.0004	-0.0018	0.0013	0.0003
	Lanewid	0.0028	-	0.0104	0.0104
	Outshwd	<0.0001	-	0.0004	0.0004
vehtype	Pass_car				
	Light_trk	0.0083	0.0520	-0.0444	0.0076
rururb	Urban				
	Rural	-0.0014	-	-0.0016	-0.0016
no_lanes	2				
	4	-0.0044	-	0.0002	0.0002
spd_limt	45				
	50	0.0031	-	0.0072	0.0072
	55	-0.0048	-	-0.0025	-0.0025
light	Daylight				
	Nighttime	0.0108	0.0971	-0.0828	0.0143
weather	Clear				
	Inclement	0.0317	-0.1110	0.1348	0.0238
drvsex	Female				
	Male	0.0284	0.0145	0.0148	0.0294
drvage	Youngers				
	Middle_aged	0.0043	0.1154	-0.1038	0.0115
	Olders	-0.0149	0.2104	-0.2273	-0.0169
median	None				
	Flush	0.0014	-	-0.0015	-0.0015
	Raised	0.0067	-	0.0005	0.0005
outshtp	Non_curb				
	Curb	0.0122	-0.1310	0.1363	0.0053
phour	Yes				
	No	-0.0066	0.0162	-0.0206	-0.0044
surfyear	Pre_2000				
	Post_2000	-0.0009	-	-0.0010	-0.0010
frst_loc	On_pave				
	Off_pave	-0.0569	-0.2679	0.2021	-0.0658
crashtype	Animal				
	Object	-0.1192	-0.1367	0.0504	-0.0863
	Overturn	-0.2116	-0.5706	0.3430	-0.2277

Table 3.5 Marginal effects on minor injury, incapacitating and fatality

Parameter		Minor injury		Incapacitating		fatality	
		P(y=1)		P(y=2)		P(y=3)	
		OP	ZIOP	OP	ZIOP	OP	ZIOP
continuous variables	AADT	-0.0003	-0.0002	-0.0001	-0.0002	<0.0001	<0.0001
	Lanewid	-0.0019	-0.0068	-0.0008	-0.0032	-0.0001	-0.0004
	Outshwd	<0.0001	-0.0002	<0.0001	-0.0001	<0.0001	<0.0001
vehtype	Pass_car						
	Light_trk	-0.0057	-0.0066	-0.0022	-0.0010	-0.0003	<0.0001
rururb	Urban						
	Rural	0.0010	0.0010	0.0004	0.0005	0.0001	0.0001
no_lanes	2						
	4	0.0031	-0.0001	0.0012	-0.0001	0.0002	<0.0001
spd_limt	45						
	50	-0.0021	-0.0047	-0.0008	-0.0022	-0.0001	-0.0003
	55	0.0033	0.0016	0.0013	0.0008	0.0002	0.0001
light	Daylight						
	Nighttime	-0.0074	-0.0123	-0.0029	-0.0019	-0.0004	<0.0001
weather	Clear						
	Inclement	-0.0219	-0.0119	-0.0086	-0.0102	-0.0012	-0.0017
drvsex	Female						
	Male	-0.0196	-0.0195	-0.0077	-0.0087	-0.0011	-0.0012
drvage	Youngers						
	Middle_aged	-0.0029	-0.0111	-0.0012	-0.0006	-0.0002	0.0002
	Olders	0.0103	0.0043	0.0040	0.0106	-0.0006	0.0020
median	None						
	Flush	-0.0010	0.0010	-0.0004	0.0005	<0.0001	0.0001
	Raised	-0.0046	-0.0003	-0.0018	-0.0002	-0.0003	<0.0001
outshtp	non_curb						
	curb	-0.0084	0.0007	-0.0033	-0.0050	-0.0005	-0.0010
phour	Yes						
	No	0.0046	0.0023	0.0018	0.0018	0.0003	0.0003
surfyear	Pre_2000						
	Post_2000	0.0007	0.0006	0.0003	0.0003	<0.0001	<0.0001
Frst_loc	On_pave						
	Off_pave	0.0393	0.0513	0.0154	0.0133	0.0022	0.0013
crashtype	Animal						
	Object	0.0823	0.0604	0.0323	0.0230	0.0046	0.0029
	Overturn	0.1461	0.1659	0.0573	0.0553	0.0082	0.0064

shows that crashes occurred on roadways with the presence of curbed outside shoulder are less likely to be firm safe as compared to non-curb outside shoulders. The latter marginal effect indicates that curbs were associated with a relatively higher likelihood of no injury compared with other outside shoulder types, conditional on being *injury prone*. As illustrated in the last a few sections, the ZIOP model is based on the assumption that zero observations come from two distinct sources of *injury free* and no injury conditional on being *injury prone*. This assumption allows for the existence of opposite decisions during the prediction process. Therefore, the opposing effects of curb in this study compensate each other and results in an extremely small increase (0.53%) on the probability of observing no injury, which is close to the parameter estimate from the OP model (1.22%).

In addition, the resulting marginal effect on the unconditional probabilities of each injury level also comes from two sources. Table 3.5 reveals to us that curb as opposed to non-curb, were associated with 0.7% increase in the probability of minor injury, but 0.05% and 0.1% decrease in the probability of incapacitating injury and fatality, respectively. However, with one latent variable, tradition OP model shows a monotonic positive effect in the probabilities of three categories of injury involved crashes.

The marginal effects of other covariates presented in Table 3.4 and Table 3.5 are not illustrated in detail because of the space limit. Readers are suggested to read these tables with the references to Table 3.3, if interested.

3.7 Model Evaluation

Table 3.6 presents the performances of both the ZIOP model and the traditional OP model. Both models are built on the identical 46159 observations. The result shows that both the AIC and BIC values of the ZIOP model are less than that of the OP model, indicating that the ZIOP model outperformed the OP model. The result of the Vuong's Test is 6.4896, which is much greater than the critical value 1.96, also suggesting that the ZIOP model is superior. In sum, in terms of AIC, BIC and the Vuong's Test results, the ZIOP model appears to offer a clear improvement in the overall prediction of injury severity in comparison with the traditional OP model.

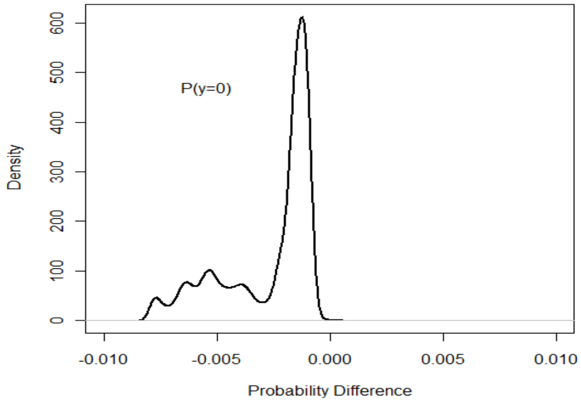
Refocusing on our original question, is it necessary to drop the speed limit on roadways with curbs installed on outside shoulders? To address this the authors consider roadways with curbs installed and compute the change in predicted probability of each injury level when the speed limit is changed.. This computation is based on the estimated coefficients from Table 3.3. It can be achieved by setting up "outshtp=curb" in both processes of the ZIOP model, then calculating the differences of the predicted probability of each injury level for each observation when assigning "spd_limt=45 mph" and "spd_limt=55 mph" separately.

Table 3.6 Performance of the ZIOP model and the OP model

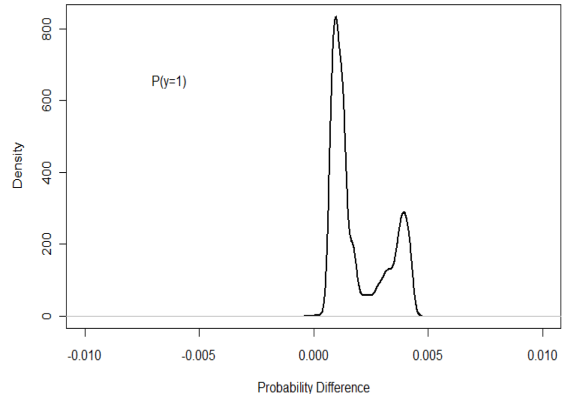
	ZIOP Model	OP Model
Log-likelihood	-15320.27	-15432.16
AIC	30673.54	30885.32
BIC	30994.95	31089.86
Vuong's test (ZIOP/OP)	6.4896	

For each injury level, Fig. 3.3 presents the distribution of the probability differences across observations when speed limit is 55 mph minus that when speed limit is 45 mph, with all the other covariates held fixed. The distribution of predicted probability differences is shown in Figure 3.3. The figure shows that the distribution of estimated probability differences in the "no injury" case is negatively skewed indicating that accidents occurring at the slower 45mph speed have a very slightly higher probability of no injury than accidents occurring at 55mph. Note that a mode estimate of this difference may be as small as .001 while the mean difference is slightly larger. The positive skew in the other figures is to be expected and indicates that higher speed leads to higher estimated probability of injury at each severity level. However, similar to above, the difference if estimated by the mode of the distribution is at most on the order of .001 and only slightly higher if the mean is used. In particular for the "incapacitating injury" and "fatality", Figure 3.3 shows extremely high frequencies of zeros with very minor right skewness, implying that the probabilities for these two injury levels between 45 mph and 55 mph are not noticeably different.

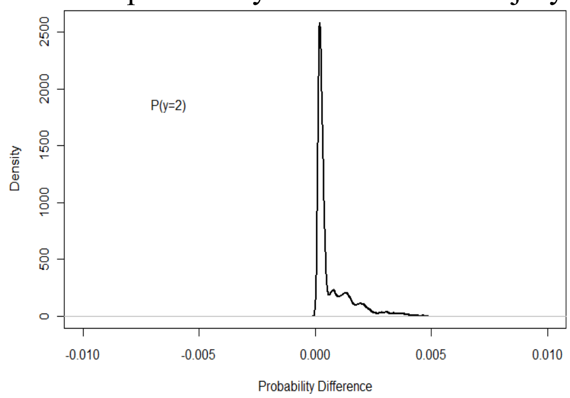
As an alternative approach to interpretation, for the overall 46159 crashes, the prediction result shows that the mean probability differences in injury levels are -0.00267(no injury), 0.00193(minor injury), 0.00067(serious injury), and 0.00007(fatality). Assuming that all 46159 single-vehicle crashes happened on curb installed roadways but that all other accident features remained constant, the model predicts that increasing speed limits from 45 to 55 mph, would lead to a decrease of 123 in no injury accidents, and increases of 89 minor, 31 incapacitating, and 3 fatal accidents.



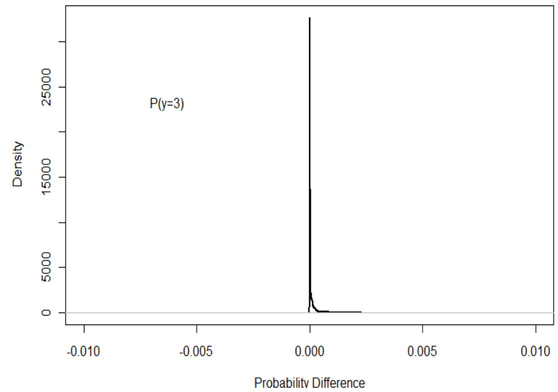
a. probability difference of no injury



b. probability difference of minor injury



c. probability difference of incapacitating



d. probability difference of fatality

Figure 3.3 Probability difference of injury severity between 55 mph and 45 mph with the presence of curbed shoulders

3.8 Conclusion

There have been numerous efforts to investigate the outcome of crashes as related to roadway design features, environmental conditions, drivers' characteristics and traffic features. However, very few of them have specifically estimated the influence of curbed outside shoulders on high-speed roadways, as well as the possible effects that a speed limit change might produce given curbs have been installed. On the basis of crash data collected from Illinois Highway Safety Information database from 2003 to 2007, the influences of curbed outside shoulders, speed limit, as well as other traditional traffic safety factors, on the injury severity of single-vehicle crashes have been estimated using the ZIOP model.

The results suggest that on one hand, crashes occurring on roadways with curbed outside shoulders are more likely to be *injury prone*. On the other hand, a curb is more likely to be associated with lower levels of injury compared with no curb, conditional on being *injury prone*. In the other word, the presence of a curb is negatively related to the propensity of injury, but is positively related to the severity of injury given that the crash is injury prone. Overall, the presence of curbs along outside shoulders are likely to have higher probability of no injury and minor injury crashes, but lower likelihood of incapacitating injury and fatality crashes. The increase of speed limit from 45 mph to 55 mph adds only a small impact to the severity of injury, given that a single-vehicle crash has occurred on roadways with curbs installed.

Other factors exhibiting statistically significant influence on injury severity include:

- The increase of AADT and peak hour are both insignificant in injury propensity prediction, but are likely to be associated with lower injury severity given that crashes fall in the *injury prone* category;
- Light trucks as opposed to passenger cars and night time as opposed to day time were both associated with significantly lower likelihood of *injury prone* crashes, but were not related to injury severity conditional on being *injury prone*;
- Inclement weather conditions were associated with higher likelihood of *injury prone* crashes as opposed to clear weather, but also appears to decrease the chance of severe injury conditional on being *injury prone*;
- Male drivers were less likely to suffer severe injuries as opposed to females. Middle aged and older drivers were less likely to be involved in *injury prone* crashes, but older drivers had remarkably higher chances to be severely injured conditional on being *injury prone*;
- Off-pavement crashes versus on-pavement crashes, impacts with objects, and overturn crashes versus collision with animals crashes, were more likely to be *injury prone*, and associated with more severe injuries conditional on being *injury prone*.

Methodologically, the traditional OP model has accounted for the inherent order of injury severities, but does not allow the effects of factors to vary across different severity levels. In addition, the application of the traditional OP model in dealing with the preponderance of zeros (no injury category) in traffic crash data leaves much to be desired. This part of research introduces the ZIOP model into the traffic safety literature for the first time in order to address these deficiencies. After performing the comparison against the traditional OP model, the ZIOP model was found to be significantly superior in this example. Employing the ZIOP model as a superior alternative in future studies of traffic injury severity is strongly suggested.

Even though a great improvement in parameter estimates has been brought up by the ZIOP model, the ability of this model in explaining underreported crashes is still unknown. Future work to evaluate this model in dealing with underreported crashes is desired. In addition, as an alternative to the Maximum Likelihood Estimating method, the Bayesian inferences can also be introduced into the ZIOP model. The authors expect that a even greater improvement could be achieved by employing Bayesian ZIOP models. Last but not the least, this part of research is also limited by the incompleteness of the database so that a few typical variables were not accounted for in this research, such as type and geometric design of curbs, pavement conditions, vertical alignment and so on.

References

Akaike, H., 1973. Information theory and an extension of the maximum likelihood principle. In B.N. Petrov and F.Csaki (Eds.), Second international symposium on information theory, pp. 267-281.

Abdel-Aty, M.A., Abdelwahab, H.T., 2001. Development of artificial neural network models to predict driver injury severity in traffic accidents at signalized intersections. Transport. Res. Rec. 1745, 6-13.

Abdel-Aty, M., 2003, Analysis of driver injury severity levels at multiple locations using ordered probit models. Journal of Safety Research, 34 (5), pp. 597-603

AASHTO, 2005. A policy on geometric design of highways and streets, Washington, D.C.

Bligh, R.P., Mak, K.K., 1999. Crashworthiness of roadside features across vehicle platforms. Transport. Res. Rec. 1690, pp. 68-77.

Baker, S. P., Braver, E. R., Chen, L.H., Li, G., and Williams, A. F, 2002. Drinking Histories of Fatally Injured Drivers. Injury Prevention, 8 (3), pp. 221-226.

Bedard, M., Guyatt, G. H., Stones, M. J., and J. P. Hirdes, 2002. The Independent Contribution of Driver, Crash, and Vehicle Characteristics to Driver Fatalities. Accident Anal. Prev., 34 (6), pp. 717-727.

Baek, J., Hummer, J., Williams, B.M., Cunningham, C.M., 2006. Reasonable speed limits on suburban multilane highways with curbs. Transport. Res. Rec. 1969, pp. 10-11.

Baek, J., Hummer, J., 2008. Collision models for multilane highway segments to examine safety of curbs. Transport. Res. Rec. 2083, pp. 128-136.

Council, F.M., Mohamedshah, Y.M., 2009. Guidebook for the Illinois state datafiles, FHWA, U.S. Department of Transportation.

Evans, L., 1988. Risk of fatality from physical trauma versus sex and age. J. Trauma 28 (3), pp.368-378.

Evans, L., 2001. Female compared with male fatality risk from similar physical impacts. J. Trauma 50 (2), pp.281-288.

Finison, K. S., Dubrow, R. B., 2002. A comparison of maine crashes involving older drivers using CODES (Crash Outcome Data Evaluation System) linked data (NHTSA

Report Report No. DOT HS 809 407). Washington, DC: National Highway Traffic Safety Administration.

Greene, W., 1994. Accounting for Excess Zeros and Sample Selection in Poisson and Negative Binomial Regression Models, Working Paper, Department of Economics, New York University.

Harries, M.N., Zhao, X.Y., 2007. A zero-inflated ordered probit model, with an application to modeling tobacco consumption. Journal of Econometrics 141 (2007), pp. 1073-1099

Jiang, X.M., Yan, X.D., Huang, B.S., Richards, S., 2011. Influence of Curbs on Traffic Crash Frequency on High-Speed Roadways. Traffic Injury Prevention 12 (4), pp. 412-421

Khattak, A.J., P.K. Kantor, F.M., Council, 1998, Role of adverse weather in key crash types on limited-access roadways implications for advanced weather systems. Transport. Res. Rec, 1621, pp.10-19.

Krull, K., Khattak, A., Council, F., 2000. Injury Effects of Rollovers and Events Sequence in Single-Vehicle Crashes. Presented at the 80th Annual Meeting of the Transportation Research Board, Washington, D.C..

Keall, M., Frith, W., Patterson, T., 2004. The influence of alcohol, age and number of passengers on the night-time risk of driver fatal injury in New Zealand. Accident Anal. Prev. 36 (1), pp. 49-61.

Lyman, S., Ferguson, S. A., Braver, E. R., & Williams, A. F., 2002. Older driver involvements in police reported crashes and fatal crashes: trends and projections. Injury Prevention, 8, pp. 116-120.

Maddala, G.S., 1983. Limited Dependent and Qualitative Variables in Econometrics. Cambridge University Press, Cambridge, UK.

National Highway Traffic Safety Administration (NHTSA), 1997. Traffic safety facts 1997. US Department of Transportation. National Highway Traffic Administration.

O'Donnell, C. J., Connor, D.H., 1996. Predicting the severity of motor vehicle accident injuries using models of ordered multiple choice, Accident Anal. Prev., 28(6), pp. 739-753.

Plaxico, C.A., Ray, M.H., Weir, J.A., Orengo, F., Tiso, P., 2005. Recommended guidelines for curb and curb-barrier installations. Transport. Res. Board, NCHRP Report 537.

Quddus, M., Wang, C., Ison, S., 2010. Road Traffic Congestion and Crash Severity: An Econometric Analysis Using Ordered Response Models. Forthcoming in the Journal of Transportation Engineering.

Renski, H., Khattak, A. J., Council, F.M., 1999. Effect of Speed Limit Increases on Crash Injury Severity: Analysis of Single-Vehicle Crashes on North Carolina Interstate Highways, Transport. Res. Rec. 1665, pp. 100-108

Sawa, T., 1978. Information criteria for discriminating among alternative regression models. Econometrica 46, pp. 1273-1282.

Smink, B.E., Ruiter, B., Lusthof, K.J., de Gier, J.J., Uges, D.R.A., Egberts, A.C.G, 2005. Drug use and the severity of a traffic accident. Accident Anal. Prev., 37 (3), pp. 427-433.

Shimamura, M. Y., Fujita, G., 2005. Method to evaluate the effect of safety belt use by rear seat passengers on the injury severity of front seat occupants, Accident Anal. Prev. 37(1), pp. 5-17.

Savolainen, P.T., Ghosh, I., 2008. Examination of Factors Affecting Driver Injury Severity in Michigan's Single-Vehicle—Deer Crashes. Transport. Res. Rec. 2078, pp. 17-25.

Schneider, W.H., Savolainen, P.T., Zimmerman, K., 2009. Driver Injury Severity Resulting from Single-Vehicle Crashes Along Horizontal Curves on Rural Two-Lane Highways. Transport. Res. Record 2102, pp. 85-92.

Savolainen, P.T., Mannering, F.L., Lord, D., Quddus M.A., 2011. The Statistical Analysis of Highway Crash-Injury Severities: A Review and Assessment of Methodological Alternatives. Accident Anal. Prev., 43(5), pp. 1666-1676.

Vuong, Q.H., 1989. Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses, Econometrica 57(2), pp. 307-333.

Valent, F., Schiava, F., Savonitto, C., Gallo, T., Brusaferro, S., and Barbone, F., 2002. Risk Factors for Fatal Road Traffic Accidents in Udine, Italy. Accident Anal. Prev., 34 (1), pp. 71-84.

Washington, S., Karlaftis, M., Mannering, F., 2003. *Statistical and Econometric Methods for Transportation Data Analysis.* Chapman and Hall/CRC, Boca Raton, Fla..

Wang, X., Abdel-Aty, M., 2008. Analysis of left-turn crash injury severity by conflicting pattern using partial proportional odds models. Accident Anal. Prev., 40(5), pp. 1674-1682.

Zajac, S., Ivan, J., 2003. Factors influencing injury severity of motor vehicle crossing pedestrian crashes in rural Connecticut. Accident Anal. Prev. 35 (3), pp. 369-379.

Zhao, X., Harris, M., 2004. Demand for marijuana, alcohol and tobacco: participation frequency, and cross equation correlation. Economic Record 80 (251), pp. 394–410.

PART 4 ESTIMATING SAFETY EFFECTS OF PAVEMENT MANAGEMENT FACTORS UTILIZING BAYESIAN RANDOM EFFECT MODELS

4.1 Abstract

Previous study of pavement management factors that relate to the occurrence of traffic-related crashes are rare. Traditional research mostly employed summary statistics of bidirectional pavement quality measurements in extended longitudinal road segments over a long time period, which may lose important information and result in biased parameter estimates. This research presented in this part of research is focusing on estimating the effects of rear-time pavement management factors on the occurrence of crashes. The crash data and corresponding pavement quality data in Tennessee State Route highways from 2004 to 2009 were employed.

The potential correlations among observations in the same road segments across time, and among road segments in the same route caused by unobserved factors were considered. Overall six models were built accounting for no correlation, temporal correlation only and both the temporal and spatial correlations. These models include: Poisson, Negative Binomial (NB), One Random Effect Poisson and Negative Binomial (OREP, ORENB), Two Random Effect Poisson and Negative Binomial (TREP and TRENB) models. Bayesian method was employed to construct these models. The inference is based on the posterior distribution from the MCMC simulation.

The models were compared using the Deviance Information Criterion. The result suggests that the temporal correlation is significant, while the spatial correlation is weak. In addition, it was shown that the ORENB model outperforms the other models. Analysis of the posterior distribution of model parameters indicates that pavement management

factors indexed by Present Serviceability Index (PSI), Pavement Distress Index (PDI) have significant association with the occurrence of crashes, whereas the variable Rutting depth is not significant. Among other factors, lane width, median width, type of terrain and posted speed are significantly related to crash frequency.

4.2 Introduction

In view of the growing amount of losses caused by traffic-related crashes, traffic safety issues have gained considerable attention among scholars in recent decades. Numerous traffic crash studies have attempted to determine the influence of road geometric design, traffic attributes, environment features, as well as driver behavior on the happening of crashes. The major attributes that have been studied include roadway location, cross section design (Hadi, et al., 1995; Jiang, et al. 2011), intersection features (Guo et al., 2010; Wang and Abdel-Aty, 2006), horizontal and vertical alignment (Donnell and Mason, 2006; Zhang and Ivan, 200), posted speed limit (Patterson et al. 2002; Kweon and Kockelman, 2005), as well as traffic volume (Lord, 2002; Lord et al. 2005). However, relatively few studies have taken into account pavement management factors, such as pavement roughness, rutting, cracking and other distress, as possible causes of traffic-related crashes.

According to previous studies in the Long-Term Pavement Performance (LTPP) Program, numerous pavements have been experiencing wear off rapidly, which further leads to structurally deficient and functionally obsolete of pavements (Özbay and Laub, 2001). High level of pavement distress may either lead drivers to lose control of their vehicles or cause vehicles to change trajectories, hence increasing the dangers of triggering crashes (Quinn and Hildebrand, 1973; Wambold et al., 1984). In addition, observable defects such as potholes and rutting may distract drivers and encourage drivers to shy away from them, which may result in vehicles' collisions or running out of driveways (Tighe et al., 2000).

Despite the significance of pavement management factors in affecting either driver or vehicle behavior, statistically evaluation of their effects on the occurrence of crashes has been relatively rare. Jacobs (1976) and Kamel and Garshore (1982) reported that the increase in road roughness would increase crash experience on rural roads. Recent research conducted by Anastasopoulos et al. (2008; 2012), Anastasopoulos and Mannering (2009) identified the identical effects of roughness on urban interstates in Indiana. On the contrary, Cleveland (1987) reported that the occurrences of crashes increased slightly but significantly with a smoother pavement surface on two-lane rural roads having the AADT range from 1000 to 8000 vehicles per day. Moreover, Al-Masaeid (1997) found that rougher roads would decrease the single-vehicle crash rate but increase the multiple-vehicle crash rate, which is consistent with the research conducted by Tighe et al. (2000).

Strat et al. (2004) conducted a study to quantify how pavement rutting affects crash rates. The rut depth measurements were average values of both directions of 1.8 km (1.1 mile) in Wisconsin. The results indicated that the defined rut-related crash rate begins to increase at a significantly greater rate as Rutting Depth (RD) exceeds 7.6 mm (0.3 in.).

Similarly, Anastasopoulos et al. (2008) employed the Tobit model to identify factors that affect the highway crash rate using 5-year data from urban interstates in Indiana. They also found that lower rutting depth was associated with lower crash rate. Their successive research utilizing the random-parameter model (Anastasopoulos and Mannering, 2009; Anastasopoulos et al., 2012) consolidated this finding. Moreover, Chan et al. (2010) employed one year of Interstate crashes in Tennessee to estimate the effects of pavement quality on various types of crashes. The results suggested that high levels of rutting depth and roughness produce significantly adverse effects on the occurrence of crashes on Interstate highways during nighttime and under rainy weather conditions.

Previous studies have contributed a certain amount to the understanding of pavement management factors as related to the occurrence of crashes. Traditional studies of pavement management factors mostly considered before-and- after data in pavement resurfacing, while not the actual measurement of pavement roughness and distress. Among those with actual measurement, the majority have employed summary statistics (such as mean, median, min, max) of bidirectional measurement in long road segments over a long time. However, the employment of bidirectional mean statistics in a long longitudinal distance over a long time period may lose important information and result in biased parameter estimates. Moreover, to use the bidirectional mean statistics in the Interstate highways may lead to even worse inference, because Interstate highways are mostly divided thus the pavement quality in each direction may vary significantly.

In order to eliminate information lose in explanatory variables, data are often considered in small time intervals and short homogenous road sections. Hence, when crash frequencies for a number of time periods were obtained in each road segment, a longitudinal panel data is formed. The major issue in analyzing panel data is the potential correlations among observations. In general, there are two levels of correlations in the panel data:(1)temporal correlations among observations in a specific segments across time, and (2)spatial correlations among observations of different segments within a certain geological region in the same time period. In the presence of temporal and spatial correlations, traditional models based on the independence assumptions of unobserved error terms will produce biased parameter estimates (Greene, 1999). Bearing this concern, a few models accounting for temporal and spatial correlations were developed. The mostly used models include the fixed and random effect Poisson and Negative Binomial models, as well as the random parameter model.

Many studies have employed the fixed and random effect models to adjust serial correlations in traffic safety study. Hausman, Hall and Griliches (1984) developed fixed-effect Poisson (FEP), random-effect Poisson (REP), fixed-effect Negative Binomial (FENB) and random-effect Negative Binomial (RENB) models to analyze panel data. Cameron and Trivedi (1998) have described the computation and application of these models in much detail. According to their research, the random-effect and fixed-effect models are conceptually different. On one hand, the random-effect models assume that the individual-specific factor is identical and independent distributed, implying that the unobserved random effects are uncorrelated with observed regressors, while the fixed-effect models do not rely on this assumption. On the other hand, the random-effect models can accommodate random slop parameters and intercept more easily, while for the fixed-effect models, the coefficients of time-invariant variables cannot be identified in that they are absorbed into the individual-specific factor. Overall, the fixed-effect models are more applicable in explaining the existed sample, while the random-effect models are more appropriate in doing inference on the population based on the sample.

Shankar et al. (1998) employed both the Random-effect Negative Binomial (RENB) and the cross-sectional Negative binomial (considering location and time as covariate) to estimate factors that affect median crossover accidents in Washington. They found that the RENB outperformed the NB model when spatial and temporal effects are totally unobserved, which is reasonable because geometric and traffic variables are likely to have location-specific effects. Their successive work (Ulfarsson and Shankar, 2003) explored the use of the Negative Multinomial (NM, known as Random Effect Poisson) model to form a predictive model of median crossover crash frequencies. By comparing with the former research, they reported that the NM model outperformed both the NB model and the RENB model in terms of the Likelihood. Chin and Quddus (2003) built RENB model to identify the elements that affect intersection safety. By treating the data in time-series cross-section panels, the RENB model explicitly accounted for the unobserved time and space effects. Caliendo et al. (2007) employed the NM model to estimate the safety effects of factors for Italian multilane roads. Their research suggests that the NM distribution has a decidedly higher explanatory power than the Negative Binomial distribution, thus supporting the hypothesis that the latter model is inappropriate when multiple observations for the same road section at different years are analyzed.

A few studies have attempted to apply the random parameter models in traffic crash research. Anastasopoulos and Mannering (2009) tried to use the random parameter models instead of the fixed parameter models in estimating traffic safety effects of a few factors. The results suggest that the random parameter models are significantly superior than the fixed parameter models, and the negative binomial model provided better overall fit in relative to the random and fixed parameters Poisson models. Moreover, Anastasopoulos et al. (2012) employed the random-parameters Tobit regression to account for the unobserved heterogeneity in the study of motor-vehicle accident rates using 9 year urban interstate data in Indiana. The results indicate that the random parameters Tobit model outperforms its fixed parameters counterpart and has the potential to provide a fuller understanding of the factors determining accident rates on specific roadway segments. However, the employment of the random parameter models always require much more computations. The precision that the random parameter improved may not worth the extremely longer time it takes as compared to the fixed parameter models. In addition, the parameter estimates obtained from the random parameter models can hardly be applied to the samples from other populations.

In recent years, a growing number of researchers start to incorporate the Bayesian inference into traffic safety analysis (Miaou et al., 2003; Guero-Valverde and Jovanis, 2006; Quddus, 2008). Previous studies suggest that the Bayesian method can serve as a great alternative to the Maximum Likelihood Estimates. For example, Xie et al (2009) performed a series of studies on the prediction of severity of crash injuries utilizing both the Ordered Probit (OP) model and the Bayesian Order Probit (BOP) model with various priors. The result shows that the BOP model outperforms the OP model in both the parameter estimates and the predictive capabilities, especially for small data sample. Guo et al. (2010) conducted a traffic safety study with 170 signalized intersections in the state of Florida. This study employed both the non-Bayesian and Bayesian Poisson and Negative Binomial models. The results indicate that the Bayesian models with non-informative priors can provide similar results as the non-Bayesian models, and the flexibility of the Bayesian method allows sophisticated models to be constructed.

The objective of this paper is to explore the effects of pavement management factors on the occurrence of traffic-related crashes. The real time (two years period) measurement of pavement roughness and distress in each 0.1 mile (0.16 km) longitudinal distance were employed in this study. Two levels of potential correlations were considered. At the first level, crash frequencies for the same road segments across time are supposed to be correlated, which is so called the temporal correlation. At the second level, crash frequencies for road segments in the same route might be correlated, which is so called the spatial correlation. Six alternative models were built with Bayesian method: Poisson, Negative Binomial (NB), One Random Effect Poisson (OREP), One Random Effect Negative Binomial (ORENB), Two Random Effect Poisson (TREP), and Two Random Effect Negative Binomial (TRENB) models. Crashes occurred on Tennessee State Route highways from 2004 to 2009 were employed in this study.

4.3 Data Preparation

The primary source of traffic crash data is the Tennessee Roadway Information Management System (TRIMS). The data were obtained in computer-ready forms, which included coded information on reported crashes that occurred on state highways in Tennessee. The coded information for each crash contains important attributes describing the conditions contributed to the collision and the outcome. The information regarding pavement management status was obtained from the Pavement Management System (PMS) maintained by the Tennessee Department of Transportation (TDOT). The crash data occurred on State Route roadways from 2004 to 2009 and corresponding traffic and geometric features can be linked to the pavement management factors through the common variable: id_number, which is a combination of county, county sequence, route type, and route number.

Pavement management factors considered in this paper include Present Serviceability Index (PSI), Pavement Distress Index (PDI) and Rutting Depth (RD). PSI is a measure of the roughness of roadways on a scale of 0-5, representing worst to perfect; PDI is specified as a scale from 0-5 representing worst to perfect of the overall pavement quality including the sensitivity of drivers to each type of pavement distress; RD is the original measurement of rutting depth in inch.

In the State of Tennessee, roughness, rutting, cracks and other pavement defects are measured once every two years on State Route highways. Measurements are recorded for each 0.1 mile (0.16 km) pavement section. The pavement management data from 2004 to 2009 were collected, screened and treated according to the following criteria:

- Records with PSI and PDI less than 1.00 were excluded because they may subject to missing value or in a short section where there may be a railroad crossing, intersection, or other discontinuity;
- Records with mean RD greater than 1.00 inch (25.40 mm) were removed since they are abnormal in terms of the overall pavement quality for the State Route in Tennessee;
- PDI, PSI and RD records in every successive two years were combined to form the complete pavement quality measurement for each unit time period (2 years). For example, the pavement quality information for 2004 and 2005 are combined and denoted by 2004-05, so do the pavement quality records for 2006 and 2007, and for 2008 and 2009.

The geometric and traffic factors considered in the present study include: Annual Average Daily Traffic (AADT), posted speed limit, lane width, median width, grade of horizontal alignment, types of terrain(flat, rolling, mountain). Thousand of AADT (thAADT) was employed in the regression modeling as an exposure variable, which was calculated through dividing AADT by 1000, since the change of injury severity with increment by one vehicle would be too trivial.

Records of crashes, roadway traffic and geometric features are mapped to aforementioned 0.1 mile (0.16 km) segment by ID_number, log-mile and years. Interstate highways are excluded from this study because both the pavement quality measurements and the traffic crash records were not direction specified in the original database. As mentioned earlier, pavement quality measurements on Interstate highways in one direction can hardly represent that in the other direction since they are mostly divided.

This study is restricted to the state routes with the total number of lanes equals or less than 4, and median width equals or less than 16 feet (4.88m). The authors assume that the one directional measurement of pavement quality in these roads can represent the average statistics throughout the cross section. Thus, each of these 0.1 mile (0.16 km) segment can be deemed to be homogenous, i.e. pavement roughness and distress, traffic and geometric features are identical along each segment. The total number of crashes is calculated for each road segment during the unit time period. Consequently, there are three crash frequencies for each segment, indicating the number of crashes occurred in 2004-05, 2006-07 and 2008-09.

The combined dataset was further cleaned according to the following criteria:

- Intersection related segments were removed since the crash patterns are different from continuous roadways;
- Pedestrian, motorcycle, and non-traditional types of vehicles involved crashes were screened out because they have different injury mechanisms;
- Road segments with crash frequency greater than 30 were excluded as outliers because it is rare and extremely large in terms of 2 years, 0.1 mile (0.16 km) segments;

- The observations with missing values either in the dependent variable or in each of the predictors were excluded for the purpose of completeness;
- Speed limits of 30 mph (48 km/hr),35 mph (56 km/hr),40 mph (64 km/hr),45 mph (72 km/hr),50 mph (80 km/hr) and 55 mph (88 km/hr) were selected for this study since they are quite common for Tennessee State Route highways;

Through the preliminary treatment, the original traffic crash, roadway traffic, geometric and pavement quality data was split into three time periods: 2004-05, 2006-07, 2008-09. Each time period has 93,783 homogeneous segments, and all these segments have equal length of 0.1 mile (0.16 km). The cleaned dataset applied in this paper includes 121,525 total crashes occurred in 851 state routes. The statistics of the crash frequency, roadway features is presented in Table 4.1.

4.4 Methodology

4.4.1 Poisson and Negative Binomial model

Poisson model is known as the simplest and most common model for count data regression analysis and has been widely used in crash frequency modeling studies (Jovanis et al. 1990, Joshua et al. 1993, Miaou et al. 1993). Let y_{ijt} denotes the observed number of crashes occurring in road segment$i$, on route j, during time t. Assume y_{ijt} follows a Poisson distribution, then the probability mass function of y_{ijt} can be expressed as Equation 4.1.

$$P\left(Y_{ijt} = y_{ijt} \middle| \mu_{ijt}\right) = \frac{e^{-\lambda_{ijt}} * \lambda_{ijt}{}^{y_{ijt}}}{y_{ijt}!} \tag{4.1}$$

Table 4.1 Summary statistics of the overall factors

Variables	Description	Year	Min	Max	Mean	Std
Count	Crash frequency in each 0.1 mile segment	2004-05	0.00	30.00	0.46	1.27
		2006-07	0.00	29.00	0.42	1.19
		2008-09	0.00	28.00	0.41	1.14
PDI	Pavement distress index(0-5 Indicating worst to perfect	2004-05	1.00	5.00	4.77	0.49
		2006-07	1.00	5.00	4.66	0.53
		2008-09	1.02	5.00	4.49	0.60
PSI	Pavement serviceability index (0-5 indicating worst to perfect)	2004-05	1.00	4.37	3.24	0.54
		2006-07	1.00	4.38	3.25	0.54
		2008-09	1.00	4.45	3.28	0.53
RD	Rutting depth in inch	2004-05	0.00	0.74	0.05	0.06
		2006-07	0.00	0.86	0.11	0.08
		2008-09	0.00	0.78	0.10	0.07
ThAADT	Annual Average Daily Traffic in 1000 vehicles per day	2004-05	0.08	74.07	4.70	5.87
		2006-07	0.06	69.36	4.67	5.88
		2008-09	0.05	71.72	4.55	5.70
tyterrain	Type of terrain(1 for flat, 2 for rolling, 3 for mountain)	2004-09	1	3	1.99	0.37
spdlimt	Speed limits		30	55	47.98	7.58
pct_grde	Grade of horizontal alignment		0.00	14.80	1.60	1.95
lanewid	Lane width in feet		8.00	12.00	10.96	0.90
medwid	Median width in feet		0.00	16.00	0.13	1.25

where the parameter λ_{ijt} is the mean number of events. We can write the mean as a function of the covariate vector $X_{ijt}^{'}$ through the exponential function

$$\lambda_{ijt} = Exp(X_{ijt}^{'}\beta) \tag{4.2}$$

where β is a $(k+1)\times 1$ parameter vector (β_0 is the coefficient for intercept, and β_1, β_1, $\beta_2 \ldots \beta_k$ are for k regressors). This ensures that the mean will be positive for any set of covariates and parameters β .

In Poisson regression, the conditional variance of the response variable is equal to the conditional mean

$$V(y_{ijt}|X_{ijt}) = E(y_{ijt}|X_{ijt}) = \lambda_{ijt} \tag{4.3}$$

so that the Variance to Mean Ratio (VMR) equals one. Empirical analysis shows that vehicle accident data don’t satisfy this feature, typically having a larger variance relative to its mean, a phenomenon known as overdispersion. In this situation, the common Poisson regression model is inappropriate as it can result in bias and inconsistent parameter estimates. The negative binomial (NB) distribution, a more flexible probability distribution for counts, has become popular in modeling crash count data in the last few years because it can address the issue of overdispersion (Milton et al. 1998, Shankar et al. 1995, Carson et al. 2001).

The NB regression model can be derived from the Poisson model by assuming the mean λ_{ijt} follows the gamma distribution, as shown in Equation 4.4.

$$\begin{cases} y_{ijt} \sim Poisson(\lambda_{ijt}) \\ \lambda_{ijt} \sim Gamma(\mu_{ijt}, \delta) \\ \mu_{ijt} = Exp(X'_{ijt}\beta) \end{cases} \tag{4.4}$$

where δ is constant both across road segments and across times. The mean and variance of λ_{ijt} are then $E[\lambda_{ijt}] = e^{X'_{ijt}\beta}/\delta$ and $V(\lambda_{ijt}) = e^{X'_{ijt}\beta}/\delta^2$. Thus, the mean and variance for y_{ijt} are $E[y_{ijt}] = e^{X'_{ijt}\beta}/\delta$ and $V(y_{ijt}) = e^{X'_{ijt}\beta}(1+\delta)/\delta^2$. Consequently, the VMR equals to $(1+\delta)/\delta > 1$, which indicates that the NB model allows for overdispersion.

However, the NB model shares a common δ across road sections and across time. This implies that λ_{ijt} is independent for a given road section over time. In addition, the VMR is constant across road segments, which may not represent the truth because it is expected that VMR grows with λ_{ijt} in each road segment. In order to account for potential correlations among observations, and allow the VMR to grow with crash frequency random effect (RE) models are developed.

4.4.2 Random effect models

There are two levels of inherent correlations in this particular data. First, crashes happening in one specific road segment across years should be similar. These so called temporal correlations are due to unobserved variables unique to each intersection. In order to account for this level of correlations, a segment specific random effect is added to both the original Poisson model and the NB model. The One Random Effect Poisson (OREP) model can be expressed as

$$\begin{cases} y_{ijt} \sim \text{Poisson}(\lambda_{ijt}) \\ \lambda_{ijt} = \text{Exp}(x'_{ijt}\beta + \varepsilon_i) = \mu_{ijt} * \text{Exp}(\varepsilon_i) \\ \varepsilon_i \sim Normal(0.0, \sigma^2) \end{cases} \tag{4.5}$$

, where ε_i is a road segment specific random effect. This quantity is assumed to follow normal distribution with mean 0 and variance σ^2. The mean and variance for y_{ijt} are $E[\,y_{ijt}] = \mu_{ijt} * e^{\frac{1}{2}\sigma^2}$ and $V(\,y_{ijt}) = \mu_{ijt} * e^{\frac{1}{2}\sigma^2}(1 + \mu_{ijt}(e^{\sigma^2} - 1) * e^{\frac{1}{2}\sigma^2}\,)$. The VMR is $1 + \mu_{ijt}(e^{\sigma^2} - 1) * e^{\frac{1}{2}\sigma^2} > 1$. It can be seen that the OREP model allows for overdispersion, as well as a road segment specific VMR, which grows with μ_{ijt} as expected.

There are also possible correlations among road sections in the same route, which is known as spatial correlation. Hence, it is desired to add a route specific random effect into the OREP model. In this part of research, for simplicity, the authors assume that the spatial correlation is firm for each two segments in the same route, i.e., the spatial correlation does not rely on the distances between segments. The Two Random Effect Poisson model (TREP) with both temporal correlation and spatial correlation can be expressed as

$$\begin{cases} y_{ijt} \sim \text{Poisson}(\lambda_{ijt}) \\ \lambda_{ijt} = \text{Exp}(x'_{ijt}\beta + \varepsilon_{ijt1} + \varepsilon_{ijt2}) = \mu_{ijt} * \text{Exp}(\varepsilon_i) * \text{Exp}(\varepsilon_j) \\ \varepsilon_i \sim Normal(0.0, \sigma_1^2) \\ \varepsilon_j \sim Normal(0.0, \sigma_2^2) \end{cases} \tag{4.6}$$

The shortcoming of both the OREP and TREP models is that λ_{ijt} for road sections are fixed over time given that the X_{ijt} is constant. To allow λ_{ijt} to vary across time, the Random Effect Negative Binomial (RENB) models are also considered in this part of research.

To consider temporal correlations, the road segment specific random effect was added to the NB regression model. The One Random Effect Negative Binomial (ORENB) model can be derived as

$$\begin{cases} y_{ijt} \sim Poisson(\lambda_{ijt}) \\ \lambda_{ijt} \sim Gamma(\mu_{ijt}, \delta_i) \\ \mu_{ijt} = Exp(X'_{ijt}\beta) \\ \delta_i = beta(1,1) \end{cases} \tag{4.7}$$

where δ_i is road segment specific random effect. The mean and variance of λ_{ijt} are $E[\lambda_{ijt}] = e^{X'_{ijt}\beta}/\delta_i$, $V(\lambda_{ijt}) = e^{X'_{ijt}\beta}/\delta_i^2$. Note that even if X_{ijt} remains constant for a road section over time, λ_{ijt} can still vary. The mean and variance for y_{ijt} are then $E[y_{ijt}] = e^{X'_{ijt}\beta}/\delta_i$ and $V(y_{ijt}) = e^{X'_{ijt}\beta}(1+\delta_i)/\delta_i^2$. Therefore, the VMR equals to

$(1+\delta_i)/\delta_i$>1, which allows for overdispersion, as well as road specific VMRs which traditional NB does not.

The Two Random Effect Negative Binomial (TRENB) can be achieved by adding a route specific random effect into the ORENB, which can be expressed as

$$\begin{cases} y_{ijt} \sim Poisson(\lambda_{ijt}) \\ \lambda_{ijt} \sim Gamma(\mu_{ijt}, \delta_i / e^{\mu_j}) \\ \mu_{ijt} = Exp\left(X'_{ijt}\beta\right) \\ \mu_j \sim Normal(0.0, \sigma_1^2) \\ \delta_i = beta(1,1). \end{cases} \tag{4.8}$$

The TRENB model has all of those advantages as the ORENB model has, and also explains the inherent correlation among road segments in the same route caused by the identical attributes they might have.

All together, six full Bayesian models considering no correlation, temporal correlation only and both temporal and spatial correlations were built in this paper: (1) Poisson regression model; (2) NB model; (3) OREP model accounting for temporal correlation; (4) ORENB model accounting for temporal correlation; (5) TREP accounting for both temporal and spatial correlations; (6) TRENB model accounting for both temporal and spatial correlations.

These Models were estimated using the open source software WinBUGS® 3.0.2 (Windows Bayesian Inference Using Gibbs Sampling). For the Poisson, NB, OREP and

TREP models, 2,000 iterations were discarded as burn-in, and the 10,000 iterations that followed were used to obtain summary statistics of the posterior inference. For the TREP and TRENB models, 4,000 iterations were discarded and the following 10,000 iteration were accepted to obtain posterior inference. Convergence was assessed by visual inspection of the Markov chains for the parameters. Furthermore, the number of iterations was selected so that the Monte Carlo error is less than 0.05 for each parameter.

4.5 Results

4.5.1 Model comparison

The performance of these models was evaluated to identify the model that provides the best fitting for the data. The model comparison was conducted using the Deviance Information Criterion (DIC), which was introduced by Spiegelhalter et al.(2002). It is given by the expression,

$$\text{DIC} = \bar{D} + pD \,,$$

$$pD = \bar{D} - \hat{D}$$

where $\bar{D}$ is the posterior mean of the deviance. This deviance is defined as -2 * log(*likelihood*): *likelihood* is defined as $p(y|\theta)$, where y comprises all response observations, and θ comprises the explanatory variables upon which the distribution of y depends. $\hat{D}$ is a point estimate of the deviance obtained by substituting θ with the posterior mean $\bar{\theta}$, thus $\hat{D}$= -2 * log(p($y|\bar{\theta}$)). pD is the effective number of parameters.

In general, $\overline{D}$ decreases with the increased number of explanatory variables. However, the employment of pD in the DIC equation will punish the preponderance of explanatory variables. Therefore, the DIC value adjusts for both the fitting and the complexity of the model. The model with the smallest DIC is estimated to be the model that would best predict a replicate dataset of the same structure as that currently observed. A difference of DIC value greater than 5 can be deemed to be significant. The results of the DIC computation are presented in Table 4.2.

It can be observed from Table 4.2 that the performance of these models can be ranked from best to worst as: ORENB, NB, TREP, TRENB, OREP, Poisson. The Poisson model performs the worst as expected because it does not allow for overdispersion as discussed earlier. With one random effect, the DIC value decreases from 456427.00 in the Poisson model to 391441.00 in the OREP model, which implies that the segment specific effect is extremely significant. To add the second random effect, i.e., route specific effect, the DIC value drops to 385405.00, which is far less than that the drop from Poisson to OREP. The difference of DIC values change suggests that the route specific effect contributes relatively less than the segment specific effect to the model fitting, i.e., the correlation among segments in the same route are not as strong as the correlation among observations in the segment across time.

The NB model outperforms all three Poisson models, which might be attributed to the hierarchical structure of the NB model, that is even if X_{ijt} remains constant for a road segment over time, λ_{ijt} can still vary. The ORENB model is superior to the NB model,

Table 4.2 The performance of models

Models	$\bar{D}$	$\hat{D}$	pD	DIC
Poisson	456410.00	456393.00	16.00	456427.00
NB	312312.00	257820.00	54493.00	366805.00
OREP	361181.00	330921.00	30260.00	391441.00
TREP	360251.00	335097.00	25154.00	385405.00
ORENB	284339.00	218395.00	65945.00	350284.00
TRENB	343741.00	299059.00	44682.00	388423.00

with the DIC value drops remarkably from 366805.00 to 350284.00. This once again implies the significance of the segment specific random effect. Unfortunately, the TRENB model performs even worse than either the NB model or the ORENB model. This implies that the route specific random effect is redundant, which to some extent supports the findings of model comparisons for the Poisson models.

In view of the discussion above, the route specific correlation among segments is weak. The incorporating of the spatial correlation may produce adverse effects to the posterior inference. Therefore, the following sections will only illustrate the modeling results from the Poisson, NB, OREP and ORNB models in detail.

4.5.2 Posterior inference

The model fitting was conducted using MCMC simulation and the posterior distributions were constructed from the simulation output. Table 4.3 presents the posterior means, standard deviations, and 95% posterior intervals (PI) of the Poisson, NB, ORENB, OREP models. The 95% PIs are obtained from the 0.025 and 0.975 posterior quantiles, which describes the information about the location of the true value of parameter estimates. The 95% PI also indicates the significance of covariates: when the 95% PI includes zero, the corresponding factor is not significant at the 95% level and vice versa.

Looking at the NB model, the value of δ (0.9997) indicates that the VMR for this dataset is $(1 + 0.9997)/0.9997 \approx 2.0$. This is a strong proof of overdispersion, which provides a good explanation of the significant improvement that the NB model produced as

Table 4.3 Posterior statistics of classical and one random effect models

Factors	Poisson		NB		OREP		ORENB	
	Mean	Std	Mean	Std	Mean	Std	Mean	Std
Intercept	-2.919	0.060	-2.797	0.035	-3.267	0.075	-3.232	0.091
	(-3.050,-2.810)[a]		(-2.852,-2.740)		(-3.388,-3.177)		(-3.590,-3.440)	
PSI	-0.318	0.005	-0.225	0.006	-0.155	0.008	-0.152	0.006
	(-0.327,-0.307)		(-0.236,-0.210)		(-0.170,-0.142)		(-0.167,-0.141)	
PDI	0.136	0.006	0.121	0.007	0.084	0.006	0.089	0.004
	(0.125,0.147)		(0.110,0.137)		(0.072,0.096)		(0.081,0.098)	
RD	0.249	0.036	0.121	0.053	-0.175	0.049	-0.007	0.052
	(0.173,0.352)		(0.013,0.223)		(-0.273,-0.078)		(-0.118,0.092)	
Pct_grde	-0.011	0.002	-0.001	0.002	0.008	0.003	0.006	0.002
	(-0.015,-0.008)		(-0.005,0.003)		(0.003,0.050)		(0.002,0.011)	
Lanewid	0.182	0.004	0.146	0.007	0.134	0.006	0.125	0.014
	(0.173,0.191)		(0.130,0.154)		(0.125,0.014)		(0.098,0.146)	
Medwid	-0.022	0.001	-0.017	0.002	-0.015	0.003	-0.011	0.002
	(-0.025,-0.020)		(-0.021,-0.013)		(-0.021,-0.009)		(-0.016,-0.007)	
ThAADT	0.077	0.000	0.071	0.000	0.097	0.001	0.073	0.001
	(0.076,0.077)		(0.070,0.072)		(0.095,0.098)		(0.071,0.074)	
Flat	-	-	-	-	-	-	-	-
Rolling	0.307	0.031	0.348	0.030	0.401	0.020	0.374	0.029
	(0.174,0.297)		(0.286,0.394)		(0.360,0.443)		(0.308,0.395)	
Mountain	0.097	0.021	0.162	0.035	0.266	0.030	0.207	0.034
	(0.056,0.139)		(0.093,0.220)		(0.209,0.326)		(0.132,0.254)	
Spdlmt_30	-	-	-	-	-	-	-	-
Spdlmt_35	0.018	0.014	-0.017	0.024	-0.097	0.027	-0.076	0.027
	(-0.010,0.045)		(-0.066,0.032)		(-0.150,-0.046)		(-0.126,-0.029)	
Spdlmt_40	-0.112	0.012	-0.130	0.022	-0.254	0.023	-0.201	0.024
	(-0.135,-0.089)		(-0.176,-0.089)		(-0.299,-0.211)		(-0.244,-0.162)	
Spdlmt_45	-0.347	0.011	-0.320	0.021	-0.392	0.020	-0.336	0.022
	(-0.369,-0.325)		(-0.364,-0.281)		(-0.431,-0.354)		(-0.378,-0.299)	
Spdlmt_50	-0.541	0.015	-0.445	0.024	-0.467	0.025	-0.410	0.025
	(-0.570,-0.512)		(-0.495,-0.402)		(-0.516,-0.419)		(-0.457,-0.368)	
Spdlmt_55	-0.689	0.012	-0.544	0.018	-0.581	0.020	-0.523	0.018
	(-0.713,-0.666)		(-0.582,-0.508)		(-0.621,-0.543)		(-0.558,-0.487)	
σ^2	-		-		0.696	0.008	-	
	-		-		(0.680,0.712)		-	
δ	-		0.9997	2.9E-4	-		-	
	-		(0.9989,1.0000)		-		-	

([a] 95% confidence interval)

compared to the Poisson model. Overall, the original NB model and the Poisson model provide quite similar posterior statistics for most of these explanatory variables, except for the factor horizontal alignment (pct_grde). The posterior inference of pct_grde in the Poisson model is significantly negative, with the mean -0.011, but is not significant (-0.005,0.003) in the NB model. Furthermore, it can be observed that both the OREP and ORENB models provide significant positive posterior inference for pct_grde with mean 0.008 and 0.006 separately, which are quite opposite to that obtained from the Poisson model. This is reasonable because the better fitting model is more likely to provide reliable results. Based on this consideration, the insignificance inference of pct_grde in the NB model is a transition from lower quality model (Poisson model) to the higher quality models (OREP and ORENB models).

Comparing the two one random effect models, the posterior distribution of the OREP model and the ORENB model are quite similar, except for the variable RD. The OREP model provides significant negative posterior inference for RD, while the ORENB model indicates the effect of RD is not significant.

The variance for the segment specific random effect is estimated to be 0.696. This implies the existence of correlation for each segment across time. It is noticeable that the intercept for OREP and ORENB model is smaller than that in the Poisson and NB model. This is expected because in the OREP model, the mean of crash frequency is $E[\,y_{ijt}] = \mu_{ijt} * \mathrm{e}^{\frac{1}{2}\sigma^2} = e^{X_{ijt}^{'}\beta+\frac{1}{2}\sigma^2}$. The intercept in the OREP model (-3.267), adding up the variance term $\frac{1}{2}\sigma^2$ (0.348) equals to -2.919, which is close to the intercepts in both the Poisson and NB models. The same reasoning applies to the ORENB model, where the segment specific random effect parameters δ_i are not presented due to the space considerations (around 20,000 values).

These models quantify average effects of pavement management factors on the occurrence of crashes. All four models present significantly negative posterior statistics for PSI, indicating that higher PSI value are associated with lower crash frequency. In the other word, crashes are less likely to occur on smooth roads as compared to rough roads. PSI is a measure of pavement roughness, which is defined as the deviation of a pavement surface from a true planar surface with characteristic dimensions that affect vehicle dynamics, ride quality, dynamic loads, and drainage. Hence, the negative effect of roughness is expected. This result is consistent to many earlier findings (Jacobs ,1976; Kamel and Garshore, 1982; Anastasopoulos et al., 2008&2012; Anastasopoulos and Mannering, 2009).

To the authors surprise, these models all provide significantly positive posterior statistics for PDI: the higher PDI values were associated with higher crash frequency. PDI is a evaluation of pavement distresses including: Fatigue, Rutting, Longitudinal Cracks In the Wheel Path, Patching, Block Cracking, Raveling, Transverse Cracks, Longitudinal Cracks (Non-Wheel Path), & Longitudinal Cracks In the Lane Joints. The modeling results indicate roadways with high level of distress were associated with lower numbers of crashes as compared to those with low levels of distress. This might be attributed to unobserved factors that might affect the occurrence of crashes. For example, PDI measures the overall pavement quality including the sensitivity of drivers to each type of pavement distress. High PDI value means there were distresses that drivers are strongly sensitive to. When exposed to high level of perceivable distress, drivers may slow down and drive more carefully, thereby decreasing the possibility of crashes.

As mentioned earlier, the one random effect models present inconsistent posterior statistics for the variable RD, one is significantly negative and the other is not significant. However, both the original Poisson and NB model show significantly positive posterior inference for the variable RD. The difference between models with and without temporal correlation suggests that the ignorance of potential correlation in the panel data could result in biased parameter estimates. Taking the ORENB model as the superior model in terms of the DIC value, it can be concluded that the association between rutting depth and crash frequency is not significant. This finding is consistent to the research conducted by Chan et al. (2010), which further consolidates the explanation for the variable PDI that drivers may slow down and drive more carefully when they notice apparent defects on the roads.

Turning to other factors, all four models indicate that the increase of AADT is more likely to result in higher crash frequency. This is expected because higher traffic volume increases the chance of collision. Higher width of median width was associated with lower probability of crashes, which is consistent with the literature (Abdel-Aty et al., 2000; Knuiman et al., 1993). Roadways with wider lane width were associated with higher number of crashes. This contradicts previous research results (Hadi et al., 1995; Andrew, 1998; Garber et al., 2000). There are a few possible reasons. First of all, the width of lanes considered in this paper ranges from 8.00 ft to 12.00 ft, with the mean of 10.96 ft. The relatively lower proportion of narrow lanes might have biased the posterior inference. Secondly, narrow lanes are normally associated with lower traffic volume in comparison to wide lanes, which may lead to lower crash frequency. Last but not least, drivers may drive more carefully on narrow lanes, thus decreases the chances of crashes.

The increase in the grade of horizontal curves is associated with higher crash frequency, as indicated by the OREP and the ORENB model. Nevertheless, the original Poisson model generates the exactly opposite result, and the NB model shows that the effect of horizontal curve is not significant. This once again supports the statement that the ignorance of temporal correlations in the panel data may lead to biased posterior inference. Rolling and mountain terrain were associated with higher number of crashes as compared to flat terrain. This is expected because rolling and mountain terrain have more vertical curves with higher grades as compared to flat terrain, which were reported to increase the chances of crashes (Ma and Damien, 2008).

All four models show that the speed limit of 40, 45, 50 and 55 mph were associated with lower crash frequencies as compared to the speed limit of 30 mph. The OREP and the ORENB models additionally show that the speed limit of 35 mph was associated with significantly lower crash frequency in relative to that of 30 mph. This result supports the conclusions of many other studies (Jiang, et al. 2011). This is reasonable in that higher posted speed limits are mostly related to highways in more rural area, with better driving conditions.

4.6 Conclusions

Previous studies of pavement management factors as related to crash frequency normally employed summary statistics of bidirectional pavement quality measurements in long longitudinal road segments over a long time period, which may lose important information and result in biased parameter. This research presented in this paper is focusing on estimating the effects of rear-time pavement management factors on the happening of crashes. The crash data and corresponding pavement quality data in Tennessee State Route highways from 2004 to 2009 were employed, and mapped into every 0.1 mile by each two years.

The potential correlations among observations in the same road segments across time, and among road segments in the same route caused by unobserved factors were considered. Overall six models were built accounting for no correlation, temporal correlation only and both the temporal and spatial correlations. These models include: Poisson, NB, OREP, ORENB, TREP and TRENB models. Bayesian inference was employed to conduct the analysis. The inference is based on the posterior distribution from the MCMC simulation.

The performances of these models were evaluated in terms of the DIC values. The model comparison result suggests that the temporal correlation is significant, while the spatial effect is weak. In addition, it was shown that the ORENB model outperforms the other models, the NB model next, followed by the TREP, TRENB, and then OREP, Poisson model. The ignorance of the temporal correlation was found to result in biased posterior inference for several variables.

The posterior distribution in these models show significant association between pavement management factors and the occurrence of crashes:

- Roadways of high roughness were associated with higher crash frequency as compared to smooth roads;
- Roadways with high level of perceivable distresses were associated with lower number of crashes;
- The effect of rutting depth is not significant based on the posterior inference in the ORENB model;

The effects of traffic and geometric features include:

- Roadways with wide lane width were associated with higher crash in comparison with those with narrow lane width;
- The increase of median width is related to lower probability of crashes;
- Rolling and mountain terrain were associated with higher likelihood of crashes.
- Roadways with high posted speed limit were associated with lower crash frequencies as compared to those with low posted speed limit.

There are several possible future developments for this research. First of all, the research presented in this paper assume the pavement quality information in state route undivided highways are homogeneous throughout the transverse lanes in both direction. This assumption is reasonable but not perfect. Hence, the future study using direction specific pavement quality information and crash data are desired. Secondly, the overall pavement quality in Tennessee state route is very good. The future study utilizing proportional roadways with median to bad pavement quality is encouraged; Last but not the least, the ability of random effect models in dealing with preponderance of zeros are not agreeable. The future study to incorporate zero-inflated models in the random effect structures are desired.

References

Al-Masaeid, H. R, 1997. Impact of Pavement Condition on Rural Road Accidents. Canadian Journal of Civil Engineering, 24 (4), pp.523-532.

Abdel-Aty, M.A., Radwan, A.E., 2000. Modeling traffic accident occurrence and involvement. Accid. Anal. Prev. 32, pp. 633–642.

Andrew, V., Bared, J., 1998. Accident Models for Two-Lane Rural Segments and Intersections, paper presented at the 77th Annual Meeting of the Transportation Research Board.

Anastasopoulos, P.Ch., Mannering, F.L., 2009. A note on modeling vehicle-accident frequencies with random parameter count models. Accident Analysis and Prevention 41 (1), 153–159.

Anastasopoulos, P.Ch., Tarko, A.P., Mannering, F.L., 2008. Tobit analysis of vehicle accident rates on interstate highways. Accident Analysis and Prevention 40 (2),768–775.

Anastasopoulos, P.Ch., Mannering, F.L., Shankarc, V. N. Haddockd, J.E., 2012. A study of factors affecting highway accident rates using the random-parameters tobit model. Accident Analysis and Prevention 45, 628-633.

Chin, H.C., Quddus, M.A., 2003. Applying the randomeffect negative binomial model to examine traffic accident occurrence at signalized intersections. Accident Analysis and Prevention 35 (2), 253-259.

Caliendo, C., Guida, M., Parisi, A., 2007. A crash-prediction model for multilane roads. Accident Analysis and Prevention 39 (4), 657-670.

Chan, C.Y., Huang, B., Yan, X., and Richards, S., 2010. Investigation effects of asphalt pavement conditions on traffic accidents in Tennessee based on the pavement management system (PMS). J. Adv. Transp. 44(3). P. 150-161.

Cleveland, D.E., 1987. Effect of resurfacing on highway safety. In Relationship between safety and key highway features. Edited by E.T. Crump. State of the Art Report 6, National Research Council, Washington, D.C., pp. 78–95.

Cameron, A. C., and P. K. Trivedi. 1998. Regression Analysis of Count Data. Cambridge, U.K.: Cambridge University Press.

Donnell, E.T., Mason Jr., J.M., 2006. Predicting the frequency of median barrier crashes on Pennsylvania interstate highways. Accident Analysis and Prevention 38, pp. 590-599.

Guo, F., Wang, X.S., Abdel-Aty, M.A., 2010. Modeling signalized intersection safety with corridor-level spatial correlations. Accident Analysis and Prevention 42, pp. 84-92.

Greene, W. Econometric Analysis, 4th Ed. Prentice Hall, 1999.

Garber, N. J., Ehrhart, A. A., 2000. Effect of Speed, Flow, and Geometric Characteristics on Crash Frequency for Two-Lane Highways. Transport. Res. Record 1717, pp. 76–83.

Hausman, J.A., Hall, B.H., Griliches, z., 1984. Econometric models for count data with an application to the patents-R&D relationship. Econometrica 52, pp. 909-938.

Hadi, M.A., Aruldhas, J., Chow, L.F., Wattleworth, J.A., 1995. Estimating safety effects of cross-section design for various highway types using negative binomial regression. Transportation Research Record 1500. pp.169-177.

Jacobs, G.D, 1976. A study of accident rates on rural roads in developing countries. Transport and Road Research Laboratory, Crowthorne, United Kingdom, TRRL Laboratory Report 732.

Jiang, X.M., Yan, X.D., Huang, B.S., Richards, S., 2011. Influence of Curbs on Traffic Crash Frequency on High-Speed Roadways. Traffic Injury Prevention 12 (4), pp. 412-421.

Kamel, N., Garshore, T., 1982. Ontario's wet pavement accidents reduction program. In Pavement characteristics and materials. STP 763. Edited by C.M. Hayden. American Society for Testing and Materials, Philadelphia, Pa., pp. 98–117.

Knuiman, M.W., Council, F.M., Reinfurt, D.W., 1993. Association of median width and highway accident rates. Transp. Res. Rec. 1401, pp. 70-82.

Kweon, Y.J., Kockelman, K.M., 2005. Safety Effects of Speed Limit Changes Use of Panel Models, Including Speed, Use, and Design Variables. Transportation Research Record 1908. pp.148-158.

Lord, D., 2002. Issues related to the application of accident prediction models for the computation of accident risk on transportation networks. Transportation Research Record 1784, 17-26.

Lord, D., Manar, A., Vizioli, A., 2005. Modeling crash-flow-density and crash-flow-v/c ratio for rural and urban freeway segments. Accident Analysis and Prevention 37 (1), 185–199.

Ma, J., Kockelman, K.M., Damien, P., 2008. A multivariate Poisson-lognormal regression model for prediction of crash counts by severity, using Bayesian methods. Accident Analysis and Prevention 40, 964–975.

Özbay, K., Laub,R., 2001. Models for pavement deterioration using LTPP: final report. Federal Highway Administration, Report No.: FHWA-NJ-1999-030.

Patterson, T. L., Frith, W. J., Povey, L. J., Keall, M. D., 2002. The Effect of Increasing Rural Interstate Speed Limits in the United States. Traffic Injury Prevention, Vol. 3, pp. 316–320.

Quinn, B.E., Hilderbrand, S.E., 1973. Effect of Road Roughness on Vehicle Steering, Highway Res. Rec., 47, pp.62-75.

Rigden P.J. *Skid resistance of roads and streets.* CSIR Special Report Pad 64, Pretoria, 1988.

Strat, M.R., Kim, J., Berg, W.D., 2004. Potential Safety Cost-Effectiveness of Treating Rutted Pavements; Transport. Res. Record 1629, pp.208-213.

Shankar, V.N., Albin, R.B., Milton, J.C., Mannering, F.L., 1998. Evaluating Median Crossover Likelihoods with Clustered Accident Counts: An Empirical Inquiry Using the Random Effects Negative Binomial Model. Transport. Res. Record 1635, pp.44-48.

Tighe, S., Li, N., Falls, L. C., Haas, R., 2000. Incorporating Road Safety into Pavement Management. In Transportation Research Record 1699, TRB, National Research Council, Washington, D.C., pp. 1-10.

Ulfarsson, G.F., Shankar, V.N., 2003. Accident Count Model Based on Multiyear Cross-Sectional Roadway Data with Serial Correlation. Transport. Res. Record 1840, pp.193-197.

Wang, X.S., Abdel-Aty, Mohamed, 2006. Temporal and spatial analyses of rear-end crashes at signalized intersections. Accident Analysis and Prevention 38, pp. 1137-1150.

Wambold, J.C., Zimmer, R.A., Ivey, D.L., Sicking, D.L., 1984. Roughness, holes and bumps. In State of the art report 1, TRB, National Research Council, Washington, D.C..

Zhang, C., Ivan, J.N, 2005. Effects of Geometric Characteristics on Head-On Crash Incidence on Two-Lane Roads in Connecticut. Transportation Research Record 1908. pp.159-164.

PART 5 TWO-VEHICLE INJURY SEVERITY MODELS BASED ON INTEGRATION OF PAVEMENT MANAGEMENT AND TRAFFIC ENGINEERING FACTORS

5.1 Abstract

The severity of traffic-related injuries has been studied by many researchers in recent decades. A growing body of literature has been focusing on the effects of traffic engineering factors on the outcome of crashes. However, the evaluation of many factors is still in dispute and until this point, few studies have taken into account pavement management factors as points of interest. The objective of this part of research is to evaluate the combinative influences of pavement management factors and traditional traffic engineering factors on the severity of two-vehicle crash injuries.

In this study, three types of two-vehicle crashes occurred on the state routes of Tennessee from 2004 to 2008 are studied, respectively. They are Rear-End, Sideswipe, and Angle collisions. Both the traditional Ordered Probit (OP) model and Bayesian Ordered Probit (BOP) model with weak informative prior are fitted for each collision types. The performances of these models are evaluated based on the parameter estimates and deviances.

The results indicate that pavement management factors played identical roles in all three collision types. Pavement Serviceability (PSI) is positively related to the severity of injuries, i.e., rougher pavements are associated with less severe injuries. Pavement Distress Index (PDI), Rutting Depth (RD) and Rutting Depth Difference between right and left wheels (RD_df) are not significant in affecting injury severity in any of these three collision types. The effects of traffic engineering factors vary across collision types:

some are consistently significant in all three collision types, such as AADT, rural urban location, speed limit, peaking hour, and light condition; some are significant in one or two of these collision types, such as median width, weather condition, driver's sex and age, as well as vehicle's type; the others are not significant in any of these collision types, such as lane width, vertical alignment and type of terrain.

The study also found that the BOP model with weak informative prior can be taken as a great alternative, but is not superior to traditional OP model in terms of the overall performance.

5.2 Introduction

Along with the increased number of automobiles in use all over the world, the traffic safety issue has become more critical than ever before. Numerous traffic crash studies have attempted to determine the influences of traffic engineering factors on both the occurrence and outcome of crashes. Among them, a certain number of studies have specifically focused on the effects of variables on the severity of traffic-related injuries. However, relatively few studies have taken into accounts of pavement management factors, such as pavement cracking, rutting and roughness, as possible causes of traffic-related crashes or the severity of corresponding injuries.

According to previous studies in the Long-Term Pavement Performance (LTPP) Program, numerous pavements have been experiencing rapidly wear off, which further leads to structurally deficient and functionally obsolete of pavements (Özbay and Laub, 2001).

High level of pavement distress may either lead drivers to lose control of their vehicles or cause vehicles to change trajectories, hence increase the dangers of triggering crashes (Quinn and Hildebrand, 1973; Wambold et al., 1984). In addition, observable defects such as pothole and severe rutting, may distract drivers and encourage drivers to shy away from them, which may result in vehicles' collisions or running out of driveways (Tighe et al., 2000). Treat et al. (1979) conducted a three level (baseline data collection, on site investigation, multidisciplinary in depth analysis) study on the causes of traffic crashes. The results of the study indicated that road environmental factors of all kinds, including slick road were causes for at least 12.4% of crashes, and were probable causes for about 33.8%-34.9% of crashes investigated.

Despite the significance of pavement qualities in affecting either driver or vehicle's behavior, statistically study of their impacts on traffic safety has been rare. Jacobs (1976) and Kamel and Garshore (1982) reported that the increase in road roughness would increase crash experience on rural roads. Recent researches conducted by Anastasopoulos et al. (2008; 2012), Anastasopoulos and Mannering (2009) found the similar effects of roughness on urban Interstate highways. On the contrary, Cleveland (1987) reported that the occurrences of crashes increased slightly but significantly with a smoother pavement surface on two-lane rural roads having the AADT range from 1000 to 8000 vehicles per day. Moreover, Al-Masaeid (1997) found that rougher roads would decrease the single-vehicle crash rate but increase the multiple-vehicle crash rate, which is consistent to the research conducted by Tighe et al. (2000).

Strat et al. (2004) conducted a study to quantify how pavement rutting affects crash rates. The results indicated that the defined rut-related crash rate begins to increase at a significantly greater rate as Rutting Depth (RD) exceeds 7.6 mm (0.3 in.). This study also reported that rutting is more hazardous in wet weather when water accumulates in the rut path and leads to hydroplaning. Similarly, Anastasopoulos et al. (2008) employed tobit model to identify how factors affecting highway crash rate using 5-year data from urban interstates in Indiana. They also found that lower rutting depth were associated with lower crash rate. Their successive research utilizing random-parameter model (Anastasopoulos and Mannering, 2009; Anastasopoulos et al., 2012) consolidated this finding. Moreover, Chan et al. (2010) suggested that high level of rutting depth and roughness produce significantly adverse effects on the occurrence of crashes occurred on Interstate highways during nighttime and under rainy weather conditions.

Unfortunately, when the authors investigated into literatures, no study has accounted for pavement management factors as related to traffic injury severity. Previous studies have primarily focused on the influences of traffic engineering factors on traffic-related injury severity. These traffic engineering factors generally include roadway features (number of lanes, lane, median and shoulder width, horizontal and vertical alignment, et al.), environmental factors (weather and light conditions , traffic condition, et al.), driver conditions (age, gender, alcohol et al.), vehicle attributes (passenger car, single-unit truck, multi-unit truck et al.), as well as collision manners (rollover, sideswipe, angle, et al.).

A large number of crash injury studies have attempted to identify driver factors that have contributed to the traffic injury severity, such as drinking and driving (Baker et al., 2002; Smink et al., 2005; Lee and Abdel-Aty, 2005), restraint device use (Valent et al., 2002; Bedard, 2002；Abdel-Aty, 2003), as well as age and gender (Lyman et al., 2002; Zhang et al., 2000; Boufous et al., 2008). Zajac and Ivan (2003) and Keall et al. (2004) reported that drinking and driving can significantly increase the risk of fatal crashes. Khattak et al. (1998) found that older drivers relative to younger drivers have a greater likelihood of severe injury, and male drivers are likely to be more severely injured than females. Similarly, Finison and Dubrow (2002) revealed that the risk of hospitalization and death for older drivers increases by 3.5% for every year's increase in age. However, a few other scholars (Chipman et al., 1992; Kockelman and Kweo, 2001; Bauer et al., 2003) indicated that older drivers generally display safe driving behaviors, including lower speed, wearing seatbelts, no alcohol and so forth, which may lead to less severe crashes.

Many researchers considered environmental conditions as potential factors that may affect traffic-related injury severity (Brodsky and Hakkert, 1998; Finison and Dubrow, 2002; Baker et al., 2003; Awadzia et al., 2008). Adams (1985) concluded that practical speeds decrease in adverse weather condition, which leaded to less severe crashes. Similarly, Khattak et al. (1998) performed a study to explore the role of adverse weather in key crash types on limited-access roadways. They concluded that adverse weather significantly decreases crash severity. In contrast, the National Transportation Safety Board (1980) reported that the risk of a fatal crashes in the United States as a whole was 3.9 to 4.5 time greater on wet than on dry pavements. On the other hand, Krull et al.

(2002) analyzed driver injury severity involved in single-vehicle crashes. The results indicated that driver injury severity increased in daylight conditions relative to nighttime. Conversely, Abdel-Aty (2003) concluded that dark lighting conditions contributed to higher probability of injury on roadway sections.

Additionally, there are quite a few other factors that have been proved to be significant in impacting the severity of traffic-related injuries. For example, Shankar et al. (1996) performed a study to investigate the impact of a broad range of variables on the severity of crash injuries. They indicated that the increased number and percent grade of horizontal curves are more likely to result in higher level injuries. Noland and Oh (2002) conducted a study on the effect of infrastructure changes on traffic fatalities and crashes based on Illinois county-level data. The results showed that increased number of lanes, lane widths and outside shoulder widths were associated with higher traffic-related fatalities. In addition, vehicle type (Chang and Mannering, 1999; Kockelman and Kweon, 2001), collision manner (Rifaat and Chin, 2007; Rifaat and Tay 2009; Lee and Mannering, 2002), driver and vehicle action (Chang and Wang, 2006), traffic volume (Khattak et al., 1998; Chang and Wang, 2006), speed limit (Kockelman and Bottom, 2006; Malyshkina and Mannering, 2008), roadway location (Lee and Mannering, 2002; Rifaat and Tay 2009), cause of accidents (Hutchinson, 1986; Al-Ghamdi,2001), were also found to be critical in determining the severity of injuries caused by traffic crashes.

Methodologically, a large variety of statistical models have been used to evaluate the influence factors on the severity of traffic-related injuries (Savolainen et al., 2011).

Among them, ordered logit and probit models have been frequently employed (O'Donnell and Connor, 1996; Shimamura and Fujita, 2005). Traditional ordered logit and probit models are based on Maximum Likelihood (ML) estimation, which is a classic method for model fitting. However, ML estimation is very sensitive to the quality of data. If the data cannot fully represent the characteristics of each population, the result is likely to be biased. In response, Bayesian inference was introduced into the analysis for the severity of traffic-related injuries. For example, Xie et al (2009) performed a series of studies on the prediction of severity of crash injuries utilizing both the Ordered Probit (OP) model and the Bayesian Order Probit (BOP) model with various priors. The result shows that the BOP model outperformed the OP model in both the parameter estimates and the predictive capabilities, especially for small data sample.

As has been summarized above, numerous traffic engineering factors related to traffic injury severity have been investigated and the results generated so far have made considerable contribution to the improvement of roadway safety. However, few studies have specifically focused on two-vehicle collisions, and even less of them have identified the effects of factors separately by different types of collisions, such as Rear-End, Sideswipe and Angle collisions. In addition, the conclusions of previous studies are not consistent and until this point, few studies have taken into account pavement management factors as points of interest to predict the severity of traffic related injuries. Hence, a re-evaluation of traditional traffic engineering factors under different two-vehicle collision types, incorporating pavement management factors is desired.

The objective of this part of research is to explore the effects of pavement management and traffic engineering factors on the outcome of two-vehicle crashes given that the collisions have occurred. The database used in this study is based on crashes on Tennessee State Route highways from 2004 to 2008. Three types of two-vehicle collisions are studied including Rear-End collisions, Sideswipe collisions, Angle collisions. Both the OP models and the BOP models were established and the performance of the BOP models as compared to the OP models is evaluated based on their parameter estimates and deviances.

5.3 Methodology

5.3.1 Data Preparation

The primary source of traffic crash data is the Tennessee Roadway Information Management System (TRIMS). The data were obtained in computer-ready forms, which included coded information on reported crashes that occurred on state highways in Tennessee. The coded information for each crash contains important attributes describing the conditions contributed to the collision and the outcome. The information regarding pavement management status was obtained from the Pavement Management System (PMS) maintained by the Tennessee Department of Transportation (TDOT). The crash data occurred on State Route roadways from 2004 to 2008 and corresponding traffic engineering data were linked to the pavement management factors through the common variable: id_number, which is a combination of county, county sequence, route type, and route number.

Pavement management factors considered in this part of research include Present Serviceability Index (PSI), Pavement Distress Index (PDI), Rutting Depth (RD), as well as Rutting Depth Difference (RD_df). RD_df represents the difference of RD between right and left side of vehicles when measured. PSI is a measure of the roughness of roadways on a scale of 0-5, representing worst to perfect; PDI is specified as a scale from 0-5 representing worst to perfect of the overall pavement quality including the sensitivity of drivers to each type of pavement distress. Distresses that are evaluated in PDI include fatigue, rutting, longitudinal cracks, patching, block cracking, raveling, transverse cracks; RD is the original measurement of rutting depth in inch.

In the State of Tennessee, roughness, rutting, cracks and other pavement defects are measured once every two years on State Route highways. Measurements are recorded for each 0.1 mile (0.16 km) pavement section. The pavement management data from 2004 to 2009 were collected, screened and treated according to the following criteria:

- Records with PSI and PDI less than 1 were excluded because they may subject to missing value or in a short section where there may be a railroad crossing, intersection, or other discontinuity;
- Records with RD in either left wheel or right wheel greater than 1.0 inch (25.4mm) were removed since they are abnormal in terms of the overall pavement quality for the State Route in Tennessee.
- PDI, PSI, RD and RD_df records in each year were substituted by the combination of the original records in the current year and the records of next year because the pavement quality indices were measured every two years as

mentioned earlier. For example, the pavement quality information for 2004 was a combination of those measured in 2004 and 2005, and the pavement quality information for 2005 was a combination of those measured in 2005 and 2006, and so on…

The pre-treated pavement management observations from 2004 to 2008 were linked to traffic accidents by both road section and the year, i.e. crashes happened in each year were linked to the pavement quality condition records measured in the same year. The overall statistics of the pavement management factors for Tennessee State Routes after preliminarily processing are shown in Table 5.1.

The traffic engineering factors considered in the present study include: types of terrain, rural or urban, lane width, median width, Annual Average Daily Traffic (AADT), speed limit, grade of horizontal alignment, peak hour, light condition, weather condition, vehicle type, driver's age, as well as driver's gender. Thousand of AADT (thAADT) was employed in the regression modeling as an exposure variable, which was calculated through dividing AADT by 1000, since the change of injury severity with increment by one vehicle would be too trivial. Some of these factors were categorized into subclasses from the original records as shown in Table 5.2.

The combined dataset of pavement management information, traffic engineering data and collision records was pre-treated according to the following criteria:

Table 5.1 Distribution of overall pavement management factors

Variables	Min	Max	Median	Mean	Std
PDI	1.00	5.00	5.00	4.62	0.58
PSI	1.00	4.50	3.34	3.25	0.57
RUT	0.00	0.94	0.07	0.09	0.08
RUT_df	0.00	0.99	0.03	0.05	0.06

Table 5.2 Variables classification criteria

Variables	Description	Categories & Classification Criteria	
PDI	Pavement Distress Index	No_dfct	PDI=5
		Minor_dfct	PDI ≥4 & PDI<5
		Severe_dfct	PDI < 4
PSI	Present Serviceability Index	Good	PSI > 3
		Fair/Bad	PSI ≤ 3
RD	Rut Depth	Shallow	rut depth ≤ 0.2''
		Med/Deep	rut depth > 0.2''
RD_df	Rut Depth difference	Small	rut depth difference ≤ 0.2''
		Large	rut depth difference > 0.2''
tyterrain	type of terrain	Flat	flat region
		Rolling	rolling region
		Mountain	mountain region
rururb	rural or urban	Rural	rural area
		Urban	urban area
light	light condition	Daylight	daylight
		Nighttime	dark lighted and unlighted, dawn and dusk
weather	weather condition	Clear	Clear
		Inclement	rainy, foggy, snowy, etc.
phour	peak hour	No	10:00-16:00 & 19:00-6:00
		Yes	6:00-10:00 & 16:00-19:00
vehtype	vehicle type	Pass_Pass	passenger car[a] to passenger car
		Lgt_Pass	light truck[b] to passenger car
		Lgt_Lgt	light truck to light truck
		Hvy_inv	heavy truck[c] involved
drvage	driver's age	Yng_Yng	Youngers[d] to youngers
		Midage_Yng	Middle-aged[e] to youngers
		Midage_Midage	Middle-aged to middle aged
		Old_inv	older drivers[f] involved
drvsex	driver's gender	M_M	male to male
		M_F	male to female
		F_F	female to female
spd_lmt	speed limit	30, 35, 40, 45, 50, 55 mph	

(a. convertible, hatchback, sedan hardtop, etc. ; b. pickup, SUV, minivan, light truck, etc.; c. large van, bus, heavy truck, etc; d. driver's age in between 16 and 24; e. driver's age in between 25 and 64; f. driver's age ≥ 65.)

- Rear-End, Sideswipe and Angle crashes were selected for this study since they are the most common types of two-vehicle crashes;
- All the intersection related collisions were removed since the crash patterns are different from those occurred in continuous roadways;
- The observations with missing values either in the dependent variable or in each of the predictors were excluded for the purpose of completeness;
- Speed limits of 30 mph (48 km/hr),35 mph (56 km/hr),40 mph (64 km/hr),45 mph (72 km/hr),50 mph (80 km/hr) and 55 mph (88 km/hr) were selected for this study since they are quite common for Tennessee State Route highways;
- Road sections with 1, 2, 3 or 4 lanes in each direction were investigated in details in that they are typical for State Route;
- Pedestrian, motorcycle, and non-traditional types of vehicles involved accidents were screened out because they have different injury mechanisms.

The treated dataset applied in this part of research includes 50,908 total two-vehicle crashes that occurred on state routes within the 5 years from 2004 to 2008. In this study, the severity level of each crash was determined by the injury level of the worst-injured occupant in all the vehicles involved. Table 5.3 provides the distribution of injury severity for some key factors, as well as summarized statistics of continuous variables considered in this study. The overall crashes were further grouped into three classes according to the collision types in each crash: Rear-End collisions, Sideswipe collisions, and Angle collisions. As a result, the numbers of observations selected for two-vehicle Rear-End collisions, Sideswipe collisions and Angle collisions are 27456, 8911, and

Table 5.3 Injury severity distribution on key variables and the description of continuous variables

Distribution of injury severity by categorical variables					
Variables	Categories	Injury Severity			
		1 (No injury)	2 (Minor injury)	3 (Incapacitating)	4 (Fatal)
PDI	No_dfct	77.77	20.57	1.35	0.31
	Minor_dfct	78.71	19.79	1.28	0.22
	Severe_dfct	79.71	19.16	1.05	0.09
PSI	Good	75.66	22.27	1.67	0.40
	Fair/Bad	80.98	18.03	0.90	0.09
RD	Shallow	78.25	20.18	1.31	0.26
	Med/Deep	79.66	19.18	1.04	0.12
RD_df	Small	78.38	20.11	1.27	0.24
	Large	81.34	17.57	1.03	0.06
Light	Daylight	79.17	19.50	1.15	0.18
	Nighttime	75.04	22.61	1.84	0.50
weather	Clear	78.33	20.16	1.29	0.22
	Inclement	79.45	19.13	1.12	0.31
phour	No	77.57	20.75	1.40	0.28
	Yes	79.67	19.07	1.08	0.18
tyterrain	Flat	80.41	18.26	1.14	0.19
	Rolling	78.42	20.09	1.26	0.24
	Mountain	73.19	23.43	2.90	0.48
rururb	Rural	70.37	26.05	2.63	0.95
	Urban	79.68	19.13	1.06	0.13
spd_lmt	30 mph	82.19	16.89	0.90	0.02
	35 mph	81.01	17.84	1.10	0.05
	40 mph	80.24	18.90	0.79	0.07
	45 mph	78.05	20.37	1.37	0.21
	50 mph	78.40	19.80	1.22	0.58
	55 mph	70.44	26.17	2.54	0.85
vehtype	Pass_Pass	76.75	21.97	1.14	0.14
	Lgt_Pass	78.91	19.57	1.27	0.26
	Lgt_Lgt	80.34	18.13	1.27	0.26
	Hvy_inv	78.26	19.51	1.77	0.46
drvage	Yng_Yng	82.56	16.19	1.05	0.20
	Midage_Yng	78.45	20.34	1.07	0.14

Table 5.3 (continued)

Distribution of injury severity by categorical variables					
Variables	Categories	Injury Severity			
		1 (No injury)	2 (Minor injury)	3 (Incapacitating)	4 (Fatal)
	Midage_Midage	77.90	20.52	1.35	0.23
	Old_inv	78.38	19.74	1.45	0.42
drvsex	M_M	79.92	18.35	1.37	0.36
	M_F	77.89	20.64	1.26	0.21
	F_F	77.75	21.02	1.12	0.12

Statistical description of continuous variables						
Variables	Label	Min	Max	Mean	Median	Standard deviation
thAADT	Thousand of AADT	0.08	131.14	23.32	22.15	14.43
lanewid	Lane width (ft)	8.00	17.00	11.66	12.00	0.76
medwid	Median width (ft)	0.00	60.00	5.42	0.00	11.27
pct_grde	Percent grade of vertical alignment	0.00	10.00	1.17	0.30	1.56

14541, respectively.

5.3.2 The OP and BOP Model

The OP model is commonly used for analyzing data sets when the response variable is inherently ordered categorical data. A typical example is a survey where the respondents are asked to rate the quality of service on a scale of 1 to 5. The central idea for the OP model is that there is a latent continuous metric underling the ordinal responses. Let $x_i = \{ x_{i1}, \cdots, x_{ij}, \cdots, x_{im}\}^{\mathrm{T}}$ represents variables that may affect the injury severity, where i indicates the ith crash, j indicates the jth variable and m indicates the total number of explanatory variables. The latent variable z_i is assumed to be expressed as:

$$z_i = x_i^T \beta + \varepsilon_i \qquad i = 1, \cdots, n \tag{5.1}$$

where $\beta = \{\beta_1, \cdots, \beta_j, \cdots, \beta_m\}$, is a vector of parameters to be estimated and ε_i is a random error term (assumed to follow an independent and identically distributed standard normal distribution). n is the total number of crashes.

Let y_i be a categorical random variable with C categories that represents the injury severity. In the current work the response has 4 categories: no injury, minor injury, incapacitating injury, and fatality. The observed variable $Y = \{y_1, \cdots y_i \cdots y_n\}$ can be connected to the latent variable $Z = \{z_1, \cdots z_i \cdots z_n\}$ through a function $g(Z)$:

$$y_i = g(z_i) = \begin{cases} 1 & if -\infty = \gamma_0 < z_i \le \gamma_1 \\ 2 & if \quad \gamma_1 < z_i \le \gamma_2 \\ \vdots & \vdots \\ c & if \quad \gamma_{c-1} < z_i \le \gamma_c = +\infty \end{cases} \tag{5.2}$$

where $\gamma = \{\gamma_0, \gamma_1, \cdots, \gamma_{c-1}, \gamma_c\}$, is the threshold value for all categories, wherein $\gamma_0 = -\infty, \gamma_c = +\infty$, and the remaining threshold values are subjected to the constraint $\gamma_1 \le r_2 \le \cdots \le \gamma_{c-1}$. The function $g(Z)$ is taken to be non-decreasing, so that small and large values of z_i can be interpreted as corresponding to small and large values of y_i. This also means that the sign of a regression coefficient β_j indicates whether Y is increasing or decreasing with x_j.

Given the value of x_i, the probability that the injury severity of individual i belongs to each category is

$$P(y_i = 1) = \Phi(\gamma_1 - x_i^T \beta)$$

$$P(y_i = 2) = \Phi(\gamma_2 - x_i^T \beta) - \Phi(\gamma_1 - x_i^T \beta)$$

$$P(y_i = c) = 1 - \Phi(\gamma_{c-1} - x_i^T \beta) \tag{5.3}$$

where $\Phi(\cdot)$ stands for the cumulative probability function of the standard normal distribution.

To calculate β and γ, one restriction is applied to the threshold values to make γ_1=0 (Mckelvey and Zavoina, 1975). In the classical OP model, the values of β and $\gamma^* =$

$\{\gamma_2, \cdots \gamma_{c-1}\}$ can be determined by the ML estimation method. Based on the knowledge in the above text, the likelihood function can be presented as Equation 5.4.

$$L(\beta, \gamma^*|y) = \prod_{i=1}^{n} \prod_{k=1}^{c} [\Phi(\gamma_k - x_i^T \beta) - \Phi(\gamma_{k-1} - x_i^T \beta)]^{I(y_i=k)} \tag{5.4}$$

Where $k = (1,2,\cdots,c)$, and $I(y_i = k)$ is an indicator function: if $y_i = k$, then $I(y_i = k)$ is 1, otherwise 0. The parameters β and γcan be determined by maximizing$L(\beta, \gamma^*|y)$.

However, there are several limitations with the method. The most critical issue is that the parameter estimation results depend completely on the data, which may bring about bias in the estimated parameters when the data cannot represent the population. In addition, the maximization process is a nonlinear optimization problem, which does not guarantee a convergence to a global optimal solution (Mckelvey and Zavoina, 1975).

Due to the limitations of ML estimation method, Bayesian inference was introduced to estimate the coefficients, denoted as BOP model. If a normal prior distribution was applied to β and γ, the joint posterior distribution of {β, γ_1, …,γ_{c-1}, z_1,…, z_n} given Y=($y_1, y_2, \dots, y_n$) can be approximated using a Gibbs sampler. Gibbs sampling for the BOP model is based upon a set of full conditional distributions which are described as below.

- Conditional distribution of β

Given $Y = y$, $Z = z$, and $\gamma = (\gamma_1, r_2, \cdots, \gamma_{c-1})$, the conditional distribution of β depends only on z and satisfies $p(\beta|y, z, \gamma) \propto p(\beta)p(z|\beta)$. Assuming β follows a multivariate normal distribution as shown in Equation 5.5,

$$\beta \sim multivariate\ normal(0, n(X^T X)^{-1}) \tag{5.5}$$

then, the posterior distribution $p(\beta|z)$ is multivariate normal with

$$var[\beta|z] = \frac{n}{n+1}(X^T X)^{-1}, \text{ and}$$

$$E[\beta|z] = \frac{n}{n+1}(X^T X)^{-1} X^T z \tag{5.6}$$

- Conditional distribution of Z

Under the sampling model, the conditional distribution of z_i given β is $z_i \sim normal(\beta^T x_i\,, 1)$. Given γ, observing $Y_i = y_i$ indicates that z_i must lie in the interval $(\gamma_{y_i - 1}, \gamma_{y_i})$. Letting a=$\gamma_{y_i - 1}$ and b=γ_{y_i}, the full conditional distribution of z_i given {β,y, γ} is

$$p(z_i|\beta, y, \gamma) \propto \mathrm{dnorm}(z_i, \beta^T x_i, 1) \times \delta_{(a,b)}(z_i) \tag{5.7}$$

where dnorm() represents the probability density function of a normal distribution. The full conditional distribution of Z is a constrained normal distribution: a $normal(\beta^T x_i\,, 1)$ distribution constrained to the interval (a, b).

- Full conditional distribution of γ

Suppose the prior distribution for γ has an arbitrary density $p(\gamma)$. Given $Y = y$, $Z = z$, it can be observed from Equation 5.2 that γ_k must be higher than all $z_i's$ for which $y_i = k$ and lower than all $z_i's$ for which $y_i = k + 1$. Letting $a_k = \max\{z_i: y_i = k\}$ and $b_k = \min\{z_i: y_i = k + 1\}$, the full conditional distribution of γ is then proportional to $p(\gamma)$ but constrained to the set $\{\gamma: a_k < r_k < b_k\}$.

The algorithm employed is discussed in depth by Cowles (1996). Both the OP and BOP model was constructed using the *Zelig* package (Goodrich and Lu, 2007) in R software (2011) which calls the function MCMCoprobit from the MCMCpack package (Andrew et al., 2011). The normal prior with both the mean and precision parameter for each coefficient equal to zero were employed in the BOP model, which is known as a weak informative prior. The corresponding posterior distribution of $\{\beta, z, \gamma\}$ was approximated with a Gibbs sampler consisting of 17,000 scans. The first 2,000 iterations were removed to meet stationary status of simulation. The samples were saved every 15th scan and resulted in 1,000 values for each parameter with which to approximate the posterior distribution.

In the following sections, both the OP and BOP models were built to quantitatively estimate the influence of each covariate on traffic-related injury severity, in three types of two-vehicle collisions. The performance of BOP models was evaluated as compared to the corresponding OP models based on the parameter estimates and deviances

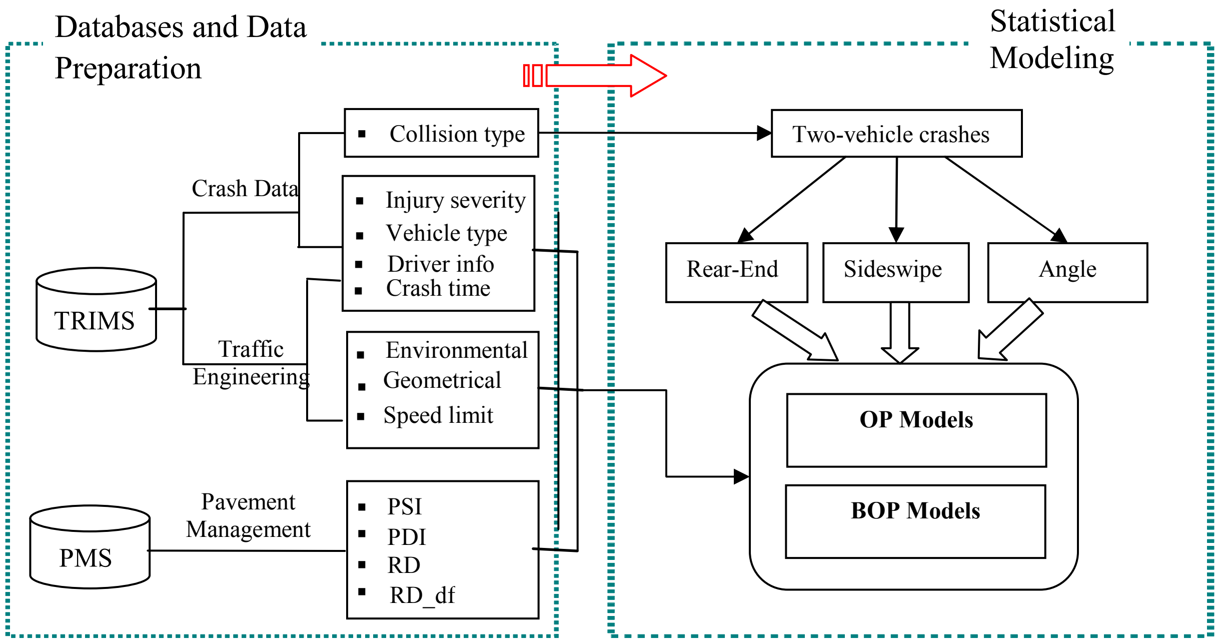


Figure 5.1 Flow chart of methodology

(-2*log*Likelihood*). The smaller the deviance is, the better the model fits. Figure 5.1 illustrates the overall structure of the methodology.

5.4 RESULTS

Both the OP model and BOP model were built for each type of two-vehicle collisions in this section. The parameter estimates and corresponding p-values of Type 3 Chi-Square tests are provided in the OP models. Type 3 Chi-Square tests evaluate the significance of the variables in the model individually. A P-value less than 0.05 indicates significance of this variable in this model, and vice versa. The means and corresponding 2.5%, 97.5% quantiles of posterior estimates computed from Markov Chain Monte Carlo simulation are presented in the BOP models. The 2.5% and 97.5% quantiles construct the 95% confidence interval (CI), indicating where the posterior estimate for each parameter locates: if it does not contain 0, i.e., both the 2.5% quantile and 97.5% quantile are negative or positive, it can be deemed that this attribute is significant in affecting the injury severity as compared to the reference attribute; if it does contain 0, then it can be concluded that this attribute does not produce a significant effect on the injury severity.

5.4.1 Two-vehicle Rear-End Collisions

Modeling results from both the OP model and the BOP model for two-vehicle Rear-End collisions are provided in Table 5.4.

It can be observed from Table 5.4 that the OP model and the BOP model produce quite identical parameter estimates. In addition, the 95% CI obtained from the BOP model is

Table 5.4 Parameter estimates for two-vehicle Rear-End collisions

Variable	Classes	OP model		BOP model		
		Parameter estimates	Pr > γ^2	Mean	2.5%	97.5%
PSI	Good					
	Fair/Bad	-0.0881	<.0001	-0.0887	-0.1258	-0.0528
PDI	No_dfct					
	Minor_dfct	0.01724	0.3718	0.017	-0.0204	0.0558
	Severe_dfct	0.0084	0.7174	0.0085	-0.0399	0.0554
RD	Shallow					
	Med/Deep	0.0286	0.2408	0.0291	-0.0191	0.0754
RD_df	Small					
	Large	-0.0648	0.1961	-0.0668	-0.1642	0.0314
thAADT	NA	-0.0042	<.0001	-0.0042	-0.0055	-0.0030
lanewid	NA	0.0051	0.6724	0.0051	-0.0198	0.0285
medwid	NA	-0.0025	0.0011	-0.0025	-0.0041	-0.0011
pct_grde	NA	0.0083	0.1307	0.0083	-0.0024	0.0189
rururb	Urban					
	Rural	0.1092	0.0002	0.1083	0.0537	0.1679
spdlmt	30 mph					
	35 mph	0.0490	0.2212	0.0457	-0.0382	0.1206
	40 mph	0.0329	0.3482	0.0317	-0.0382	0.1019
	45mph	0.0750	0.0316	0.0724	0.0035	0.1404
	50 mph	-0.0057	0.9023	-0.0101	-0.1030	0.0790
	55 mph	0.2699	<.0001	0.2691	0.1885	0.3475
tyterrain	Flat					
	Rolling	0.0469	0.2365	0.0442	-0.0289	0.1208
	Mountain	-0.0655	0.5523	-0.0620	-0.2795	0.1506
weather	Clear					
	Inclement	-0.0810	0.0010	-0.0803	-0.1274	-0.0318
light	Daylight					
	Nighttime	0.1930	<.0001	0.1912	0.1441	0.2364
phour	Yes					
	No	0.0898	<.0001	0.0899	0.0556	0.1241
drvsex	M_M					
	M_F	0.0640	0.0014	0.0646	0.0257	0.1056
	F_F	0.0912	0.0002	0.0922	0.0456	0.1395
drvage	Yng_Yng					
	Midage_Yng	0.2392	<.0001	0.2396	0.1742	0.3108
	Midage_Midage	0.3099	<.0001	0.3105	0.2409	0.3775
	Old_inv	0.2976	<.0001	0.2983	0.2264	0.3747
vehtype	Pass_Pass					
	Lgt_Pass	-0.0745	0.0002	-0.0743	-0.1126	-0.0356
	Lgt_Lgt	-0.1232	<.0001	-0.1236	-0.1736	-0.0749
	Hvy_inv	0.0622	0.0974	0.0617	-0.0089	0.1325
gamma1	NA	1.1577	<.0001	1.1536	0.8481	1.4606
gamma2	NA	2.7452	<.0001	2.7405	2.3906	3.0917
gamma3	NA	3.5842	<.0001	3.5842	3.1611	4.0144
Deviance	NA	30703.79		30703.80		

fully consistent with the p-value obtained from the OP model, in terms of the significance of each variable. The deviance of the OP model (30703.79) is almost the same to that of the BOP model (30703.80). The similarity of these two models might be attributed to two reasons: 1, the weak informative prior was employed in the BOP model, which does not bring too much information to the posterior estimates; 2, the sample size employed in these models is very large. When sample size is large enough, Bayesian and the MLE methods will generally produce similar results (Xie et al; 2009).

To the authors' point of interest, both the OP and BOP models show significant effects ($p<0.0001$) of PSI on the severity of injures for two-vehicle Rear-End crashes. The "Fair/Bad" indicator of PSI is associated with less severe injuries than "Good" PSI. In other words, Rear-End crashes occurring on rougher roads are less likely to be severe injuries. This might contradict to the common sense because driving on rough roads are less comfortable than driving on smooth roads. However, rough roads may lead drivers to operate relatively lower speed and pay more attention on driving. In addition, roughness has been proved to have the ability to reduce braking distance (Reul and Winner, 2009), which may also serve to relieve the severity of injuries when crashes have occurred, particularly in Rear-End crashes. When pros outperform cons, it is highly possible to see the positive effects of roughness on the severity of injuries.

These models also reveal to us that PDI, rutting depth (RD), and difference between RDs on right and left wheel are not associated with the outcome of two-vehicle Rear-End collisions. This is understandable. In contrast to PSI, the indices PDI, RD and RD_df are measures of pavement distress, especially observable defects such as rutting, cracking and pot hole. On one hand, drivers may slow down and pay more attention to their driving when they notice road distress. On the other hand, the severe distress may cause vehicles' loss of control when braking or turning, or lead drivers to shy away from them and thus hitting other vehicles or fixed objects, running out of the lane, or even rolling over. The insignificance of these variables implies that the existence of both advantages and disadvantages cancel each out for two-vehicle Rear-End crashes. The other possible reason for this is that the sampled road sections on average have very high PDI (indicating no defect), low RD and RD_df, which may not be enough to bring about any remarkable effect on the severity of injuries.

With regard to traffic engineering factors, Table 5.4 shows that two-vehicle crashes occurred on highways with higher AADT are associated with less severe injuries. This is reasonable because high traffic volume restricts driving speeds, thus decreasing the possibility of severe injuries. An increase in the speed limits from 30 (48 km/hr) to 35mph (56 km/hr), 40 mph (64 km/hr) and 50 mph (80 km/hr) does not jeopardize traffic safety in terms of injury severity; whereas, increasing the speed limit from 30 (48km/h) to 45mph(72 km/hr) 55 mph (88 km/hr) appears to result in a notable increase in injury severity. Non-peak hours were found to have higher probability of server injuries as compared to peak hours. This finding matches to many previous researches. For example, Duncan, et al. (1998) claimed that high level of congestion is suggested to significantly reduce injury severity. Rural locations are more likely to be related to server injuries comparing to urban areas, which is consistent to the results from several published studies (Krull et al., 2000; Abdel-Aty, 2003). This might be due to the higher speeds that drivers maintain in rural highways than in urban highways.

Moving to roadway geometric features, it is found that the wider median width are likely to have crashes with less severe injuries. This is understandable in that drivers have more space to take actions thus prevented serious injuries from happening. For example, when the rear vehicle was about to hit the front vehicle, he can turn the wheel towards the median to prevent from full contacting, which may help to decrease the possibility of severe injuries. Lane width (lanewid), vertical slope (pct_grde) and type of terrain did not produce any significant influence to the severity of injuries.

Looking at environment variables, inclement weather were more likely to be associated with less severe injury in relative to clear weather. This result supports the findings of Khattak et al. (1998) and Krull et al. (2000), which can be explained by the drop in practical speed in bad weather conditions (Adams, 1985). Moreover, people driving in bad weather, particularly under raining condition, have the potential to avoid the spray and splash caused by the front vehicles. Therefore, they are supposed to maintain longer distance with front vehicles for both safety and comfort concern, which may also help to decrease the possibility of severe injuries when Rear-End crashes happened.

Night time condition appears to increase the likelihood of severe injuries in comparison to daytime, which is consistent to the finding of Abdel-Aty (2003). There are two possible reasons. First of all, This may be due to the fact that drivers have poor sight to discern unexpected vehicles and other objects during nighttime and also cannot accurately estimate the gap with other vehicles, thus resulting in improper following distance. Secondly, night time is associated with low traffic volume, which may encourage drivers to exceed the speed limits, thus lead to severe injuries when crashes happened.

As for factors regarding drivers and vehicles, middle-aged and older drivers involved two-vehicle Rear-End crashes are found more likely to have severe injuries as compared to younger drivers only crashes. This may be due to the weaker physical status and longer reaction time of middle-aged and older drivers. Light truck involved Rear-End collisions were associated with lower probability of severe injuries as compared to passenger car only crashes. This is to some extent consistent with the result reported by Krull et al. (2000), where they indicate that injury severity increases with passenger cars as opposed to pick-up trucks. The lower injury propensity of light truck involved crashes might be attributed to the relatively higher rigidity of a light truck as opposed to a passenger car. Heavy truck involved collisions do not show significant difference in injury severity as compared to passenger car only collisions, which might be due to the insufficient observations of heavy truck involved Rear-End crashes in the sampled dataset.

5.4.2 Two-vehicle Sideswipe Collisions

Modeling results from both the OP model and the BOP model for two-vehicle Sideswipe collisions are provided in Table 5.5.

Table 5.5 Parameter estimates for two-vehicle Sideswipe collisions

Variable	Classes	OP model		BOP model		
		Parameter estimates	Pr > γ^2	Mean	2.5%	97.5%
PSI	Good					
	Fair/Bad	-0.1150	0.0035	-0.1164	-0.1922	-0.0421
PDI	No_dfct					
	Minor_dfct	0.0164	0.6844	0.0164	-0.0596	0.0903
	Severe_dfct	-0.0216	0.6627	-0.0209	-0.1183	0.0736
RD	Shallow					
	Med/Deep	0.0483	0.3559	0.0462	-0.0533	0.1428
RD_df	Small					
	Large	-0.1327	0.2392	-0.1399	-0.3598	0.0803
thAADT	NA	-0.0039	0.0047	-0.0039	-0.0068	-0.0013
lanewid	NA	-0.0156	0.4976	-0.0161	-0.0605	0.0270
medwid	NA	-0.0058	0.0004	-0.0058	-0.0091	-0.0025
pct_grde	NA	0.0108	0.3226	0.0106	-0.0094	0.0309
rururb	Urban					
	Rural	0.2293	<.0001	0.2250	0.1185	0.3265
spdlmt	30 mph					
	35 mph	0.0907	0.2894	0.0912	-0.0717	0.2635
	40 mph	0.0665	0.3873	0.0666	-0.0896	0.2255
	45mph	0.1806	0.0176	0.1809	0.0314	0.3332
	50 mph	0.3295	0.0009	0.3338	0.1535	0.5213
	55 mph	0.5206	<.0001	0.5244	0.3578	0.6915
tyterrain	Flat					
	Rolling	0.0241	0.7566	0.0253	-0.1208	0.1879
	Mountain	0.1686	0.2655	0.1661	-0.1314	0.4520
weather	Clear					
	Inclement	-0.0205	0.7055	-0.0196	-0.1215	0.0950
light	Daylight					
	Nighttime	0.1083	0.0107	0.1078	0.0290	0.1937
phour	Yes					
	No	0.1142	0.0017	0.1148	0.0424	0.1892
drvsex	M_M					
	M_F	0.0001	0.9979	-0.0005	-0.0737	0.0746
	F_F	-0.0563	0.2992	-0.0561	-0.1557	0.0435
drvage	Yng_Yng					
	Midage_Yng	0.1095	0.1942	0.1103	-0.0380	0.2801
	Midage_Midage	0.1430	0.0836	0.1463	-0.0099	0.3142
	Old_inv	-0.0251	0.7748	-0.0238	-0.1907	0.1583
vehtype	Pass_Pass					
	Lgt_Pass	-0.0658	0.1341	-0.0670	-0.1587	0.0212
	Lgt_Lgt	-0.1055	0.0595	-0.1051	-0.2238	0.0048
	Hvy_inv	-0.0435	0.4978	-0.0425	-0.1704	0.0794
gamma1	NA	1.2525	<.0001	1.2498	0.6980	1.8580
gamma2	NA	2.4823	<.0001	2.4799	1.8542	3.1680
gamma3	NA	3.0611	<.0001	3.0723	2.3710	3.8257
Deviance	NA	6808.1690		6808.2490		

Table 5.5 also shows identical parameter estimates and deviances in the OP model and the BOP model, indicating that these two models perform equally well, which further support the findings that the employment of the BOP model does not improve the model as compared to the OP model when weak informative prior was used and the sample size is large.

Looking first at the pavement management factors, the results show exactly the same effects of these factors as in the Rear-End crashes: PSI is significantly negative ($p=0.0035$), and others are not significant. The possible reasons might be similar to that has been discussed in the Rear-End crashes subsection.

With regard to traffic engineering factors, AADT, median width, rural urban location, light condition, peaking hour play the identical roles as they do in the Rear-End collisions. Similar to that in the Rear-End collisions, increasing the speed limit from 30 mph (48km/h) to 45 mph(72 km/hr), 50 mph (80 km/hr) and 55 mph (88 km/hr) lead to a notable increase in injury severity. In contrast to that in Rear-End collisions, weather, driver's sex, age and vehicle's type are not significant in Sideswipe crashes. It is known that on one hand, drivers may slow down in bad weather condition. On the other hand, in bad weather, especially raining conditions, the spray and splash caused by the front vehicle may blur the vision of drivers and distract their focus. In addition, surface water may also cause vehicles' hydroplaning when braking. The insignificance of weather in Sideswipe collisions indicates that two sides of weather's effect play equally important role and cancel each out when crashes happened.

5.4.3 Two-vehicle Angle Collisions

Modeling results from both the OP model and the BOP model for two-vehicle Angle collisions are provided in Table 5.6.

Once again, it can be observed from Table 5.6 that the parameter estimates from the OP model is quite close to the mean posterior estimates from the BOP model. In addition, the deviances of the OP model (19338.64) and the BOP model (19338.68) indicates that these two models perform equally well.

Table 5.6 shows that the pavement management factors play the same roles as they do in Rear-End and Sideswipe crashes: the rougher the pavement, the less likelihood of severe injuries. PDI, rutting depth and rutting depth difference between right and left wheels are not significant in predicting severity of injuries.

With regard to traffic engineering factors, AADT, rural urban location, light condition, peaking hour play the same role as they do in both the Rear-End collisions and Sideswipe collisions. Similar to that in the Rear-End collisions. In addition to the speed limit change from 30 mph (48km/h) to 45 mph(72 km/hr), 50 mph (80 km/hr) and 55 mph (88 km/hr), the increase of speed limit from 30 mph (48km/h) to 40 mph (64 km/hr) also led to a notable increase in injury severity. It is interesting that the severity of injuries in Angle collisions are more sensitive to speed limit change than in Rear-End and Sideswipe collisions. This makes sense because the outcome of Angle collisions highly relies on the speed of hitting vehicles, while in Rear-End and Sideswipe crashes, the outcome is

Table 5.6 Parameter estimates for two-vehicle Angle collisions

Variable	Classes	OP model		BOP model		
		Parameter estimates	Pr > γ^2	Mean	2.5%	97.5%
PSI	Good					
	Fair/Bad	-0.1907	<.0001	-0.1917	-0.2381	-0.1446
PDI	No_dfct					
	Minor_dfct	-0.0070	0.7804	-0.0066	-0.0535	0.0432
	Severe_dfct	-0.0164	0.5919	-0.0160	-0.0786	0.0412
RD	Shallow					
	Med/Deep	0.0131	0.6859	0.0136	-0.0487	0.0780
RD_df	Small					
	Large	-0.0278	0.6739	-0.0297	-0.1598	0.1064
thAADT	NA	-0.0043	<.0001	-0.0044	-0.0064	-0.0025
lanewid	NA	0.0060	0.6776	0.0064	-0.0219	0.0339
medwid	NA	-0.0015	0.2020	-0.0014	-0.0036	0.0009
pct_grde	NA	0.0005	0.9419	0.0005	-0.0138	0.0137
rururb	Urban					
	Rural	0.2088	<.0001	0.2088	0.1274	0.2839
spdlmt	30 mph					
	35 mph	0.1633	0.0014	0.1629	0.0657	0.2572
	40 mph	0.2387	<.0001	0.2384	0.1546	0.3223
	45mph	0.3200	<.0001	0.3188	0.2344	0.4068
	50 mph	0.4331	<.0001	0.4333	0.3141	0.5574
	55 mph	0.5379	<.0001	0.5370	0.4274	0.6488
tyterrain	Flat					
	Rolling	0.0444	0.3839	0.0451	-0.0527	0.1427
	Mountain	0.2121	0.1058	0.2068	-0.0436	0.4607
weather	Clear					
	Inclement	-0.0294	0.3677	-0.0301	-0.0958	0.0335
light	Daylight					
	Nighttime	0.1435	<.0001	0.1446	0.0903	0.1989
phour	Yes					
	No	0.0925	<.0001	0.0926	0.0449	0.1345
drvsex	M_M					
	M_F	0.1077	<.0001	0.1082	0.0596	0.1580
	F_F	0.0977	0.0027	0.0994	0.0384	0.1591
drvage	Yng_Yng					
	Midage_Yng	0.0227	0.6530	0.0248	-0.0749	0.1198
	Midage_Midage	0.0440	0.3775	0.0456	-0.0499	0.1460
	Old_inv	0.0766	0.1405	0.0779	-0.0202	0.1749
vehtype	Pass_Pass					
	Lgt_Pass	-0.0358	0.1538	-0.0354	-0.0868	0.0160
	Lgt_Lgt	-0.1321	0.0002	-0.1318	-0.2025	-0.0573
	Hvy_inv	-0.0331	0.5009	-0.0329	-0.1280	0.0646
gamma1	NA	0.9846	<.0001	0.9910	0.6305	1.3514
gamma2	NA	2.3162	<.0001	2.3234	1.9213	2.7263
gamma3	NA	2.9869	<.0001	2.9984	2.5628	3.4330
Deviance	NA	19338.6400		19338.6800		

mostly determined by the trajectory of two vehicles after the impacting instead of the impacting itself. Light vehicle to light vehicle Angle crashes are associated with less severe injuries, which is consistent to that in Rear-End Collisions, might be attributed to the more rigidity of light trucks than passenger cars.

5.5 CONCLUSIONS

With the growing rate of pavement deterioration in the United States, pavement management has become a major concern to pavement and traffic engineers. The research presented in this part of research investigated the effects of pavement management factors and traditional traffic engineering factors, on the severity of three types of two-vehicle crash injuries.

On the basis of the crashes occurred on Tennessee State Route highways from 2004 to 2008, both the OP and BOP models were fitted for two-vehicle Rear-End, two-vehicle Sideswipe and two-vehicle Angle collisions separately. The results indicate that PSI, as a measure of pavement roughness, is significant in all three collision manners studied in this part of research: the rough roads were associated with less severe injuries as compared to smooth roads. PDI (a measure of Pavement distress), rutting depth, rutting depth difference between left and right wheels are not significant in affecting the severity of injuries when any of these three types of two-vehicle collisions occurred. These findings suggest that high quality pavement does not necessary improve traffic safety in terms of the severity of crash injuries.

The traffic engineering factors exhibited considerable influences to injury severity include:

- The increase of AADT (in thousands) is positively related to the injury severity level of all three types of two-vehicle collisions;
- Wider median is likely to has less severe injuries in Rear-End crashes, while not significant in Sideswipe and Angle crashes;
- Raising speed limits by a certain magnitude appears to increase the likelihood of serious injury in all of these three collision types;
- Rural area and non-peak hours were associated with higher probability of severe injuries exclusively for all three collision types studied;
- Night time light condition increases the likelihood of more severe injuries as compared to daytime light condition in all three collision types;
- Inclement weather is more likely to be associated with less serious injuries in Rear-End crashes, while not significant in other two collision manners;
- Female drivers involved Rear-End and Angle crashes were associated with higher probability of severe injury as compared to male drivers only crashes;
- Middle-aged and older drivers involved Rear-End crashes are more likely to have severe injuries as compared to younger drivers only crashes;
- Light trucks involved Rear-End and Angle collisions are less likely to have severe injuries as compared to passenger cars only crashes.

By comparing parameter estimates and deviances, the authors found that the BOP model with weak informative prior and the traditional OP model based on the MLE method performs equally well. To employ Bayesian inference with non-informative or weak informative priors does not necessary improve the fitness of models, especially when a large size of sample data is available. However, the employment of the BOP model does provide a good alternative to the traditional OP model in necessary.

This part of research is also subject to its limitations. There are a few variables not accounted for in this research because of the absence or incompleteness in the database such as driver's drinking condition, use of seat belts, horizontal alignment, type of shoulder, and shoulder width. The pavement management information considered here is basically a general reflection of the real roughness and distress, while specific recording of pavement distress like cracking, potholes, and friction is lacking. The further study of each specific distress on injury severity is desired.

References

Adams, J.G.U, 1985. Risk and freedom-the record of road safety regulations. Transport Publishing Projects, London.

Al-Masaeid, H.R., 1997. Impact of Pavement Condition on Rural Road Accidents. Canadian Journal of Civil Engineering, 24 (4), pp.523-532.

Al-Ghamdi, A.S., 2001. Using logistic regression to estimate the influence of accident factors on accident severity. Accid. Anal. Prev., 34(6), pp.729-741.

Abdel-Aty, M., 2003, Analysis of driver injury severity levels at multiple locations using ordered probit models. Journal of Safety Research, 34 (5), pp. 597-603

Awadzia, K.D., Classena, S., Hallb, A., Duncanb, R.P., Garvan, C.W., 2008. Predictors of injury among younger and older adults in fatal motor vehicle crashes. Accident Anal. Prev. 40(6), pp.1804-1810.

Andrew, D.M., Quinn, K.M., Park, J.H., 2011. MCMCpack, version 1.0-9. http://CRAN.R project.org/package=MCMCpack.

Brodsky, H., Hakkert, A.S., 1998. Risk of a road accident in rainy weather. Accident Anal. Prev, 20 (3), pp. 161-176.

Baker, S. P., Braver, E. R., Chen, L.H., Li, G., and Williams, A. F, 2002. Drinking Histories of Fatally Injured Drivers. Injury Prevention, 8 (3), pp. 221-226.

Bedard, M., Guyatt, G. H., Stones, M. J., and J. P. Hirdes, 2002. The Independent Contribution of Driver, Crash, and Vehicle Characteristics to Driver Fatalities. Accident Analysis and Prevention, 34 (6), pp. 717-727.

Bauer, M.J., Adler, G., Kuskowski, M.A., Rottunda, S., 2003. The influence of age and gender on driving patterns of older adults. Journal ofWomen & Aging 15, pp. 3-16.

Baker, T.K., Falb, T., Voas, R., Lacey, J., 2003. Older women drivers: fatal crashes in good conditions. Journal of Safety Research 34, pp. 399-405.

Boufous, S., Finch, C., Hayen, A., Williamson, A., 2008. The impact of environmental, vehicle and driver characteristics on injury severity in older drivers hospitalized as a result of a traffic crash. Journal of Safety Research 39, pp. 65-72.

Chan, C.Y., Huang, B., Yan, X., and Richards, S., 2010. Investigation effects of asphalt pavement conditions on traffic accidents in Tennessee based on the pavement management system (PMS). J. Adv. Transp. 44(3). P. 150-161.

Chang, L.Y., Mannering, F.L., 1999. Analysis of injury severity and vehicle occupancy in large-truck and non-large-truck-involved accidents. Accident Anal. Prev., 31(5), pp. 579-592.

Chang, L.Y, Wang, H.W., 2006. Analysis of traffic injury severity: An application of non-parametric classification tree techniques. Accident Anal. Prev. 38(5), pp. 1019-1027.

Chipman, M. L., MacGregor, A. G., Smiley, A. M., Gosselin, M. L., 1992. Time vs. distance as measures of exposure in driving surveys. Accident Anal. Prev., 24 (6), pp. 679-684.

Cleveland, D.E., 1987. Effect of resurfacing on highway safety. In Relationship between safety and key highway features. Edited by E.T. Crump. State of the Art Report 6, National Research Council, Washington, D.C., pp. 78–95.

Cowles, M.K., 1996. Accelerating Monte Carlo Markov Chain Convergence for Cumulative-Link Generalized Linear Models. Statistics and Computing 6, pp. 101-111.

Duncan, C., Khattak, A., Council, F., 1998. Applying the ordered probit model to injury severity in truck-passenger car rear-end collisions. Proc., 77th Annual Meeting of the Transportation Research Board, Transportation Research Board, Washington, D.C..

Efroymson, M.A., 1960. Multiple regression analysis. In Ralston, A. and Wilf, HS, editors, Mathematical Methods for Digital Computers. Wiley.

Finison, K. S., Dubrow, R. B., 2002. A comparison of maine crashes involving older drivers using CODES (Crash Outcome Data Evaluation System) linked data (NHTSA

Report Report No. DOT HS 809 407). Washington, DC: National Highway Traffic Safety Administration.

Goodrich, B., Lu, Y., 2007. "Oprobit.Bayes: Bayesian Ordered Probit Regression" in Kosuke Imai, Gary King, and Olivia Lau, "Zelig: Everyone's Statistical Software," URL: http://gking.harvard.edu/zelig.

Hutchinson, T.P., 1986. Statistical modelling of injury severity, with special reference to driver and front seat passenger in single-vehicle crashes. Accident Anal. Prev., 18, pp. 157-167.

Jacobs, G.D, 1976. A study of accident rates on rural roads in developing countries. Transport and Road Research Laboratory, Crowthorne, United Kingdom, TRRL Laboratory Report 732.

Kamel, N., Garshore, T., 1982. Ontario's wet pavement accidents reduction program. In Pavement characteristics and materials. STP 763. Edited by C.M. Hayden. American Society for Testing and Materials, Philadelphia, Pa., pp. 98–117.

Khattak, A.J., P.K. Kantor, F.M., Council, 1998, Role of adverse weather in key crash types on limited-access roadways implications for advanced weather systems. Transport. Res. Rec, 1621, pp.10-19.

Krull, K., Khattak, A., Council, F., 2000. Injury Effects of Rollovers and Events Sequence in Single-Vehicle Crashes. Presented at the 80th Annual Meeting of the Transportation Research Board, Washington, D.C..

Kockelman, K.M., Kweon, Y.J., 2001. Driver injury severity: an application of ordered probit models. Accident Anal. Prev., 34(3), pp. 313-321.

Keall, M., Frith, W., Patterson, T., 2004. The influence of alcohol, age and number of passengers on the night-time risk of driver fatal injury in New Zealand. Accident Anal. Prev. 36 (1), 49-61.

Kockelman, K., Bottom, J., 2006. Safety Impacts and Other Implications of Raised Speed Limits on High-Speed Roads. National Cooperative Research Program, Project 17-23, Washington, D.C..

Lee, J., Mannering, F.L., 2002 Impact of roadside features on the frequency and severity of run-off-roadway accidents: an empirical analysis. Accident Anal. Prev. 34(2), pp. 149-161.

Lyman, S., Ferguson, S. A., Braver, E. R., & Williams, A. F., 2002. Older driver involvements in police reported crashes and fatal crashes: trends and projections. Injury Prevention, 8, 116-120.

Lee, C., Abdel-Aty, 2005, M. Comprehensive Analysis of Vehicle-Pedestrian Crashes at Intersections in Florida. Accident Analysis and Prevention, 37(4), pp. 775-786.

Mckelvey, R. D., Zavoina, W., 1975. A statistical model for the analysis of ordinal level dependent variables, Journal of Mathematical Sociology 4. pp. 103-120.

Malyshkina, N.V., Mannering, F.L., 2008. Effect of Increases in Speed Limits on Severities of Injuries in Accidents. Transportation Res. Rec., 2083, pp. 122-127.

National Transportation Safety Board, 1980. Fatal Highway Accidents on Wet Pavement-The Magnitude, Location and Characteristics, HTSB-HSS-80-1. National Technical Information Service. Springfield, VA.

Noland, R.B., L., Oh, 2002. The effect of infrastructure and demographic change on traffic related fatalities and crashes: A case study of Illinois county-level data. paper presented at the 77th Annual Meeting of the Transportation Research Board. Washinton, DC.

O'Donnell, C. J., Connor, D.H., 1996. Predicting the severity of motor vehicle accident injuries using models of ordered multiple choice, Accident Anal. Prev., 28(6), pp. 739-753.

Özbay, K., Laub,R., 2001. Models for pavement deterioration using LTPP: final report. Federal Highway Administration, Report No.: FHWA-NJ-1999-030.

Quinn, B.E., Hilderbrand, S.E., 1973. Effect of Road Roughness on Vehicle Steering, Highway Res. Rec., 47, pp.62-75.

Rifaat, S. M., Chin, H. C., 2007. Accident Severity Analysis Using Ordered Probit Model. Journal of Advanced Transportation, 41(1), pp. 91-114.

Rifaat, S.M., Tay,R., 2009. Effects of street patterns on injury risks in two-vehicle crashes, Transportation Research Record: Journal of the Transportation Research Board, No. 2102, Transportation Research Board of the National Academies, Washington, D.C., pp. 61–67.

Reul, M., Winner, H., 2009. Enhanced Braking Performance by Integrated ABS and Semi-Active Damping Control, The 21ST (ESV) international technical conference on the enhanced safety of vehicles, Germany.

R Development Core Team, 2011. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

Shankar, V., Mannering, F.L., Barfield, W.,1996. Statistical analysis of accident severity in rural highways, Accident Anal. and Prev., 28(3), pp.391-401.

Strat, M.R., Kim, J., Berg, W.D., 2004. Potential Safety Cost-Effectiveness of Treating Rutted Pavements; Transport. Res. Record 1629, pp.208-213.

Smink, B.E., Ruiter, B., Lusthof, K.J., de Gier, J.J., Uges, D.R.A., Egberts, A.C.G, 2005. Drug use and the severity of a traffic accident. Accident Anal. Prev., 37 (3), pp. 427-433.

Shimamura, M. Y., Fujita, G., 2005. Method to evaluate the effect of safety belt use by rear seat passengers on the injury severity of front seat occupants, Accident Anal. Prev. 37(1), pp. 5-17.

Savolainen, P.T., Mannering, F.L., Lord, D., Quddus M.A., 2011. The Statistical Analysis of Highway Crash-Injury Severities: A Review and Assessment of Methodological Alternatives. Accident Anal. Prev., 43(5), pp. 1666-1676.

Treat, J.R., Tumbas, N.S., McDonald, S.T., Shinar, D., Hume, R.D., Mayer, R.E., Stansifer R.L., Castellan, N.J., 1979. Tri-level study of the causes of traffic accidents executive summary. NHTSA, Washington, D.C.

Tighe, S., Li, N., Falls, L. C., Haas, R., 2000. Incorporating Road Safety into Pavement Management. In Transportation Research Record 1699, TRB, National Research Council, Washington, D.C., pp. 1-10.

Valent, F., Schiava, F., Savonitto, C., Gallo, T., Brusaferro, S., and Barbone, F., 2002. Risk Factors for Fatal Road Traffic Accidents in Udine, Italy. *Accident Analysis and Prevention,* 34 (1), pp. 71–84.

Wambold, J.C., Zimmer, R.A., Ivey, D.L., Sicking, D.L., 1984. Roughness, holes and bumps. In State of the art report 1, TRB, National Research Council, Washington, D.C..

Xie, Y.C., Zhang, Y.L., Liang, F.M., 2009. Crash injury severity analysis using Bayesian Ordered Probit Models. Journal of Transportation Engineering, 135(1), pp. 18-25.

Zhang, J., Lindsay, J., Clarke, K., Robbins, G., Mao, Y., 2000. Factors affecting the severity of motor vehicle traffic crashes involving elderly drivers in Ontario. Accident Anal. Prev., 32 (1), pp.117-125.

Zajac, S., Ivan, J., 2003. Factors influencing injury severity of motor vehicle crossing pedestrian crashes in rural Connecticut. Accident Anal. Prev. 35 (3), 369–379.

PART 6 CONCLUSIONS AND RECOMMENDATIONS

6.1 Conclusions

In this dissertation, the influences of outside shoulder curbs and pavement management factors on both the occurrence and the outcome of traffic-related crashes were studied. Based on the analysis in the last four sections, the primary conclusions can be summarized as follows:

- the presence of curbed outside shoulders in high-speed roads is not likely to have a higher crash frequency compared to other types of outside shoulders. The decrease of speed limit from 55 to 50 or 45 mph will not have a positive effect on highway safety when curbs are installed;
- the presence of curbs along outside shoulders were associated with higher probability of no injury and minor injury crashes, but lower likelihood of incapacitating injury and fatality crashes as compared to non-curb shoulders. The increase of speed limit from 45 mph to 55 mph adds only a small impact to the severity of injury, given that a single-vehicle crash has occurred on roadways with curbs installed;
- roadways of high roughness were associated with higher crash frequency as compared to smooth roads. Roadways with high level of perceivable distresses were less likely to have crashes. There is no apparent relationship between rutting depth and crash frequency;
- for two-vehicle rear-end, sideswipe and angle collisions, the rough roads were associated with less severe injuries as compared to smooth roads. PDI (a measure

of Pavement distress), rutting depth, rutting depth difference between left and right wheels are not significant in affecting the severity of injuries.

A few findings were identified for the methodologies:

- the BOP model with weak informative prior and the traditional OP model based on the MLE method performs equally well. To employ Bayesian inference with non-informative or weak informative priors does not necessary improve the fitness of models, especially when a large size of sample data is available. However, the employment of the BOP model does provide a good alternative to the traditional OP model in necessary;
- the ZIOP model outperforms the traditional OP model in several means: (1) the ability to explain the preponderance of zero observations; (2) it allows for the influence of the explanatory variables varying across the levels of injury severity; (3) the overall goodness of fitting in terms of the Vuong's statistics;
- incorporating random effect factors in the count models may serve to improve the overall fitting when panel data was employed, and the ignorance of the potential temporal or spatial correlation may result in biased parameter estimates.

6.2 Recommendations

Based on the collective research findings, it is recommended that TDOT develop an alternative design standard for use on "transitional roadways" as defined in this study. The alternative standard would allow the use of design speeds and posted speed limits in excess of 45 mph, along with the use of appropriate sloping curbs, on this new category of roadways.

The research herein provides a few good references for highway design engineers and pavement management engineers:

- it is recommended that AASHTO and state DOTs develop an alternative design standard for use on highways with curbs installed in suburban area. The alternative standard would allow the use of design speeds and posted speed limits in excess of 45 mph, along with the use of appropriate sloping curbs, on this new category of roadways.
- to maintain a low level of pavement roughness is recommended for safety concern;
- employing the ZIOP model as a superior alternative in future studies of traffic injury severity is strongly suggested;
- incorporating Bayesian methods in traditional count models are encouraged because the flexibility of the Bayesian method allows sophisticated models to be constructed.

Future researches are recommended in both the improvement of methodologies and the completeness of the database:

- the studies presented in this dissertation was limited by the incompleteness of the database so that a few typical variables were not accounted for in fitted models. The future study with more complete variables are desired;
- the research presented in this dissertation assumes the pavement quality information in state route undivided highways are homogeneous throughout the transverse lanes in both direction. This assumption is reasonable but not perfect. Hence, the future study using direction specific pavement quality information and crash data are desired.
- the overall pavement quality in Tennessee state route is very good. The future study utilizing proportional roadways with median to bad pavement quality is encouraged;
- the pavement management information considered here is basically a general reflection of the real roughness and distress, while specific recording of pavement distress like cracking, potholes, and friction is lacking. The further study of each specific distress on injury severity is desired.
- even though a great improvement in parameter estimates has been brought up by the ZIOP model, the ability of this model in explaining underreported crashes is still unknown. Future work to evaluate this model in dealing with underreported crashes is desired;
- as an alternative to the Maximum Likelihood Estimating method, the Bayesian inferences can also be introduced into the ZIOP model. The authors expect that a even greater improvement could be achieved by employing Bayesian ZIOP models;

- the ability of random effect models in dealing with preponderance of zeros are not agreeable. The future study to incorporate zero-inflated models in the random effect structures are desired.

APPENDICES

Appendix A: Original SAS code for the Zero-Inflated Negative Binomial Model

```
proc countreg data=Il.regmodel1019 outest=Il.zinb_out1019 covout
printall;*zero inflated negative binomial;
model Tot_n=medtyp_n medtyp_f medtyp_b medtyp_m nolane_2
            nolane_4 spdlimt_45 spdlimt_50 spdlimt_55
            outshtp_c outshtp_s outshtp_h Thaadt
                             seg_lng outshwd medwid/dist=zinb method=nra
                                                                    covb
                                                                   corrb;

zeromodel Tot_n~ Thaadt seg_lng outshwd medwid spdlimt_45 spdlimt_50
spdlimt_55 outshtp_c outshtp_s outshtp_h/link=logistic;

output out=Il.zinb_out10192 xbeta=xb pred=predtot_n prob=pr zgamma=zga
probzero=pzero;
ods output parameterestimates=Il.pe;
run;
```

Appendix B: Original R code for the Zero-Inflated Ordered Probit Model

```
crashSR1=read.csv(file="f://HSIS Data/logistic regression/sveh0305.csv",header=TRUE)

####assing variables####
sevlev=crashSR1$sevrty
spdlimt=crashSR1$spd_limt
nlane=crashSR1$no_lanes
prob_sev=crashSR1$prob_sev
lanewid=crashSR1$lanewid
aadt=crashSR1$aadt
vehtype=crashSR1$veh_type
rururb=crashSR1$rur_urb
light=crashSR1$light_cond
weather=crashSR1$weather_cond
drvsex=crashSR1$gender
drvage=crashSR1$drvage
outshtp=crashSR1$outshtp
medtype=crashSR1$medtype
phour=crashSR1$Phour
outshwd=crashSR1$outshwd1
surfyear=crashSR1$surfyear
frstloc=crashSR1$frst_loc
crashtype=crashSR1$crashtype

####coding dummy variables########
spdlimt.f=factor(spdlimt)
nlane.f=factor(nlane)
weather.f=factor(weather)
light.f=factor(light)
phour.f=factor(phour)
phour.f=relevel(phour.f,"Yes")
drvage.f=factor(drvage)
drvage.f=relevel(drvage.f,"Youngers")
drvsex.f=factor(drvsex)
drvsex.f=relevel(drvsex.f,"Male")
vehtype.f=factor(vehtype)
vehtype.f=relevel(vehtype.f,"Pass_car")
rururb.f=factor(rururb)
rururb.f=relevel(rururb.f,"Urban")
medtype.f=factor(medtype)
medtype.f=relevel(medtype.f,"none")
outshtp.f=factor(outshtp)
outshtp.f=relevel(outshtp.f,"non_curb")
surfyear.f=factor(surfyear)
```

```
surfyear.f=relevel(surfyear.f,"Pre_2000")
frstloc.f=factor(frstloc)
frstloc.f=relevel(frstloc.f,"On_pave")
crashtype.f=factor(crashtype)
tmp=glm(sevlev~aadt+lanewid+outshwd+vehtype.f+rururb.f+nlane.f+spdlimt.f+light.f+
weather.f+drvsex.f+drvage.f+medtype.f+outshtp.f+phour.f+outshtp.f+surfyear.f+frstloc.f
+crashtype.f)
x=model.matrix(tmp)

####construting dataframe####
X=x[,-1]
y1=prob_sev
y2=sevlev
Y=cbind(y1,y2,X)
crashsr1=data.frame(Y)
n=dim(Y)[1]
p=dim(Y)[2]

###obtain initial value for beta1, binomial probit model###

x1=cbind(Y[,3],Y[,6],Y[,11:15],Y[,18:19],Y[,21:23])
y1=data.frame(cbind(Y[,1],x1))
p11=matrix(0,n,1)
p10=matrix(0,n,1)
beta1=rep(0,12)
beta1=matrix(beta1,nrow=1,ncol=12)

###obtain intitial value for beta2, probit model with 4 categories of y value###
z.out<-
zelig(as.factor(y2)~aadt+lanewid+outshwd+vehtype.fLight_trk+rururb.fRural+nlane.f4+
            spdlimt.f50+spdlimt.f55+light.fNighttime+weather.fInclement+drvsex.fFemale+
           drvage.fMiddle_aged+drvage.fOlders+medtype.fflush+medtype.fraised+
           outshtp.fcurb+ phour.fNo+surfyear.fPost_2000+frstloc.fOff_pave+
           crashtype.fOverturn+crashtype.fFix_objt+ crashtype.fOther_objt+
           crashtype.fOther_noncoll, model="oprobit", data=crashsr1)
x2=Y[,3:23]
y2=data.frame(cbind(Y[,2],x2))

p20=p21=p22=p23=matrix(0,n,1)
beta2=c(-0.002,-0.018,-0.0004,0.052,-0.015,0.0227,-0.0095,0.04,-0.068,-0.224,-0.194,
        0.123,0.017,-0.002,-0.056,-0.046,-0.056,-0.002,-0.303,-0.44,0.886)
beta2=matrix(beta2,nrow=1,ncol=21)

r=c(1.85,2.76)
theta=c(beta1,beta2,r)
```

```
pr0=pr1=pr2=pr3=matrix(0,n,1)

###Construct function of loglikelihood####
LLhood=function(theta,x1,x2,y2) {
y=as.matrix(y2[,1])
LL=0
x1=as.matrix(x1)
x2=as.matrix(x2)
n=dim(y2)[1]
for (i in 1:n)
  {

m1 = sum(x1[i,]*(theta[1:12]))
m2 = sum(x2[i,]*(theta[13:33]))
p11[i]=pnorm(m1)
p10[i]=1-p11[i]

p20[i]=pnorm(-m2)
p21[i]=pnorm(theta[34]-m2)-p20[i]
p22[i]=pnorm(theta[35]-m2)-pnorm(theta[34]-m2)
p23[i]=1-p20[i]-p21[i]-p22[i]

pr0[i]=p10[i]+p11[i]*p20[i]
pr1[i]=p11[i]*p21[i]
pr2[i]=p11[i]*p22[i]
pr3[i]=p11[i]*p23[i]

Li=log(pr0[i])*(y[i]==0)+
log(pr1[i])*(y[i]==1)+log(pr2[i])*(y[i]==2)+log(pr3[i])*(y[i]==3)
LL=LL+Li
}
answer=-LL
print(answer)
}

###optimize LLhood###
thetamax=optim(theta,LLhood,gr=NULL,x1,x2,y2,method=c("BFGS"),hessian=TRUE)

###computation of standard error####
SE=sqrt(diag(solve(thetamax$hessian)))
```

Appendix C: Original WinBUGS code for the OREP Model

```
model
{
    for (i in 1:N)
{
    count[i]~dpois(theta[i])
    log(theta[i])<-
beta[1]+beta[2]*psi[i]+beta[3]*pdi[i]+beta[4]*rut[i]+beta[5]*pct.grde[i]+
beta[6]*lane.wid[i]+ beta[7]*med.wid[i]+beta[8]*rolling[i]+beta[9]*mount[i]+
beta[10]*thaadt[i]+beta[11]*spdlmt.35[i]+beta[12]*spdlmt.40[i]+beta[13]*spdlmt.45[i]+
beta[14]*spdlmt.50[i]+beta[15]*spdlmt.55[i]+mu1[id.segmnt2[i]]
     }

   for (j in 1:93783) {
   mu1[j]~dnorm(0.0,tau1)
    }

 tau1~dgamma(0.001,0.001)

for (m in 1:15) {
beta[m]~dnorm(0.0,0.001)
                 }

sigma1<-1/tau1
}
```

Appendix D: Original WinBUGS code for the ORENB Model

```
model
{
    for (i in 1:N)  {
    count[i]~dpois(lambda[i])
    lambda[i]~dgamma(gama[i], theta[i])
    log(gama[i])<-
        beta[1]+beta[2]*psi[i]+beta[3]*pdi[i]+beta[4]*rut[i]+beta[5]*pct.grde[i]+
        beta[6]*lane.wid[i]+ beta[7]*med.wid[i]+beta[8]*rolling[i]+beta[9]*mount[i]+
        beta[10]*thaadt[i]+beta[11]*spdlmt.35[i]+beta[12]*spdlmt.40[i]+
        beta[13]*spdlmt.45[i]+beta[14]*spdlmt.50[i]+beta[15]*spdlmt.55[i]

     theta[i] <- delta[id.segmnt2[i]]
  }

   for (j in 1:93783) {
   delta[j]~dbeta(1,1)
    }

for (m in 1:15) {
beta[m]~dnorm(0.0,0.001)
                  }
}
```

VITA

Ximiao Jiang was born in China in 1984. He entered China University of Geosciences, Beijing, China in 2002 and received his Bachelor's Degree in the Water Resource and Environment Department in 2006. He studied as a master student in the Department of Geotechnical Engineering in Tongji University, Shanghai, China from 2006 to 2009. Mr. Jiang entered the Department of Civil and Environmental Engineering in The University of Tennessee (UTK) at Knoxville to pursue the Ph.D. degree in August 2009. During his stay in UTK, he extended his research area to traffic safety analysis and statistics modeling. He is also active in incorporating pavement quality information into traffic safety studies.