MD-03-SP 208B46

STATE HIGHWAY ADMINISTRATION

RESEARCH REPORT

THE RELATIONSHIP BETWEEN CONGESTION LEVELS AND ACCIDENTS

UNIVERSITY OF MARYLAND, COLLEGE PARK

MD-03-SP 208B46
FINAL REPORT

July 2003

Technical Report Documentation Page

1. Report No. MD-03-SP 208B46	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle The Relationship between Congestion Levels and Accidents		5. Report Date August 21, 2003
		6. Performing Organization Code SP 208B46
7. Author/s *Dr. Gang-Len Chang*, Professor, gang@eng.umd.edu *Hua Xiang*, Research Assistant		8. Performing Organization Report No.
9. Performing Organization Name and Address Department of Civil Engineering University of Maryland College Park, MD 20742		10. Work Unit No. (TRAIS)
		11. Contract or Grant No. SP 208B46
12. Sponsoring Organization Name and Address Maryland State Highway Administration Office of Policy & Research 707 North Calvert Street Baltimore MD 21202		13. Type of Report and Period Covered Final Report
		14. Sponsoring Agency Code
15. Supplementary Notes		

16. Abstract

This study was conducted to investigate the relationship between congestion and accidents with a specific emphasis on the impact of traffic volume levels on accident frequency, rate, and severity. The accident data from five freeway (I-495, I-695, I-95, I-270, and US50) and five arterials (MD2, MD355, US1, MD410, and MD97) were analyzed wi multivariate statistical methods to evaluate the widespread belief among traffic safety professionals that an increase in congestion levels often result in more but less severe accidents on freeways and/or local arterials. However, the impact of congestion on the accident rate tends to vary between freeways and arterials, and differs significantly across peak and off-peak periods. The estimation results, based on the available sample data, reveal that accident rates on local arterials tend to decrease with an increase in traffic volume. In contrast, accident rates on freeway segments during peak hours indicate a positive correlation with traffic volume per lane. Additionally, freeway accident rates during off-peak periods appear to be random in nature, and not necessarily correlated to any specific factors.

17. Key Words Accident frequency, accident rate, accident severity, volume, Poisson regression, negative binomial regress, ordered probit regression.	18. Distribution Statement: No restrictions This document is available from the Research Division upon request.		
19. Security Classification (of this report) *None*	20. Security Classification (of this page) *None*	21. No. Of Pages 111	22. Price

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1
INTRODUCTION

1.1 Motivation

Relieving traffic congestion and improving roadway safety are clearly top priorities for most state highway agencies. These two issues have grown to become very dependent on one another as substantial improvements to one could result in significant impacts on the other. For example, an increase in the congestion level is likely to cause a higher number of less severe accidents. This relationship seems to exist in the freeway accident data recorded by the Maryland State CHART program (Chang, 2002).

There is also a widespread belief that similar relationship between congestion levels and accidents may also exist on major arterials and/or streets. The severities of certain types of crashes in the statewide arterial network tend to decrease as congestion levels increase. However, rigorous studies conducted to analyze the complex relationship between congestion and accidents (including frequency, rate, and severity) on freeways or arterials have not yet been published in the transportation literature.

1.2 Research Objectives

In response to the aforementioned needs, this study intends to achieve the following objectives:

- Better understanding the relationship between congestion levels and the frequency, rate, and severity of accidents on freeways and arterials;
- Developing statistical models for assessing the impacts of traffic congestion on the frequency, rate, and severity of accidents;
- Identifying key factors that may have an impact on frequency, rate, and severity of accidents that occur at various levels of congestion.

This study is based on a sample dataset from the Year 2000 accident information record of the *Maryland Automated Accident Reporting System* (MAARS) from the

Maryland State Highway Administration (SHA), including a total of 9944 accidents that occurred on five primary commuting freeways and five major arterials. In addition, to illustrate the highway geometric features of each accident analyzed, this study also refers to the SHA highway information system (including the traffic monitoring system and the roadway geometry database).

1.3 Organization and Summary

Subsequent chapters of this report are organized as follows: Chapter 2 provides a comprehensive review of related literature, and includes the following three sections: review of accident frequency modeling, review of accident rate modeling, and review of accident severity modeling. In addition, a review of literature on identification of contributing variables and the definition of accident rate has also been included.

Chapter 3 presents the relationship between accident frequency and congestion levels based on associated research findings. A graphical illustration and statistical test results are provided in the exploratory analysis section. The exploratory analyses suggests that the higher the level of congestion, the greater the probability that there will be a higher level of accident frequency. Based on the preliminary findings from exploratory analyses, this chapter further investigates the relationship between accident frequency and congestion by examining the impacts of several factors using advanced statistical methods, such as Poisson and Negative Binomial (NB) regression methods. This chapter will illustrate that the surrogate variable, volume per lane, increases the frequency of accidents on arterials and freeways. In addition, median type (divided roadway or not), intersection density (number of intersections per unit length on a link), and the number of through lanes have all been identified as significant variables contributing to the accident frequency model for arterials. Median width, auxiliary lane ratio (ratio between the length of auxiliary lanes and the link length), and the number of through lanes were identified as significant variables for frequency models.

Chapter 4 presents the relationship between the accident rate and congestion levels based on three different analyses: a comparison of the average accident rate between peak and off-peak periods; a comparison of the accident rate among sampled roadway segments experiencing different levels of congestion; and a bivariate correlation analysis between the accident rate and the congestion levels. These analyses are intended to examine whether highways with higher congestion levels yield a lower accident rate.

Subsequent to the exploratory analysis results, Poisson and Negative Binomial regression methods were used to develop the accident rate model. The results indicate that the accident rate on arterials tends to decrease with the volume per lane. Additionally, the accident rate for freeways during off-peak hours appears to be random, exhibiting no systematic relationship with the traffic volumes. However, during the peak period, accident rates appear to increase significantly with traffic volumes. In addition, median type (divided roadway or not), intersection density (number of intersections per unit length on a link), and the total number of through lanes have all been identified as significant variables in the accident rate model for arterials. In contrast, the median width was the only variable identified that had significant impact on the accident rate model for freeways.

Chapter 5 presents the relationship between accident severity and congestion levels. This chapter begins with an exploratory analysis that intends to identify factors that may be associated with accident severity (e.g. accident location, roadway geometric features, and driver conditions). An aggregated analysis of the relationship between the number of accidents at various levels of severity and congestion levels on sample freeways and arterials was conducted. Subsequently, other identified key factors were used as explanatory variables and an Ordered Probit regression model was applied to estimate severity models for arterial and freeway accidents. The estimation results indicated that accidents that occurred on more congested freeways and arterials were more likely to happen at a lower level of severity, however, levels of severity may vary when introduced to other contributing factors (e.g. at intersection or on roadway segment, driver condition, median type, and weather condition).

Chapter 6 summarizes major findings of this study and offers additional recommendations for consideration for future research in areas that could potentially have an impact on traffic safety.

CHAPTER 2
LITERATURE REVIEW

2.1 Introduction

Despite the wealth of information available on accident analyses and congestion monitoring, most of the existing research focuses on the two key issues (congestion and traffic safety) separately and does not provide a concise examination of interrelationship. The potential relationship between congestion and accidents (e.g. the impacts of peak and off-peak traffic volumes on the accident rate or severity) has not been fully explored. This chapter provides an overview of some of the research findings related to this subject, and includes an analysis of the relationship between congestion and accident frequency, the impact of congestion on accident severity, and the variation of accident rate at different levels of congestion.

This literature review is divided into the three sections. Recent studies and research methods for modeling accident frequency is summarized in Section 2.2. Section 2.3 summarizes related studies on accident rates. Section 2.4 examines the state-of-the-art research related to accident severity along with key research results. Finally, conclusions and research findings are reported in Section 2.5.

2.2 Congestion level and accident frequency

Among a large body of recent literature in accident frequency analysis, some studies have made unique contributions and are summarized hereafter. For example, Shankar, Mannering and Barfield (1995) performed a study on a 61 km portion of I-90 located about 48 km east of Seattle. To minimize potential heteroskedasticity problems (see Greene 2000, pp 499-524) and to maximize the estimation efficiency, they partitioned the test portion of I-90 into ten fixed-length sections. A monthly time-series accident frequency data set was constructed, and the estimated model included solely the geometric variables (e.g., number of horizontal curves in a section and maximum horizontal curve radius in a section) and weather condition variables (e.g., number of raining days in a month and maximum daily rainfall in a month). No examination of the relationship between accident frequency and congestion levels was conducted.

Shankar, Milton and Mannering (1997) developed an accident frequency model for local arterials in Washington State where they defined roadway sections by their homogeneous features such as number of lanes, roadway width, shoulder width, Annual Average Daily Traffic (AADT), speed, and peak hour factors. One of the primary findings of this study indicated that accident frequency increases with the AADT per lane.

With respect to the estimation method, a significant number of studies have been conducted using Poisson and Negative Binomial (NB) regressions to model accident frequency (Miaou, 1994), which is due to the discrete and non-negative nature of accident data. For example, Shankar, Mannering and Barfield (1995) used a NB regression to develop the I-90 accident frequency model. However, in a later study (Shankar, Milton, and Mannering, 1997), the criteria for defining sections result in a large number of sections with short length and having zero accident frequency. To contend with this data constraint, Shanker et al modeled accident frequencies as zero-altered probability processes, and used the zero-inflated Poisson (ZIP) and the zero-inflated negative binomial (ZINB) models to account for links without accidents.

In a related study, Persaud and Dzbik (1993) explored the nonlinear relationship between accident frequency and volume. In their conclusion it was noted that on congested roadways there was a higher occurrence of accidents than on uncongested roadways with comparable volume levels. In addition, Abdel-Aty and Radwan (2000) used both Poisson and negative binomial regressions to model traffic accident occurrence and involvement on a sample freeway. They also used the likelihood ratio test to evaluate the over-dispersion of the Poisson model and re-estimate their models with Negative Binomial (NB) regression when over-dispersion was detected. The results indicated that an increase in AADT per lane also increases the likelihood of higher accident frequency. Greibe (2002) used generalized linear Poisson regression to establish the accident prediction models for urban roads. The AADT was found to be the most significant variable in the prediction of accident frequency.

Abbas (2003) developed a number of statistical models based on the accident data over 10 years in Egypt. These models were based on the assumption that the number of accidents, injuries, fatalities and casualties are a function of exposure represented with AADT and AAVK (annual average vehicle kilometers). Five functional forms were evaluated in the study conducted by Abbas, they include linear, power, logarithmic, exponential and quadratic polynomial. The model, however, includes only AADT and AAVK as explanatory variables.

Note that in all of the aforementioned studies AADT per lane was always used as a surrogate variable of congestion. Besides AADT, only a small set of geometric and weather condition variables were used in the model specification. The weather conditions were accounted by variables such as number of rainy days and the maximum daily rainfall in a month.

The results of additional studies on accident frequency seem to share a common finding that accident frequency is more likely to increase with the volume per lane. It is also important to note that Poisson and NB regressions are recognized as appropriate methods for accident related analysis (Miaou, 1994, and Shankar, Mannering and Barfield, 1995).

2.3 Congestion level and accident rate

Studies on congestion level and accident rate indicate that the accident rate is defined as the ratio between the number of accidents and associated volumes. This implies that there is a linear positive correlation between the accident frequency and volumes. As mentioned in the previous section, the accident data are discrete and non-negative in nature. Therefore, it is appropriate to use Poisson or Negative Binomial regressions to analyze the accident-related data. For example, in a recent study Mayora and Rubio (2003) combined a multivariate Negative Binomial regression model and an Empirical Bayes procedure to predict the accident rate. However, they did not examine the relationship between accident rate and traffic volumes in their research.

Karlaftis and Golias (2002) adopted a non-parametric statistical methodology, known as the hierarchical tree-based regression (HTBR), to model the accident rate with rural road geometric characteristics and traffic volumes. Traffic volumes were not included as an independent variable in their regression model, and although the functional form needs not to be specified in advance, the estimation for HTBR requires a large sample size to form the hierarchical tree.

Regarding independent variable selection, Knuiman et al (1993) explored various methods for associating the median width with the highway accident rate, including using both a categorical variable and a continuous variable to represent the median width. The research findings indicated that accident rates decreased with an increased median width, and there was insignificant decrease in accident rates for medians less than 20 to 30 ft in width.

Zhou and Sisiopiku (1997) examined the general relations between hourly accident rate and hourly traffic volume/capacity (*v/c*) ratios. With a U-shaped graph their study revealed that the accident rate decreases rapidly with an increase in the *v/c* ratio until *v/c* falls in the range of 0.55 to 0.65, at which time the rates gradually increases with the *v/c* ratio. Qin et al. (2003) and Kam (2002) both made some scaling operations to transform the relationship between "accident rate" and "exposure" into a linear from. Qin et al. (2003) used the estimated zero-inflated Poisson model to recalculate risk-oriented crash rates (e.g. the normalized crash rate). Kam (2002) used a disaggregated approach by matching accident records to a defined travel corridor to derive an induced exposure. His results revealed the existence of a polynomial function of a cubic order when crash rates were plotted against age groups. It was distinctly different with the U-shaped curve generated using the conventional approach. Both of the above approaches are also used to observe the relationship between accident rate and traffic volume. Martin (2002) explored the relationship between crash rate and annual average hourly volume on French interurban motorway networks. It was determined that such a relationship varies based on the number of through lanes on a roadway and the number of vehicles involved in accidents.

In summary, very few of the existing studies have examined the relationship between accident rate and traffic volume. The results of studies on the accident rate seem to share a common conclusion that the relationship between the accident rate and traffic volume cannot be fully captured using a linear relation, and either the definition of accident rate or the functional form of the relationship between accident rate and volume should be further evaluated.

2.4 Congestion level and accident severity

The severity of an accident is often measured by the level of injury of the most-seriously injured vehicle occupant (Chang and Mannering, 1999). Thus, the severity level has a discrete outcome and this nature of response data tends to suggest the use of a logistic regression in model development (e.g., Shankar and Mannering 1996; Chang and Mannering, 1999; Carson and Mannering, 2001). Accident severity can also be indexed using a binary variable such as a fatal or non-fatal indicator. In fact, this method was applied by Al-Ghamdi (2002) and it was determined that the following variables are most associated with the accident severity: location, accident type, vehicle type, license status, collision type, and accident time.

In a study conducted by Lee and Mannering (2000), a nested logit model was used to isolate a wide range of factors that significantly influence the severity of run-off-roadway accidents. In the work by Amoros (2002), severity was measured by the ratio between fatal and injury accidents, which corresponds to the probability of a binomial setting. In addition to the logistic regression methods, some researchers (Kockelman and Kweon, 2001, and O'Donnell and Connor, 1996) have adopted a multi-class crash analysis with the Ordered Probit models for accident severity analyses. Yau (2003) used stepwise logistic regression models to identify the risk factors associated with each vehicle type and indicated that weekday indicator and time-of-day are important variables that may affect the severity of injuries.

In the literature on modeling accident severity, very few studies have attempted to address the relationship between the road traffic flow and crash occurrence. Among these, it was the work of Martin (2002) that has explored the relationship between accident severity and hourly traffic flow. Martin's analysis of this relationship was implemented in two steps. First he addresses the probability of observing a crash and the number of vehicles exposed to the accident. Then he used a logistic regression to model the probability that a vehicle involves in injury-crashes. The explanatory variables used were day-night difference, traffic volumes, and the interaction between these two factors. Martin did not reach any conclusion with respect to the relationship between the crash severity and traffic volumes.

2.5 Summary

Based on the literature review, it can be determined that traffic volume, as a surrogate variable of congestion, plays a significant role in accident frequency, rate, and severity analyses. Some significant relationships were identified including the relationship that a higher traffic volume usually results in higher accident frequency and that there is likely a U-shaped relationship between traffic volume and the accident rate. Although key factors affecting the accidents have been extensively studied, the complex relationship between congestion and accident, especially the impact of the traffic volume on accident severity, has not been sufficiently investigated. For example, the relationship between congestion and accident (rate or severity) may vary with time of day (e.g. peak or off-peak hours), and differs significantly between arterials and freeways. In addition, this relationship may also change with the roadway environment and weather conditions.

CHAPTER 3

ACCIDENT FREQUENCY AND CONGESTION LEVEL

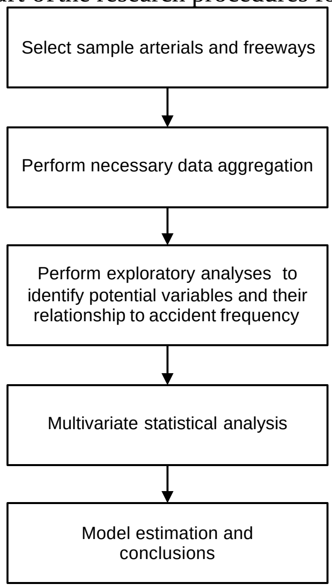
3.1 Introduction

This Chapter examines research results related to the relationship between accident frequency and congestion level on both sample freeways and arterials. It will also examine accident frequency during peak and off-peak hours and the potential factors that may contribute to an increase in accident frequency during congestion. The primary focus of this chapter is to test the hypothesis that accident frequency on either freeways or arterials will increase with congestion level.

To begin a comparison of average accident frequency (per hour per mile) between peak and off-peak periods is examined. This examination is based on the assumption that average accident frequency during peak hours is generally higher than average accident frequency during off-peak periods. The results of the comparison along with the data from five freeways and five local arterials are presented in Section 3.3. In addition to the exploratory analysis is a comparison of accident frequency between sampled roadway segments experiencing different levels of congestion, and a bivariate aggregate correlation analysis between accident frequency and congestion level. It is expected that highways with higher levels of congestion yield a higher accident frequency.

Based on the preliminary findings from the exploratory analyses, this study further investigates the target relationship between accidents and congestion under the compound impacts of various contributing factors using advanced statistical methods such as Poisson and Negative Binomial regression models. The estimation results with respect to freeways and arterials are presented in Section 3.4 and Section 3.5. A brief description of the research procedures is presented in a flowchart in Figure 3-1.

Figure 3-1 A flowchart of the research procedures for accident frequency analysis



3.2 Data Set Available for Analysis

In organizing a sample dataset for analysis, all accidents on each roadway link were converted into the following definition of accident frequency per mile:

$$\text{Accident frequency} = \frac{\text{Number of accidents on a link}}{\text{The link length}}$$

In addition, the data collected for analysis also includes accident nature, traffic flows, and roadway features in detail. Primary information associated with accidents and congestion was obtained from the highway information system and the *Maryland Automated Accident Reporting System* (MAARS) from Maryland State Highway Administration (SHA). The first database contains a list of roadway segments and associated traffic and geometric characteristics. The second database includes the location of accidents and related information. A careful integration of these two databases yielded the initial sample dataset that consists of five arterials and five freeway segments (see Table 3-1). The main reasons of choosing these sampled roadways are that they have complete geometric and traffic information in two databases, and they are the major arterials/ freeways in the Washington/Baltimore Area.

Table 3-1 Sample arterials and freeway segments for accident frequency analysis

Index	Arterials		Freeway Segments	
	Road name	Segment location	Road name	Segment location
1	US1	Between Baltimore City Line and Washington DC Line	I-495	Between Virginia State Line and I-95 Exit 27
2	MD2	The entire length	I-270	The entire length
3	MD97	The entire length	I-695	The entire length
4	MD355	The entire length	I-95	Between Baltimore City Line and Virginia State Line
5	MD410	The entire length	US50	Between Washington DC Line and Bay Bridge
Sample Accidents	4542		5402	
Sample Year	Year 2000		Year 2000	

To minimize the potential sampling bias and partially account for the stochastic nature of the accident distribution, this study aggregated short but interconnected links with common features as long links. The criteria used for link aggregation are summarized in Table 3-2.

Table 3-2 Criteria for link aggregation and the results

	Arterials		Freeway segments	
Main variables for clustering	AADT level		AADT level	
	Median type (divided or not)		Median width	
	Number of through lanes		Number of through lanes	
Clustering Results	US1	29 Links	I-495	18 Links
	MD2	32 Links	I-270	39 Links
	MD97	25 Links	I-695	59 Links
	MD355	25 Links	I-95	49 Links
	MD410	18 Links	US50	14 Links

Indicators of congestion levels

Since a rigorous definition of "congestion" is beyond the scope of this study and is one of the on-going research issues in the traffic community, the remaining analyses intend to use the "volume per lane" as the surrogate variable for congestion. Although it does not accurately reflect the actual congestion level on a given link, it should be sufficient for comparison purposes.

3.3 Exploratory Analyses

The following exploratory analysis intends to investigate whether or not the accident frequency increases with congestion level using three different comparisons, which include:

- A comparison between peak-hour (7-9AM and 4-6PM) and off-peak-hour accident frequencies, using the hypothesis that on most highway segments the average peak-hour accident frequency should be higher than off-peak-hour accident frequency, if a higher level of congestion is more likely to cause more frequent accidents.

- The second analysis performed a cross-section comparison of accident frequencies on five sample arterials and freeways to evaluate whether highways with higher levels of congestion yield more accidents.
- The third analysis was conducted to evaluate the correlation between accident frequency and volumes per lane, which was used as a surrogate variable representing congestion level.

Comparison of accident frequency during peak and off-peak hours

Figure 3-2 through Figure 3-6 illustrates the differences between peak-hour and off-peak-hour accident frequencies for five sample arterials. Figure 3-7 through Figure 3-11 illustrates the same comparison for five sample freeways. As reflected in graphical illustrations, the average accident frequency during peak hours is higher than the accident frequency during off-peak hours on all sample arterials and freeways. Results of statistical tests (see Table 3-3) and an econometric method (see Table 3-4) have further confirmed this relationship.

Figure 3-2 A comparison of the accident frequency on MD2 between peak hours and off-peak hours

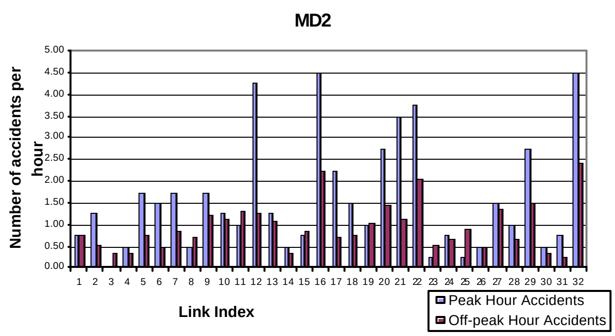


Figure 3-3 A comparison of the accident frequency on MD355 between peak hours and off-peak hours

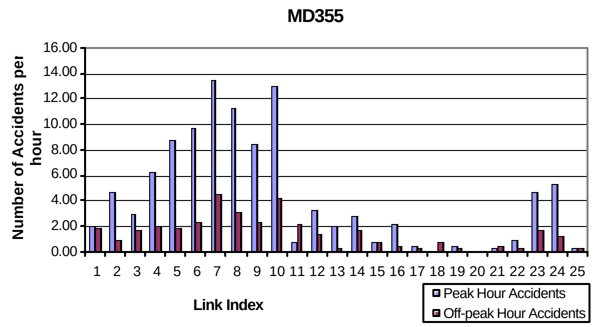


Figure 3-4 A comparison of the accident frequency on US1 between peak hours and off-peak hours

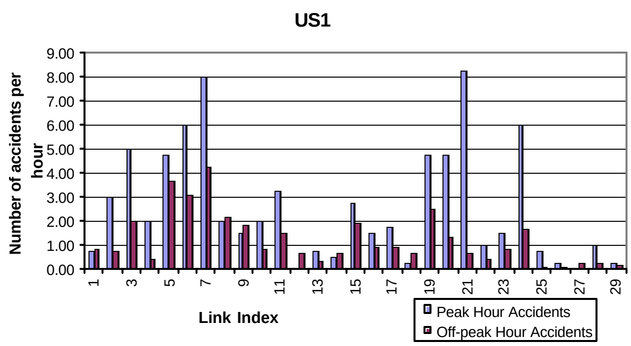


Figure 3-5 A comparison of the accident frequency on MD410 between peak hours and off-peak hours

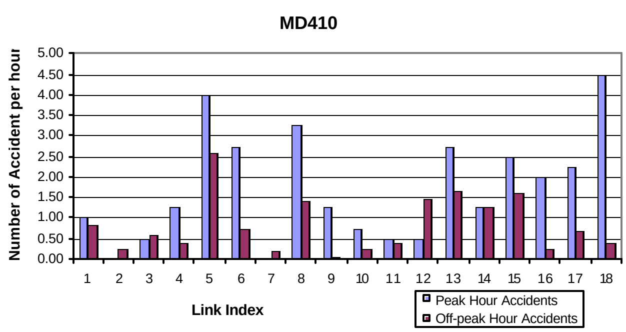


Figure 3-6 A comparison of the accident frequency on MD97 between peak hours and off-peak hours

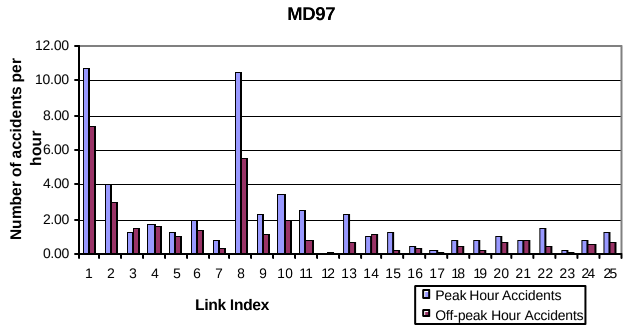


Figure 3-7 A comparison of the accident frequency on I-495 between peak hours and off-peak hours

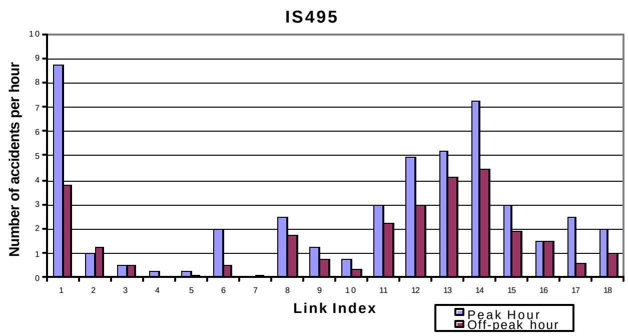


Figure 3-8 A comparison of the accident frequency on US50 between peak hours and off-peak hours

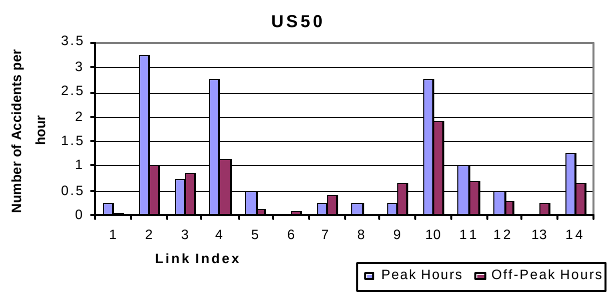


Figure 3-9 A comparison of the accident frequency on I-695 between peak hours and off-peak hours

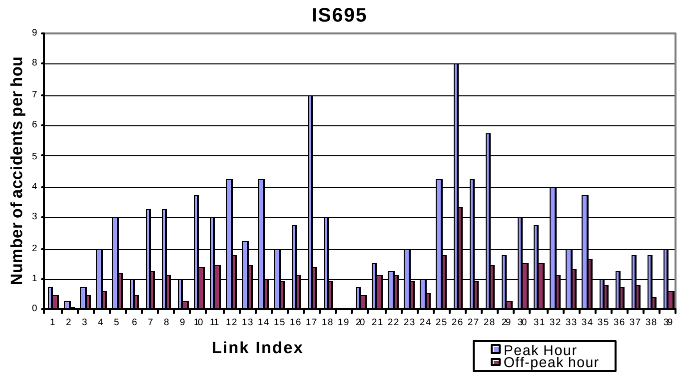


Figure 3-10 A comparison of the accident frequency on I-270 between peak hours and off-peak hours

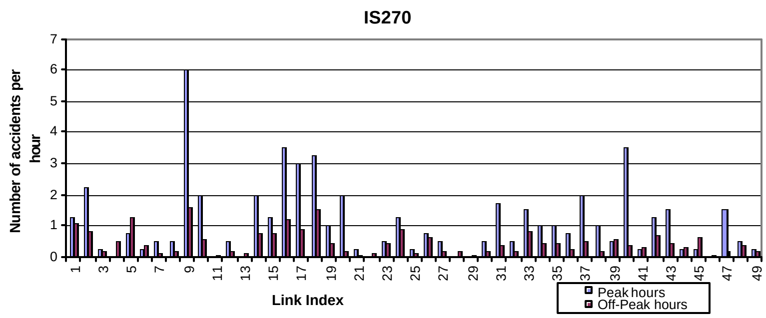
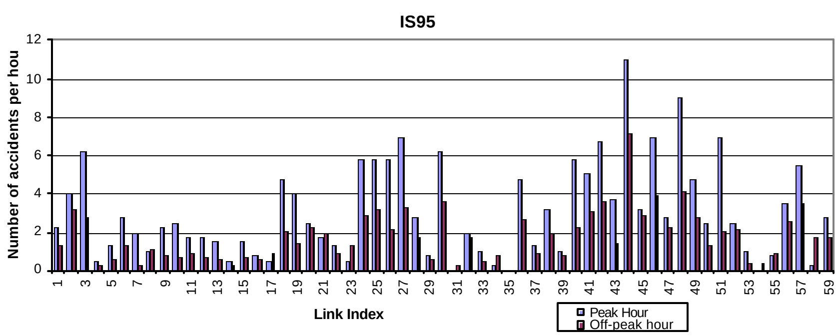


Figure 3-11 A comparison of the accident frequency on I-95 between peak hours and off-peak hours



A statistical test was performed to examine the equality of average accident frequency during the peak and off-peak periods for all sampled roadway segments. The test results are listed in Table 3-3.

Table 3-3 Mean Equality tests and results

Hypothesis	The average accident frequency during peak period is **equal** to the average accident frequency during off-peak periods among all five sample local arterials				
Data used	Accident frequency during peak hours (7-9AM and 4-6PM) Accident frequency during off peak hours				
Test results of five sample surface streets					
Route Name	MD2	MD355	US1	MD410	MD97
Sample Size (n)	32	25	29	18	25
F-ratio	6.509	9.344	7.467	6.233	1.681
$F_{1,2(n-1)}$ [0.975]	3.996	4.043	4.013	4.130	4.043
Conclusion	Reject	Reject	Reject	Reject	**Accept**
Test results of five sample freeway segments					
Route Name	I-495	I-695	I-95	I-270	US50
Sample Size (n)	18	39	59	49	14
F-ratio	2.345	28.084	12.300	12.838	1.482
$F_{1,2(n-1)}$ [0.975]	4.130	3.967	3.923	3.940	4.225
Conclusion	**Accept**	Reject	Reject	Reject	**Accept**

From the test results in Table 3-3, it was determined that the means of accident frequency during the peak and off-peak periods are significantly different for MD2, MD355, US1, MD410, I-695, I-95, and I-270. To further investigate the hypothesis that peak periods generally have a higher accident frequency than off-peak periods, the dummy variable method was used (Greene, 2000) to evaluate the target relationship. The test results are summarized in Table 3-4.

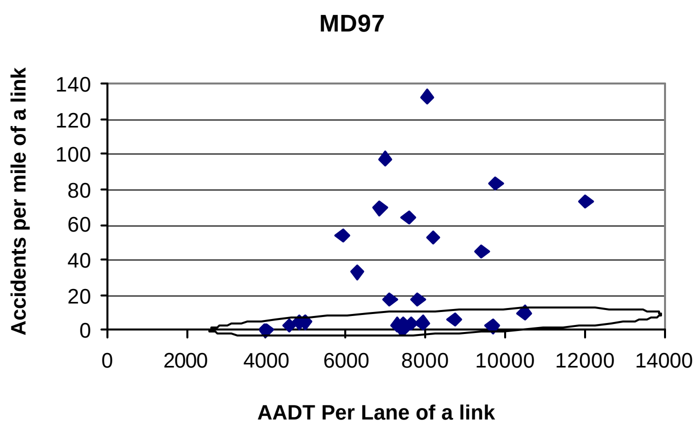
Table 3-4 Procedures and results of the dummy variable method (Greene, 2000)

Hypothesis	H_1: Accident frequency (peak hour) > Accident frequency (off-peak hour) i.e. H_0: d = 0; H_1: d > 0				
Test procedures	$y_i = m + d * Dum + e_i$ Set *Dum* = 1, if the sample is in peak hour = 0, otherwise				
Data used	Accident frequency during peak hours (7-9AM and 4-6PM) Accident frequency during off peak hours				
Test results of five sample surface streets					
Route Name	MD2	MD355	US1	MD410	MD97
Sample Size (n)	32	25	29	18	25
T-statistic of *Dum* coefficient	2.551	3.057	2.733	2.497	1.297
$T_{2n-2}(0.95)$	1.669	1.676	1.672	1.688	1.676
Conclusion	Reject H_0 Accept H_1	Reject H_0 Accept H_1	Reject H_0 Accept H_1	Reject H_0 Accept H_1	**Accept H_0**
Test results of five sample freeway segments					
Route Name	I-495	I-695	I-95	I-270	US50
Sample Size (n)	18	39	59	49	14
T-statistic of *Dum* coefficient	1.531	5.299	3.507	3.583	1.217
$T_{2n-2}(0.95)$	1.688	1.665	1.658	1.661	1.701
Conclusion	**Accept H_0**	Reject H_0 Accept H_1	Reject H_0 Accept H_1	Reject H_0 Accept H_1	**Accept H_0**

The results for MD97, I-495 and US50 are the only areas that do not support the hypothesis that average accident frequency during peak hours is higher than the average accident frequency during off-peak hours.

Further analysis of the relationships between accident frequency per mile per link and the AADT per lane per link on MD97 (as shown in Figure 3-12) indicated that there were some data points (represented in the circled area, located in Carroll County and up to Pennsylvania State Line) that caused unexpected results. In fact, this segment of MD97 is quite a distance away from any urban areas and has no significant work-related peak-hour traffic. Therefore, it is reasonable to expect that the peak-hour accident frequency does not vary significantly from the off-peak-hour accident frequency.

Figure 3-12 The relationship between accident frequency and AADT per lane on MD97



On I-495 and US50, the failure to accept the hypothesis that accident frequency differs between peak periods and off-peak periods can potentially be attributed to two factors: both freeways have a high volumes throughout the peak and off-peak periods, and factors other than congestion may contribute significantly to an increase in accident frequency on those freeways.

Cross-section comparison of the accident frequency

The following analyses are designed to further test the hypothesis that highways with higher levels of congestion should experience a higher accident frequency if there is a high correlation between accident frequency and congestion. The focus of this analysis is to compare the mean of the accident frequencies per mile between sample roadway segments.

Table 3-5 summarizes the results of the ANOVA tests for both the sample freeways and arterials. The conclusion from this test found that the average accident frequency of sample arterials during off-peak hours exhibits no significant difference among sampled arterials. However, a distinct difference does exist during the peak hour accident frequency among the sampled arterials and in the accident frequency on freeways during both peak and off-peak hours. A plausible explanation for the test results is that all sample arterials experience little congestion during off-peak periods and as a result, accident frequencies are more random in nature, and not correlated with factors such as traffic volume. One may also assume that the inconsistency in peak-hour accident frequency is due to substantial differences in congestion levels as evidenced in the peak-hour volume per lane for sample arterials shown in Figure 3-13.

Table 3-5 ANOVA tests and results

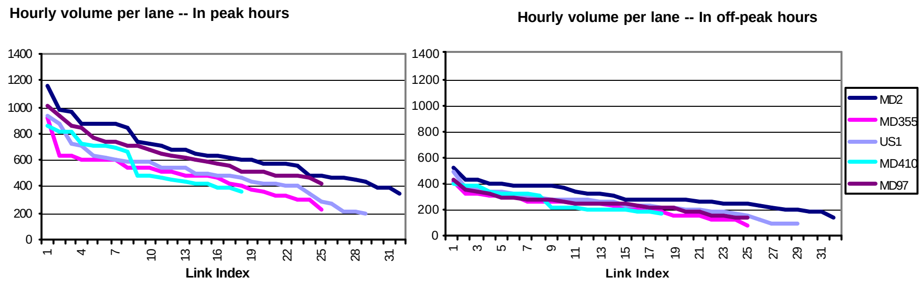
Balanced ANOVA test for arterials	
Hypothesis	The means of accident frequency are statistically equal across all five arterials
ANOVA test parameters	The number of factor levels (treatment groups): $k = 5$ The number of observations within each factor level: $n = 18$
Data used	Accident frequency in peak hours (7-9AM and 4-6PM) Accident frequency in off peak hours

Test results on the arterial dataset				
	Dependent Variable Y	*F*	$F_{4,85}^{0.975}$	*Conclusion*
1	*Peak-hour accident frequency*	*2.64*	*2.48*	***Reject***
2	*Off-peak accidents frequency*	*1.96*	*2.48*	*Accept*

Unbalanced ANOVA test for freeways	
Hypothesis	The means of accident frequency are statistically equal across five freeways
ANOVA test parameters	The number of factor levels (treatment groups): $k = 5$ The number of observations within each factor level: $n_i = \{18, 39, 59, 49, 14\}$
Data used	Accident frequency in peak hours (7-9AM and 4-6PM) Accident frequency in off peak hours

Test results on the freeway dataset				
	Dependent Variable Y	*F*	$F_{4,85}^{0.975}$	*Conclusion*
1	*Peak-hour accident frequency*	*6.29*	*2.42*	***Reject***
2	*Off-peak accidents frequency*	*4.89*	*2.42*	***Reject***

Figure 3-13 The hourly volume per lane on five arterials



Bivariate correlation test between the accident frequency and volume per lane

Figures 3-14 through 3-23 presents the relationship between volume per lane and the resulting accident frequency on each link for both sample arterials and freeways. These graphical relationships reveal the following critical information:

- Some approximate linear relationship between accident frequency and volume per lane exists.
- Other factors may contribute to an increase in accident frequency as evidenced in the variance of the linear trend.

As a result of these, the remaining multivariate analysis between accident frequency and main contributing factors will be based on the Poisson and Negative Binomial models, rather than the multiple linear regression.

Figure 3-14 Accident frequency versus volume for MD2

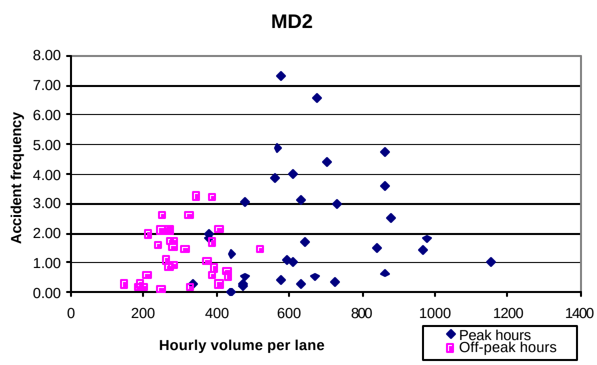


Figure 3 -15 Accident frequency versus volume for MD355

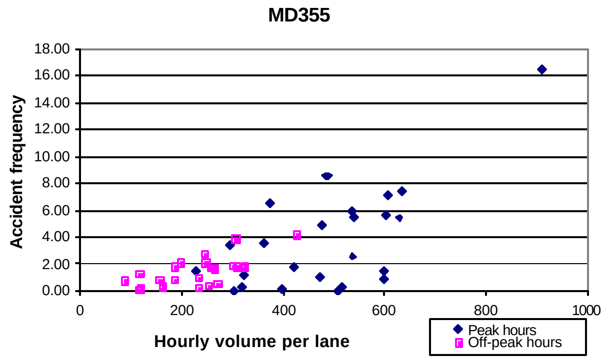


Figure 3 -16 Accident frequency versus volume for US1

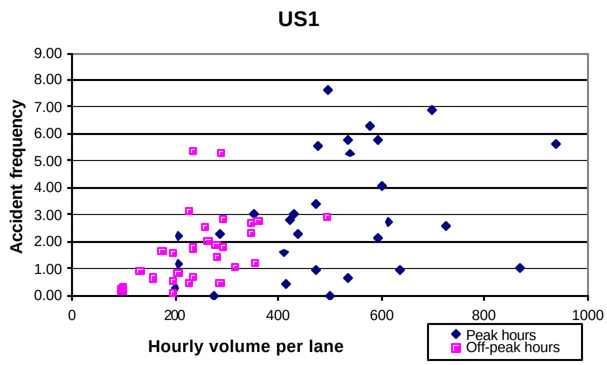


Figure 3 -17 Accident frequency versus volume for MD97

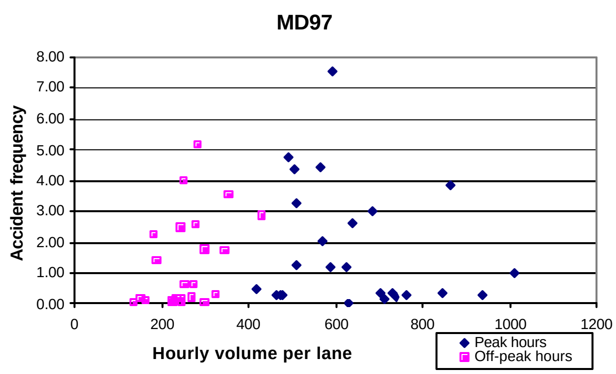


Figure 3-18 Accident frequency versus volume for MD410

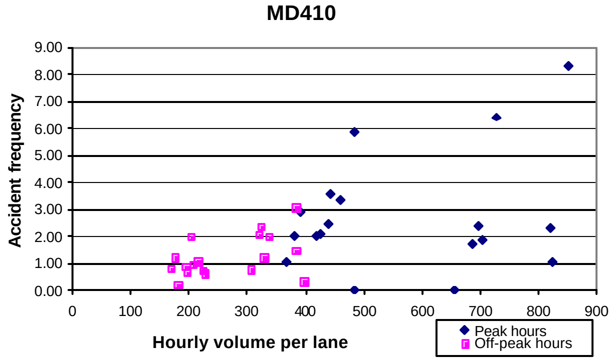


Figure 3-19 Accident frequency versus volume for I-270

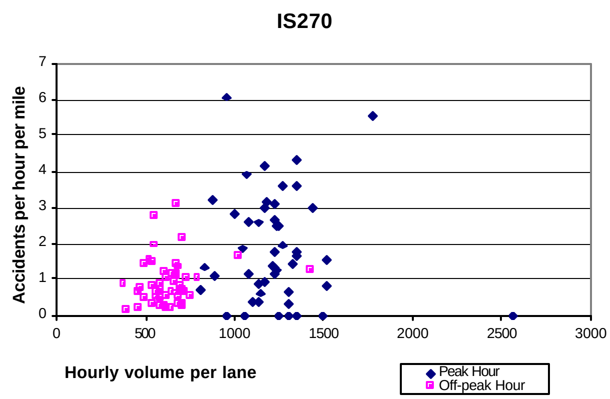


Figure 3-20 Accident frequency versus volume for I-95

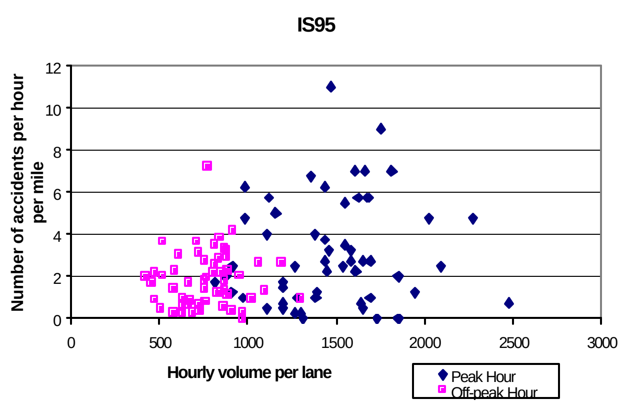


Figure 3-21 Accident frequency versus volume for I-695

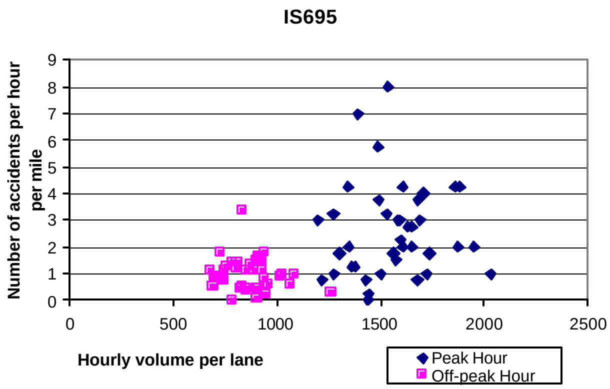


Figure 3-22 Accident frequency versus volume for I-495

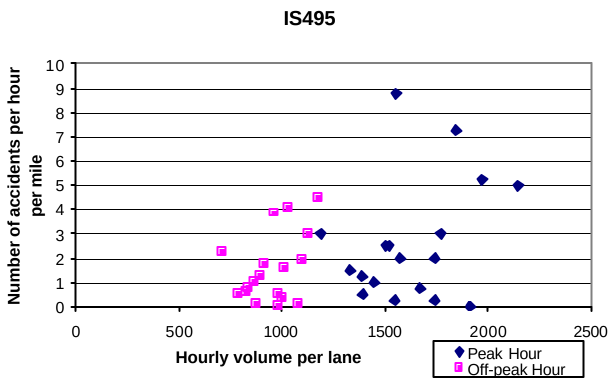
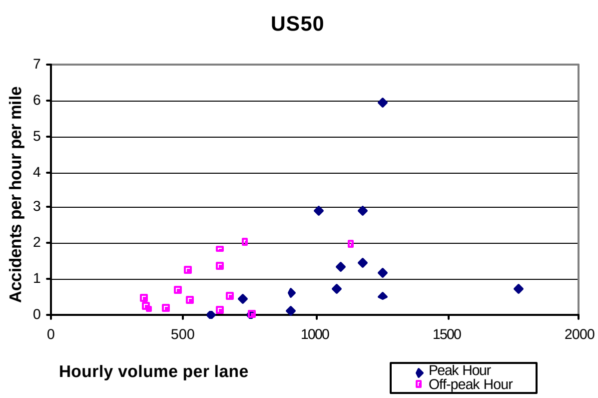


Figure 3-23 Accident frequency versus volume for US50



3.4 Model Estimation for Arterials

There are a variety of factors that may contribute to an increase in accident frequency. The previous exploratory analysis indicates that a single factor may not completely explain the relationship between accident frequency and congestion level. Therefore, this study further employed multivariate statistical methods to investigate such a relationship.

Based on the results in Section 3.3 and the information found in the literature review, the set of variables for inclusion in the analyses are listed below:

- The dependent variable of the accident frequency model is accidents per mile during peak or off-peak hours.
- The set of independent variables available for model development are:
  - x_1: Annual average peak hour volume and off-peak hour volume – "*volume*"
  - x_2: Median type (divided or not) – "*median*"
  - x_3: Number of intersections per unit length on a link – "*intdensity*"
  - x_4: Section length – "*length*"
  - x_5: Number of through lanes – "*thruln*"

From the correlation matrix of the independent variable (Table 3-6), it becomes clear that the number of intersections and the section length of road links are highly correlated. Therefore, the remaining analysis uses intersection density (the number of intersections divided by the section length) instead of the number of intersections directly. In addition, there is a high correlation between the number of through lanes and median type or section length, resulting from the design properties and the link-clustering operations. For example, divided-median roadway links are usually associated with a higher number of through lanes as opposed to undivided-median roadway links, explaining the positive correlation between these two variables.

Table 3-6 Correlation matrix for candidate variables

Y	Accidents per mile
x_1	Volume per lane
x_2	Divided median or not
x_3	Number of intersections
x_4	Length of the roadway link
x_5	Number of through lanes

Correlation Matrix

	Y	X_1	X_2	X_3	X_4
Y	1.00000				
X_1	0.33999	1.00000			
X_2	0.18828	-0.01131	1.00000		
X_3	0.23544	0.05562	-0.07749	1.00000	
X_4	-0.31710	-0.02707	-0.20435	**0.41339**	1.00000
X_5	0.43603	-0.01816	**0.46278**	0.00981	**-0.43354**

Estimation method

As is well recognized, Poisson regression is one of the most effective methods for modeling accident occurrence. A concise presentation of the Poisson regression algorithm can be found in Appendix-1. When using Poisson regression, it is important that the Lagrange Multiplier Test for over-dispersion also be conducted. Under the hypothesis of the Poisson distribution, the limiting distribution of LM statistics is Chi-Squared with one degree of freedom. If the over-dispersion is significant in the model either the Type I Negative Binomial or Type II Negative Binomial models should be used.

- Type I Negative Binomial model assumes the following relationship between mean and variance:

$$E[y] = \exp(X * b) = \mu$$
$$\text{Variance}[y] = \mu * (1 + a)$$

- Type II Negative Binomial model assumes the following relationship between mean and variance:

$$E[y] = \exp(X * b) = \mu$$
$$\text{Variance}[y] = \mu + a * \mu^2$$

Model estimation results

Of the 15 models in Table 3-7 that had different combinations of independent variables, the best model for arterials yielded the estimation results in Table 3-8.

Table 3-7 List of all models being evaluated for arterials

Functional form	Number of models	Estimation method
$Y = b_0 + b_1 X_1$	1	Poisson
$Y = b_0 + b_1 X_1 + b_2 X_i$, i=2,…,5	4	Poisson
$Y = b_0 + b_1 X_1 + b_2 X_i + b_3 X_j$, i=2,…,4, j=2,…,5, $i<j$	5	Poisson
$Y = b_0 + b_1 X_1 + b_2 X_i + b_3 X_j + b_4 X_k$, i=2,3, j=2,3,4, k=2,…,5, $i<j<k$	4	Poisson
$Y = b_0 + b_1 X_1 + b_2 X_2 + b_3 X_3 + b_4 X_4 + b_5 X_5$	1	Poisson
Total	15	

Table 3-8 Estimation results of the best arterial model with Poisson regression

Parameter	Estimate	t-statistic	P-value
C	2.960	11.891	[.000]
x_1 (Volume per lane)	.160	8.198	[.000]
x_2 (Median Indicator)	.152	1.702	[.089]
x_3 (Intersection density)	.021	9.543	[.000]
x_4 (Link length)	-.217	-3.361	[.001]
x_5 (Number of thru lanes)	.354	9.310	[.000]

Over-dispersion test result	
Chi-Squared statistics	P-value
1.097	[.295]

The estimation results shown in Table 3-8 illustrate the following conclusions for arterials:

- Accident frequency on arterials increases with the congestion level.
- The coefficient of the median indicator is positive and significant, which suggests that divided roadway links usually exhibit higher accident frequencies than undivided roadway links at the same volume levels. This may be attributed to relatively high speeds on the divided roadway links or limitations of the selected sample datasets.
- Accident frequency on arterials increases with intersection density (number of intersections per unit length of the roadway link).
- Arterials with a high number of through lanes are more likely to have a higher frequency of accidents.

To assess the potential impact of data aggregation on the estimation results, in this study Poisson regression was performed with the original dataset. In addition, since the last two explanatory variables has high correlation with median type and intersection density, only the first three explanatory variables are included in the new estimation with the original dataset. The estimated results are presented in Table 3-9, where the parameters for volume per lane, median, and intersection density are significant and have the same sign as the results using the aggregated database.

Table 3-9 Estimation results with Poisson regression for the original arterial links

Parameter	Estimate	t-statistic	P-value
C	4.02932	34.9559	[.000]
VOLUME	.126751	6.71211	[.000]
MEDIAN	.547611	5.01301	[.000]
INTDENSITY	.028548	9.23268	[.000]
Over-dispersion test result			
Chi-Squared statistics		P-value	
5.88179		[.015]	
Stability test results			
Number of coefficients: K = 4			
Number of observations in subset-1: $n_1 = 670$			
Number of observations in subset-2: $n_2 = 696$			
Residual sum of squares (scaled by 10^4):			
$\sum e_p^2 = 184695539$; $\sum e_1^2 = 64050659$; $\sum e_2^2 = 119047128$			
The resulting F statistics is $2.96 < F_{0.99}(4, 1358) = 3.34$			

Since the results on Table 3-9 illustrate the existence of over-dispersion, the Negative Binomial model was estimated and the results (NB2 model) are illustrated in Table 3-10. The estimated relationship between accident frequency and its primary explanatory variables, including volume, median, and intersection density, appeared to be consistent regardless of the differences among the datasets or the estimation algorithm used (see Tables 3-9 and 3-10).

Table 3-10 Estimation results with NB2 regression for the original arterial links

Parameter	Estimate	t-statistic	P-value
C	3.27443	17.8700	[.000]
VOLUME	.190534	5.44034	[.000]
MEDIAN	.816652	6.21659	[.000]
INTDENSITY	.054366	7.56145	[.000]
ALPHA	5.86524	24.7020	[.000]

To ensure that the estimated parameter signs are independent of the difference in the sample size, a parameter stability test was performed (refer to Appendix-2 for details). The test results (see Table 3-9) indicated that the reported relationship between accident frequency and its key factors are stable and will not vary with the available sample size.

3.5 Model Estimation for the Freeway Segment Dataset

Using the same estimation algorithm, this section explores the relationship between accident frequency and congestion level on freeways. The variables to be included in our model are listed below:

- The dependent variable: the accident frequency during peak or off-peak periods.
- The independent variables:
  - x1 : Volume per lane
  - x2 : Median width
  - x3 : Auxiliary lane ratio
  - x4 : Link length

$$\text{Auxiliary lane ratio} = \frac{\text{the total length of auxiliary lanes on a link}}{\text{the length of the link}}$$

  - x5 : Number of through lanes

The model estimation results are listed in Table 3-11, and all models were estimated using Poisson regression. The estimation results of the most consistent model are listed in Table 3-12.

Table 3-11 List of all models being evaluated for freeways

Functional form	Number of models	Estimation method
$Y = b_0 + b_1 X_1$	1	Poisson
$Y = b_0 + b_1 X_1 + b_2 X_i$, i=2,…,5	4	Poisson
$Y = b_0 + b_1 X_1 + b_2 X_i + b_3 X_j$, i=2,…,4, j=2,…,5, $i<j$	5	Poisson
$Y = b_0 + b_1 X_1 + b_2 X_i + b_3 X_j + b_4 X_k$, i=2,3, j=2,3,4, k=2,…,5, $i<j<k$	4	Poisson
$Y = b_0 + b_1 X_1 + b_2 X_2 + b_3 X_3 + b_4 X_4 + b_5 X_5$	1	Poisson
Total	15	

Table 3-12 Estimation results for freeways with Poisson regression

Parameter	Estimate	t-statistic	P-value
C	1.420	7.256	[.000]
x_1 (volume per lane)	.957E-03	9.107	[.000]
x_2 (median width)	-.2246E-02	-2.062	[.039]
x_3 (auxiliary lane ratio)	.126	1.501	[.133]
x_5 (number of thru lanes)	.058	3.133	[.002]
Over-dispersion test result			
Chi-Squared statistics		P-value	
.105		[.746]	

The over-dispersion test statistic is significant for this Poisson model, and requires additional analysis using the Negative Binomial regression. Tables 3-13 and 3-14 present the estimation results using NB1 and NB2 models. It was observed that the relationship between accident frequency and volumes per lane, median width, and the number of through lanes are all consistent regardless of the differences in the estimation algorithm (see Tables 3-12, 3-13, and 3-14). It is important to note that the parameters of variables observed also exhibited significant statistical stability as evidenced in the results of model stability test (see Table 3-14).

Table 3-13 Estimation results for freeways with NB1 regression

Parameter	Estimate	t-statistic	P-value
C	1.779	9.479	[.000]
x_1 (volume per lane)	.748E-03	8.819	[.000]
x_2 (median width)	-.298E-02	-3.845	[.000]
x_3 (auxiliary lane ratio)	.059	.938	[.348]
x_5 (number of thru lanes)	.052	2.760	[.006]
a	10.848	10.691	[.000]

Table 3-14 Estimation results for freeways with NB2 regression

Parameter	Estimate	t-statistic	P-value
C	1.176	5.068	[.000]
x1 (volume per lane)	.104E-02	9.694	[.000]
x2 (median width)	-.147E-02	-1.977	[.048]
x3 (auxiliary lane ratio)	.170	1.946	[.052]
x5 (number of thru lanes)	.070	2.837	[.005]
a	.576	12.053	[.000]
Stability test results			
Number of coefficients: K = 4 Number of observations in subset-1: n_1 = 181 Number of observations in subset-2: n_2 = 177			
Residual sum of squares (scaled by 10^8): $\sum e_p^2 = 98110 ; \sum e_1^2 = 43981 ; \sum e_2^2 = 52745$			
The resulting F statistics is 1.25 < F 0.95(5, 348) = 2.21 Therefore, the NB2 model is stable			

Based on the above stable significant results, the following can be concluded for freeways:

- Accident frequency on freeways tends to increase along with an increase in the congestion level.
- Wider medians can significantly reduce accident frequency on freeways.
- Accident frequency on freeways increases along with an increase auxiliary lane ratio, which is associated with potential lane-changing movements.
- Accident frequency on freeways increases along with an increase the number of through lanes.

3.6 Summary and Conclusions

This chapter investigated the relationship between accident frequency and congestion levels on sampled freeways and arterials, and includes exploratory analyses and multivariate statistical estimation using Poisson and Negative Binomial regressions. The research results were found to be consistent with previous assumptions, which are summarized below.

- Accident frequency on both freeways and arterials tends to increase with an increase in the congestion level.
- Divided arterial links exhibit higher accident frequencies compared to undivided arterial links at the same volume levels.
- Accident frequency on arterials increases along with the increase in intersection density (number of intersections per unit length of the arterial link).
- Wider medians can significantly reduce accident frequency on freeway links.
- Accident frequency on freeways increases with auxiliary lane ratio (the ratio of total length of auxiliary lanes on a link to its link length).
- Accident frequency increases with the increase in the number of through lanes for both freeway and arterial links.

CHAPTER 4
ACCIDENT RATE AND CONGESTION LEVEL

4.1 Introduction

This Chapter presents the research results for the analysis of the relationship between accident rate and congestion on both sample freeways and arterials. Included in this chapter is an exploratory analysis of accident rate during peak and off-peak hours and the potential factors that may contribute to changes in the accident rate under various traffic conditions. Overall, the primary focus of this chapter is to determine if a systematic pattern between accident rate and congestion on either freeways or arterials exists.

To begin, an exploratory analysis comparing the average accident rate between peak and off-peak periods was performed. It is expected that the peak hour accident rate will be generally lower than the average accident rate during the off-peak period if a negative correlation between the accident rate and congestion level exists. The comparison results, based on the data from five freeways and five local arterials, are presented in Section 4.3. In addition, the exploratory analysis includes a comparison of accident rate among sampled roadway segments experiencing different levels of congestion, and a bivariate correlation analysis between the accident rate and congestion levels. These analyses were performed to examine if highways with higher levels of congestion yield a lower accident rate.

Based on the findings of exploratory analyses, this study further investigated the target relationship between accident rate and congestion under the compound impacts of various contributing factors using Poisson and Negative Binomial regressions. The estimation results for freeways and arterials are presented in Section 4.4 and 4.5.

4.2 Data Set Available for Analysis

The accident data set used in this chapter is the same set of data used to examine accident frequency analysis, and includes the same link aggregations and the same surrogate variable for congestion. In the remaining sections and presentations the accident rate is defined as follows:

$$\text{Accident rate} = \frac{\text{Number of accidents on a link}}{\text{AADT} \times \text{The length of the link}}$$

4.3 Exploratory Analyses

The following exploratory analysis intends to investigate whether the accident rate decreases with the congestion level from three different perspectives, which include:

- A comparison between peak-hour and off-peak-hour accident rates, to determine if congestion has an impact on the resulting accident rate. As reported in the previous chapter, congestion on freeways and arterials exhibits a positive correlation with accident frequency, however the relationship with accident rate will be tested in this chapter.
- A cross-section comparison of accident rates on five sample local arterials and freeway segments was performed to evaluate whether roadways with higher levels of congestion yield lower accident rates.
- Testing the potential correlation between accident rate and volumes per lane, which is used as the surrogate variable for congestion level.

The results of above three exploratory analyses are presented in sequence below.

Comparison of accident rate in peak hours and in off-peak hours

Figure 4-1 presents the differences between peak and off-peak hour accident rates on five sample arterials, and Figure 4-2 illustrates the results for five sample freeway segments. On 75 of the 129 arterial links shown in Figure 4-1, accident rates during off-peak hours are higher than accident rates during peak hours. On the remaining links, accident rates during off-peak hours are equal to or lower than accident rates during peak hours. On more congested links (e.g. links 4-10 of MD335), accident rates during peak hours are significantly higher than accident rates during off-peak hours.

In comparison, it was also observed on 85 of the 179 freeway links, that the accident rates during off-peak hours were higher than accident rates during peak hours. On the remaining links, accident rates during off-peak hours were equal to or lower than accident rates during peak hours. For example, on 13 of the first 21 links of I-270 and on 28 of the 39 links of I-695, the accident rate during peak hours is higher than the accident rate during off-peak hours.

These two observations imply that critical factors other than congestion may have a significant impact on the accident rate and that the volume per lane may not be sufficient to fully capture the impact of congestion on accident rate. It is also likely that the relationship between accident rate and volume per lane may vary with volume level. For example, the relationship may differ from peak to off-peak hours on either freeways or arterials. In summary, the results of this analysis offer no definitive answer to the relationship between accident rate and congestion; however, it does establish the basis for further explorations in the ensuing sections.

Figure 4-1 A comparison of accident rate on five arterials during peak and off-peak hours

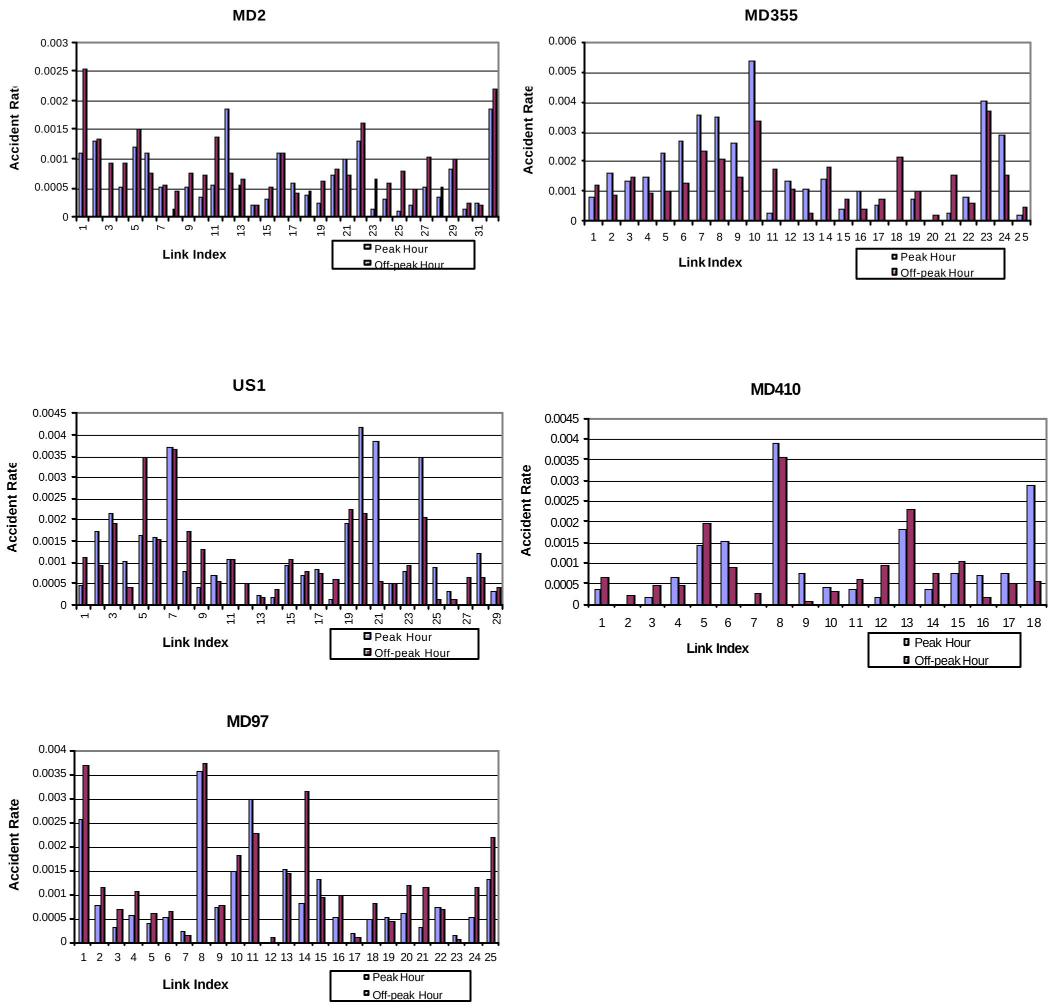
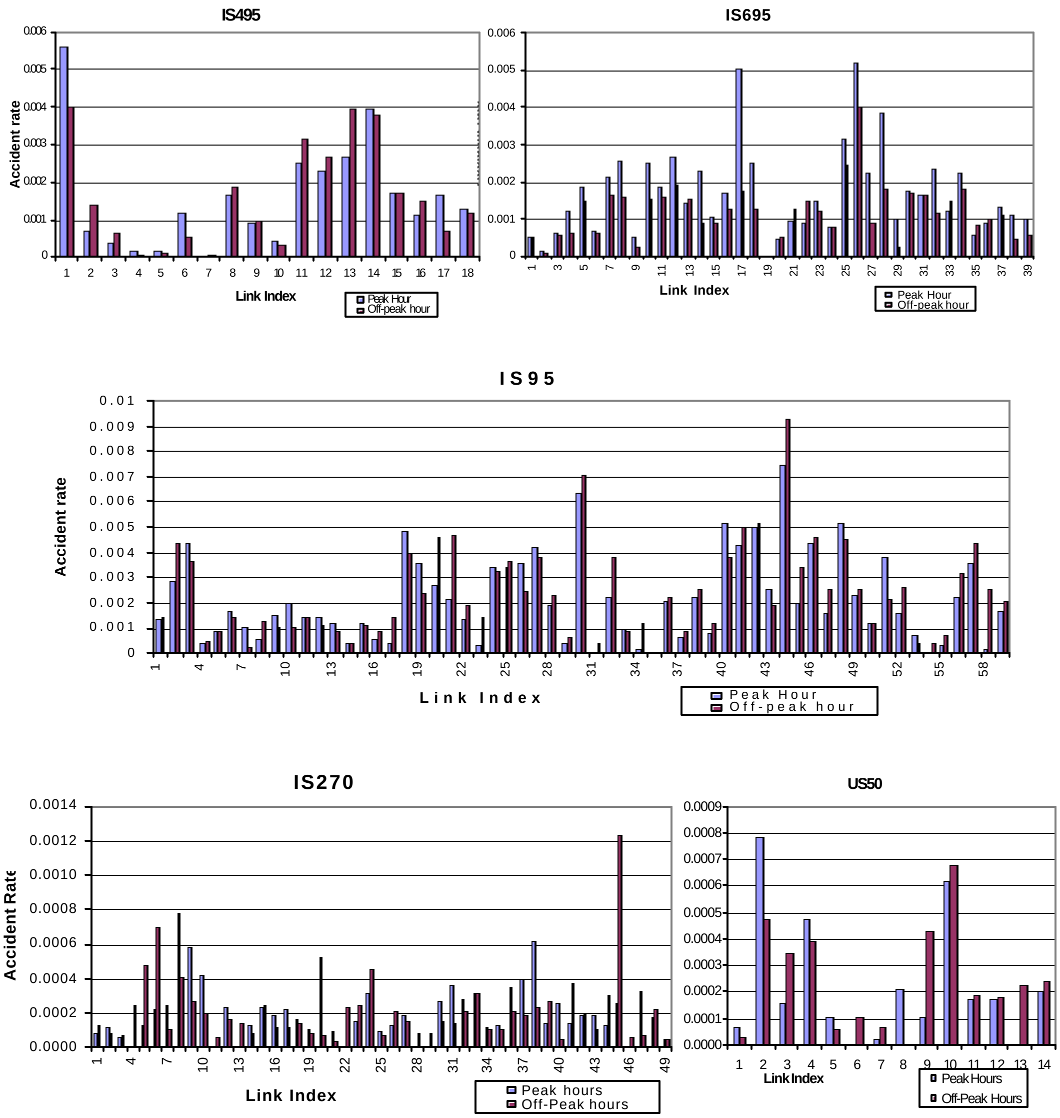


Figure 4-2 A comparison of hourly accidents on freeways during peak and off-peak hours



To further compare the peak-period accident rate with the off-peak accident rate for all sample freeways and arterials, this study has employed the following test (Greene, 2000) on the available dataset and the results are shown in Table 4-1.

Table 4-1 Procedures and results of the dummy variable method (Greene, 2000)

Hypothesis	H_1: Accident rate (peak) > Accident rate (off-peak) i.e. H_0: d = 0; H_1: d > 0				
Test procedures	$y_i = m + d * Dum + e_i$ Set *Dum* = 1, if the sample is in peak period = 0, otherwise				
Data used	Accident rate during peak hours (7-9AM and 4-6PM) Accident rate during off-peak hours				
Test results of five sample surface streets					
Route Name	MD2	MD355	US1	MD410	MD97
Sample Size (*n*)	32	25	29	18	25
T-statistic of *Dum* coefficient	-1.750	0.745	0.396	0.203	-1.166
T_{2n-2} (0.95)	1.669	1.676	1.672	1.688	1.676
Conclusion	Accept H_0	Accept H_0	Accept H_0	Accept H_0	Accept H_0

With respect to all five sampled arterials, Table 41 illustrates that it cannot be concluded that the average accident rate during the peak period is higher than the accident rate during the off-peak period. The same conclusions for freeways can be reached basedon the test results reported in Table 4-2. For example, negative parameters for I-695 tend to indicate that the average accident rate during peak hours is lower than the average accident rate during off-peak hours.

Table 4-2 Results of the dummy variable test for freeways

Test results of five sample freeway segments					
Route Name	I-495	I-695	I-95	I-270	US50
Sample Size (n)	18	39	59	49	14
T-statistic of *Dum* coefficient	0.071	-2.427	0.854	-0.778	0.286
T_{2n-2} (0.95)	1.669	1.665	1.658	1.661	1.701
Conclusion	Accept H_0	Accept H_0	Accept H_0	Accept H_0	Accept H_0

In summary, the inconclusive results illustrated in Table 4-2 indicate that additional factors need to be considered when exploring the complex interactions between accident rate and congestion.

Cross-section comparison of the accident rate

As stated previously, the following analyses were designed to test whether more congested highways experience lower accident rate and whether a systematic relationship between accident rate and congestion exists. Since all five freeways and arterials experienced different levels of congestion, one may expect that they should exhibit significantly different average accident rates.

Table 4-3 summarizes the results of ANOVA tests for the sampled arterials and freeways. As expected, the accident rate varies significantly among the five arterials with different volumes per lane during peak and off-peak periods. With respect to freeways, the test results indicate that a distinct difference during peak-hour accident rate among five sample freeways exists. However, the test also indicates that the off-peak accident rate does not vary significantly among the sampled freeways.

Table 4-3 ANOVA tests and results

Balanced ANOVA test for arterials	
Hypothesis	The average accident rates for all five arterials are statistically equal.
ANOVA test parameters	The number of factor levels (treatment groups): $k = 5$ The number of observations within each factor level: $n = 18$
Data used	Accident rate in peak hours (7-9AM and 4-6PM) Accident rate in off peak hours

Test results from the arterial dataset				
	Dependent Variable Y	F	$F_{4,85}^{0.975}$	*Conclusion*
1	*Peak-hour accident rate*	*3.06*	*2.48*	*Reject*
2	*Off-peak accident rate*	*5.91*	*2.48*	*Reject*

Unbalanced ANOVA test for freeways	
Hypothesis	The average accident rates for all five freeways are statistically equal.
ANOVA test parameters	The number of factor levels (treatment groups): $k = 5$ The number of observations within each factor level: $n_i = \{18, 39, 59, 49, 14\}$
Data used	Accident rate during peak hours (7-9AM and 4-6PM) Accident rate during off peak hours

Test results from the freeway dataset				
	Dependent Variable Y	F	$F_{4,85}^{0.975}$	*Conclusion*
1	*Peak-hour accident rate*	*4.11*	*2.42*	*Reject*
2	*Off-peak accidents rate*	*0.93*	*2.42*	***Accept***

Bivariate correlation test between the accident rate and congestion level

Figures 4-3 and 4-4 present the relationship between volume per lane and the resulting accident rate on each link for the sample freeways and arterials. These graphical relationships reveal two important factors:

- The accident rate does not exhibit any distinct trend with volumes per lane.
- Increasing the variance of the exhibited data patterns does not support the use of linear multivariate regression for further analyses.

Based on this information, the use of Poisson and Negative Binomial regressions for accident rate analysis should be examined, as these statistical models are more accurate when there is a better account for the non-linear and non-negative nature of accident rate data.

Figure 4-3 A graphical illustration of accident rate versus corresponding volume for arterials

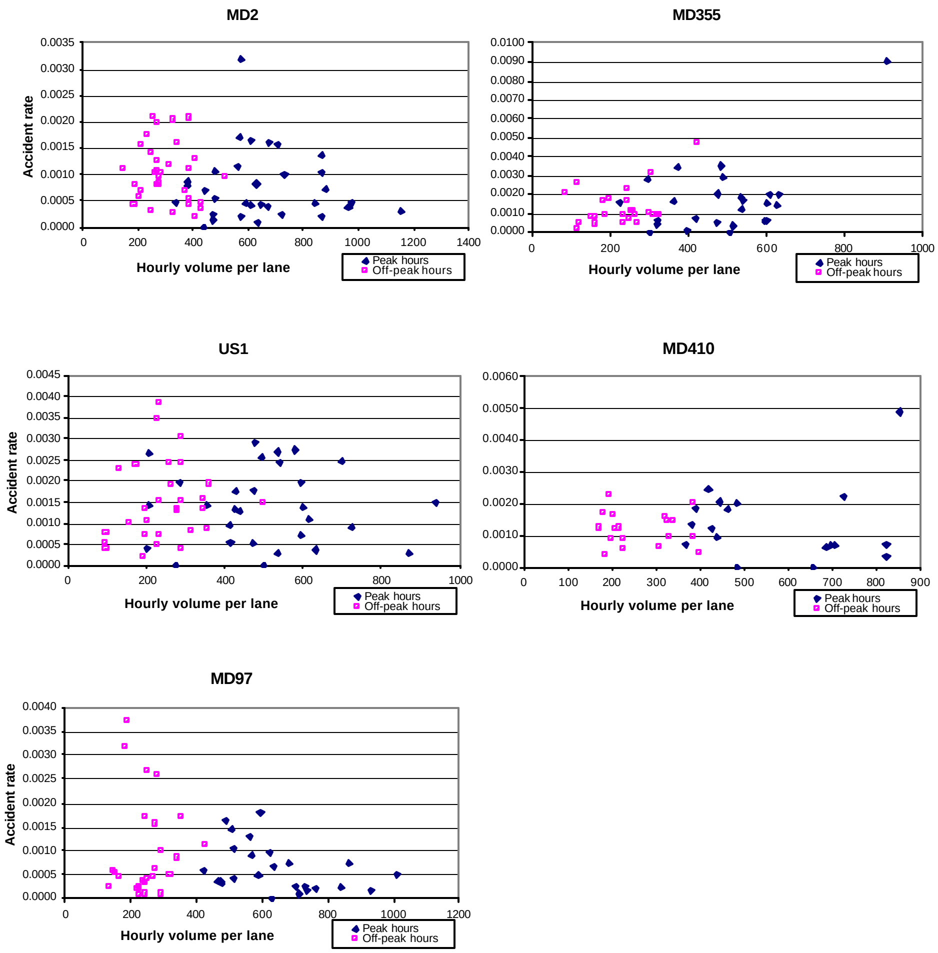
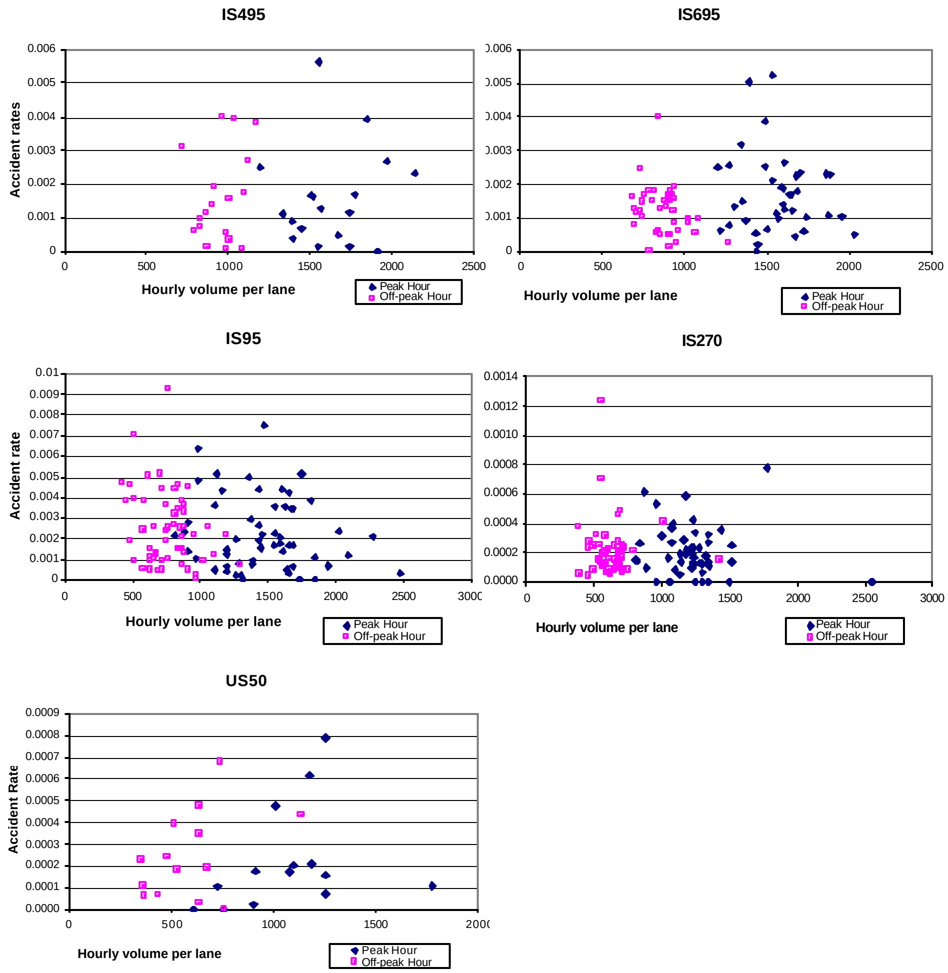


Figure 4 -4 A graphical illustration of accident rate versus corresponding volume for freeways



4.4 Model estimation for arterials

There are a variety of factors that may contribute to the variation of the accident rate. The previous exploratory analysis using a single factor may not be sufficient to render an "unbiased" picture of the relationship between the accident rate and congestion level. Therefore, this section attempts to further investigate any potential relationships using multivariate statistical methods, including Poisson and Negative Binomial regressions. The set of variables to be included in the model estimation are listed below:

- The dependent variable: the accident rate in peak or off-peak hours.
- The set of independent variables:
  - x_1: Annual average volume per hour during peak and off-peak periods
  - x_2: Median type (divided or not)
  - x_3: Intersection density = Number of intersections / Link length
  - x_4: Link length
  - x_5: Number of through lanes per link

The correlation matrix of the independent variable (see Table 4-4) makes it clear that the number of intersections and the length of roadway links are highly correlated (the correlation coefficient is 0.414). The analysis uses the intersection density (i.e., the number of intersections divided by the link length) as one of the explanatory variables in the model estimation. There is also a high correlation between the number of through lanes and median type or section length. These correlations come from the design properties and the link-clustering operations. For example, divided-median roadway links are usually associated with a higher number of through lanes than undivided-median roadway links, which explains the positive correlation between these two variables.

Table 4-4 Correlation matrix for candidate variables

Y	Accident rate
x_1	Volume per lane
x_2	Divided median or not
x_3	Number of intersections
x_4	Length of the roadway link
x_5	Number of through lanes

Correlation Matrix

	Y	X_1	X_2	X_3	X_4
Y	1.00000				
X_1	-0.08589	1.000000			
X_2	0.26387	-0.011310	1.00000		
X_3	0.16687	0.055620	-0.07749	1.00000	
X_4	-0.38058	-0.027073	-0.20435	0.41339	1.00000
X_5	0.57783	-0.018155	0.46278	0.00981	-0.43354

Estimation method

The Poisson regression is recognized as one of the most effective methods for examining accident related data. A concise presentation of the Poisson regression algorithm can be found in Appendix-1. When using Poisson regression, it is important that the Lagrange Multiplier Test for over-dispersion also be conducted. If the over-dispersion is found to be significant in the estimated results, it is suggested in the literature that Type-I Negative Binomial model or Type-II Negative Binomial model should be used. The fundamental assumptions for Type-I and Type-II Negative Binomial models are summarized below:

- Type I Negative Binomial model assumes the following relationship between mean and variance: $E[y] = \exp(X * b) = \mu$
  $\text{Variance}[y] = \mu * (1 + a)$
- Type II Negative Binomial model assumes the following relationship between mean and variance: $E[y] = \exp(X * b) = \mu$
  $\text{Variance}[y] = \mu + a * \mu^2$

Model estimation results

A total of 15 model specifications were estimated with Poisson regression (see Table 4-5). Among those, the specification shown in Table 4-6 best illustrates the relationship between accident rate and congestion.

Table 4-5 List of estimated models

Functional form	Number of models	Estimation method
$Y = b_0 + b_1 X_1$	1	Poisson
$Y = b_0 + b_1 X_1 + b_2 X_i$, i=2,…,5	4	Poisson
$Y = b_0 + b_1 X_1 + b_2 X_i + b_3 X_j$, i=2,…,4, j=2,…,5, $i<j$	5	Poisson
$Y = b_0 + b_1 X_1 + b_2 X_i + b_3 X_j + b_4 X_k$, i=2,3, j=2,3,4, k=2,…,5, $i<j<k$	4	Poisson
$Y = b_0 + b_1 X_1 + b_2 X_2 + b_3 X_3 + b_4 X_4 + b_5 X_5$	1	Poisson
Total	15	

Table 4-6 Model estimation results for arterials

Parameter	Estimate	t-statistic	P-value
C	7.268	35.844	[.000]
x_1 (Volume per lane)	-.048	-2.438	[.015]
x_2 (Median Indicator)	.139	1.733	[.083]
x_3 (Intersection density)	.020	6.995	[.000]
x_4 (Link length)	-.225	-4.274	[.000]
x_5 (Number of thru lanes)	.325	10.321	[.000]

Over-dispersion test result	
Chi-Squared statistics	P-value
0.348	[.555]

From the estimation results listed in Table 4-6, the following conclusions were reached:

- The accident rate for local arterials tends to decrease with the volume level.
- The coefficient of the median indicator is positive and significant, which suggests that divided roadway links exhibit higher accident rates than undivided roadway links with the same volume levels. This may be due to the higher speed on the divided roadway links or other factors that have not been identified.
- The accident rate increases with intersection density (number of intersections per unit length of the roadway link).
- The accident rate on arterial links increases with the number of through lanes. More through lanes indicate that potential lane-changing maneuvers on the roadway link may contribute to an increase in the accident frequency and rate at the same volume levels.

To eliminate the potential biases due to the link partitioning process, this study also used the original (not clustered) link dataset to perform the model estimation. Estimation results of the Poisson regression model with the original (not aggregated) dataset (1366 links in total) are listed in Table 4-7, note that the volume per lane is a significant variable and has a negative coefficient. The divided median has a significant positive coefficient, which implies that the accident rate tends to be higher on a divided roadway link than on an undivided roadway link.

Table 4-7 Poisson model for the original arterial links

Parameter	Estimate	t-statistic	P-value
C	4.61215	36.7155	[.000]
VOLUME	-.085300	-3.37066	[.001]
MEDIAN	.327695	2.96343	[.003]
INTDENSITY	.027944	8.52762	[.000]

Over-dispersion test result	
Chi-Squared statistics	P-value
.502	[.478]

Stability test results
Number of coefficients: K = 4 Number of observations in subset-1: $n_1 = 670$ Number of observations in subset-2: $n_2 = 696$
Residual sum of squares (scaled by 10^{10}): $\sum e_p^2 = 82181205$; $\sum e_1^2 = 38229845$; $\sum e_2^2 = 43787362$
The resulting F statistics is 0.68 < F 0.95(4,1358) = 2.37 Therefore, the Poisson model is stable

Since the over-dispersion test statistics are not significant for the previous models (see Table 4-7), it is not necessary to perform the Negative Binomial model estimation.

To ensure that all estimated parameter signs are independent of the difference in sample size, a standard parameter stability test was also performed. The primary procedures are summarized in Appendix-2, and test results (see Table 47) clearly indicate that the estimated relationship between accident rate and key factors is stable, and will not vary with sample size. f the same analysis is conducted using a larger dataset the results conclusion should be identical to those reported in this section.

4.5 Model estimation for freeway segments

Using the same estimation algorithm, this section explores the relationship between the accident rate and the congestion level on freeways. Variables to be included in the model estimation are as follows:

- The dependent variable (Y): the accident rate during peak or off-peak hours.
- The set of independent variables:
- x_1 : Volume per lane
- x_2 : Median width
- x_3 : Auxiliary lane ratio

$$\text{Auxiliary lane ratio} = \frac{\text{the total length of auxiliary lanes on a link}}{\text{the length of the link}}$$

- x_4 : Link length
- x_5 : Number of through lanes

Estimation results

A total of 15 model specifications with different variables were tested using Poisson regression (see Table 4-8).

Table 4-8 List of estimated models

Functional form	Number of models	Estimation method
$Y = b_0 + b_1 X_1$	1	Poisson
$Y = b_0 + b_1 X_1 + b_2 X_i$, i=2,…,5	4	Poisson
$Y = b_0 + b_1 X_1 + b_2 X_i + b_3 X_j$, i=2,…,4, j=2,…,5, $i<j$	5	Poisson
$Y = b_0 + b_1 X_1 + b_2 X_i + b_3 X_j + b_4 X_k$, i=2,3, j=2,3,4, k=2,…,5, $i<j<k$	4	Poisson
$Y = b_0 + b_1 X_1 + b_2 X_2 + b_3 X_3 + b_4 X_4 + b_5 X_5$	1	Poisson
Total	15	

With both peak hour data and off-peak hour data, the most consistent estimation results from these Poisson models are listed in Table 4-9. Note that for volume per lane the variable was not significant but did have a positive coefficient.

Table 4-9 Estimation results with Poisson regression for freeways

Parameter	Estimate	t-statistic	P-value
C	5.81252	28.0219	[.000]
x_1 (volume per lane)	.1140E-03	1.08253	[.279]
x_2 (median width)	-.1358E-02	-1.40470	[.160]
x_3 (auxiliary lane ratio)	.142474	1.87529	[.061]
x_5 (# of thru lanes)	-.064002	-3.12061	[.002]

Over-dispersion test result	
Chi-Squared statistics	P-value
.167	[.683]

To further investigate the impact volume per lane has on the accident rate, this study divided the freeway dataset into peak and off-peak subsets, and estimated each independently using Poisson regression. Table 4-10 and Table 4-11 present the most consistent estimation results of the Poisson models from these two subsets. Note that although the parameter for volume per lane remains insignificant in both models, the statistical significance level indicates that the impact of peak hour volume on the accident rate should be examined further.

Table 4-10 Estimation results with Poisson regression for the peak-hour freeway dataset

Parameter	Estimate	t-statistic	P-value
C	5.49095	16.0450	[.000]
x_1 (volume per lane)	.29446E-03	1.56221	[.118]
x_2 (median width)	-.25313E-02	-1.78660	[.074]
x_3 (auxiliary lane ratio)	.182054	1.66505	[.096]
x_5 (# of thru lanes)	-.057942	-2.25588	[.024]

Over-dispersion test result	
Chi-Squared statistics	P-value
.131	[.718]

Table 4-11 Estimation results with Poisson regression for the off-peak-hour freeway dataset

Parameter	Estimate	t-statistic	P-value
C	5.74487	14.7920	[.000]
x_1 (volume per lane)	.026450	.858914	[.390]
x_2 (median width)	-.5371E-03	-.401647	[.688]
x_3 (auxiliary lane ratio)	.084449	.783483	[.433]
x_5 (# of thru lanes)	-.064892	-1.95222	[.051]

Over-dispersion test result	
Chi-Squared statistics	P-value
.006	[.940]

The results also show that the t-statistic of volume per lane in Table 4-10 is 1.56 and is close to the significance boundary. In addition, the mean and variance of accident rate (scaled by 10^5) are 24.9 and 428.18 (a ratio of 0.058), which suggests the need to use the Negative Binomial regression models.

In Table 4-12 the estimation results with NB1 regression were based on the off-peak data. The variable of volume per lane remained insignificant. However, it is important to note that the volume per lane that represents the congestion level exhibits a significant sign when the NB1 model is estimated with the peak-hour dataset (see Table 4-13).

Table 4-12 Estimation results with NB1 for the off-peak-hour freeway dataset

Parameter	Estimate	t-statistic	P-value
C	3.27640	11.1825	[.000]
x_1 (volume per lane)	.028692	1.16502	[.244]
x_2 (median width)	-.9382E-03	-1.14488	[.252]
x_3 (auxiliary lane ratio)	.066380	.783986	[.433]
x_5 (# of thru lanes)	-.040057	-1.50911	[.131]
a	12.0517	8.11631	[.000]

Table 4-13 Estimation results with NB1 for the peak-hour freeway dataset

Parameter	Estimate	t-statistic	P-value
C	3.01253	9.11135	[.000]
x_1 (volume per lane)	**.037621**	**2.31598**	**[.021]**
x_2 (median width)	-.442367E-02	-3.41930	[.001]
x_3 (auxiliary lane ratio)	.062311	.648298	[.517]
x_5 (# of thru lanes)	-.031555	-1.08130	[.280]
a	16.3193	7.62379	[.000]

Table 4-13 summarizes the estimation results for the NB1 model based on the peak-hour freeway accident data where both the volume per lane and the median width had significant impacts on the accident rate. In contrast, none of these candidate variables in the off-peak model revealed any significant signs. Therefore, the following tentative conclusions regarding the relationship between accident rate and congestion can be reached:

- Accident rate on freeways seems to increase with traffic volume during peak hours.
- Wider medians can significantly reduce the accident rate on freeways.

- Accident rate on freeways during off-peak hours tends to be independent of traffic volume levels.

As shown in Table 4-13 (NB1 model), the relationship was evaluated using the stability test. The stability test results indicate that the reported relationship between accident rate and key factors is stable and does not vary with sample size.

4.6 Summary and Conclusions

Chapter 4 attempted to explore the relationship between accident rate and congestion level on sampled freeways and arterials. The investigation of this relationship includes exploratory analyses and multivariate model development using the Poisson and Negative Binomial regressions. The results indicate the following:

- For arterials, the accident rate tends to decrease as the volume increases.
- For freeway segments, the accident rate during off-peak hours appears to be quite random, exhibiting no systematic relation with the traffic volume.
- During the peak-congested period, accident rates on freeways appear to increase significantly with traffic volume.
- Divided arterial links generally exhibit higher accident rates than undivided arterial links at the same volume level.
- Wider medians on freeways can significantly reduce the accident rate.
- The accident rate on arterials is likely to increase with intersection density.
- An increase in the number of through lanes may cause a significant increase in the accident rate on arterials.

CHAPTER 5
ACCIDENT SEVERITY AND CONGESTION LEVEL

5.1 Introduction

This chapter examines the relationship between accident severity and congestion level. The focus of this chapter is to examine the common belief of many traffic safety professionals that accident severity on freeways or arterials decreases with congestion because of the high traffic volumes and reduced flow speed.

An exploratory analysis was conducted to identify potential factors associated with accident severity such as accident location, roadway geometry, and driver conditions. This was followed by an aggregated analysis of the relationship between the number of accidents in each severity level and congestion level on sample freeways and arterials. Based on the results of the exploratory analysis, a final examination using the Ordered Probit models was conducted to estimate the relationship between accident severity and the identified key variables. It was anticipated that hourly volume per lane, the surrogate variable for the congestion level, would have a negative coefficient in the estimation results if more accidents occurred during periods with higher congestion levels.

Chapter 5 provides a description of the sample data and the severity classifications in Section 5.2 and 5.3. Section 5.4 presents the exploratory analysis results for the arterial and freeway databases. Section 5.5 provides the aggregated analysis results illustrating the relationship between accident severity and congestion levels. Applications of the multivariate model for estimating the relationship between accident severity and all associated factors, along with the research results, are presented in the last two sections.

5.2 Data Available for Analysis

The accident information records from the MAARS contain a comprehensive set of accident data for analysis, including the injury severity of the drivers/occupants, the number of persons injured, weather conditions, visibility condition, road surface condition, collision type, and the location and the time of each accident. However, the traffic volume during the time of each accident is not recorded in the MAARS database.

As described in earlier chapters, the highway information system from Maryland State Highway Administration provides traffic and geometric information associated with most arterials and freeway segments. After integrating these two databases with accident location information, it becomes easier to obtain the AADT and geometric information related to each recorded accident.

The analyses reported in the remainder of this chapter are based on the individual accident data points recorded in Year 2000 for five arterials and five freeway segments (see Table 5-1).

Table 5-1 Accident dataset for analysis

	4542 accident data from the surface street dataset		**5402** accident data from the freeway segment dataset	
Sample Year	2000		2000	
Source Index	Road name	Segment location	Road name	Segment location
1	US1	Between Baltimore City Line and Washington DC Line	I-495	Between Virginia State Line and I-95 Exit 27
2	MD2	The entire length	I-270	The entire length
3	MD97	The entire length	I-695	The entire length
4	MD355	The entire length	I-95	Between Baltimore City Line and Virginia State Line
5	MD410	The entire length	US50	Between Washington DC Line and Bay Bridge

Severity classification

Table 5-2 presents two classifications for accident severity by the Maryland State Highway Administration (SHA). In analyzing the severity data, this study employed the five-level accident severity classification by the SHA, which includes property damage only, possible injury, capacitating injury, non-capacitating injury, and fatal. Since the category of "possibly injured" has not been clearly defined, this study has also explored the impact of grouping "possibly injured," either with "injured" or "not injured" on the estimated results. The models reported in later sections were estimated to determine the relationship between accident severity and congestion under various scenarios of data aggregation and classification. Also note that the probable under-reporting of property damage only accidents exists due to the decreasing police response to such accidents.

Table 5-2 Severity classification

For Accidents	For Drivers/Occupants/ Pedestrians
1. Property Damage Only 2. Injury 3. Fatal	1. Not injured (Property damage only) 2. Possibly injured 3. Injured (Capacitating injury) 4. Disabled (Non-capacitating injury) 5. Fatal

5.3 Exploratory Analysis for the arterial database

An exploratory analysis was conducted to identify variables for further econometric model development, and includes a comparison of accident frequency at various severity levels for each potential contributing factor. Based on the differences in key characteristics, a list of candidate exploratory variables was classified into the following three groups.

- Roadway geometric and weather condition variables:
    - Median type (divided or not)
    - Number of through lanes
    - Intersection or not
    - Work zone or not
    - Weather conditions (e.g., rain, snow)

- Traffic condition variables:
    - Traffic composition
    - Annual Average Hourly Volume (per lane)
- Driver condition variables:
    - Drinking alcohol or using drugs

Preliminary comparisons of accident severity distribution, classified with the above list of critical variables are presented in the sequence below:

Peak hours/ off-peak hours (for arterials)

Figure 5-1 and Table 5-3 summarize the comparison results for accident severity distributions between peak and off-peak hours. The accident severity distribution exhibits a similar pattern between peak hours and off-peak hours, where the percentage of accidents decreases with the severity level.

Figure 5-1 The accident severity distribution in peak and off-peak periods on arterials

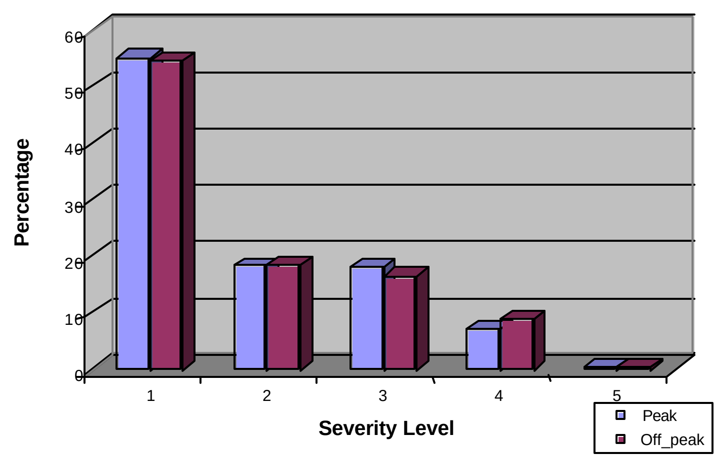


Table 5-3 Distribution of arterial accidents by severity in peak and off-peak periods

Aggregated by *Peak hour or not*		Severity					Total
		Level-1	Level-2	Level-3	Level-4	Level-5	
Peak	# of accidents	648	217	216	85	6	1172
	Percentage	*55.3*	*18.5*	*18.4*	*7.3*	*0.5*	*100*
Off-Peak	# of accidents	1850	630	562	307	21	3370
	Percentage	*54.9*	*18.7*	*16.7*	*9.1*	*0.6*	*100*

At intersection/ not at intersection

Similar to the previous analysis, this comparison was performed to evaluate the severity distribution of accidents that occurred at intersections with those at roadway segments. The primary concern is to identify if the location, such as intersection, plays any significant role in the resulting severity the accident. Figure 5-2 and Table 5-4 present the comparison results of the accident severity distribution between accidents that occurred at intersections and those that occurred at non-intersection locations.

Figure 5-2 A comparison of the severity distribution of accidents that occurred at intersections and non-intersection locations

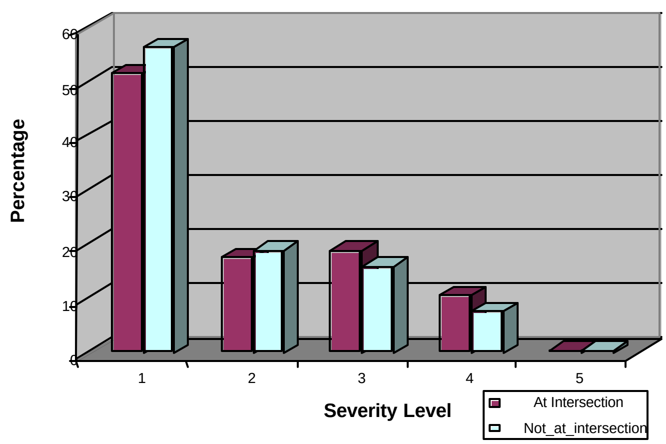


Table 5-4 Distribution of accidents by severity at intersections or non-intersection locations

Aggregated by *At-intersection or not*		Severity					Total
		Level-1	Level-2	Level-3	Level-4	Level-5	
At Intersection	# of accidents	802	275	295	164	8	1544
	Percentage	*51.9*	*17.8*	*19.1*	*10.6*	*0.5*	*100*
Not at Intersection	# of accidents	1696	572	483	228	19	2998
	Percentage	*56.6*	*19.1*	*16.1*	*7.6*	*0.6*	*100*

Accordingly, given that an accident already happened, the probability of having the accident at different levels of severity are summarized as follows:

	Property damage only	injury	fatality
Intersection	0.697	0.297	0.005
Non-intersection	0.757	0.237	0.006

The probability of having an accident that results in "injury" at intersections is about 0.297 and higher than the probability at non-intersection locations. This is consistent with the finding that accidents occurred at non-intersection locations are more likely to be at the level of property damage only (0.757 vs. 0.697).

Weather conditions

Figure 5-3 and Table 5-5 present the impact of weather conditions on the distribution of accident severity.

Figure 5-3 The severity distribution of arterial accidents under various weather conditions

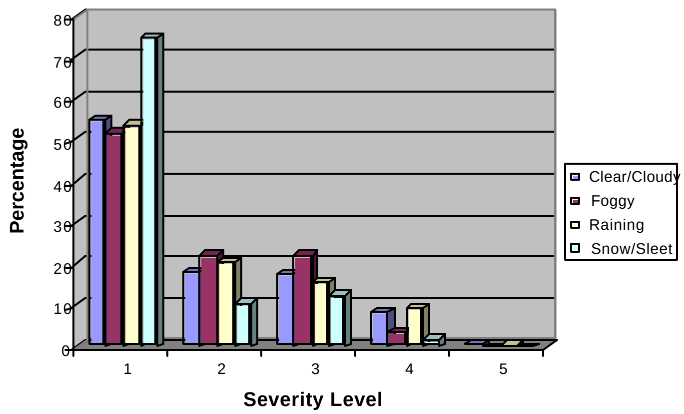


Table 5-5 Distribution of arterial accidents by severity under various weather conditions

Aggregated by *Weather condition*		Severity					Total
		Level-1	Level-2	Level-3	Level-4	Level-5	
Clear/ Cloudy	# of accidents	1964	653	627	305	23	3572
	Percentage	*55.0*	*18.3*	*17.6*	*8.5*	*0.6*	*100*
Foggy	# of accidents	14	6	6	1	0	27
	Percentage	*51.9*	*22.2*	*22.2*	*3.7*	*0.0*	*100*
Raining	# of accidents	471	182	138	82	3	876
	Percentage	*53.8*	*20.8*	*15.8*	*9.4*	*0.3*	*100*
Snow/ Sleet	# of accidents	42	6	7	1	0	56
	Percentage	*75.0*	*10.7*	*12.5*	*1.8*	*0.0*	*100*

By using the clear/cloudy condition as a base for comparison, the rain condition has no distinct impact on accident severity; however, the snow condition often results in more accidents at lower severity levels.

In work-zone or not (for arterials)

Figure 5-4 and Table 5-6 illustrate the comparison results for accident distribution by severity in and not in work-zones. Since work-zone safety has long been a primary concern in traffic operations, it is essential to know if work-zone operations have an effect on the severity of accidents.

Figure 5-4 The distribution of arterial accidents by severity for those in work-zones or non-work-zone locations

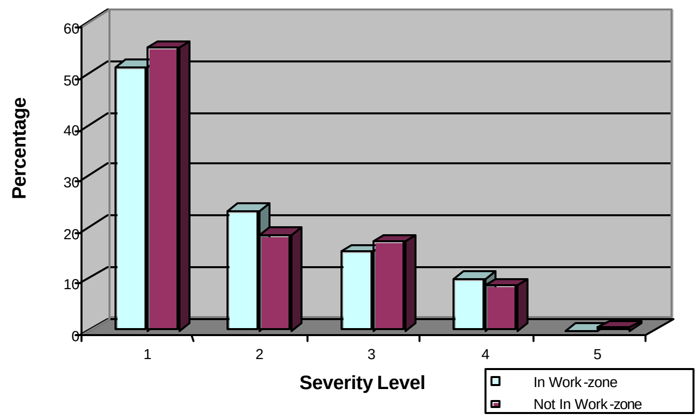


Table 5-6 Distribution of arterial accidents by severity in work-zones or non-work-zone locations

Aggregated by *In work-zone or not*		Severity					Total
		Level-1	Level-2	Level-3	Level-4	Level-5	
In Work-zone	# of accidents	57	26	17	11	0	111
	Percentage	*51.4*	*23.4*	*15.3*	*9.9*	*0.0*	*100*
Not in Work-zone	# of accidents	2441	821	761	381	27	4431
	Percentage	*55.1*	*18.5*	*17.2*	*8.6*	*0.6*	*100*

The preliminary statistics indicate that work-zone operations may not be a significant contributor to accident severity. For example, the probability that an accident will occur at the level of "property damage only" is 0.514 in work-zones, compared to 0.551 when in non-work-zone locations. The total percentage of accidents at the first two severity levels is 53.8% in work-zones and 53.6% for those in non-work-zone locations.

Median type

Figure 5-5 and Table 5-7 present the distribution of accidents by severity level for those that occurred on divided and undivided highway segments. The analysis indicates that the existence of highway medians may not reduce accident frequency, but do contribute to the improvement of safety.

Figure 5-5 The severity distribution of arterial accidents on arterials with various median types

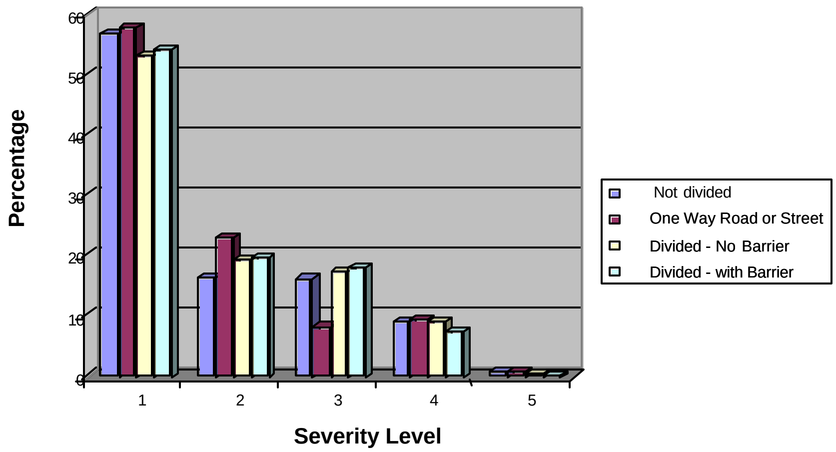


Table 5-7 Distribution of arterial accidents by severity on arterials with various median types

Aggregated by *Median type*		Severity					Total
		Level-1	Level-2	Level-3	Level-4	Level-5	
Not Divided	# of accidents	946	293	264	142	14	1659
	Percentage	*57.0*	*17.7*	*15.9*	*8.6*	*0.8*	*100*
Divided	# of accidents	1465	529	488	236	12	2730
	Percentage	*53.7*	*19.4*	*17.9*	*8.6*	*0.4*	*100*

The statistics summarized in Table 5-7, however, do not provide a definitive answer regarding the impact of median type on the accident severity, which indicates the need to explore the compound impacts of other contributing factors.

Driver conditions

Figure 5-6 and Table 5-8 illustrate the results of a comparison of the number of accidents and their distributions at different severity levels under the following three categories of driver conditions: apparently normal, had been drinking, and other abnormal conditions (e.g. using drugs or having physically defects).

Figure 5 -6 The severity distribution of arterial accidents for drivers under various conditions

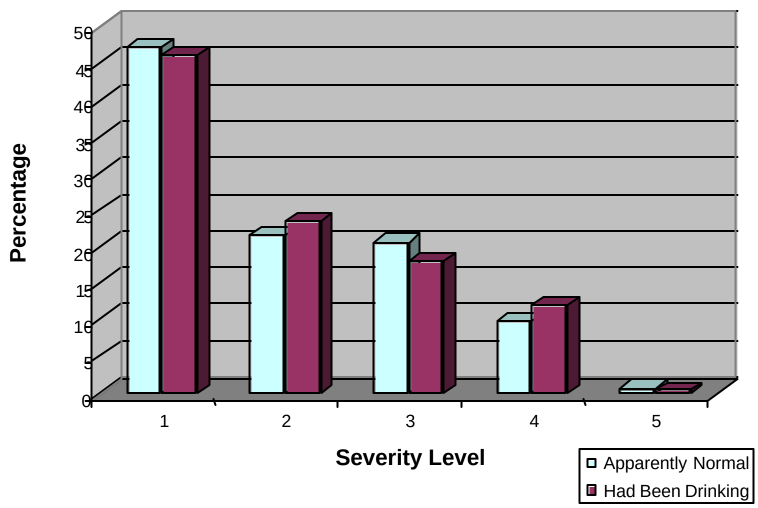


Table 5-8 Distribution of arterial accidents by severity for drivers under various conditions

Aggregated by *Driver condition*		Severity					Total
		Level-1	Level-2	Level-3	Level-4	Level-5	
Apparently Normal	# of accidents	1926	881	839	400	25	4071
	Percentage	*47.3*	*21.6*	*20.6*	*9.8*	*0.6*	*100*
Had been drinking	# of accidents	120	61	47	31	1	260
	Percentage	*46.2*	*23.5*	*18.1*	*11.9*	*0.4*	*100*
Other abnormal	# of accidents	19	8	5	12	6	50
	Percentage	*38.0*	*16.0*	*10.0*	*24.0*	*12.0*	*100*

Based on above statistics, one may reach a tentative conclusion that drivers are more likely to experience severe accidents if they are under the influence of alcohol or are affected by other abnormal variables (e.g. drugs, physical defects).

Similar exploratory analyses were conducted using visibility condition (daylight, dawn/dusk, dark-lights on, dark- no lights), the number of through lanes, and collision type (head on, rear end, sideswipe, etc.), as exploratory variables. However, the preliminary results indicated that none of these factors exhibited a significant impact on accident severity. Therefore, the following factors were included in the estimation of the relationship between congestion and accident severity on arterials:

- Median type (divided or not)
- At an intersection or not
- In a work zone or not
- Weather conditions (Snow/Sleet and Fog)

5.4 Exploratory Analysis for the freeway database

A preliminary set of variables for analysis are summarized below:

- Roadway geometric and weather condition variables:
  - Number of through lanes
  - Work zone or not
  - Auxiliary lane ratio

$$\text{Auxiliary lane ratio} = \frac{\text{the total length of auxiliary lanes on a link}}{\text{the length of the link}}$$

  - Weather (rain, snow, or other conditions)
- Traffic condition variables:
  - AADT, and Peak-hour volume
  - Annual Average Hourly Volume (per lane)
- Driver condition variables:
  - Drinking alcohol or using drugs

Peak hours/ off-peak hours (for freeways)

Table 5-9 summarizes the distribution of accidents by severity during peak and off-peak hours. The severity distribution patterns for peak and off-peak periods exhibit no significant differences for selected freeways.

Table 5-9 Distribution of freeway accidents by severity in peak and off-peak periods

Aggregated by *Peak hour or not*		Severity					Total
		Level-1	Level-2	Level-3	Level-4	Level-5	
Peak	# of accidents	897	283	203	117	6	1506
	Percentage	*59.6*	*18.8*	*13.5*	*7.8*	*0.4*	*100*
Off-Peak	# of accidents	2264	654	595	353	30	3896
	Percentage	*58.1*	*16.8*	*15.3*	*9.1*	*0.8*	*100*

Weather conditions

A preliminary exploration of potential weather impacts on the severity of accidents is presented in Table 5-10. Based on the results reported in the statistical summary, snow conditions tend to cause less severe accidents despite the fact that it is likely that during these conditions more accidents occur. For example, an accident that occurs on a day when it is snowing has a 0.738 probability to reach Level-1 severity (property damage only).

Table 5-10 Distribution of freeway accidents by severity under various weather conditions

Aggregated by *Weather condition*		Severity					Total
		Level-1	Level-2	Level-3	Level-4	Level-5	
Clear/ Cloudy	# of accidents	2559	768	660	397	34	4418
	Percentage	57.9	17.4	14.7	9.0	0.8	100
Foggy	# of accidents	16	3	6	3	0	38
	Percentage	57.1	10.7	21.4	10.7	0.0	100
Raining	# of accidents	472	148	121	61	1	803
	Percentage	58.8	18.4	15.1	7.6	0.1	100
Snow/ Sleet	# of accidents	96	14	11	9	0	130
	Percentage	73.8	10.8	8.5	6.9	0.0	100

In work-zone or not in work-zone

Table 5-11 summarizes the differences between accident distribution by severity for accidents that occurred in work-zones and on normal freeway segments. The preliminary statistics reported tend to offer no definitive conclusion regarding the potential impacts of work-zone on accident severity.

Table 5-11 Distribution of freeway accidents by severity within and beyond work-zones

Aggregated by *In work-zone or not*		Severity					Total
		Level-1	Level-2	Level-3	Level-4	Level-5	
In Work-zone	# of accidents	68	22	19	18	0	127
	Percentage	53.5	17.3	15.0	14.2	0.0	100
Not in Work-zone	# of accidents	3093	915	779	452	36	5275
	Percentage	58.6	17.3	14.8	8.6	0.7	100

Driver conditions

Similar to the analysis of accidents on arterials, Table 5-12 presents the distribution of accidents by severity and by driver condition. In the MAARS accident database, all drivers involved in accidents are classified in one of the following groups: normal, had been drinking, and other abnormal states such as using drugs. The distinct differences in the resulting severity are proven by the statistics illustrated in Table 5-12. For example, drivers under abnormal conditions, excluding those that "had been drinking," have about a 0.138 probability to be involved in an accident that results in fatalities (Level-5). This is compared to only 0.005 of drivers who have accidents under normal conditions. Therefore, it is likely that driver conditions are a significant factor in the severity of accidents and should be included in further statistical analyses to determine the relationship between congestion and accident severity.

Table 5-12 Distribution of freeway accidents by severity and driver conditions

Aggregated by *Driver condition*		Severity					Total
		Level-1	Level-2	Level-3	Level-4	Level-5	
Apparently Normal	# of accidents	2471	989	866	400	25	4751
	Percentage	52.0	20.8	18.2	8.4	0.5	100
Had been drinking	# of accidents	205	49	38	34	4	330
	Percentage	62.1	14.8	11.5	10.3	1.2	100
Other abnormal	# of accidents	46	2	5	3	9	65
	Percentage	*70.8*	*3.1*	*7.7*	*4.6*	*13.8*	*100*

The following factors were also examined to determine their impacts on the distribution of freeway accidents by severity.

- Visibility condition (daylight, dawn/dusk, dark-lights on, dark- no lights)
- The number of through lanes
- The number of vehicles involved
- Collision type (head on, rear end, sideswipe, etc.)

The existing dataset provides no indication of the impacts these factors have on the accident severity distribution. Therefore, they are not included in the advanced statistical estimation provided in Section 5.6 and 5.7.

5.5 Relationships between AADT and accident severity

This section analyzes the aggregated relationship between AADT and accident severity, to evaluate whether the percentage of severe accidents reveals a decreasing trend with the level of congestion (represented with the volume per lane), if more congested traffic conditions will result in less severe accidents as expected. Figures 5-7 and 5-8 present accident distribution by severity under different AADT levels on local arterials and freeways.

Figure 5 -7 Percentage of accidents at each severity level vs. AADT per lane from the local arterial dataset

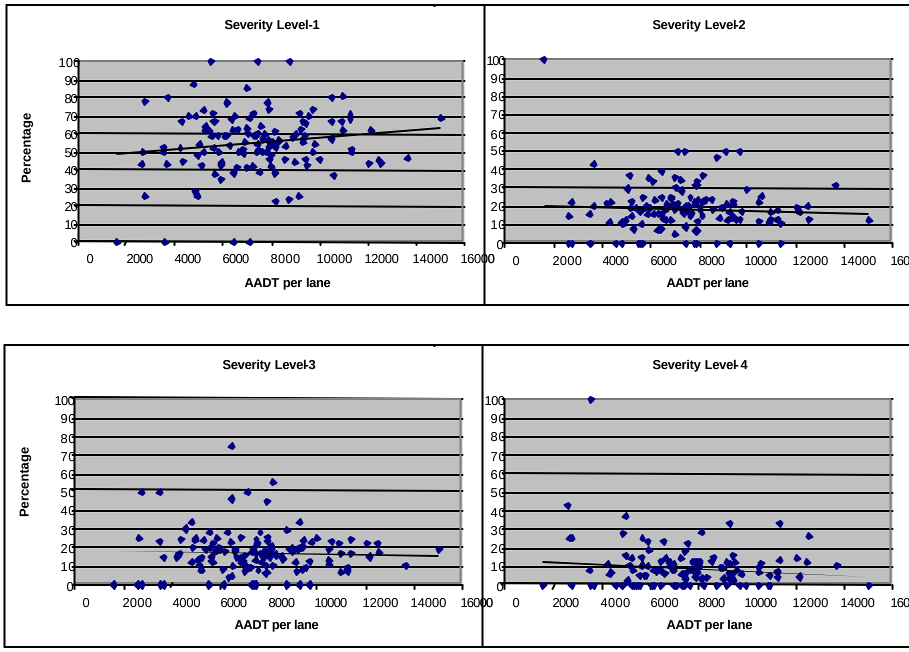
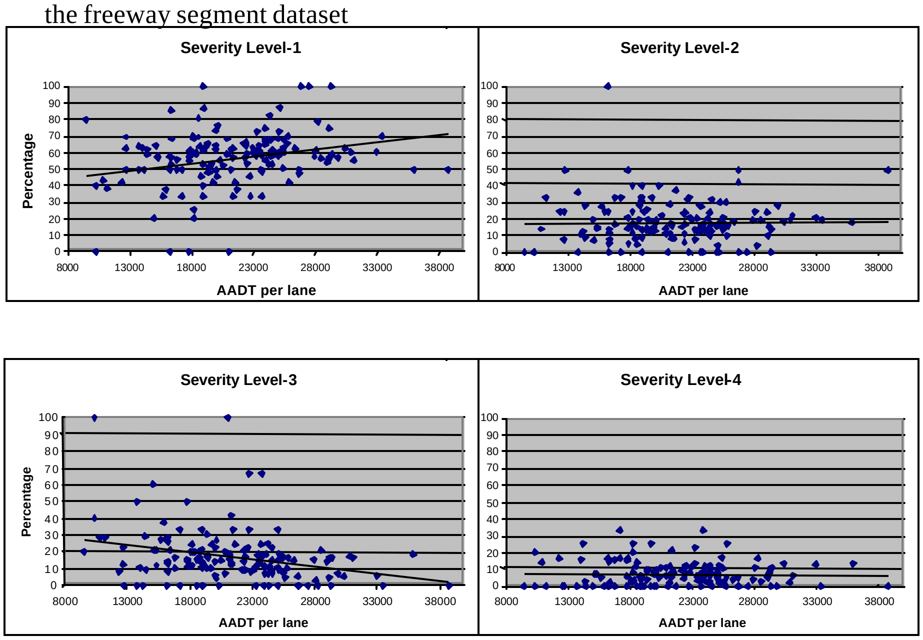


Figure 5 -8 Percentage of accidents at each severity level vs. AADT per lane from the freeway segment dataset



The statistical trends illustrate that the percentage of accidents at the lowest severity level tend to increase with the AADT per lane on both freeways and arterials. In contrast, the percentage of accidents at severity levels 2-4 exhibits a decreasing trend with the AADT per lane. These results, despite their preliminary nature, tend to offer supporting evidence regarding the general perception that accidents that occur during more congested traffic conditions tend to be at a less severe level.

Since Level-2 severity, “possible injuries”, is not rigorously defined, the following multivariate statistical estimation was used to explore the effectiveness of reclassifying the severity level with different data aggregations. A list of these candidate data sets after reclassification is presented below:

- Property damage only, possible injury, injury, disabled, and fatal. (5 levels)
- Property damage only, injury, and fatal. (3 levels)
    - 1, 2+3+4, 5;
    - 1+2, 3+4, 5.
- Injury, disabled, and fatal. (3 levels)
    - 2, 3+4, 5;
    - 2+3, 4, 5;
    - 2, 3, 4, 5.

Sections 5.6 and 5.7 present the investigation results regarding the relationship between accident severity and contributing factors on both freeways and arterials using the previous datasets.

5.6 Model Estimation for Arterials

This section presents the statistical method used to estimate the relationship between accident severity on arterials and primary contributing factors, especially the volume per lane that is used as the surrogate variable for congestion.

The dependent variable (i.e. severity level) is discrete and inherently ordered in nature, therefore, the traditional discrete choice models, such as multinomial logit or probit models will not be sufficient to account for the embedded ordinal relationship. Thus, the remaining estimation uses the Ordered Probit Model (Greene, 2000) to explore the relationships of interest. Both the arterial dataset (containing 4518 individual accident cases) and the freeway dataset (containing 4868 individual accident cases) are sufficiently large for assuming that the disturbance terms are jointly normally distributed.

Core concepts of the ordered probit model

The ordered probit model is grounded on the following latent regression:

$$y^{*} = b^{'}x + e$$

Where, y* is unobserved. What we do observe is:

y = 1 if y* <= 0

= 2 if 0 < y* <= µ1

= 3 if µ1 < y* <= µ2

= 4 if µ2 < y* <= µ3

= 5 if µ3 < y*

µ1, µ2, and µ3 are the unknown parameters to be estimated with ß

Prob(y=1) = cnorm(0 – ß'x) – 0

Prob(y=2) = cnorm(µ1- ß'x) - cnorm(0 – ß'x)

Prob(y=3) = cnorm(µ2- ß'x) - cnorm(µ1- ß'x)

Prob(y=4) = cnorm(µ3- ß'x) - cnorm(µ2- ß'x)

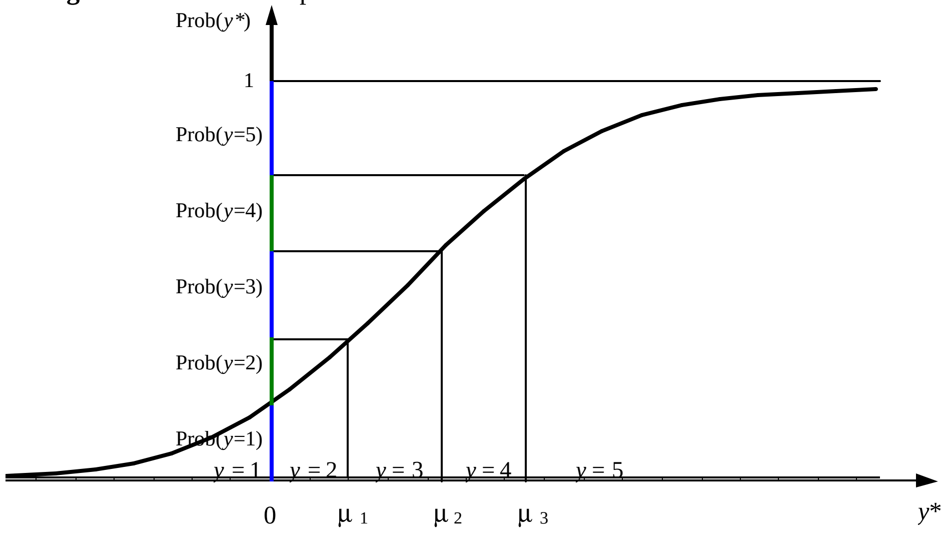
Prob(y=5) = 1- cnorm(µ3- ß'x)

For all the probabilities to be positive, we must have

0 < µ1<µ2< µ3

Figure 5-9 shows the implications of the structure.

Figure 5-9 Cumulative probabilities in the Ordered Probit Model



One can construct the log-likelihood function and compute its derivatives with standard methods.

Model estimation results

Table 5-13 summarizes the list of probit models estimated using the TSP4.5 software (Hall and Cummins, 1999).

Table 5-13 A list of estimated severity models for arterials

	Severity Levels Modeled (Y)	Dataset	Independent variables
Model-1	(1, 2, 3, 4, 5)	All available data	AADT per lane, peak-hour indicator, intersection indicator (X_2), weather (X_3), median type indicator (X_4)
Model-2	(1, 2, 3, 4, 5)	All available data	Hourly volume per lane, X_2, X_3, X_4
Model-3	(1, 2, 3, 4, 5)	Weekday data	Hourly volume per lane, X_2, X_3, X_4
Model-4	(1, 2+3+4, 5)	All available data	Hourly volume per lane, X_2, X_3, X_4
Model-5	(1, 2+3+4, 5)	Weekday data	Hourly volume per lane, X_2, X_3, X_4
Model-6	(1+2, 3+4, 5)	All available data	Hourly volume per lane, X_2, X_3, X_4
Model-7	(1+2, 3+4, 5)	Weekday data	Hourly volume per lane, X_2, X_3, X_4
Model-8	(2, 3, 4, 5)	All available data	Hourly volume per lane, X_2, X_3, X_4
Model-9	(2, 3, 4, 5)	Weekday data	Hourly volume per lane, X_2, X_3, X_4
Model-10	(2+3, 4, 5)	All available data	Hourly volume per lane, X_2, X_3, X_4
Model-11	(2+3, 4, 5)	Weekday data	Hourly volume per lane, X_2, X_3, X_4
Model-12	(2, 3+4, 5)	All available data	Hourly volume per lane, X_2, X_3, X_4
Model-13	(2, 3+4, 5)	Weekday data	Hourly volume per lane, X_2, X_3, X_4

Of the 13 types of data aggregations presented in Table 5-13, the following 4 models illustrate a better consistency on the relationship between severity level and the associated variables.

- Model-2: All five levels
- Model-6: 3 levels (1+2, 3+4, 5)
- Model-8: 4 levels (2, 3, 4, 5)
- Model-10: 3 levels (2+3, 4, 5)

Model-2: All five levels

Table 5-14 presents the estimation results for Model-2, where hourly volume is not a significant variable. For these results, "*Intersection*" and "*driver_other*" have significant positive coefficients, while "*weather_snow*" has a significant negative coefficient.

Table 5-14 Ordered Probit Model-2 for arterial accidents

Severity classification	Level-1, Level-2, Level-3, Level-4, Level-5		
Data points	4518 accidents From MD2, MD355, MD97, MD410, and US1		
Model estimation results			
Parameter	Estimate	t-statistic	P-value
C	-.170268	-4.40369	[.000]
HOURLY_VOLUME	-.840E-02	-1.00432	[.315]
INTERSECTION	**.137688**	**3.82537**	**[.000]**
WEATHER_SNOW	**-.524007**	**-2.98789**	**[.003]**
WEATHER_FOG	-.024564	-.112186	[.911]
DRIVER_DRINKING	.098716	1.24698	[.212]
***DRIVER_OTHER**	**.323975**	**2.94150**	**[.003]**
WORKZONE	.015619	.141684	[.887]
MEDIAN_DIVIDED	.015272	.422901	[.672]
μ_3	.508958	31.6692	[.000]
μ_4	1.20611	45.8941	[.000]
μ_5	2.40500	34.7019	[.000]

**Note: DRIVER_OTHER refers to the involved drivers who are under some abnormal conditions other than had-been-drinking.*

Model-6: 3 levels (1+2, 3+4, 5)

Table 5-15 presents the estimation results of Model-6, where hourly volume becomes a significant variable and other relationships remain unchanged. This model specification is intended to explore the impact of reclassifying severity into three distinct levels on the estimated relationships.

Table 5-15 Ordered Probit Model-6 for arterial accidents

Severity classification	Level-1+Level-2, Level-3+Level-4, Level-5		
Data points	4518 accidents From MD2, MD355, MD97, MD410, and US1		
Model estimation results			
Parameter	Estimate	t-statistic	P-value
C	-.641253	-13.0495	[.000]
HOURLY_VOLUME	**-.023771**	**-1.72515**	**[.085]**
INTERSECTION	**.167542**	**4.01417**	**[.000]**
WEATHER_SNOW	**-.417863**	**-2.01275**	**[.044]**
WEATHER_FOG	-.034805	-.134809	[.893]
DRIVER_DRINKING	.131483	1.44084	[.150]
DRIVER_OTHER	**.399042**	**3.25205**	**[.001]**
WORKZONE	-.088992	-.669264	[.503]
MEDIAN_DIVIDED	-.010797	-.256871	[.797]
μ_5	1.89868	27.5385	[.000]

Model-8: 4 levels (2, 3, 4, 5)

Table 5-16 presents the estimation results of Model-8.

Table 5-16 Ordered Probit Model-8 for arterial accidents

Severity classification	Level-2, Level-3, Level-4, Level-5		
Data points	2032 (excluding the property-damage-only accidents) From MD2, MD355, MD97, MD410, and US1		
Model estimation results			
Parameter	Estimate	t-statistic	P-value
C	.342533	4.88236	[.000]
HOURLY_VOLUME	**-.051202**	**-2.41656**	**[.016]**
INTERSECTION	**.155234**	**3.00065**	**[.003]**
WEATHER_SNOW	-.174407	-.580374	[.562]
WEATHER_FOG	-.256906	-.815444	[.415]
DRIVER_DRINKING	.124494	1.08203	[.279]
DRIVER_OTHER	**.557906**	**3.67382**	**[.000]**
WORKZONE	-.185056	-1.15257	[.249]
MEDIAN_DIVIDED	**-.114395**	**-2.17070**	**[.030]**
μ_4	1.04953	32.4510	[.000]
μ_5	2.47255	31.4301	[.000]

As reported in the exploratory analysis, the relation between the Level-1 severity and congestion is quite different from other severity levels as most accidents in the level-1 severity involve property-damage-only accidents. Many minor accidents tend to occur during snow or poor weather conditions. This is proven by the significant parameter for the variable of "*Weather_snow*". Model-2 and Model-6 consistently indicate that "*Intersection*" and "*Driver conditions*" are two significant factors affecting the results for accident severity. For both models, congestion (volume per lane) does not exhibit a strong significant impact due most likely to the difference in its relationship to accidents at the severity level of property damage only and personal injury. Therefore, Model-8 focuses on estimating the target relationship without the Level-1 severity accident data. After excluding the Level-1 severity data, the variable for snow days becomes insignificant and hourly volume emerges as one of the more significant factors. This fact is consistent with the belief that a large volume of Level-1 accidents (i.e. property damage only accidents) exists.

Model-10: 3 levels (2+3, 4, 5)

Table 5-17 presents the estimation results of Model-10.

Table 5-17 Ordered Probit Model-10 for arterial accidents

Severity classification	Level-2+Level-3, Level-4, Level-5		
Data points	2032 (excluding the property-damage-only accidents) From MD2, MD355, MD97, MD410, and US1		
Model estimation results			
Parameter	Estimate	t-statistic	P-value
C	-.589618	-6.29549	[.000]
HOURLY_VOLUME	**-.091365**	**-2.82283**	**[.005]**
INTERSECTION	**.135151**	**2.09398**	**[.036]**
WEATHER_SNOW	-.652049	-1.25695	[.209]
WEATHER_FOG	-.636695	-1.22499	[.221]
DRIVER_DRINKING	.066267	.464270	[.642]
DRIVER_OTHER	**.632713**	**3.71988**	**[.000]**
WORKZONE	-.112528	-.551702	[.581]
MEDIAN_DIVIDED	**-.146471**	**-2.23421**	**[.025]**
μ_5	1.42960	18.7947	[.000]

Table 517 presents the estimation results using the same model specification as Model-8 but integrating Level-2 (possibly injured) with Level-3 (injured). The estimation results indicate that the integration of Level-2 and Level-3 accidents does not affect the relationship between accident severity and the identified significant variables.

Based on the estimation results of Tables 5-14 through 5-17, it can be concluded that the relationship between accident severity and key associated variables are as follows:

- **Congestion level** (volume per lane): Accidents occurring on more congested arterials are more likely to be less severe. This is proven by the negative and significant parameters for volume per lane.
- **Intersection or not**: Accidents occurring at intersections are more likely to be more severe. This may be attributed to the fact that there are more head-on collisions occurring at intersections than at roadway links. In addition, head-on collisions usually result in higher personal injury severity than other types of collisions, such as rear-end collisions.
- **Driver condition**: The estimation results illustrate that if drivers involved in accidents are affected by abnormal conditions such as using drugs or having physical defects, the resulting severity is likely to be higher than for drivers under normal driving conditions.
- **Divided median type**: The existence of median seems to contribute significantly to the reduction in the resulting accident severity, as evidenced in its significant and negative coefficient.
- **Weather conditions**: When Level-1 severity of accidents is included in the sample dataset, the estimation results indicate that the snowing weather condition is a significant variable. This is proven by a larger number of Level-1 accidents (i.e., property-damage-only accidents). The estimation results of Models 2 and 6 further suggest that the accidents that occurred in snow conditions tend to be at a lower severity level.

5.7 Model Estimation for Freeway Segments

Using the same procedures and estimation algorithms, this section investigates the relationship between congestion and accident severity levels on freeways. Variables to be included in model specifications are listed below:

- x1 : Hourly volume per lane
- x2 : Auxiliary lane ratio

$$\text{Auxiliary lane ratio} = \frac{\text{the total length of auxiliary lanes on a link}}{\text{the length of the link}}$$

- x3 : Weather_snow
- x4 : Driver conditions
- x5 : In work-zone or not

Model estimation results

Table 5-18 illustrates the list of model specifications explored in this section and the associated datasets used for estimation. Of the 12 experimental specifications listed in Table 5-18, Models 1, 5, and 9 present a consistent relationship between accident severity and key associated variables (see Tables 5-19, 5-20, and 5-21).

Table 5-18 A complete list of estimated severity models for freeways

	Severity Levels Modeled (Y)	Dataset	Independent variables
Model-1	(1, 2, 3, 4, 5)	All available data	Hourly volume per lane, auxiliary lane ratio (X_2), weather_snow (X_3), driver condition (X_4), work-zone indicator (X_5)
Model-2	(1, 2, 3, 4, 5)	Weekday data	Hourly volume per lane, X_2, X_3, X_4, X_5
Model-3	(1, 2+3+4, 5)	All available data	Hourly volume per lane, X_2, X_3, X_4, X_5
Model-4	(1, 2+3+4, 5)	Weekday data	Hourly volume per lane, X_2, X_3, X_4, X_5
Model-5	(1+2, 3+4, 5)	All available data	Hourly volume per lane, X_2, X_3, X_4, X_5
Model-6	(1+2, 3+4, 5)	Weekday data	Hourly volume per lane, X_2, X_3, X_4, X_5
Model-7	(2, 3, 4, 5)	All available data	Hourly volume per lane, X_2, X_3, X_4, X_5
Model-8	(2, 3, 4, 5)	Weekday data	Hourly volume per lane, X_2, X_3, X_4, X_5
Model-9	(2+3, 4, 5)	All available data	Hourly volume per lane, X_2, X_3, X_4, X_5
Model-10	(2+3, 4, 5)	Weekday data	Hourly volume per lane, X_2, X_3, X_4, X_5
Model-11	(2, 3+4, 5)	All available data	Hourly volume per lane, X_2, X_3, X_4, X_5
Model-12	(2, 3+4, 5)	Weekday data	Hourly volume per lane, X_2, X_3, X_4, X_5

Table 5-19 Ordered Probit Model-1 for freeway accidents

Severity classification	Level-1, Level-2, Level-3, Level-4, Level-5			
Data points	4868 accidents From I-495, I-695, I-95, I-270, and US50			
Model estimation results				
Parameter		Estimate	t-statistic	P-value
C		-.090075	-1.82659	[.068]
AUX_RATIO		**-.089372**	**-2.43609**	**[.015]**
HOURLY_VOLUME		**-.017528**	**-2.58344**	**[.010]**
WEATHER_SNOW		**-.341361**	**-2.93543**	**[.003]**
DRIVER_DRINKING		**.133097**	**2.01522**	**[.044]**
***DRIVER_OTHER**		**.183071**	**2.44868**	**[.014]**
WORKZONE		.119781	1.15409	[.248]
μ_3		.483448	31.2991	[.000]
μ_4		1.09561	44.3738	[.000]
μ_5		2.24982	35.7829	[.000]

**Note: DRIVER_OTHER refers to the involved drives who are in some abnormal conditions other than had-been-drinking.*

Table 5-20 Ordered Probit Model-5 for freeway accidents

Severity classification	Level-1+Level-2, Level-3+Level-4, Level-5
Data points	4868 accidents From I-495, I-695, I-95, I-270, and US50

Model estimation results			
Parameter	Estimate	t-statistic	P-value
C	-.432557	-7.71280	[.000]
AUX_RATIO	**-.171304**	**-3.88531**	**[.000]**
HOURLY_VOLUME	**-.034780**	**-4.41406**	**[.000]**
WEATHER_SNOW	**-.294648**	**-2.15177**	**[.031]**
DRIVER_DRINKING	.091651	1.20733	[.227]
DRIVER_OTHER	**.225370**	**2.67454**	**[.007]**
WORKZONE	.081391	.680242	[.496]
μ_5	1.77113	28.3265	[.000]

Table 5-21 Ordered Probit Model-9 for freeway accidents

Severity classification	Level-2+Level-3, Level-4, Level-5
Data points	1995 (excluding the property-damage-only accidents) From I-495, I-695, I-95, I-270, and US50

Model estimation results

Parameter	Estimate	t-statistic	P-value
C	-.371140	-4.21757	[.000]
AUX_RATIO	**-.239855**	**-3.43718**	**[.001]**
HOURLY_VOLUME	**-.050965**	**-4.07057**	**[.000]**
WEATHER_SNOW	.047629	.205046	[.838]
DRIVER_DRINKING	**.215036**	**1.94036**	**[.052]**
DRIVER_OTHER	**.316210**	**2.57182**	**[.010]**
WORKZONE	.138299	.798666	[.424]
μ_5	1.41237	20.2250	[.000]

It is important to note that when severity level-1 data are excluded from the estimation the variable of snow condition becomes insignificant similar to the estimation results for arterials. This is proven in the Model-9 results.

Based on preliminary statistical results, Model-1 was selected for further estimation and the estimation results are reported in Table 5-22.

Table 5-22 Final Ordered Probit Model for freeway accidents

Severity classification	Level-1, Level-2, Level-3, Level-4, Level-5			
Data points	4868 accidents From I-495, I-695, I-95, I-270, and US50			
Model estimation results				
Parameter		Estimate	t-statistic	P-value
C		-.084126	-1.71567	[.086]
AUX_RATIO		-.091038	-2.48349	[.013]
HOURLY_VOLUME		-.017851	-2.63363	[.008]
WEATHER_SNOW		-.344029	-2.95859	[.003]
DRIVER_DRINKING		.133093	2.01529	[.044]
DRIVER_OTHER		.182727	2.44436	[.015]
Stability test results				
Number of coefficients: K = 6 Number of observations in subset-1: $n_1 = 2383$ Number of observations in subset-2: $n_2 = 2485$				
Residual sum of squares: $\sum e_p^2 = 5278; \sum e_1^2 = 2533; \sum e_2^2 = 2746$ The resulting F statistics is 1.04 < F 0.95(6, 4856) = 2.10 Therefore, the final Ordered Probit model is stable.				

Table 5-22 presents the estimation results using the same model specification as Model-1 but only including the significant exploratory variables. To ensure that all estimated parameter signs are independent to the differences in the sample size, a standard parameter stability test was also preformed. The test results are illustrated in Table 5-22 and clearly indicate that the estimated relationship between accident severity and key factors is stable and will not vary with the selected sample size.

It can be concluded from Tables 5-19 through 5-22 that the relationship between accident severity on freeways and key associated variables is as follows:

- **Congestion level** (volume per lane): Accidents that occurred on more congested freeways are more likely to be less severe. This is proven by the negative and significant parameters for volume per lane.
- **The auxiliary lane ratio**: Accidents that occurred on roadway links with higher auxiliary lane ratios are more likely to be less severe. This is proven by the negative and significant parameters for the auxiliary lane ratio.
- **Snowing weather conditions**: Accidents that occur under snow conditions are more likely to be less severe. This may be caused by lower speeds and longer headways maintained by the drivers. The effect of rainy weather conditions is not statistically significant.
- **Driver conditions**: The estimation results indicate that if drivers involved in accidents are under the influence of alcohol or subject to other abnormal conditions, the resulting severity will be higher than for drivers under normal driving conditions. This may be attributed to a decrease in human response and/or less attention to the presence of other vehicles or obstacles.

5.8 Summary and Conclusions

This chapter has investigated the relationship between accident severity and congestion levels on both sample freeways and arterials. It includes an exploratory analyses and multivariate statistical estimation using Ordered Probit regression.

The research results, consistent with general beliefs, are summarized below:

- Accidents occurring on more congested freeways and arterials are more likely to happen at a lower severity levels.
- Accidents occurring at intersections are more likely to happen at higher severity levels.
- Accidents on both freeways and arterials are more likely to occur at lower severity levels during snow conditions.

- If drivers involved in accidents are under the influence of alcohol or subject to other abnormal conditions, the resulting severity will be higher than those under normal driving conditions.
- Accidents occurring on a freeway link with higher auxiliary lane ratio are more likely to be at a lower severity level.
- The presence of medians tends to contribute significantly to the reduction in the level of accident severity on arterials.

CHAPTER 6
CLOSING AND FUTURE RESEARCH

6.1 Closing

This research investigated the relationship between congestion and accidents with a specific emphasis on the impact various volume levels have on the resulting accident frequency, rate, and severity. The work presented here consists of two primary phases; Phase-1 explored the discrepancies of accident characteristics under various conditions (e.g. peak and off-peak periods, work-zones and normal highway segments, weather conditions, and presence of medians); and based on the preliminary results from Phase-1, Phase-2 focused on estimating the impacts of congestion and other primary factors on the distribution of traffic accidents on both freeways and arterials.

As a result of the stochastic nature of the accidents, this study used Poisson and Negative Binomial regressions to estimate various continuous multivariate models to determine the relationship between congestion and accident frequency, and congestion and accident rate. In view of the inherently discrete and ordered relations among different severity levels, this study also explored the use of an Ordered Probit model to determine the compound impacts of traffic volume and associated factors on accident severity. To ensure the statistical stability of the estimated relationships, a rigorous stability test for the parameters of all significant variables was performed before conclusions were formulated.

Based on the available sample freeway and arterial accident data from Year 2000, this study has yielded the following research findings:

Accident frequency vs. congestion and other associated key factors

Both the exploratory analyses and NB2 models established for arterials and freeways confirmed the following relationships:

- Accident frequency on both freeways and arterials tends to increase with the congestion levels.

- Divided arterial links exhibit higher accident frequencies than undivided arterial links with the same volume levels.
- Accident frequency on arterials generally increases with intersection density.
- Wider medians can significantly reduce accident frequency on freeways.
- Accident frequency on both freeways and arterial links reveals an increasing trend with the total number of through lanes.

Accident rate vs. congestion and other associated key factors

With the Poisson accident rate model estimated for arterials and the NB1 peak-hour accident rate model for freeways, the following conclusions on the relationship between congestion and accident rate were identified.

- The accident rate for arterials tends to decrease as volume increases.
- The accident rate on freeways during off-peak hours appears to be random, exhibiting no systematic relationship with traffic volume.
- During peak-congestion periods, accident rates tend to increase significantly with the volumes per lane.
- Divided arterial links tend to exhibit higher accident rates than undivided arterial links with the same volume levels.
- Wider medians can significantly reduce accident rates on freeway links.
- Accident rate on arterials generally increases with intersection density.
- An increase in the total number of through lanes may contribute to a higher level of accident rate on arterials but not on freeways.

Accident severity vs. congestion and other associated key factors

The Ordered Probit accident severity models were successfully established for the relationship between accident severity and congestion on both arterials and freeways. These research findings are summarized below.

- Accidents occurring on more congested freeways and arterials are more likely to be at lower severity levels.
- Accidents occurring at intersections are more likely to happen at higher severity levels than those occurring at roadway segments.

- Accidents occurring during snow conditions on freeways and arterials are more likely to be at lower severity levels than those occurring during normal conditions.
- If drivers involved in accidents are under the influence of alcohol or subjected to any abnormal conditions, the severity of accidents is likely to be higher than those occurring under normal driving conditions.
- Accidents occurring on freeway links with higher auxiliary lane ratios are more likely to be at lower severity levels.
- The presence of medians tends to contribute significantly to the reduction in the resulting accident severity on arterials.

6.2 Future Research Needs

Although this study provide an in-depth analysis of the relationship between congestion and accidents, further investigation on the impacts of congestion on traffic safety is necessary. Recommendations for future research areas include:

- The relationship between accident rate and intensity of lane-changing movements that is likely to be correlated to congestion levels.
- The relationship between accidents and other indicators of the congestion level such as *v/c* ratio and speed reduction.
- The impacts of highway geometric features (e.g. horizontal curvatures, and vertical gradients) on accident severity at various congestion levels.
- The effects of congestion on behavior of accident-prone drivers (e.g. changing lanes when there is no sufficient length of gaps, failure to maintain a safety distance to the leading vehicle).
- The impact of congestion on the secondary incident rate during the response and management of primary accidents.

REFERENCES

1. Abbas, K.A., (2003) "Traffic safety assessment and development of predictive models for accidents on rural roads in Egypt," *Accident Analysis and Prevention*, 935, 1-15.

2. Abdel-Aty, M.A., and Radwan, A.E., (2000) "Modeling traffic accident occurrence and involvement," *Accident Analysis and Prevention*, Vol. 32, 633-642.

3. Al-Ghamdi, A. S., (2002) "Using logistic regression to estimate the influence of accident factors on accident severity," *Accident Analysis and Prevention*, Vol. 34, 729-741.

4. Amoros, E., Martin, J.L., and Laumon, B., (2002) "Comparison of road crashes incidence and severity between some French counties," *Accident Analysis and Prevention*, 865, 1-11.

5. Carson, J., and Mannering, F., (2001) "The effect of ice warning signs on ice-accident frequencies and severities," *Accident Analysis and Prevention*, Vol. 33, 99-109.

6. Chang, G. L., and Point-du-Jour, J. Y., (2002) *Performance evaluation of CHART - the real time incident management system in Year 2000*, final report.

7. Chang, L.-Y., Mannering, F, (1999) "Analysis of injury severity and vehicle occupancy in truck- and non-truck-involved accidents," *Accident Analysis and Prevention*, Vol. 31, 579-592.

8. Greene, W., (2000), *Econometrics Analysis*, 4th Edition, Prentice Hall International.

9. Greibe, P., (2002) "Accident prediction models for urban roads," *Accident Analysis and Prevention*, 839, 1-13.

10. Hall, B., and Cummins, C., (1999), *User's Manual and Reference Manual for Time Series Processor Version 4.5*, TSP international.

11. Kam, B., (2002) "A disaggregate approach to crash rate analysis." *Accident Analysis and Prevention*, 882, 1-17.

12. Karlaftis, M.G., and Golias, I. (2002) "Effects of road geometry and traffic volumes on rural roadway accident rates," *Accident Analysis and Prevention*, Vol. 34, 357-365.

13. Knuiman, M., Council, F., and Reinfurt, D., (1993) "Association of median width and highway accident rates," *Transportation Research Record* 1401, 70-82.

Formatted: Bullets and Numbering

14. Kochelman, K. M., and Kweon, Y.-J., (2002) "Driver injury severity: an application of ordered probit models," *Accident Analysis and Prevention*, Vol. 34, 313-321.

15. Lee, J., and Mannering, F., (2002) "Impact of roadside features on the frequency and severity of run-off-roadway accidents: an empirical analysis," *Accident Analysis and Prevention*, Vol. 34, 149-161.

16. Martin, J.-L., (2002) "Relationship between crash rate and hourly traffic flow on interurban motorways," *Accident Analysis and Prevention*, Vol. 34, 619-629.

17. Mayora, J., and Rubio, R. (2003) "Relevant variables for crash rate prediction in Spain's two lane rural roads", *TRB*, 82nd Annual Meeting.

18. Miaou, S. P. (1994) "The relationship between truck accidents and geometric design of road sections: Poisson versus negative binomial regression," *Accident Analysis and Prevention,* Vol. 26, 471-482.

19. O'Donnell, C.J., and Connor, D.H., (1996) "Predicting the severity of motor vehicle accident injuries using models of ordered multiple choice," *Accident Analysis and Prevention*, Vol. 28, 739-753.

20. Persaud, B., and Dzbik, L., (1993) "Accident prediction models for freeways," *Transportation Research Record* 1401, 55-60.

21. Qin, X., Ivan, J., and Ravishanker, N., (2003) " Selecting exposure measures in crash rate prediction for two-lane highway segments," *Accident Analysis and Prevention,* 938, 1-9.

22. Shankar, V., Mannering, F., and Barfield, W., (1995) "Effect of roadway geometrics and environmental factors on rural freeway accident frequencies," *Accident Analysis and Prevention*, Vol.27, 371-389.

23. Shankar, V., and Mannering, F. (1996) "An exploratory multinomial logit analysis of single-vehicle motorcycle accident severity," *Journal of Safety Research,* 27 (3), 183-194.

24. Shankar, V., Milton, J. and Mannering, F., (1997) "Modeling Accident Frequencies as Zero-Altered Probability Processes: An Empirical inquiry," *Accident Analysis and Prevention*, Vol. 29, 829-837.

25. Yau, K. (2003) "Risk factors affecting the severity of single vehicle traffic accidents in Hong Kong," *Accident Analysis and Prevention,* article in press.

26. Zhou, M., and Sisiopiku, V. P., (1997) "Relationship Between Volume-to-Capacity Ratios and Accident Rate," *Transportation Research Record* 1581, 47-52.

BIBLIOGRAPHY

1. Abbas, K.A., (2003) "Traffic safety assessment and development of predictive models for accidents on rural roads in Egypt," *Accident Analysis and Prevention*, 935, 1-15.

Formatted: Bullets and Numbering

2. Abdel-Aty, M.A., and Radwan, A.E., (2000) "Modeling traffic accident occurrence and involvement," *Accident Analysis and Prevention*, Vol. 32, 633-642.

3. Abdel-Aty, M.A., and Abdelwahab, H., (2003) "Modeling rear-end collisions including the role of driver's visibility and light truck vehicles using a nested logit structure," *Accident Analysis and Prevention*, article in press.

4. Al-Ghamdi, A. S., (2002) "Using logistic regression to estimate the influence of accident factors on accident severity," *Accident Analysis and Prevention*, Vol. 34, 729-741.

5. Al-Ghamdi, A. S., (1993) "Comparison of accident rates using the likelihood ratio testing technique," *Transportation Research Record* 1401, 50-54.

6. Amoros, E., Martin, J.L., and Laumon, B., (2002) "Comparison of road crashes incidence and severity between some French counties," *Accident Analysis and Prevention*, 865, 1-11.

7. Bedard, M., Guyatt, G., Stomes, M, and Hirdes, J. (2002) "The independent contribution of driver, crash, and vehicle characteristics to driver fatalities," *Accident Analysis and Prevention*, Vol. 34, 717-727.

8. Bonneson, J.A., and McCoy, P.T., (1997) "Effect of Median Treatment on Urban Arterial Safety: An accident Prediction Model," *Transportation Research Record* 1581, 27-36.

9. Carson, Jodi, and Mannering, Fred, (2001) "The effect of ice warning signs on ice-accident frequencies and severities," *Accident Analysis and Prevention*, Vol. 33, 99-109.

10. Chang, G. L., and Point-du-Jour, J. Y., (2002) *Performance evaluation of CHART - the real time incident management system in Year 2000*, final report.

11. Chang, L.-Y., Mannering, F, (1999) "Analysis of injury severity and vehicle occupancy in truck- and non-truck-involved accidents," *Accident Analysis and Prevention*, Vol. 31, 579-592.

12. Cherpitel, C., Tam, T., Midanik, L., Caetano, R., and Greenfield, T. (1995) "Alcohol and non-fatal injury in the U.S. general poulation: a risk function analysis," *Accident Analysis and Prevention*, Vol. 27, 651-661.

13. Clark, D. (2003) "Effect of population density on mortality after motor vehicle collisions," *Accident Analysis and Prevention*, 915, 1-7.

14. Davis, G. A., (2002) "Is the claim that 'variance kills' an ecological fallacy?" *Accident Analysis and Prevention*, Vol. 34, 343-369.

15. Duncan, C. S., Khattak, A. and Council, F., (1998) "Applying the ordered probit model to injury severity in truck-passenger car rear-end collisions," *Transportation Research Record* 1635, 63-71.

16. Fitzpatrick, K., and Balke, K., (1995) "Evaluation of flush medians and two-way, left-turn lanes on four-lane rural highways," *Transportation Research Record* 1500, 146-152.

17. Greene, W., (2000). *Econometrics Analysis*, 4th Edition, Prentice Hall International.

18. Greibe, P., (2002) "Accident prediction models for urban roads," *Accident Analysis and Prevention*, 839, 1-13.

19. Hall, B., and Cummins, C., (1999). *User's Manual and Reference Manual for Time Series Processor Version 4.5*, TSP international.

20. Jones, B., Janseen, L., and Mannering, F., (1991) "Analysis of the frequency and duration of the freeway accidents in Seattle," *Accident Analysis and Prevention*, Vol. 23, 239-255.

21. Kam, B., (2002) "A disaggregate approach to crash rate analysis." *Accident Analysis and Prevention*, 882, 1-17.

22. Karlaftis, M.G., and Golias, I., (2001) "Effects of road geometry and traffic volumes on rural roadway accident rates," *Accident Analysis and Prevention*, Vol. 34, 357-365.

23. Karlaftis, M.G., and Tarko, A.P. (1997) "Heterogeneity considerations in accident modeling," *Accident Analysis and Prevention*, Vol. 30, 425-433.

24. Knuiman, M., Council, F., and Reinfurt, D., (1993) "Association of median width and highway accident rates," *Transportation Research Record* 1401, 70-82.

25. Kochelman, K. M., and Kweon, Y.-J., (2002) "Driver injury severity: an application of ordered probit models," *Accident Analysis and Prevention*, Vol. 34, 313-321.

26. Larsen, L., and Kines, P., (2002) "Multidisciplinary in-depth investigations of head-on and left-turn road collisions," *Accident Analysis and Prevention*, Vol. 34, 367-380.

27. Lee, J., and Mannering, F., (2002) "Impact of roadside features on the frequency and severity of run-off-roadway accidents: an empirical analysis," *Accident Analysis and Prevention*, Vol. 34, 149-161.

28. Martin, J.-L., (2002) "Relationship between crash rate and hourly traffic flow on interurban motorways," *Accident Analysis and Prevention*, Vol. 34, 619-629.

29. Mayora, J., and Rubio, R. (2003) "Relevant variables for crash rate prediction in Spain's two lane rural roads", *TRB*, 82nd Annual Meeting.

30. Mensah, A., and Hauer, E., (1998) "Two problems of averaging arising in the estimation of the relationship between accidents and traffic flow," *Transportation Research Record* 1635, 37-43.

31. Miaou, S. P. (1994) "The relationship between truck accidents and geometric design of road sections: Poisson versus negative binomial regression," *Accident Analysis and Prevention*, Vol. 26, 471-482.

32. Navon, D. (2002) "The paradox of driving speed: two adverse effects on highway accident rate," *Accident Analysis and Prevention*, 845, 1-7.

33. O'Donnell, C.J., and Connor, D.H., (1996) "Predicting the severity of motor vehicle accident injuries using models of ordered multiple choice," *Accident Analysis and Prevention*, Vol. 28, 739-753.

34. Persaud, B., and Dzbik, L., (1993) "Accident prediction models for freeways," *Transportation Research Record* 1401, 55-60.

35. Preusser, D., and Williams, A., and Ulmer, R. (1995) "Analysis of fatal motorcycle crashes: crash typing," *Accident Analysis and Prevention*, Vol. 27, 845-851.

36. Qin, X., Ivan, J., and Ravishanker, N., (2003) " Selecting exposure measures in crash rate prediction for two-lane highway segments," *Accident Analysis and Prevention*, 938, 1-9.

37. Rock, S. M., (1995) "Impact of the 65 mph speed limit on accidents, deaths, and injuries in Illinois," *Accident Analysis and Prevention*, Vol. 27, 207-214.

38. Shankar, V., Mannering, F., and Barfield, W., (1995) "Effect of roadway geometrics and environmental factors on rural freeway accident frequencies," *Accident Analysis and Prevention*, Vol. 27, 371-389.

39. Shankar, V., and Mannering, F. (1996) "An exploratory multinomial logit analysis of single-vehicle motorcycle accident severity," *Journal of Safety Research*, 27 (3), 183-194.

40. Shankar, V., Milton, J. and Mannering, F., (1997) "Modeling Accident Frequencies as Zero-Altered Probability Processes: An Empirical inquiry," *Accident Analysis and Prevention*, Vol. 29, 829-837.

41. Sullivan, J., and Flannagan, M., (2002), "The role of ambient light level in fatal crashes: inferences from daylight saving time transitions," *Accident Analysis and Prevention*, Vol. 34, 487-498.

42. Wood, G.R., (2002) "Generalized linear accident models and goodness of fit testing," *Accident Analysis and Prevention* , Vol.34, 417-427.

43. Yau, K. (2003) "Risk factors affecting the severity of single vehicle traffic accidents in Hong Kong," *Accident Analysis and Prevention*, article in press.

44. Zhou, M., and Sisiopiku, V. P., (1997) "Relationship Between Volume-to-Capacity Ratios and Accident Rate," *Transportation Research Record* 1581, 47-52.

Appendix-1: The Poisson and negative binomial regression models

As proven in the literature review, accident occurrence is a Poisson Process in nature; therefore, it is appropriate to use the Poisson regression model to explore the relationship between accident frequency and identified exploratory variables.

- Poisson Distribution

  $$p(Y = y) = \frac{e^{-l} l^k}{k!}, k = 0,1,2......; \ l > 0$$

- ? is the mean of y. The most common formulation for ? is the log-linear model

  $$\log l = b^{'} X$$

- The log -likelihood function is:

  $$\ln L = \sum_{i=1}^{n} [-l_i + y_i b^{'} X_i - \ln y_i !]$$

- Use Maximum Likelihood Method to estimate the coefficient

The assumption of the Poisson regression model is that the mean of the dependent variable is approximately equal to the variance of the dependent variable. Therefore, when this assumption is violated, the Poisson regression model will not provide a valid estimation of the relationship between accident frequency and congestion levels. The Lagrange Multiplier Test for over-dispersion is performed on every Poisson model. Under the hypothesis of the Poisson model, the limiting distribution of LM statistics is chi-squared with one degree of freedom. If the over-dispersion is significant in the model, Type I Negative Binomial and Type II Negative Binomial models are used.

- Type I Negative Binomial model assumes the following relationship between mean and variance: E [y] = exp (X * b) = μ

  Variance [y] = μ * (1 + a)

- Type II Negative Binomial model assumes the following relationship between mean and variance: E [y]= exp (X * b) = μ

  Variance [y] = $\mu + a * \mu^2$

Although a stronger assumption on the equality of the mean and variance of the dependent variable is needed for the Poisson model, it is shown to be more robust in terms of the model specification. Therefore, this study always starts with the Poisson model and whenever the over-dispersion presents negative binomial models will be employed. Furthermore, if the over-dispersion is not significant, the NB models will be estimated when the mean-variance ratio of the dependent variable is significantly different than 1.

Appendix-2: The Parameter Stability Test

The parameter stability test is carried out by the Chow test. First it estimates the regression model with the complete dataset and calculates the residual sum of squares ($\sum e_p^2$). Next, the sample dataset is randomly partitioned into two comparable sub-datasets. Third, the regression models are estimated with the resulting two sub-datasets respectively and the residual sum of squares ($\sum e_1^2$, $\sum e_2^2$) is calculated. Finally, calculate the F-statistic:

$$F = \frac{\left[\sum e_p^2 - (\sum e_1^2 + \sum e_2^2)\right] / K}{(\sum e_1^2 + \sum e_2^2)/(n_1 + n_2 - 2K)}$$

Where, K is the number of coefficients in the regression model, n1 and n2 are the number of observations in two sub-datasets.

Stability test results of the accident frequency model for arterials:

- Partition the sample dataset and test the Poisson model stability.

$$F = \frac{\left[\sum e_p^2 - (\sum e_1^2 + \sum e_2^2)\right] / K}{(\sum e_1^2 + \sum e_2^2)/(n_1 + n_2 - 2K)}$$

Where: K =4, n1 =670, n2 =696

Residual sum of squares (scaled by 10^4):

$$\sum e_p^2 = 184695539 \ ; \sum e_1^2 = 64050659 \ ; \sum e_2^2 = 119047128$$

The resulting F statistics is 2.96< F 0.99(4, 1358) =3.34

- Conclusion: the estimated Poisson model is stable.

Stability test results of the accident frequency model for freeways:

- Partition the sample dataset and test the final model stability.

$$F = \frac{\left[\sum e_p^2 - (\sum e_1^2 + \sum e_2^2)\right] / K}{(\sum e_1^2 + \sum e_2^2)/(n_1 + n_2 - 2K)}$$

Where: K = 5, n1 = 181, n2 = 177

Residual sum of squares (scaled by 10^8):

$$\sum e_p^2 = 98110\,;\sum e_1^2 = 43981\,;\sum e_2^2 = 52745$$

The resulting F statistics is 1.25 < F 0.95(5, 348) = 2.21

- Conclusion: the NB2 model is stable.

Stability test results of the accident rate model for arterials:

- Partition the sample dataset and test the Poisson model stability.

$$F = \frac{\left[\sum e_p^2 - (\sum e_1^2 + \sum e_2^2)\right] / K}{(\sum e_1^2 + \sum e_2^2)\,/(n_1 + n_2 - 2K)}$$

Where: K =4, n1 =670, n2 =696

Residual sum of squares (scaled by 10^5):

$$\sum e_p^2 = 82181205; \sum e_1^2 = 38229845; \sum e_2^2 = 43787362$$

The resulting F statistics is 0.68< F 0.95(4,1358) = 2.37

- Conclusion: the Poisson model is stable.

Stability test results of the accident rate model for freeways:

- Partition the sample dataset and test the model stability.

$$F = \frac{\left[\sum e_p^2 - (\sum e_1^2 + \sum e_2^2)\right] / K}{(\sum e_1^2 + \sum e_2^2)\,/(n_1 + n_2 - 2K)}$$

Where: K = 3, n1 = 89, n2 = 90

Residual sum of squares:

$$\sum e_p^2 = 73208\,;\sum e_1^2 = 43983\,;\sum e_2^2 = 28143$$

The resulting F statistics is 0.64 < F 0.95(3, 173) = 2.60

- Conclusion: the final model is stable.

Stability test results of the accident severity model for freeways:

- Partition the sample dataset and test the model stability.

$$F = \frac{\left[\sum e_p^2 - (\sum e_1^2 + \sum e_2^2)\right] / K}{(\sum e_1^2 + \sum e_2^2) / (n_1 + n_2 - 2K)}$$

Where: K=6, n1=2383, n2=2485

Residual sum of squares:

$$\sum e_p^2 = 5278 ; \sum e_1^2 = 2533 ; \sum e_2^2 = 2746$$

The resulting F statistics is 1.04 < F 0.95(6, 4856) = 2.10

- Conclusion: the final model is stable.