Final Technical Report
TNW2008-04

Research Project Agreement No. 61-2394

Cost Effective Safety Improvements for Two-Lane Rural Roads

Yinhai Wang Associate Professor	Ngan Ha Nguyen Graduate Research Assistant
Atli Björn E. Levy Graduate Research Assistant	Yao-Jan Wu Graduate Research Assistant

Department of Civil and Environmental Engineering
University of Washington
Seattle, Washington 98195-2700

A report prepared for

Transportation Northwest (TransNow)
University of Washington
135 More Hall, Box 352700
Seattle, Washington 98195-2700

in cooperation with

U.S. Department of Transportation
Federal Highway Administration

March 2008

TECHNICAL REPORT STANDARD TITLE PAGE

1. REPORT NO.	2. GOVERNMENT ACCESSION NO.	3. RECIPIENT'S CATALOG NO.	
TNW2008-04			
4. TITLE AND SUBTITLE Cost Effective Safety Improvements for Two-Lane Rural Roads		5.REPORT DATE February 2008	
		6. PERFORMING ORGANIZATION CODE	
7. AUTHOR(S) Yinhai Wang, Ngan Ha Nguyen, Atli Björn E. Levy and Yao-Jan Wu		8. PERFORMING ORGANIZATION REPORT NO. TNW2008-04	
9. PERFORMING ORGANIZATION NAME AND ADDRESS Transportation Northwest Regional Center X (TransNow) Box 352700, 129 More Hall University of Washington Seattle, WA 98195-2700		10. WORK UNIT NO.	
		11. CONTRACT OR GRANT NO. DTRS99-G-0010	
12. SPONSORING AGENCY NAME AND ADDRESS United States Department of Transportation Office of the Secretary of Transportation 400 Seventh St. S.W. Washington, D.C. 20590		13. TYPE OF REPORT AND PERIOD COVERED Final Research Report	
		14. SPONSORING AGENCY CODE	
15. SUPPLEMENTARY NOTES This study was conducted in cooperation with the University of Washington and the Washington State Department of Transportation			

ABSTRACT

Traffic accidents cause loss of life and property. Proper identification of accident causal factors is essential for composing countermeasures against traffic accidents and reducing related costs. However, two-lane rural roads have distinctive roadway characteristics compared with other types of roads. In order to find cost-effective countermeasures and prioritize roadway safety improvement plans for two-lane rural roadways, a better understanding of the relationship between accident risk and respective characteristics is necessary. This study focuses on accident analysis of two-lane rural roads in Washington State. Six representative state routes (SRs), SR-2, SR-12, SR-20, SR-21, SR-97 and SR-101, are selected as study routes based on their location, length, and geometric characteristics. Along with the six-year (1999~2004) accident data from the Highway Safety Information System (HSIS), roadway video image data and geographical information system data retrieved from Washington State Department of Transportation are employed in this study. Econometric modeling methods are utilized to identify accident causal factors and evaluate their impacts on accident risk at roadway segments and intersections, respectively. Results from the statistical analyses and accident risk models not only help identify accident causal factors, but also provide valuable insights for developing countermeasures against two-lane rural road traffic accidents.

17. KEY WORDS		18. DISTRIBUTION STATEMENT	
Two-lane rural roads, safety, accident modeling			
19. SECURITY CLASSIF. (of this report)	20. SECURITY CLASSIF. (of this page)	21. NO. OF PAGES	22. PRICE
None	None	113	

DISCLAIMER

TABLE OF CONTENT

TABLE OF FIGURE

TABLE OF TABLES

EXECUTIVE SUMMARY

Traffic accidents have been a huge financial burden on society. Their cost has not only been the pain and suffering of the individuals involved in them but also the economic loss to society. It is statistically shown that the fatal accident rate on rural highways is more than twice as high as that for urban roads, even though the rate for all rural highway accidents is barely half of that for urban highways. Additionally, though Washington State's two-lane rural highways account for only 25% of total yearly vehicle miles of travel, approximately 56% of fatal and disabling accidents occurred on these roads. The above statistics clearly indicate that traffic safety conditions on two-lane highways need improvement.

The goal of this study is to better understand rural roadway accident causes in Washington, in order to help find cost-effective solutions for reducing the frequency and severity of crashes on rural two-lane roadways. To achieve such a goal, traffic accident data, roadway geometric data, traffic volume data, traffic control data, and related land use data from six study routes are collected and analyzed. The six study state routes (SRs), SR-2, SR-12, SR-20, SR-21, SR-97, and SR-101are considered representative to all state routes in Washington. These six routes are selected based on their location, length, and geometric characteristics. A total of six-year data from 1999 to 2004 are collected from multiple sources, including the Highway Safety Information System (HSIS), roadway video image data (State Route Web), and geographical information systems (GIS) data retrieved from Washington State Department of Transportation (WSDOT).

Since occurrence mechanism and casual factors are very different between roadway segment and intersection accidents, this project separated intersection accidents from roadway segment accidents for modeling and statistical analyses. However, the methodologies used for the two groups of accidents are similar. Statistical analyses including t-test and ANalysis Of VAriance (ANOVA) are used to identify accident causal factors. Statistical models such as Poisson regression, negative binomial regression, and zero-inflated Poison and negative binomial models are evaluated and applied to assess the impact of explanatory variables on accident risks. Results from the statistical analyses and accident risk models provide valuable insights in developing cost-effective solutions against roadway segment and intersection accidents on two-lane rural roads.

For roadway segment accidents, we conducted regular statistical analyses and quantitatively evaluated the effects of explanatory variables on all-type accident risk (AAR) and rear-end accident risk (RAR). Based on the modeling and statistical analysis results, cost-effective measures that can be applied to reduce roadway segment accident risk are:

- Avoid frequent speed limit changes along the curvy roadway segments.
- Warn drivers before they enter a curved or steep roadway segment since degree of curvature and grade have increasing effects on both AAR and RAR. Warning signs or other pavement-based warning techniques, such as pavement markers and rumble strips, can help reduce the risk.
- Widen the surface width and add an additional passing lane in high accident rate roadway segments.
- Widen shoulder width help reduce AAR but at the cost of increasing RAR.
- Remove roadside curbs and walls.

Similarly, statistical analyses and econometric models were applied to intersection accidents. Based on the analysis results, cost-effective measures that can be applied to reduce intersection accident risk are as follows:

- Lower speed limit at intersection approaches.
- Put more signs upstream of intersection to make drivers aware of the presence of intersection.
- Remove wall(s) at the inbounds of intersections.
- Increase shoulder width (greater than 6 feet) of intersection approaches.
- Keep shoulder widths consistent along intersection sections.
- Decrease the degree of curvature at intersections.
- Minimize the change in slope between the inbound and outbound of an intersection.

CHAPTER 1: RESEARCH BACKGROUND

1.1 INTRODUCTION

1.1.1 Research Background

Traffic accidents have been a huge financial burden on society. Their cost has not only been the pain and suffering of the individuals involved but also the economic loss to society. According to statistics provided by the National Safety Council (NSC, 2005), Motor-Vehicle accidents have been the leading cause of unintentional deaths in the United States from 1969 to 2005, as shown in Figure 1-1.

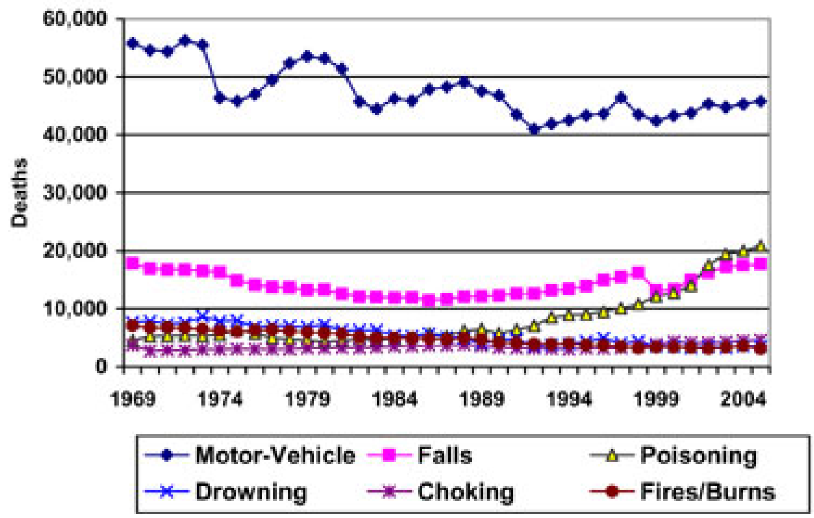

Figure 1-1 Leading causes of U-I deaths, U.S., 1969-2005
(Source: NSC, 2005)

The National Safety Council estimates the average cost of motor-vehicle accidents each year, including losses in wages, productivity, medical expenses, motor-vehicle expenses, property damages, and employers' uninsured costs (NSC, 2005). These costs reflect the impact of traffic accidents on the nation's economy. They are a measure of the amount of money spent on and the loss of potential income caused by injury or fatal accidents (NSC, 2005). This measure can be used to consider how momentous traffic safety improvement work should be. The calculable average comprehensive costs of motor-vehicle accident per injured person are estimated and shown in Table 1-1.

Table 1-1 Average comprehensive cost by injury severity

Death	$3,840,000
Incapacitating injury	$ 193,800
Nonincapacitating evident injury	$ 49,500
Possible injury	$ 23,600
No injury	$ 2,200

Source: NSC, 2005

The above figures cannot truthfully estimate the value of a person's natural desire to live longer or to protect the quality of one's life. However, they try to take into account an objective measure of the value of lost quality of life based on the results from empirical studies of people's willingness to pay for safety improvement. Therefore, improving traffic safety has been an important task as it not only relieves the weighty impact on society financially caused by traffic accidents but also helps protect the quality of people's life from being affected or taken away by those accidents.

Generally, accident rate is defined as the number of accident per million vehicle miles of travel. The fatal accident rate for rural highways was 1.32 and 1.43 respectively for year 2004 and 2005 whereas that for urban highway was 0.49 and 0.87 (WSDOT, 2004 and WSDOT, 2005). This implies that the average fatal accident rate for rural highways over the two years was more than twice as high as that for urban highways. Additionally, statistics produced by National Highway Traffic Safety Administration (NHTSA, 2004) show that 38.8% of total accidents and 74.9% of fatal accidents took place on U.S. two-lane highways. All these figures indicate that two-lane rural highway accidents are much more severe than accidents on other types of roadways.

In 2004, the total number of rural highway accident in Washington State was 10,727. It reached to 11,215 accidents in 2005, which is a 4.5% increase to that in 2004 (WSDOT, 2004 & WSDOT, 2005). Accordingly, accident rate increased from 0.95 to 0.99 accidents per million vehicle miles of travel (WSDOT, 2004 & WSDOT, 2005). Although Washington State's two-lane rural highways account for only 25% of total yearly vehicle miles of travel, approximately 56% of fatal and disabling accidents occurred on these roads (Olson and Glad, 2004). These statistics reflect a strong need for traffic safety improvements on two-lane rural highways.

Two-lane highways have a unique feature of having only one lane in each direction; therefore, driving behaviors on these roads are different from those on multiple-lane roadways. It is risky for a passing vehicle to occupy the opposing lane in order to pass a slow moving vehicle on a two-lane highway, especially when the traffic volume in the opposing lane is high. It is even riskier when roadway geometric features such as curvature, grade, etc. or roadside objects constrain the driver's line-of-sight. Moreover, two-lane roadways have limited space for vehicles that need to leave the road for emergency maneuvers.

Roadway segments and intersections have their own distinct characteristics; therefore, different accident risk models should be developed for different roadway locations and also for different types of accidents (Wang, 1998). Previous studies often address safety issues on multi-lane highways. This study concentrates accident analysis for both roadway segments and intersections on rural roads in Washington State.

1.1.2 Research Objective

The goal of this study is to better understand rural roadway accident causes in Washington, in order to help find cost-effective solutions for reducing the frequency and severity of crashes on rural two-lane roadways. Specifically, we have the following objectives for this research:

- Provide a better understanding of traffic accidents occurring on rural two-lane roads;
- Model the relationships between major accident types and causal factors quantitatively; and
- Recommend identified controllable factors in developing cost-effective solutions to improve traffic safety on rural two-lane roads.

1.2 STATE OF THE ART

The purpose of this section is to review studies focused on traffic safety improvement methods for highways in general, not just limited to those for two-lane rural highways. More specifically, this section covers some studies dedicated to traffic safety at intersections and roadway segments and a wide range of methods that have been used for accident risk modeling.

Traffic accidents have a heavy financial impact on society, and also affect the quality of life substantially. Improving traffic safety has been an important task over the past decades; thus, there has been much research done trying to find methods to reduce the frequency of accidents. Due to some of their unique features, two-lane highways are prone to fatal accidents. There have been many studies conducted to address this problem. Most of the studies such as Polus and Mattar-Habib (2004) and Lamm *et al.* (2002) focused on finding the relationship between geometric features, speed, traffic conditions, environmental characteristics, and accident rate. Other studies such as Persaud *et al.* (2004), Hickey (1997), and Washington *et al.* (2002) compared data from before and after a countermeasure were implemented to evaluate the effectiveness of the countermeasure.

Fitzpatrick *et al* (2002) performed a fairly complete review on crash treatment methods in Texas. It also discussed low-cost safety treatments and their effectiveness. According to Fitzpatrick *et al*. (2002), a crash study in Texas was conducted by following the following five steps: identifying sites and crash characteristics, gathering existing conditions, collecting additional field data, assessing the situation and selecting treatments, and implementing and evaluating. The study also identified the types of treatment being used on rural highways including rumble strips, passing improvement, two-way left-turn lanes, lane or shoulder widening, pavement edge drop-off improvements, pavement markings, mowing, skid resistance improvements, side slope flattening, recovery distance improvements, tree mitigation, culvert modifications, advance warning for horizontal curves, delineation, barrier reflectors, and animal countermeasures. Shoulder rumble strips were found effective with a relatively low cost that can reduce run-off-road crashes by 15 to 70 percent. Tree mitigation was also found to reduce 22 to 71 percent of vehicle-tree crashes with a relatively moderate cost.

Fitzpatrick *et al.* (2002) also discussed some safety treatments for rural intersections such as advance warning for intersections, approach rumble strips, left-turn bays, shoulder bypass lanes, intersection flashing beacons, signalization, high-intensity strobe lights, backplates on traffic signals, illumination, and sight obstruction reduction. Advance warning for intersections were considered having relatively low cost and effective in reducing crashes at rural intersections. They mentioned that a Federal Highway Administration (FHWA) study found that adding left-turn bays, a treatment with relatively moderate cost, could reduce crash rate by up to 48 percent. Illumination was also considered as a low-cost safety improvement method and could reduce nighttime fatal, injury, and property-damage-only crashes by up to 65, 30, and 15 percent, respectively.

Persaud and Griffith (2001) provided a complete review of current practice and research on statistical methods in highway safety analysis. They pointed out the essential types of safety analysis required to support traditional highway engineering functions such as identification of unsafe locations and development and assessment of countermeasures. The methodology used in their particular research was a survey of jurisdictions with highway engineering functions to assess current practices in highway safety analysis. It also involved gathering knowledge on the best available statistical tools and contacting leading researchers from twenty-seven state departments of transportation in the United States and five provincial transportation departments in Canada to find the most recent research on highway safety analysis (Persaud and Griffith, 2001).

The four methods that were used among the participating agencies were identification of hazardous locations (100 percent), before and after evaluations (94 percent), cost-benefit analysis in development of countermeasures (85 percent), and analysis of collision trends (85 percent). Persaud and Griffith (2001) also identified current problems and issues with practices in highway safety analysis. These problems are related to underreporting of collisions, identifying comparison sites in before and after studies, information on safety effectiveness in developing countermeasures, appropriate skills and resources needed for safety analysis, ability to link collisions and related databases. As for state of research, Persaud and Griffith (2001) affirmed that multivariate models were becoming popular in modern highway safety analysis. To relate accident experience to traffic and other roadway characteristics, multivariate models were used as regression equations by Hauer (1997), and cited by Persaud and Griffith (2001).

Persaud *et al.* (2004) conducted a before-and-after analysis on two-lane rural roadways in seven states before and after the installation of centerline rumble strips. Approximately 210 miles of treated roads were analyzed in the study. The purpose of the study was to find an engineering countermeasure for a major problem on the road involving "vehicles crossing the centerline and either sideswiping or striking the front ends of opposing vehicles" (Persaud *et al., 2004*). Rumble strips installed along the centerlines of undivided rural two-lane roads can help warn the distracted, fatigued or speeding drivers not to cross the centerlines and encroach on the opposing lane (Persaud *et al., 2004*). The study used Bayes empirical before-after method to take into account the regression to the mean, in order to normalize the differences in traffic volume and other factors between the before-and-after periods. They concluded that the installation of centerline rumble strips helps reduce the hazard of frontal and sidewipe crashes based on the results of their study, in which a 14 percent reduction for all combined injury crashes and a 25 percent reduction for frontal and sidewipe injury crashes were observed.

Also working on a solution to improve traffic safety on two-lane rural roads, Ogden (1997) did a study on the safety effect of paving rural roads' shoulders in Victoria, Australia. A before-and-after comparison, using accident data obtained from two-lane roads that had recent shoulder-paving projects, was carried out. The most common treatment for shoulder paving program involved an interim bituminous sealing treatment and a reseal in conjunction with a pavement reseal about one year later. According to the result of the Ogden (1997), this type of treatment for shoulder paving can be "statistically-significant" in reducing the frequency of injury and fatality accidents on two-lane rural highways in Victoria. Specifically, there was a 41 percent reduction in accidents per vehicle kilometer. Ogden (1997) also specified that the main accident reductions were in rear end, overtaking, and off roadway to left and off roadway to right into fixed object accidents. The study also stated that this safety improvement method was a cost effective method. The benefit to cost ratio of shoulder paving was estimated as 2.8 times the AADT in thousands (Ogden, 1997). For example, if the AADT at the location is 2000 vehicle per day, the benefit to cost ratio of shoulder paving should be estimated as 5.6.

Agent and Pigman (2001) conducted a before-and-after construction analysis on two-lane rural highways in Kentucky to study the impacts of construction on the highway safety. There were 49 roadway sections used in the study, 25 of them were upgraded to four-lane roads while the rest were realigned for wider lanes and shoulders. There was a dramatic increase in annual average daily traffic after the reconstruction on those roadway sections. Accident rates after the reconstruction were reduced significantly. For the sections that were upgraded by widening lanes and shoulders, the crash rate was reduced by 51 percent, whereas for the sections where lanes were added, there was a 56 percent reduction in the crash rate. Additionally, there was also a significant reduction in injury or fatal crash rates, which were reduced by 54 percent for realigned roads and 55 percent for upgraded roads. For both cases, the number of crashes per mile was reduced by 43 percent.

Tsyganov *et al.* (2005) researched the safety impacts of edge lines on rural two-lane highways in Texas State and performed general statistic analysis using accident data from Texas Department of Transportation (TxDOT). The study involved the compilation of rural two-lane highway data, examination of typical characteristics and dimensions of such roadways and used this information to carry out accident statistical analysis. Both roadway sections with and without edge lines were included in the research. In addition to roadway characteristic variables, other factors such as accident type, intersection presence, light condition, surface condition, severity, driver age, and driver gender were examined. The major results were that the presence of an edge line may account for up to a 26% reduction in accident frequency and the effects are stronger on curved roadway sections with lane widths of 9 to 10 ft. Also, a reduction in speed-related accidents was observed where edge-line treatment had been given to the road, a positive impact from better driver paths and speed perceptions.

In a study performed by Geurts *et al.* (2005), researchers looked at the ranking and selection criteria of dangerous crash locations in Flanders, Belgium. The underlying assumption of preceding ranking technique studies is that road accidents can be treated as random events, which means that each accident location has its mean crash rate. This approach assumes that the Poisson distribution lies behind the occurrence of accidents, which is widely accepted in numerous studies. Often, ranking of crash locations has been based on this distribution but without paying special attention to severity. In their study, they investigated the difference in results between the traditional ranking and an alternative ranking criterion. The alternative criterion gives weight to the severity of the crashes by using hierarchical Bayesian approach. The approach takes into account, for a specific time period, the number of crashes, the number of fatalities, and the number of light and severely-injured casualties for each accident site. Results showed that the alternative ranking criterion would change the selection of dangerous accident sites. It would lead to a different selection of 23.8% of a total of 800 sites. The study offers probability plots that serve as a valuable tool for prioritizing crash sites. The Bayesian ranking plots illustrate the estimated probability for a certain roadway accident location to be associated with the most dangerous sites. The authors recommended further research to include the construction cost to improve safety at different locations. If that were done, the ranking of locations could have been carried out by balancing the costs and safety benefits against each other.

Gårder (2005) analyzed head-on accidents in 2000-2002 that occurred on two-lane rural roads in Maine. The analysis, which included a total of 3136 reported head-on accidents, revealed that less than 8% of fatalities involved overtaking vehicles and only 14% of the accidents involved drivers who intentionally crossed the centerline. The accident data showed that higher speed limits led to a higher risk of fatal accidents or incapacitating injuries. The study concluded that, by keeping AADT and speed within certain limits, the severity of head-on accidents can be mitigated through narrower shoulders. According to Gårder, there are two main reasons that explain why drivers occupy the opposing traffic lane (and have head-on accidents): (a) vehicles are driven too fast for the roadway conditions and (b) vehicles occupy the opposing traffic lane unintentionally. Median barriers on head-on accident-prone roadway sections were discussed to reduce accident rates in both categories. Because of the huge financial cost of the median barrier installation, rumble strips were recommended for the remaining roadway sections, but of course rumble strips would only reduce rates in the latter category. The author also recommended a speed limit reduction for targeted high-crash sections and a more strict speed enforcement effort.

The aforementioned methods are to some extent helpful in determining appropriate countermeasures; however, they cannot fully reflect the quantitative impact of each individual causal factor on accident frequency. Accident risk models must therefore be developed to evaluate countermeasures with multiple variables to consider. There have been several studies accomplished using linear regression, Poisson regression, and negative binomial regression techniques to model accident risks.

Okamoto and Koshi (1989) used multinomial linear regression in their study and found that the random error of this method varied by the number of accidents and vehicle-kilometerage of the sections. One major problem with the linear regression model is that it may predict a negative number of accidents, while in real life accident frequency always holds a non-negative value. Poisson regression and negative binomial regression models are the two models considered to be more applicable for accident modeling.

Traffic accident data are always discrete, rare and non-negative; thus, they fit the features of a Poisson distribution. Miaou *et al.* (1992) used Poisson regression to model the truck accident data collected from the Highway Safety Information System (HSIS) of one state from 1985 to 1987. The data were assumed to be Poisson distributed. Unknown parameters were estimated using the maximum likelihood estimator. Final results from the modeling process showed that annual average daily traffic per lane, horizontal curvature, and vertical grade were robustly correlated with the truck accident frequency. However, shoulder width was found to have little correlation with the truck accident frequency. Because of the extra variation in the truck accident data and the lack of covariates, the model was considered to have the potential to improve. However, the researchers found that the improvement would not significantly change the initial findings.

Miaou *et al*. (1993) compared linear regression models and Poisson models to determine their suitability for modeling vehicle accidents and highway geometric design relationships. The linear regression model was found to lack the distributional properties to successfully describe random, discrete, and non-negative accident data. Although, Miaou et al (1993) concluded that Poisson regression models had the most appropriate statistical properties in describing traffic accident events, they also pointed out the limitation of Poisson model. Real accident data rarely has its variance equal to its means; nevertheless, the Poisson regression requires the variance of the data to be equal to the mean. If the variance of a data set is greater than the mean, the data set is considered over-dispersed. Over-dispersion will result in biased coefficients and flawed standard error if used for Poisson regression model. Miaou *et al*. (1993) suggested using negative binomial or double Poisson distributions as the solution to overcome the problem of over-dispersion.

Shankar *et al.* (1995) later used negative binomial regression to deal with the over-dispersion issue. Both Poisson model and negative binomial model were applied in their study in which the effects of roadway geometric and environmental factors on the frequency of freeway accident were explored. They modeled both the overall rural freeway accident frequency and the frequency of individual accident types such as rear end, sideswipes, fixed objects, overturns, etc. Based on the study they concluded that using negative binomial model would give a better explanation on the data set compared to the Poisson regression model and be more robust with regards to over-dispersion.

Poch and Mannering (1996) in dealing with accidents at intersections also had the same opinion as Shankar *et al.* (1995). They also found that the negative binomial model performs better than the Poisson model when dealing with over-dispersed data. They used negative binomial regression model to find the geometric and traffic related factors that affect the accident frequencies at intersections. The data was collected from more than sixty intersections in Bellevue, Washington. Four different accident-frequency models were estimated: total accident frequency, rear-end accident frequency, angle accident frequency and approach-turn accident frequency. One of their findings was that the higher accident frequency might be related to the increased left-turn traffic volume. Also, a greater number of opposing approach lanes is related to an increase in total accidents. One other interesting finding was that the intersections in the central business district (CBD) have a lower likelihood of rear-end accidents. for the authors attribute this correlation to the signal progression in CBD areas, which decreases the number of times the vehicles have to start and stop, thus decreasing the potential of rear-end accidents.

Ivan *et al.* (2006) used negative binomial regression for generalized linear models to evaluate the correlation between roadway geometric features and the incidence of head-on crashes on two-lane rural roads in Connecticut. Seven hundred and twenty roadway sections, of the same length, were used in the analysis. Two variables based on the curvature of the road segments, one variable based on the vertical grade of the segments, and speed limit had significant influence on head-on crashes. Three models were developed involving different combinations of the above-mentioned variables. The models suggested that the three geometric variables caused an increase in the number of crashes but variables such as lane and shoulder width were not found to influence the occurrence of head-on-crashes. Significant correlation was found between wet roadway surface and more severe head-on-crashes and the same applies to the latter and narrow road segments.

Wang *et al* (2003) studied the relationship between the rear-end accident frequency and the combination of lead-vehicle deceleration and the ineffective response of the following vehicle's driver to this deceleration. In this paper, accident probability was expressed as the product of the probability of the leading vehicle decelerating and the probability of the following vehicle failing to respond in time to avoid a collision. Information on traffic flow, traffic regulations, roadway geometrics, and human factors from over one hundred four-legged signalized intersection in Tokyo, Japan were used to model rear-end accident probabilities. An interesting fact about speed limit was discovered from the result of the modeling process. Speed limit has a positive impact on the probability of encountering an obstacle vehicle but it has negative impact on the probability of a driver failure. Because dual impacts of the explanatory variables had not been accounted for in previous research, this finding was one of the major highlights of this work.

Applying the microscopic approach developed by Wang (1998), Kim *et al*. (2007) built a model on the occurrence of rear-end accidents on multi-lane freeways. The probability of encountering an obstacle vehicle and the probability of driver's reaction failure were estimated in this model. The final model involved both human and non-human factors by incorporating the two probabilities together. They found that both the AADT and the truck percentage-mile-per-lane variables have dual impacts on the occurrence of freeway rear-end accidents. These two variables increase the probability of encountering an obstacle vehicle but decrease the probability of driver failure. Negative binomial regression was also statistically proven to be the right approach for modeling freeway rear-end accidents. Ten significant variables such as area type, speed limit, shoulder width etc. were found to have an effect on the accident frequency in the modified negative binomial model.

Vogt and Bared (1998) conducted a study on safety analysis on segments as well as on three- and four-legged intersections of rural two-lane roads in association with the development of the Interactive Highway Safety Design Model (IHSDM) which is a set of tools to help highway designers. The data for two States, Minnesota and Washington, used in the study included accident data (both severity and type), traffic data, lane and shoulder width data, and some alignment data collected from HSIS. Data were also obtained from photologs and construction plans. Poisson and negative binomial regressions were used for the three-legged and four-legged intersection modeling. The final model chosen was a negative binomial model for Minnesota data. The Washington data was not used for the final model because of its unreliability. The Poisson models were not chosen because of data over-dispersion issues. Some of the findings from the study were that driveways seemed to decrease accidents at three-legged intersections; that roadside hazards seemed to decrease accidents at four-legged intersections; a major road right turn lane seemed to increase accidents at three-legged intersections, and the angle effect varies from state to state and from three-legged to four-legged intersections. One of the most significant points in this study was that it pointed out the difference between the intersection models and segment models. Intersection models are based on fewer observations than the segment models and thus the relationships between accident frequency and intersection variables were not as "clear-cut" (Vogt and Bared, 1998). For this reason, p-values for intersection models should be allowed to have a much greater range than for segment models in order to identify the design variables that influence accidents and can be controlled by the designer (Vogt and Bared, 1998). The p-values used in some models have a value of 30%.

An examination of zero-altered probability processes, ZIP distribution and ZINB distribution, were included in a study carried out by Shankar *et al.* (1997). They used a counting process in order to distinguish roadway sections that can be evaluated as truly safe from those that can be evaluated as unsafe. The safe sections have accident likelihood close to zero but the unsafe sections can happen to have zero accident observations during some pre-defined time period. They claimed that this counting process works better than applications of Poisson and negative binomial accident frequency models since they do not account for this distinction and can therefore produce biased coefficient estimates when zero accident observations prevail. The authors suggested that the ZIP structure models were promising in terms of the capability of revealing roadway sections with zero accidents observations.

Lee and Mannering (2002) studied run-off roadway accidents on a 96.6-km section of highway in Washington State. The study combined a number of databases including a detailed database on roadside characteristics. They employed zero-inflated count models and nested logit models to estimate accident frequency and severity. Both empirical and methodological analyses were used to establish the relationships among roadway geometrics, roadside characteristics, and severity of run-off roadway accident frequency. The purpose of the study was to identify the cost-effective countermeasures that can be used to improve highway designs and highway safety. Treatments for roadside improvement recommended by the study include avoiding cut side slopes, decreasing the distance from the outside shoulder edge to the guardrail, decreasing the number of isolated trees along road-way sections, and increasing the distance from outside shoulder edge to light poles. The limitation of the study is that it was based solely on the run-off-roadway accidents in the northbound direction of SR 3 in Washington State.

Chayanan *et al.* (2003) explored the relationship between roadway and roadside accident rates for Washington State highways. They believed that the two accident rates for a given roadway section can be correlated though geometric, traffic, and environmental factors may have different effects on roadway and roadside accident rates. This correlation is due to unobserved effects common across the roadway and roadside (Chayanan *et al.*, 2003). They employed a logical extension of Classical Linear Regression – Ordinary Least Square (CLR-OLS) model called the Seemingly Unrelated Regression Estimation (SURE) model to systematically approach roadway and roadside accident rate modeling. According to Chayanan *et al.* (2003), the advantage of this model is that there is neither an imposition of "*a priori*" assumptions on definite linkage between roadway and roadside accident rates nor hypothetical support for such linkage.

Using the SURE model makes it more efficient to estimate the parameters when disturbances that link roadway and roadside processes become significant (Chayanan *et al.*, 2003). The study used a random sample of 500 one-mile sections from the Washington State highway system for modeling. The data sets included traffic data such as volumes, compositions, speeds, AADT, and so on, with geometric data such as lane, shoulder, median, curve, and intersection information. Historical weather data collected from the National Oceanic and Atmospheric Administration database was also included. The authors concluded that it would bring no significant efficiency improvements compared to the current state of practice in Washington State if the roadway and roadside were modeled simultaneously. Also, weather variables were found to be significant in both the roadway and roadside models. The authors, in addition, stated that data about side slopes and lengths of guardrail through-sections are essential to improve the explanatory capability of the roadway and roadside models.

In conclusion, the linear regression model was not found to be applicable to traffic accident modeling due to its lack of distributional properties to accommodate traffic accident data. Although the Poisson model is frequently used in modeling traffic accidents, it cannot handle over-dispersed accident data (Miaou *et al.* (1993), Wang *et al.* (2003), Shankar *et al.* (1995), etc.). This issue can be treated with the negative binomial regression model because it allows the variance of accident data to be greater than the mean. For this reason, negative binomial regression models will be used in this study for modeling the accident frequency at Washington two-lane highway intersections.

CHAPTER 2: STUDY ROUTES AND DATA

2.1 DATA COLLECTION PROCESS

Data used in this research was obtained from three sources: HSIS, the WSDOT Office of Information Technology, and the WSDOT online tool, State Route Web (SRweb).

HSIS is a data collection program that is operated by the University of North Carolina Highway Safety Research Center and the LENDIS Corporation. It is supported by the United States Department of Transportation (USDOT). According to the Highway Safety Information System Guidebook (Council and William, 2006), the Washington database in the HSIS is maintained by the Transportation Data Office (TDO) at WSDOT. HSIS receives data from the WSDOT TDO in the form of nine different data files including accident data, basic roadway inventory data, curve data, grade data, features data, roadway crossings and roadside facilities data, special-use lane information, railroad grade crossing index, and traffic data. The data requested for this research were extracted from those data files. These requested data sets include accident data file, roadlog file, curve file, and grade file.

The accident file contains three subfiles: the accident subfile, the vehicle subfile, and the occupant subfile. The accident data is collected statewide by all the Washington State police departments following a standard format (Council and William, 2006). The roadlog, curve, and grade files describe some basic characteristics of each homogeneous roadway segment between beginning and ending mileposts. Variables in the roadlog files include surface width, lane width and type, shoulder width and type, median information, rural/urban codes, terrain codes, and other roadway descriptors such as functional class. Variables related to AADT and Legal Speed Limit (SPD_LIMT) were extracted from other files and merged into the roadlog files (Council and William, 2006). The curve file contains variables related to angle, direction, degree and radius, length, maximum super-elevation, and legal speed limit. Approximately 70 to 80 percent of roadway sections are straight segments and do not have their degree of curvature and other variables listed (Council and William, 2006). The grade file contains information on percent grade, direction, and length. The curve and grade files' information was developed from construction drawings and straight-line diagrams.

The second data source is the WSDOT GeoData Distribution Catalog, a WSDOT distribution site for roadway Geographic Information Systems (GIS) data. These GIS data sets include: intersection location, lane information, Global Positioning Systems (GPS) route data, road log, etc. All data sets are imported into ArcGIS software to be further processed. The intersection data requested from the WSDOT Office of Information Technology were GIS data. This data set shows the locations of intersections along Washington state routes. The values of the data were obtained from the Transportation Information and Planning Support (TRIPS) database.

The last major data source is the SRweb, a WSDOT online tool. This web application provides a roadway snapshot every one hundredth of a mile on each state route. Some data, which are not provided by the other two sources, can be collected manually by this online resource. Moreover, SRweb can be used to examine the accuracy of the data obtained from HSIS such as the shoulder width and shoulder type.

2.2 ROUTES SELECTION

The data files, as listed below, obtained from HSIS consist of six separate spreadsheets for each of six consecutive years from 1999 to 2004:

- Accident File
- Occupant File
- Vehicle File
- Roadway File
- Curvature File
- Gradient File

This research is meant to explore the relationship among accident frequency, the roadway's geometric features, and traffic information at intersections and at roadway segments. Before the data is prepared for the database, it is necessary to link the roadway file and intersection file to both the curvature and gradient files based on the route and milepost (MP) attributes. Moreover, as mentioned earlier, some variables, such as the number of driveways and passing lane, have to be created by using SRweb manually. The data collection process was considered too time-consuming to explore all 141 state routes in Washington State. Therefore, representative routes were selected for the study.

The two criteria used to select study routes were (1) route length and (2) the geographic location and spatial alignment of the routes. The first factor to consider was route length. The selected routes have to be the ones that have sufficient lengths to be statistically significant for modeling purpose and to be representative geographically. Geographical location and spatial alignment is the other criterion and also a more substantial criterion for study route selection. The selected routes have to cover as much area in Washington State as possible to be geographically representative. After the selection process, the six chosen routes are SR-2, SR-12, SR-20, SR-21, SR-97, and SR-101. A GIS map showing the location of each route was developed and is illustrated in Figure 2-1 below:

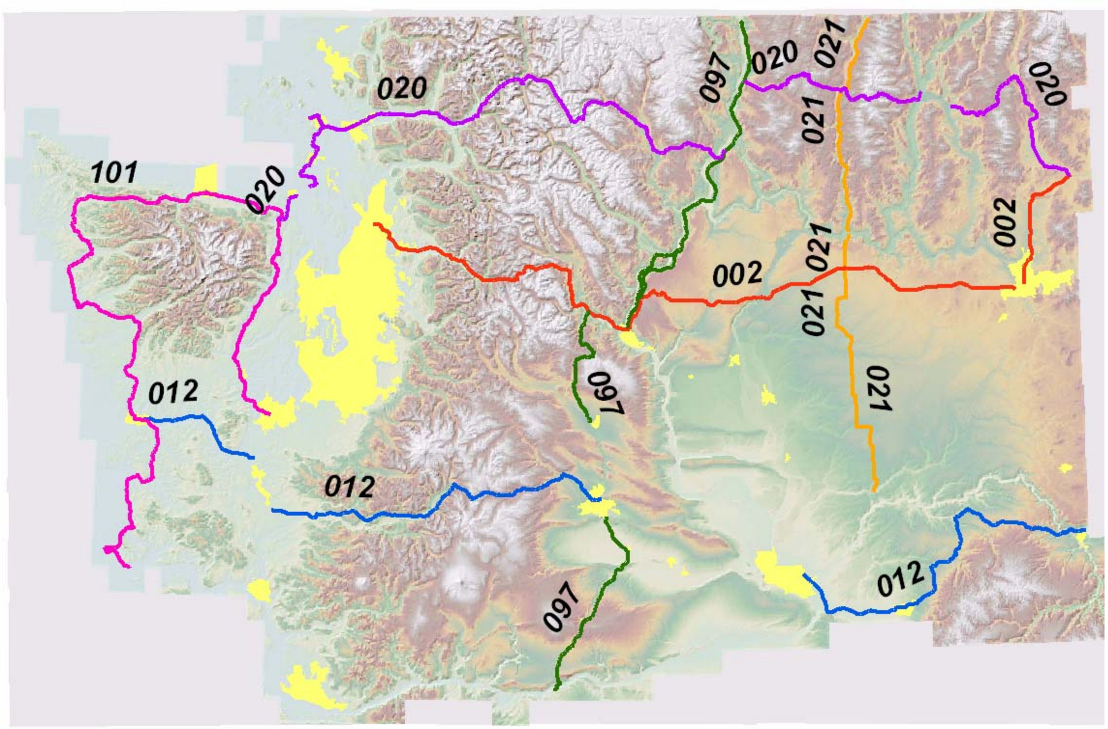

Figure 2-1 Map of six Washington State Routes used in the study

2.3 ROUTE DESCRIPTION

Three of the study routes, SR-2, SR-12, and SR-20, stretch from West to East. SR-20 covers the northern portion of Washington State whereas SR-2 and SR-12 cover the middle and the southern parts of the state, respectively. SR-20, with a length of 366.03 miles and an accident rate of 3.03 accidents per thousand vehicle-mile traveled (APTVMT), traverses Island, Skagit, Whatcom, Okanogan, Ferry, Steven, and Pend Oreille Counties. SR-2, with a length of 237.83 miles and an accident rate of 2.74 APTVMT, covers Snohomish, King, Chelan, Douglas, Lincoln, Spokane, and Pend Oreille Counties; it intersects with SR-20 on the eastside of Pend Oreille County. SR-12, with a length of 268.79 miles and an accident rate of 2.52 APTVMT, runs through Grays Harbor, Lewis, Yakima, Walla Walla, Columbia, Garfield, and Whitman Counties.

The other three study routes, SR-101, SR-97, and SR-21, stretch from North to South. SR-101 covers the western side of Washington States and runs north-south along the west coast from Olympia peninsula to the border of Oregon State. SR-97 and SR-21, respectively, represent the middle and the eastern portions of the state. SR-101 has a length of 317.86 miles and an accident rate of 2.38 APTVMT covering Pacific, Grays Harbor, Jefferson, Clallam, Mason and Thurston Counties. SR-97, with a length of 234.58 miles and an accident rate of 2.58 APTVMT, runs through Klickitat, Yakima, Kittitas, Chelan, Douglas, and Okanogan Counties. Finally, SR-21, with a length of 188.01 miles, the shortest of the six study routes, and an accident rate of 2.69 APTVMT, covers Franklin, Adams, Lincoln, and Ferry Counties.

As can be seen, these study routes geographically cover the entire state and reasonably represents traffic safety situations on two-lane rural roads in Washington.

CHAPTER 3: RESEARCH SCOPE AND METHODOLOGY

3.1 RESEARCH SCOPE

This research focuses on analysis of accidents at intersections and roadway segments on two-lane rural highways in Washington State. Data are organized using Microsoft Excel and then converted and managed by a relational database tool. Statistical analysis methods such as t-tests, ANalysis Of VAriance (ANOVA) and descriptive statistical study are applied. For accident risk modeling, Poisson regression, negative binomial regression, ZIP, and ZINB are considered. For over-dispersed data, Poisson regression is not appropriate and negative binomial regression is used instead. The goodness of fit of accident risk models is measured by several commonly used methods or statistics such as the likelihood ratio test, the sum of model deviances, and the ρ^2. The following sections provide a brief introduction on the concepts and theories applied in this study.

3.2 METHODOLOGY

3.2.1 Data Management

Most original data were received in Microsoft Excel format. Therefore, Excel is used to clean up and organize the data. Two Relational Databases, one for Roadway Segment Accidents (RSA) and the other for Intersection Accidents (IA), are designed using the Entity/Relationship (E/R) diagram method (Garcia-Molina et al., 2002). The designed E/R diagram is then converted to relational schemas. Both databases are implemented using Microsoft Structured Query Language (SQL) Server 2000. Such databases can be easily queried to support specific analysis or modeling efforts.

3.2.2 Data Organization

3.2.2.1 Data for Roadway Segments

The data needs in this study are comprehensive. Thus, multiple steps are taken to organize the data in the RSA database. The first step is to determine which variables can be used for the study. Some of the variables provided by HSIS are not complete. For example the age variable is available for every entry and the gender variable was not recorded until the beginning of 2002. Keeping the research interests and objectives in mind, numerous problematic variables are screened out from the original HSIS data files.

The second step is to verify the data accuracy and reliability, such as surface widths, lane configuration, and the existence of major cities along the routes. Road segments running through cities and towns should not be included in the database. A segment is initially considered rural (and included in the RSA database) if WSDOT's files (WSDOT, 2007) indicate the non-presence of cities and towns. However, this rural classification cannot be confirmed until the following three conditions are met:

- There are no signalized intersections present.
- There are no four-legged intersections present.
- There are no two-way turning lanes present, stretching between two or more intersections.

Where there is either a four-legged or a signalized intersection closer than five hundredths of a mile to the boundary of a given roadway segment, both that segment and the succeeding segment are excluded from the RSA database. This can be done by using the intersection data from WSDOT's GeoData and can be verified manually with SRweb. Wang (1998) classified accidents occurred within 30m from the stop bar on an intersection approach as intersection accidents, following the convention used by Tokyo Metropolitan Police Department. The 30m threshold may be appropriate at locations with prevailing speed limit of 30 km/h as in Tokyo. The average speed on the roadway segments in this study, however, is 52.85 mph (85 km/h), much higher than 18.75 mph (30 km/h). Therefore, a more appropriate distance threshold is needed to separate roadway segment accidents from intersection accidents.

In this study, we use "*Stopping sight distance*" (*SSD*) for determining if an accident belongs to a roadway segment or an intersection. *SSD* has been an important factor in roadway geometric design because it is the minimum required distance a driver should see for safe vehicle operations. *SSD* is also important for traffic safety at intersections. A driver approaching an intersection should have sufficient *SSD* so that he/she can decelerate or stop completely in time to avoid a collision. In the Traffic Detector Handbook (Kell *et al.,* 1990), an equation (Equation (3-1)) was introduced for calculating *SSD* at intersections. *SSD* calculated by this equation is the minimum distance from the stop bar that the vehicle can stop completely before the stop bar when signal is in red. If a vehicle is closer to the stop bar than *SSD*, then the vehicle may end up stopping inside the intersection.

$$SSD = V \times T + \frac{V^2}{2d} \tag{3-1}$$

- V = Approach speed, ft/sec (feet per second)
- t = Perception/reaction time (typically 1.0 sec)
- d = Constant deceleration rate in fps^2

Perception/reaction time is the summation of brake reaction time and perception time (ASSHTO, 2004). It is the time needed for a driver to see an obstacle and take an appropriate reaction such as changing speed or turning the vehicle to another direction to avoid the collision. As recommended by Kell *et al*. (1990), t=1.0 sec is used for this study. Also, we chose d=10.0 ft/sec^2 because the Institute of Transportation Engineer's (ITE) Handbook stated that it is reasonably comfortable for passenger car occupants to have the deceleration up to 10 ft/sec^2 (Pline, 1992). *SSD* depends on approach speed, perception/reaction time, and the constant deceleration rate. With t=1.0 sec and d=10.0 fps^2 for all calculations, approach speed becomes the only determinant for *SSD*. The calculated *SSD* is employed as the threshold of distance in this research to determine whether an accident is intersection-related. If an accident happened within *SSD* distance from the stop bar of an intersection approach, it is counted as an intersection-related accident. Otherwise, it is counted as a roadway segment accident.

The mean of the calculated *SSD*s is approximately 350 feet in our study, which is a reasonable distance compared with some other studies. For example, Yuan (2000) used

528 feet and Washington (2005) used 250 feet as the distance threshold for determining whether an accident is intersection-related. Although other equations are available to calculate the stopping sight distance on roadway segments or at intersections, they usually need additional information such as the friction values between tires and pavement surface, detailed intersection layouts, etc. Those types of information are not available in our data; therefore, it is considered the best to use Equation (3-1) to calculate *SSD*.

The HSIS curve file was used as a base to segment the roadways into sections. This file breaks roads down into small segments by curvature. The value of each variable, (e.g. the number of accidents) is allocated to each segment according to its beginning and ending mileposts. In cases where a variable can take multiple values on a particular roadway segment, the most extreme values are assigned to the roadway segment, e.g. the extreme values for both negative and positive grades.

3.2.2.2 Intersection Data

Exactly the same approach is used for organizing data in the IA database. The only difference is in the accident classification step. The *SSD* value calculated from Equation (3-1) is applied. All accidents that occurred inside intersections or on intersection approaches but within the corresponding *SSD* from the stop bar are considered intersection-related accidents. Using this method of classification, an accident is either included in the IA database or the RSA database. None is included in both.

Due to the lack of data on the crossing roads, this research does not consider accidents on crossing roads of the two-lane rural routes selected for this study. The intersections considered in this study are either four-legged or three-legged intersections. Data from the curve file, the grade file, and the roadway file are linked to the intersection file to provide necessary information for intersection approach sections.

3.2.3 Database Designs

3.2.3.1 Roadway Segments

Microsoft SQL Server 2000 is used for data management and query in this study. A good database design is essential for creating an effective database. The E/R diagram resulted from the design process is illustrated in Figure 3-1.

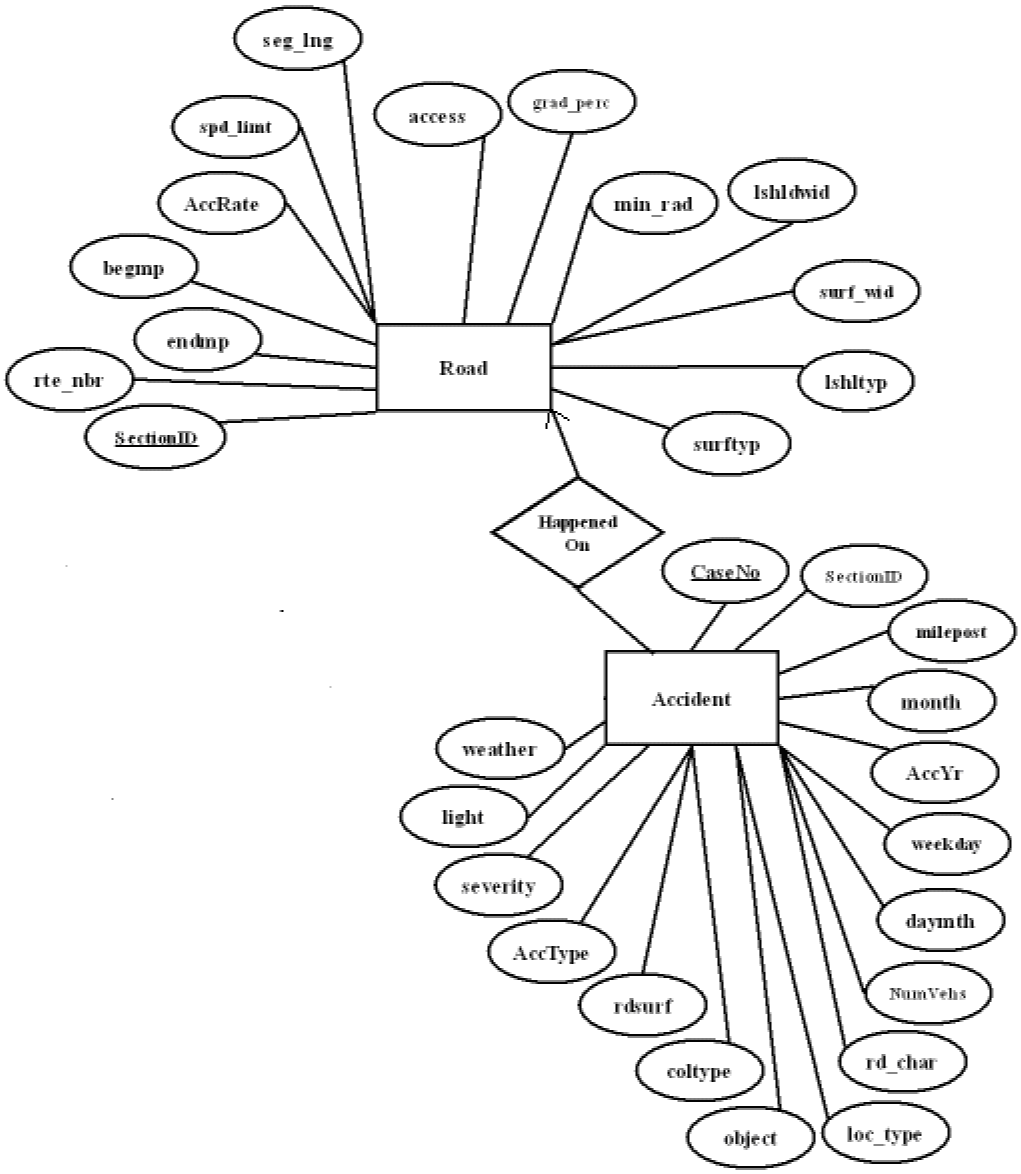

Figure 3-1 The E/R diagram for the RSA database

Relational schemas of the study database are converted from the E/R diagram design. The RSA database consists of two tables: *road* and *accident.* These relational tables contain the actual variables and each variable is referred to as an attribute. Each row in the tables is called a tuple. If an attribute has a unique value for each tuple in a table, then it is qualified to serve as a key for the table. Only one primary key can be identified for a table. The primary key attribute of each table is underlined in the E/R diagram.

Each tuple in the accident table corresponds to one roadway segment accident occurred on one of the study routes. Each accident is uniquely identified by its case number, *CaseNo*. Therefore *CaseNo* is chosen as the primary key for the accident table. Similarly, the segment ID serves as key for the road table. The relationship *HappenedOn* links together the *accident* and *road* tables. The following relational schemas are converted the E/R diagram and again the primary key attributes are underlined:

Relational Schemas

Road(SectionID, rte_nbr, begmp, endmp, AccRate, spd_limt, seg_lng , access, grad_perc, aadt, VMT, func_cls, lshldwid, lshl_typ, medwid, med_type, pop_grp, road_inv, rshlwid, rshl_typ, surf_wid, trf_cntl)

Accident(CaseNo, SectionID, rte_nbr, milepost, func_cls, rd_inv, accyr, month, daymth, weekday, acctype, severity, numvehs, rd_char1, rdsurf, loc_type, coltype, weather, light, object)

All the attributes are explained in Section 3.2.4. Following the relational schemas of the database tables, data from HSIS are re-organized and imported to Microsoft SQL Server 2000. The resulting RSA database can be used to generate new datasheets to support various statistical analysis and modeling efforts.

3.2.3.2 Intersections

Microsoft SQL Server 2000 is also used to manage and query data for intersection accidents. The IA database is designed using the E/R diagram database design method. The E/R diagram for the IA database is shown in Figure 3-2. Following the E/R diagram, relational schemas for the IA database are developed. Because *HappenedOn* is a many to one relationship from the Accident entity set to the Intersection Approach entity set, it is not converted to a standalone relation. The IA database therefore includes two tables: the Intersection Approach table and the Accident table. The primary key attribute for each entity set is underlined in the E/R diagram (Figure 3-2). Each tuple in the Intersection

Approach table corresponds to an intersection section (here an intersection section refers to the segment from *SSD* distance upstream to the stop bar of one direction to that of the opposite direction) while each tuple in the Accident table corresponds to an accident occurred on an intersection section. The primary key attribute for the Intersection Approach entity set is the intersection identification code, *InterID*, and that for the Accident entity set is the case number, *CaseNo.* The *InterID* attribute in the Accident table is a foreign key that connects the Accident table to the Intersection Approach table.

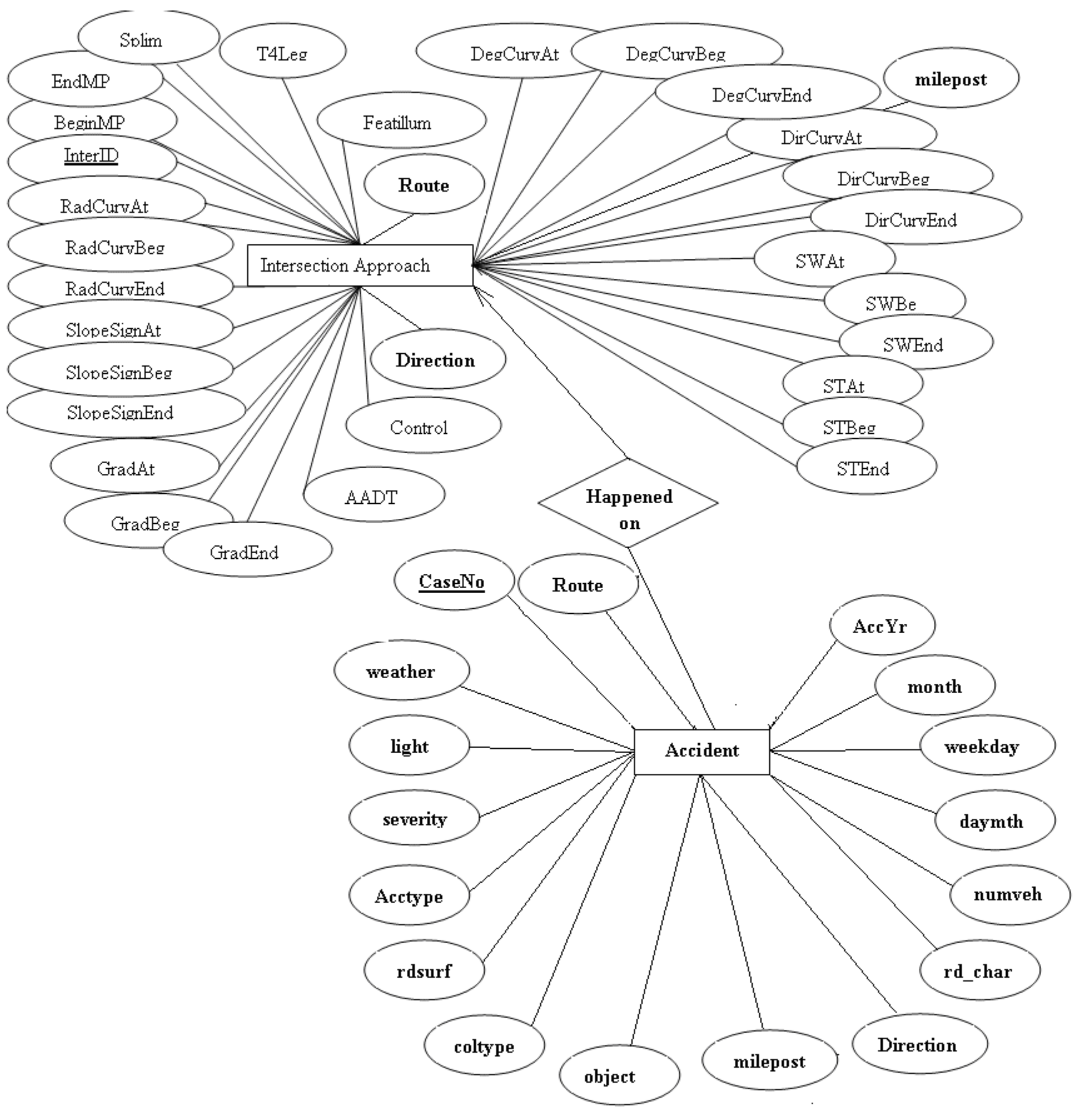

Figure 3-2 The E-R diagram for the SQL database

Relational Schemas

The following relational schemas are converted from the E/R diagram design:

Intersection Approach (InterID, BeginMP, milepost, End MP, Splim, T4Leg, Featillum, Route, DegCurvAt, DegCurvBeg, DegCurvEnd, DirCurvAt, DirCurvBeg, DirCurvEnd, SWAt, SWBe, SWEnd, STAt, STBeg, STEnd, Direction, Control, AADT, GradAt, GradBeg, GradEnd, SlopeSignAt, SlopeSignBeg, SlopeSignEnd, RadCurvAt, RadCurvBeg, RadCurvEnd)

Accident (CaseNo, route, milepost, weather, light, severity, Acctype, rdsurf, coltype, object, rd_char, numvehs, daymth, weekday, month, AccYr, Direction, InterID)

The primary key attribute for each table is underlined. Explanations of the attributes are available in Section 3.2.4.

3.2.4 Attributes Explanation

In this section, attributes of the roadway segment, intersection approach, and accident are introduced respectively.

(A) Roadway Segment Attributes

- Road Features

Milepost

Milepost refers to Accumulated Route Mileage (ARM), which is the route miles accumulated from the beginning of a state route to a specific location. The section between the beginning mile post, *BeginMP,* and the ending mile post, *EndMP,* is referred to as a roadway segment.

Route Number

The attribute *Route* indicates the route to which the roadway segment belongs. There are 6 values (2, 12, 20, 21, 97 and 101) for this attribute because there are only 6 routes considered in this study.

- Curve Features

Degree of Curvature

Degree of curvature is defined as the central angle D subtended by a chord of 100 feet as illustrated below:

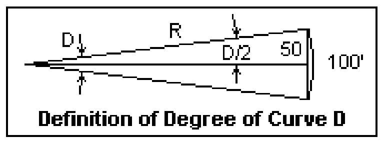

Figure 3-3 Definition of degree of curvature
(Source: Calvert, 2004)

The degree of curvature is calculated in degree using the Equation (3-2):

$$D = \frac{100' * 360^\circ}{2 * R * \pi} \qquad \textbf{(3-2)}$$

Where R is the radius of curvature. Equation (3-2) indicates that the radius of curvature is inversely proportional to degree of curvature.

Direction of Curvature

Direction of curvature is the horizontal curve direction which can be left curve, right curve, or straight segment. The direction of a curve on a roadway is in reference to the driving direction of the roadway.

Radius of Curvature

The radius of curvature is the radius of the circular curve, measured in feet. One modeling variable was created based on the *curvrad* variable.

• Grade Features

Grade

Grade, measured in percentage, is defined as the steepness of a roadway location.

• Roadway Features

Annual Average Daily Traffic

Annual Average Daily Traffic (AADT) is the annual average daily number of vehicle traveling through the intersection. The attribute *AADT* represents this information.

Shoulder and Surface Width

Shoulder and surface width is measured in feet.

Shoulder Type

Shoulder type refers to the material of the shoulder. In this study, the most common shoulder type is *Asphalt*. Besides *Asphalt*, shoulder type can also be *Bituminous*, *Portland Concrete, Gravel, Wall, or Curb*. The *Curb* shoulder type is more common in this study compared to the *Wall* shoulder type or the *Gravel* shoulder type.

(B) Intersection Approach Attributes

Intersection approach inherits attributes from roadway segment. In addition, intersection approach contains specific features as follows.

• Intersection Approach Features

Milepost

Milepost refers to the ARM of the location of the intersection. This variable together with the speed limit variable were used to calculate the stopping sight distance, which is critical for determining *BeginMP* and *EndMP,* the beginning and ending mileposts, respectively, for an intersection section. If any accident occurred on this section, it is considered intersection or intersection-related accident.

Type of Intersection

There are two types of intersection that are studied in this research: T-intersection and four-legged intersection. The attribute *T4Leg* was created using SRweb. The value of this attribute is either 0 or 1. If an intersection is a T-intersection, the value of *T4Leg* is 1. Otherwise, the value of *T4Leg* is 0.

Feature Illumination

The attribute *featillum* identifies the presence of any artificial illumination at an intersection section. The value of 1 indicates the presence of an artificial illumination at the intersection and the value of 0 indicates no illumination.

Intersection Traffic Control

The attribute *Control* identifies the presence of any type of traffic control at the intersections, such as stop sign, amber flashing, pedestrian control, red flashing, railroad signal and yield sign, etc. The value of 1 indicates the presence of traffic control(s) at the intersection and the value of 0 indicates the opposite.

- Curve Features

Degree of Curvature

DegCurvAt, DegCurvBeg, and *DegCurvEnd* were created as the degree of curvature at the intersection location, at the beginning of intersection section, and at the end of intersection section, respectively.

Direction of Curvature

The three attributes *DirCurvAt*, *DirCurvBeg*, and *DirCurvEnd* refer to the direction of curvature at the intersection location, at the beginning of intersection section, and at the end of intersection section.

Radius of Curvature

RadCurvAt, RadCurvBeg, and *DirCurvEnd,* represents the radius of curvature at the intersection location, at the beginning of intersection section, and at the end of an intersection section.

• Grade Features

Grade

Grade information was extracted from the grade file and used to create three attributes, *GradeAt*, *GradeBeg*, and *GradEnd*. which record the grade at the intersection location, at the beginning of intersection section, and at the end of intersection section, respectively.

Slope Sign

Slope sign have two values "+" (1) and "-" (0). The + value indicates that the slope at that location is positive and the - value indicates that the slope at that location is negative. *SlopeAt*, *SlopeBeg*, and *SlopeEnd* represent the sign of slope at the intersection location, at the beginning of intersection section, and at the end of the intersection approach, respectively.

• Roadway Features

Shoulder Width

Three attributes *SWA*, *SWBe*, and *SWEnd* were created to hold shoulder width in feet at the intersection area, at the beginning of intersection approach, and at the end of intersection approach.

Shoulder Type

Shoulder type data were extracted from the Roadway File and were used to create *STAt*, *STBeg*, and *STEnd*, whose values reflect the shoulder type at the intersection location area, at the beginning of intersection approach, and at the end of intersection approach.

(C) Accident Attributes

Case number

Case number is the identification code for accidents and is represented by the attribute *Caseno*.

Route

Route is the attribute used to identify which route the accident happened on.

Milepost

The *milepost* attribute identifies the ARM of the location where an accident occurred.

Weather

The *weather* attribute gives the weather information at the time when the accident happened. Possible values of this attribute are snowing, raining, fog/smog/smoke, etc.

Light

Light is the attribute used to indicate the lighting condition of a road at the time of accident. Possible values for this attribute are daylight, dawn, dusk, dark with street lights on or dark with street light off, etc.

Severity

Severity is the attribute shows the severity level of an accident. Possible values of this attribute are dead at scene, dead on arrival, died at hospital, disabling injury, possible injury, etc.

Accident type

The *Acctype* attribute represents the type of accident. There are approximately 40 types of accident. Some common types of accident are rear-end accident, overturned accident, strike-an-object accident, hit-animal-or-bird accident, or strike-other-vehicle-at-an-angle accident.

Road Surface

The *rdsurf* attribute gives the road surface condition at the time of accident. Possible values for this attribute are dry, wet, snow/slush, ice, sand/mud/dirt, standing water, etc.

Collision type

The *coltype* attribute indicates the type of collision, focusing mostly on vehicle(s) movement(s) when the accident occurred.

Object

The *object* attribute gives the information about the object presented in a collision. Possible values for this attribute are concrete median barrier wall, retaining wall, curb or raised traffic island, wood sign post, metal sign post, etc.

Road characteristic

rd_char is the attribute that shows the road characteristic of accident location. Possible values for this attribute include straight and level, straight and grade, straight at hillcrest, straight in sag, curve and level, curve and grade, curve at hillcrest, and curve in sag.

Number of vehicle involved

numvehs is the attribute that gives the number of vehicle involved in the accident.

Day of the month

daymth is the attribute records the day of the month when accident happened.

Day of the week

weekday is the attribute whose value is the day of the week when accident happened.

Month of the year

month is the attribute whose value is the month of the year when accident happened.

Year

Accyr is the attribute whose value is the year when accident occurred. In this research, all accidents happened between 1994 and 2004.

Direction

The *Direction* attribute indicates that the accident happened on the increasing milepost direction or on the decreasing milepost direction.

Intersection Approach Identification Code

InterID is the attribute whose value is the intersection approach identification code. This attribute was created using the direction attribute of the accident table, direction attribute of the intersection approach table, the milepost of accident locations, the beginning milepost of intersection section, and the ending milepost of intersection section.

3.2.5 Hypothesis Test

Hypothesis test is used to examine whether a difference in a population parameter, e.g. mean, variance, proportion, etc., between two or more groups is likely to occur by chance or whether the difference occurrs because of the impact of a certain factor (Washington *et al.*, 2003). In this study, hypothesis tests are used to evaluate the difference in means between two or more groups. Specifically, t-test is used to compare the means of two groups and ANOVA (or F-test) is used to compare means of more than two groups.

Both t-test and F-test are conducted using the statistical software SYSTAT (Version 11). SYSTAT is a software tool that can handle testing differences between two means or among three or more means of samples. The purpose of conducting the t-tests and F-tests is to find out whether certain variables have significants effect on accident frequency.

3.2.6 Accident Risk Modeling

3.2.6.1 Statistical Model Overview

Our modeling efforts focus on accidents on two-lane rural highways. Statistical models of annual accident frequency on an individual intersection approach and on an individual roadway segment are developed. Observation units are intersection approach sections or roadway segments on the selected two-lane rural routes. Each road section is either straight or uniformly curved. The dependent variable is the expected annual accident count (or called accident frequency) for each observation unit over the six-year period from 1999 through 2004.

Accident count data are discrete, non-negative, and randomly distributed. Based on previous studies, the Poisson regression model is deemed as a good fit for modeling such data. However, the foremost limitation of the Poisson regression model is that it requires the equality between the mean and the variance of the dependent variable. Accident data are often found over-dispersed (Shanker *et al*, 1995). An over-dispersed data set has its variance significantly larger than its mean. When the data set is over-dispersed, the estimated coefficients of Poisson regression models are biased. The requirement of equality between the mean and variance of data can be relaxed by using negative binomial regression. Negative binomial distribution can successfully deal with discrete, non-negative, randomly distributed, and over-dispersed data. Therefore, it is often used in modeling traffic accidents.

The frequency of zero-accident roadway sections in the data requests the significance of using ZIP and ZINB to be tested. Since the Poisson model is the base, it is discussed more thoroughly before we go to other models.

Several models introduced below use roadway sections as an example. These models work the same way with intersection approaches. Note that “roadway section” mentioned in the following means either intersection approaches or roadway segments in this study.

3.2.6.2 Poisson Regression Model

The idea of the Poisson model is to assume that the number of accidents in a given time interval on a particular roadway section follows Poisson distribution. The data from years 1999-2004 are used in the estimation of the model, which determines the time frame of the distribution. Therefore, in the Poisson regression model, the probability of having m_i accidents in a six year period at roadway section i is given by

$$P(m_i) = \frac{EXP(-\lambda_i)\lambda_i^{m_i}}{m_i!} \quad \textbf{(3-3)}$$

where

- $P(m_i)$ is the probability of section i having m_i accidents in the time frame of six years
- λ_i is the Poisson distribution parameter for roadway section i.

The Poisson regression process is to establish an estimate of the expected number of accidents, $E[m_i] = \lambda_i$. The estimate is a function of the explanatory variables such as surface width, AADT, and curvature. The explanatory variables are also called the regressors in the model. Assuming a Poisson distribution, the variance of the number of accidents on a given section during the study time period is $Var[m_i] = E[m_i]$.

The relationship between the regressors and the Poisson parameter is most commonly expressed as a log-linear relationship

$$\lambda_i = EXP(\overline{\beta X}_i) \quad \textbf{(3-4)}$$

where

- $\overline{\beta}$ is a vector of parameters being estimated
- $\overline{X}_i$ is the vector of the independent variables (regressors).

The bar notation in Equation (3-4) and in following sections of the study report represents a vector, not a single value. The most widely accepted way to estimate the parameters in

$\overline{\beta}$ is to use a Maximum Likelihood Estimation (MLE) procedure. The likelihood function can be written as

$$L(\overline{\beta}) = \prod \frac{EXP(-EXP(\overline{\beta}\overline{X}_i))(EXP(\overline{\beta}\overline{X}_i)^{m_i}}{m_i!} \qquad \textbf{(3-5)}$$

and the log likelihood function can now be derived from this equation

$$\ln L(\overline{\beta}) = \sum_{i=1}^{n} - EXP(\overline{\beta}\overline{X}_i) + m_i\overline{\beta}xi - \ln(m_i!) \qquad \textbf{(3-6)}$$

The log likelihood function is easier to manipulate than the likelihood function. This calculation of parameter estimates is carried out to find the factors that influence the count process. The Poisson parameters resulted from MLE are consistent, asymptotically normal, and asymptotically efficient.

When the Poisson parameter is estimated, the probabilities for accident observation in section *i* are given by

$$\begin{aligned} P_{0,i} &= EXP(-\lambda_i) \\ P_{j,i} &= (\frac{\lambda_i}{j})P_{j-1,i} \end{aligned} \qquad \textbf{(3-7)}$$

where

- $P_{0,i}$ is the probability that no accident occurs on section *i* in six years (*i*=1,2,3,..,6)
- *j* represents the number of accidents (*j*=1,2,3…)

The expected frequency (Poisson parameter) on roadway segment *i*, can then be written as

$$E[m_i] = \lambda_i = EXP(\overline{\beta}\overline{X}_i) \qquad \textbf{(3-8)}$$

Once the β–vector is known, the expected accident frequency can be straightforwardly calculated.

3.2.6.3 Negative Binomial (NB) Regression Model

The Poisson model has been criticized for its mean-variance equality requirement. If the variance is significantly smaller than the mean, there is no known model that can handle the situation. On the other hand, if the variance is significantly larger than the mean, the NB model is the most common alternative. For the NB model, the expected accident frequency for section i is rewritten as

$$\lambda_i = EXP(\overline{\beta X}_i + \varepsilon_i) \tag{3-9}$$

where $EXP(\varepsilon_i)$ is a gamma-distributed error term with mean 1 and variance α^2. This additional term is important because it allows the variance to differ from the mean in the following way:

$$Var[m_i] = E[m_i][1 + \alpha E[m_i]] = E[m_i] + \alpha E[m_i]^2 \tag{3-10}$$

The selection between the two models, Poisson or NB, is dependent on the value of α. As α approaches zero, the Poisson regression model is a limiting model of the NB regression model. The factor α is often referred to as the over-dispersion parameter. One of the forms the NB distribution can take is

$$P(m_i) = \frac{\Gamma((1/\alpha) + m_i)}{\Gamma(1/\alpha) m_i!} \left(\frac{1/\alpha}{(1/\alpha) + \lambda_i} \right)^{1/\alpha} \left(\frac{\lambda_i}{(1/\alpha) + \lambda_i} \right)^{m_i} \tag{3-11}$$

where $\Gamma(\cdot)$ is a gamma function. The likelihood function, based on the NB probability density function, takes the form:

$$L(\lambda_i) = \prod_{i-1}^{n} LN\left[\frac{\Gamma((1/\alpha)+m_i)}{\Gamma(1/\alpha)m_i!}\left(\frac{1/\alpha}{(1/\alpha)+\lambda_i}\right)^{1/\alpha}\left(\frac{\lambda_i}{(1/\alpha)+\lambda_i}\right)^{m_i}\right] \quad \textbf{(3-12)}$$

The MLE methods for the NB model are applied in the same way as for the Poisson regression model.

3.2.6.4 Testing for Over-Dispersion

An extended analysis can be used to test the over-dispersion in the data, i.e. whether or not the difference between the mean and variance is statistically significant. Cameron and Trivedi (1990) proposed a method to carry out an over-dispersion check. It is built on the fact that $(m_i - E[m_i])^2 - E[m_i]$ has a mean of zero in the Poisson model where $E[m_i]$ is the expected frequency. Hence, the null and alternative hypotheses are:

$$\begin{aligned} &h_0 : Var[m_i] = E[m_i] \\ &h_1 : Var[m_i] = E[m_i] + \alpha g(E[m_i]) \end{aligned} \quad \textbf{(3-13)}$$

where *g()* is a function of the expected frequency for a given model. A duplicate regression is estimated by using two different functions as *g()* and Z_i is regressed on W_i:

$$\begin{aligned} Z_i &= \frac{(m_i - E(m_i))^2 - m_i}{E(m_i)\sqrt{2}} \\ W_i &= \frac{g(E(m_i))}{\sqrt{2}} \end{aligned} \quad \textbf{(3-14)}$$

where the regression is estimated with both $g(E[m_i]) = E[m_i]$ and $g(E[m_i]) = E[m_i]^2$. If the regression $Z_i = bW_i$ reveals that b is statistically significant in either case, then H_0 is rejected.

3.2.6.5 Zero-Inflated Poisson and Negative Binomial Regression Models

ZIP and ZINB regression models have been developed to address the zero-inflated counting processes. The ZIP model for $M = (m_1, m_2, \ldots, m_n)$ accidents is

$$m_i = 0 \text{ with probability } p_i + (1 - p_i) EXP(-\lambda_i)$$

$$m_i = m \text{ with probability } \frac{p_i + (1 - p_i) EXP(-\lambda_i) \lambda_i^{\,y}}{m!}$$

where m is the number of accidents per observation unit.

The ZINB regression model has the form

$$m_i = 0 \text{ with probability } p_i + (1 - p_i) \left[\frac{1/\alpha}{(1/\alpha) + \lambda_i} \right]^{1/\alpha}$$

$$m_i = m \text{ with probability } (1 - p_i) \left[\frac{\Gamma((1/\alpha) + m) u_i^{1/\alpha} (1 - u_i)^m}{\Gamma(1/\alpha) m!} \right], \ m = 1,2,3,\ldots$$

where $u_i = (1/\alpha)/[(1/\alpha) + \lambda_i]$. For both ZIP and ZINB regression models, maximum likelihood methods are used to estimate the parameters.

To determine whether the zero-inflated or the conventional model is more appropriate to use, Vuong (1989) proposed a way to assess the appropriateness of using a zero-inflated model. The proposed test statistic is calculated for each section i

$$v_i = LN \left(\frac{f_1(y_i \mid \overline{X}_i)}{f_2(y_i \mid \overline{X}_i)} \right) \qquad \textbf{(3-15)}$$

where $f_1(y_i \mid \overline{X}_i)$ is the probability density function of model 1 and $f_2(y_i \mid \overline{X}_i)$ is the probability density function of model 2. The distributions must be specified in order to specify the equation for calculation. In this case they are either Poisson or Negative

Binomial distribution. These values are put in the following equation to obtain the Vuong's statistic

$$V = \frac{\sqrt{n}\left[\left(\frac{1}{n}\right)\sum_{i=1}^{n} v_i\right]}{\sqrt{\left(\frac{1}{n}\right)\sum_{i=1}^{n} (v_i - \bar{v})^2}} = \frac{\sqrt{n}(\bar{v})}{S_m} \qquad \textbf{(3-16)}$$

where

- $\bar{v} = \left(\frac{1}{n}\right)\sum_{i=1}^{n} (v_i)$ is the mean of the test statistic
- S_m is standard deviation and n is the sample size

The Vuong's statistic is asymptotically standard normal distributed and, therefore, it can be compared with z-values. If the calculated $|V|<V_{critical}$, the test is inconclusive and does not favor one model over the other. Positive values of V larger than $V_{critical}$ prefer model 1 over model 2. It works the same way for negative large values of V which favors model 2 over 1. For example if one would let $f_1()$ represent the NB model and $f_2()$ be the density function of the ZINB model. If the value of V is positive and larger than 1.96 (level of significance α=0.05), the test favors the traditional NB model. On the contrary, if a negative value of V is smaller than -1.96, the ZINB model should be the choice. V-values between those two critical values (-1.96<V<1.96) do not conclude anything on model choice. This test can be applied to Poisson and ZIP Poisson models following the same procedure.

3.2.6.6 Model Estimation

3.2.6.6.1 t-Statistic

Parametric hypothesis test statistics are commonly based on χ^2, t, or F tests. The t-statistic for example and its p-value (significance level) are used to tell if a variable in a model is significant. The χ^2, t, and F distributions are derived from normal distribution. The assumption of normally distributed disturbances is a base for the distributions of the above mentioned statistics. If this assumption is not valid, the statistics are dependent on the data and the parameters are not F, t, or χ^2 distributed. To evaluate the significance of the variable coefficients, the classical form of hypothesis testing is used. The null hypothesis, H_0, is opposed against the alternative hypothesis, H_1. The null hypothesis states that the estimated coefficient for the k_{th} independent variable is zero and the alternative hypothesis implies the opposite:

$$H_0\!: \ \hat{\beta}_k = 0$$

$$H_1\!: \ \hat{\beta}_k \neq 0$$

The most commonly used statistic for testing the coefficient hypothesis is the t-statistic. Assuming the above hypotheses and normal distribution of the disturbances, the t-statistic is written as

$$t_{n-K} = \frac{\hat{\beta}_k - 0}{s / \sqrt{n}} = \frac{\hat{\beta}_k}{S_{\hat{\beta}_k}} \tag{3-17}$$

where

- n is the number of observation units (roadway sections)
- K is the number of independent variables
- $n\text{-}K$ is the degree of freedom
- $S_{\hat{\beta}_k}$ is the standard error of $\hat{\beta}_k$, obtained from the standard deviation, s, and n.

The null hypothesis, H_0, is rejected if $\dfrac{|\hat{\beta}_k|}{S_{\hat{\beta}_k}} > t_{\alpha/2}$ (where α is the significance level) and the coefficient for the k_{th} independent variable can be assumed to be statistically significant. When the degrees of freedom increase, the t-distribution becomes closer to the standard normal distribution (if $(n\text{-}K)\rightarrow\infty$, $t\sim N(0,1)$). If $(n\text{-}K)>40$, the degree of freedom is generally considered high enough for the t-distribution to be approximated by a standard normal distribution. Figure 3-4 shows when the null hypothesis is rejected for a level of significance, α.

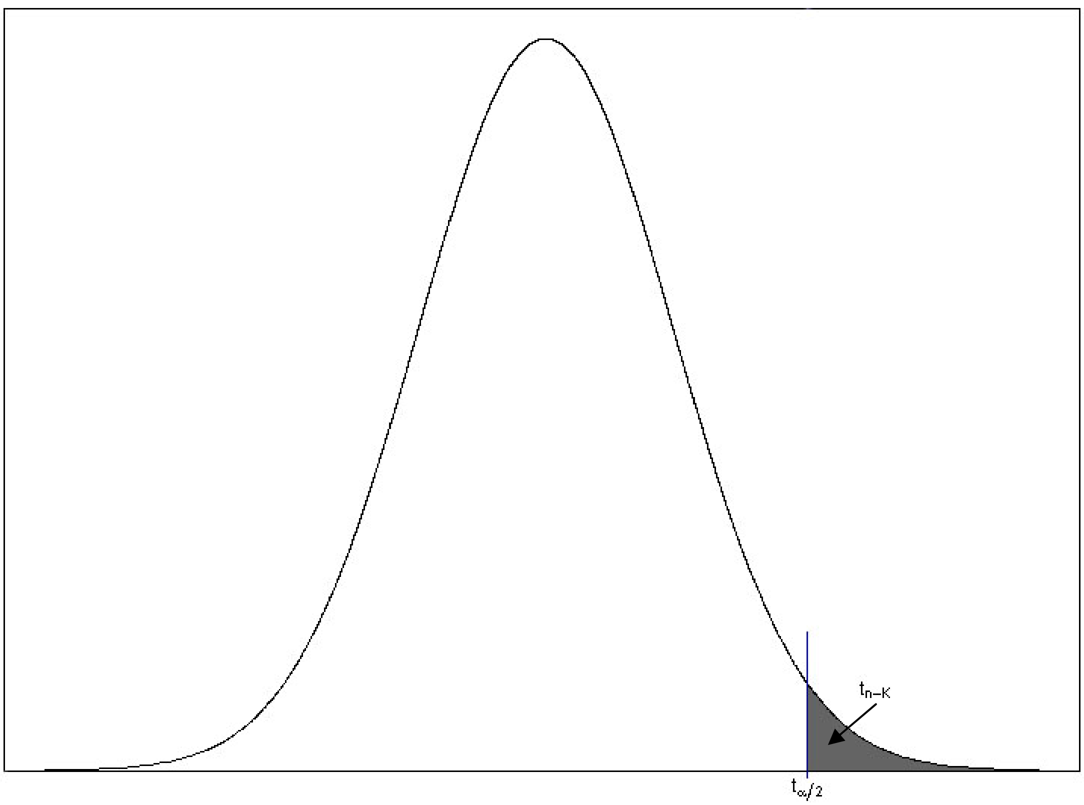

Figure 3-4 Rejection of the null hypothesis, H_0

The shaded region in the figure represents the area of rejection with $(n\text{-}K)$ degrees of freedom.

A common practice is to report the p-value, or the probability value of the test statistic. The p-value is the value that corresponds to the boundary where the null hypothesis is barely rejected. Given a value of α, the test rejects H_0 for all levels smaller than the p-value and fails to reject H_0 for all levels greater than the p-value. The smaller the p-value and greater the t-statistic, the more statistical evidence of rejecting the null hypothesis exists.

If the test is two-sided, the p-value for $\hat{\beta}_k$ is defined as

$$p(\hat{\beta}_k) = 2(1 - \Phi(\left|\hat{\beta}_k\right|)) \tag{3-18}$$

Where $\Phi(\cdot)$ is the cumulative distribution function (CDF) of the standard normal distribution. Figure 3-5 provides a visual aid on how to use the p-value in a two-sided test.

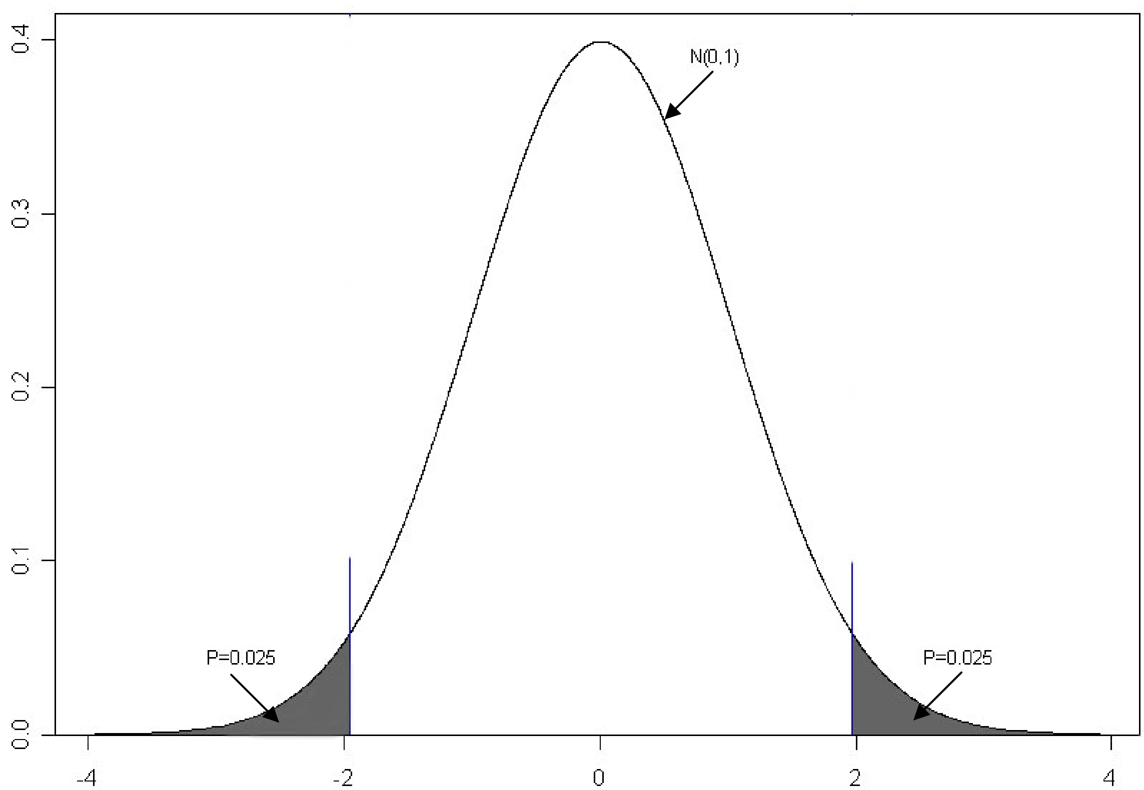

Figure 3-5 The p-value for a two-tailed test with significance level, α=0.05

3.2.6.6.2 Elasticity

In count data model estimation, the elasticity of a parameter is computed to assess the marginal impact of the regressor or the independent variable. The elasticity provides an estimate on how the variable impacts the expected frequency. They tell how heavily the expected frequency λ_i changes with a 1% change in the independent variable. The elasticity of frequency λ_i is calculated by

$$E_{x_{ik}}^{\lambda i} = \frac{\partial \lambda_i}{\lambda_i} \times \frac{x_{ik}}{\partial x_{ik}} = \beta_k x_{ik} \quad \textbf{(3-19)}$$

where

- E is the elasticity
- x_{ik} is the value of the k_{th} independent variable for roadway section i
- β_k is the estimated parameter for the k_{th} regressor and
- λ_i is the expected accident frequency on section i.

The elasticity values are computed for each roadway section but it is a popular way to compute the average of observations to represent the impact of each independent variable on the expected frequency.

Equation (3-20) is inappropriate for indicator variables and is only used for continuous variables. Indicator variables are binary variables and therefore take on values of 0 or 1. Sometimes they are called dummy variables. Dummy or indicator variables require the calculation of pseudo-elasticity which provides an estimate for the approximate elasticity of the independent variables. Pseudo-elasticity illustrates the incremental jump in frequency estimates which takes place when the indicator changes from 0 to 1. The equation for pseudo-elasticity is based on the estimated parameters of each independent variable:

$$E_{p,xik}^{\lambda i} = \frac{EXP(\beta_k) - 1}{EXP(\beta_k)} \quad \textbf{(3-20)}$$

Elasticity can tell the analyst whether an independent variable is contributing a realistic amount to the total expected frequency. In other words, how much effect it has in comparison to all other independent variables.

3.2.6.7 Maximum Likelihood Estimation Method

The method used for estimating model parameters in this study is MLE. The theory behind MLE is to identify the data generating process which stands behind an observed data sample. The MLE procedure hunts down the coefficient values that maximize the probability of the observed number of accidents. The conditional probability density function for a random variable *y*, given a set of parameters, $\overline{\theta}$ is

$$f(y_1,...,y_n|\overline{\theta}) = \prod_{i=1}^{n} f(y_i|\overline{\theta}) = L(\overline{\theta}|\overline{y}) \quad \textbf{(3-21)}$$

This joint density is applicable if the *n* independent observations are also identically distributed. The joint density is therefore the product of the individual densities and is called the likelihood function. This function is hard to manipulate mathematically and therefore the log likelihood function is introduced:

$$L(\overline{\theta}|\overline{y}) = \sum_{i=1}^{n} Ln(f(y_i|\overline{\theta}) \quad \textbf{(3-22)}$$

The MLE method calculates the derivative of the log likelihood function and sets it equal to zero. The θ values found by this method maximize both the likelihood and log likelihood functions as the example in Figure 3-6 illustrates.

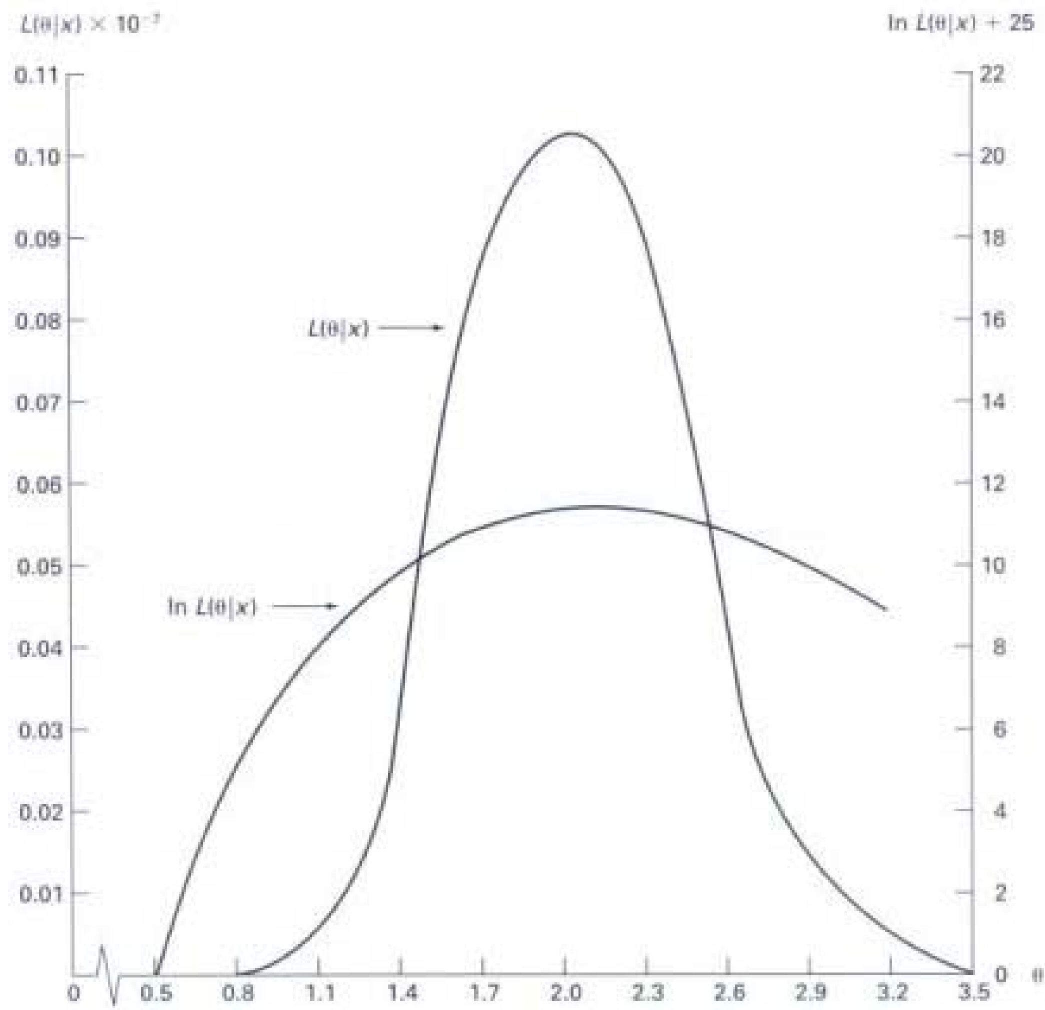

Figure 3-6 Likelihood and log likelihood functions for the Poisson distribution. [Source: Greene (2000)]

The goal of the MLE method is always to find the parameter θ that makes an observed sample most probable.

3.2.6.8 Goodness of Fit Measures

The elasticity values and maximum likelihood estimation methods do not tell how well a model fits the real accident frequency. Hence, other statistical tools are needed for that task. There are several tests to estimate the model's goodness of fit such as the likelihood ratio test, sum of the model deviances test, the ρ^2 statistic, and an equivalent statistic to the R-squared used in linear regression models.

The likelihood ratio test provides an estimate between two competing models, usually the model under consideration and a model that is restricted normally by having reduced the number of model parameters. The likelihood ratio test statistic is

$$X^2 = -2[LL(\bar{\beta}_R) - LL(\bar{\beta}_U)] \quad \textbf{(3-23)}$$

where

- $LL(\bar{\beta}_R)$ is the log likelihood at convergence of the restricted model in which all variables are set to zero,
- $LL(\bar{\beta}_U)$ is the log likelihood at convergence of the unrestricted model. The X^2 statistic is chi-squared distributed and the degrees of freedom are equal to the difference in the numbers of parameters in the restricted and the unrestricted model.
- The degree of freedom of X^2 is equal to the difference in dimension of the vectors $\bar{\beta}_R$ and $(\bar{\beta}_U)$.

As the difference between the log likelihood functions for the restricted and unrestricted gets greater, the explanatory power of the model improves. According to the same logic, the explanatory power improves as the value of X^2 gets larger.

Another measure, G^2, is the sum of model deviances. The closer the G^2 is to zero, the closer the model is to a perfect fit. This statistic is defined by

$$G^2 = 2\sum m_i LN(\frac{m_i}{\hat{\lambda}_i}) \quad \textbf{(3-24)}$$

There exists no equivalent measure in the Poisson regression model to the R^2 used in OLS linear regression. The reason is that the conditional mean, $E[m|\bar{X}]$ is nonlinear and also because of the presence of heteroscedasticity in the regression. Heteroscedasticity arises in a model when disturbances are not stable in terms of variance. Nevertheless, a like statistic is based on standardized residuals and is defined as

$$R_p^{\,2} = 1 - \frac{\sum_{i=1}^{n} \left[\dfrac{m_i - \hat{\lambda}_i}{\sqrt{\hat{\lambda}_i}} \right]^2}{\sum_{i=1}^{n} \left[\dfrac{m_i - \overline{m}}{\sqrt{\overline{m}}} \right]^2} \qquad \textbf{(3-25)}$$

where the mean accident number is expressed as $\overline{m}$. The numerator can be viewed as a sum of square errors and the denominator as a total sum of squares.

The overall model fit can be measured by the ρ^2 statistic, which is a widely used statistic for non-linear models. It uses the log likelihood values to compute it:

$$\rho^2 = 1 - \frac{LL(\overline{\beta}_U)}{LL(\overline{\beta}_R)} \qquad \textbf{(3-26)}$$

where

- $LL(\overline{\beta}_U)$ is the log likelihood at convergence with parameter vector $\overline{\beta}$,
- $LL(\overline{\beta}_R)$ is the log likelihood function with all variables set to zero and only the constant is included.

A model that predicts accident frequencies perfectly would have a likelihood function equal to one and the log likelihood would be zero which results in a ρ^2-value equal to one. The statistic is therefore between zero and one and the explanatory power of the model increases as the statistic is closer to one.

CHAPTER 4: DATA ANALYSIS

4.1 NON-PARAMETRIC ANALYSIS

4.1.1 Roadway Segments

Data used for analysis in this study include 7841 accidents which happened on 6165 roadway segments of the six study routes: SR-2, SR-12, SR-20, SR-21, SR-97 and SR-101 over a 6-year period from 1999 to 2004. Table 4-1 shows the number of accidents by type classified by HSIS for each study route:

Table 4-1 Reported accidents on roadway segments of the six study routes from 1999 to 2004

Accident Type	SR-2	SR-12	SR-20	SR-21	SR-97	SR-101	Total	Rank
Strikes other object	254	222	266	41	220	435	1438	1
Vehicle overturned	311	173	204	53	254	341	1336	2
Strikes animal or bird	196	201	125	29	212	298	1061	3
Strikes appurtenance	230	116	133	11	149	267	906	4
Strikes rear end of other vehicle	282	53	100	4	60	194	693	5
Ran into roadway ditch	81	52	90	9	34	231	497	6
Ran over embankment - no guardrail present	46	31	55	17	44	90	283	7
Strikes left side of other vehicle at angle	90	20	22	5	43	79	259	8
Sideswipes left side of other vehicle	50	25	16	2	29	53	175	9
Was struck on left side at angle by other vehicle	63	16	21	4	14	52	170	10
Strikes front end of other vehicle (not head on)	42	19	15	3	30	42	151	11
Was struck on right side at angle by other vehicle	50	16	10	1	17	45	139	12
Strikes right side of other vehicle at angle	46	7	16	2	9	32	112	13
Was struck in rear end by other vehicle	54	15	10	2	13	16	110	14
Strikes other vehicle head on	21	15	10	1	15	21	83	15
Non-collision fire	17	12	5	0	19	18	71	16
All other single vehicle involvements	14	8	10	2	16	20	70	17
Strikes or was struck by object from other vehicle	16	4	12	1	4	10	47	18
Jackknife trailer	9	2	2	0	24	4	41	19
Sideswipes right side of other vehicle	11	4	4	0	9	12	40	20
Was struck in front end by other vehicle (not head on)	11	3	3	2	3	10	32	21
Ran into river, lake, etc.	3	7	6	2	1	12	31	22
Strikes or was struck by working object	8	5	1	0	8	0	22	23
Pedestrian struck by vehicle	5	6	2	0	1	7	21	24
Was sideswiped on left side by other vehicle	7	2	3	0	3	5	20	25
Was sideswiped on right side by other vehicle	6	0	0	0	1	5	12	26
Was struck by other vehicle head on	3	1	1	2	1	0	8	27
All other multi vehicle involvements	1	1	2	0	1	2	7	28
Pedalcyclist struck by vehicle	1	0	1	0	0	3	5	29
Pedestrian strikes vehicle	0	0	1	0	0	0	1	30
Pedalcyclist strikes vehicle	0	0	0	0	0	0	0	31
Total accident from 1999 - 2004	1928	1036	1146	193	1234	2304	7841	

The 31 accident types listed above are re-classified into 12 main accident types according to the mechanism of accident occurrence. Shares of the 12 accident types are shown in Figure 4-1.

The most observed types on all the study routes are “strike other objects” (19%), “vehicle overturns” (17%), and “animals/birds” (14%). Figure 4-2 to Figure 4-7 show the shares of accident type for each study route over the six-year period. Noticeably, the “strike other objects” and “vehicle overturns” types are within the top three on any study route. The rear-end accident is the leading type on SR-2 but it is not among the top types on all other study routes.

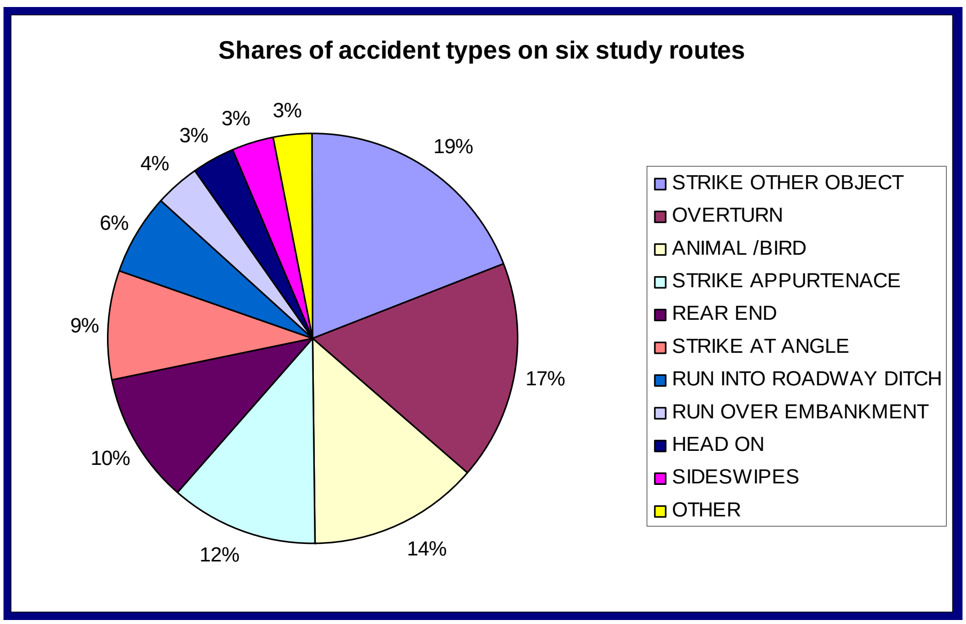

Figure 4-1 Shares of accident types on six study routes

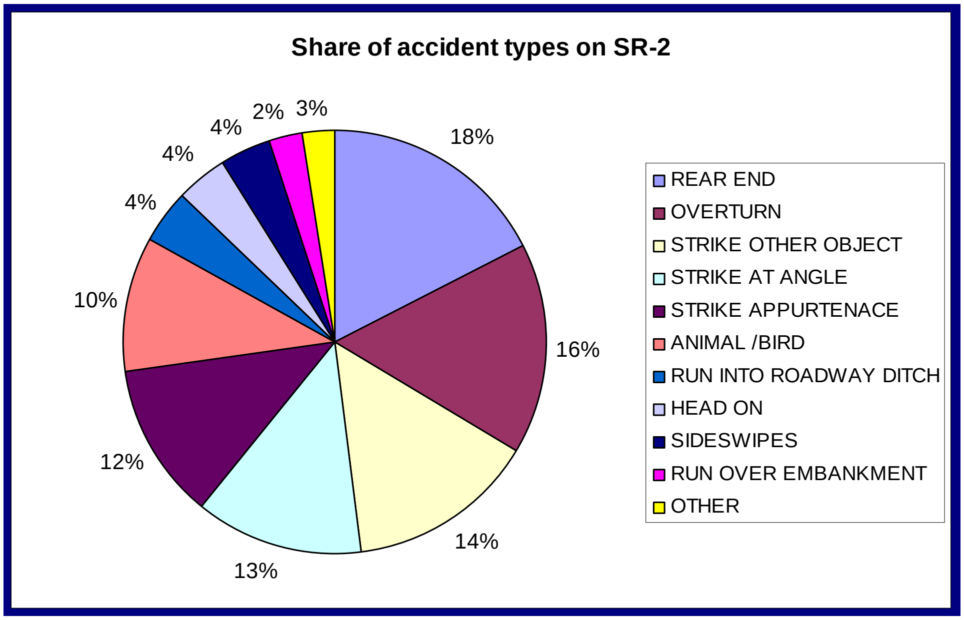

Figure 4-2 Shares of accident types on SR-2

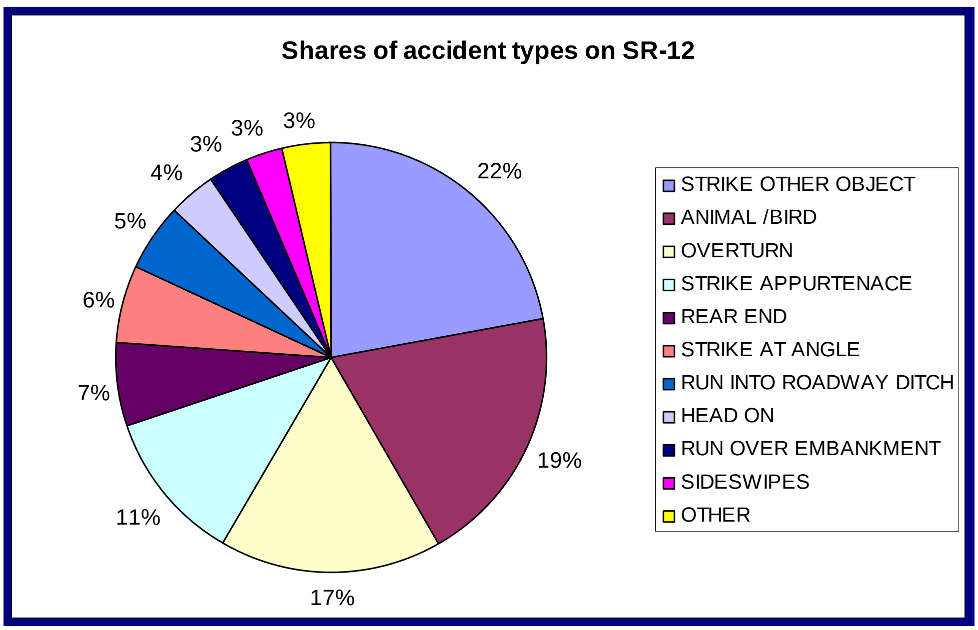

Figure 4-3 Shares of accident types on SR-12

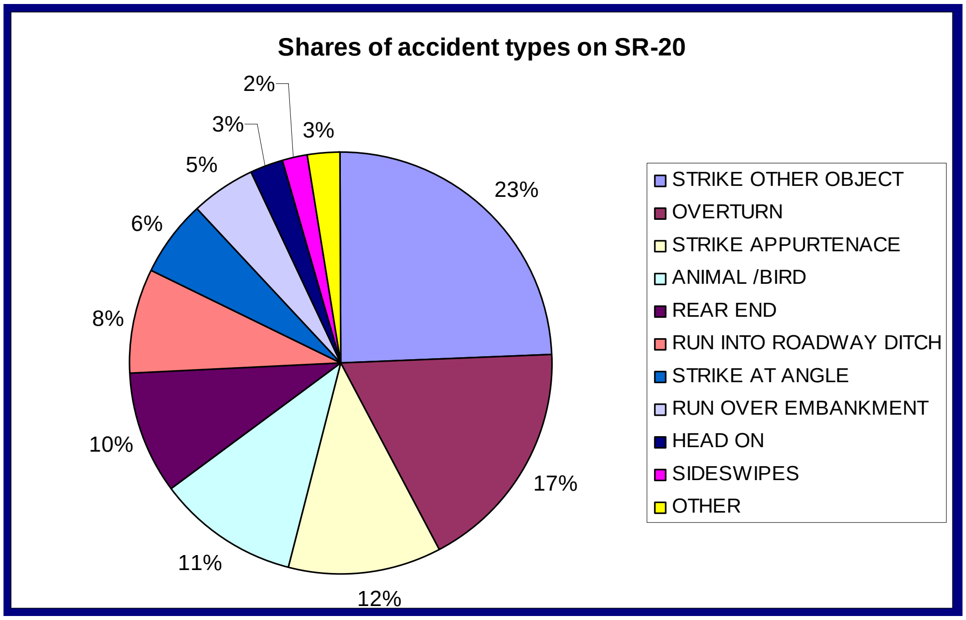

Figure 4-4 Shares of accident types on SR-20

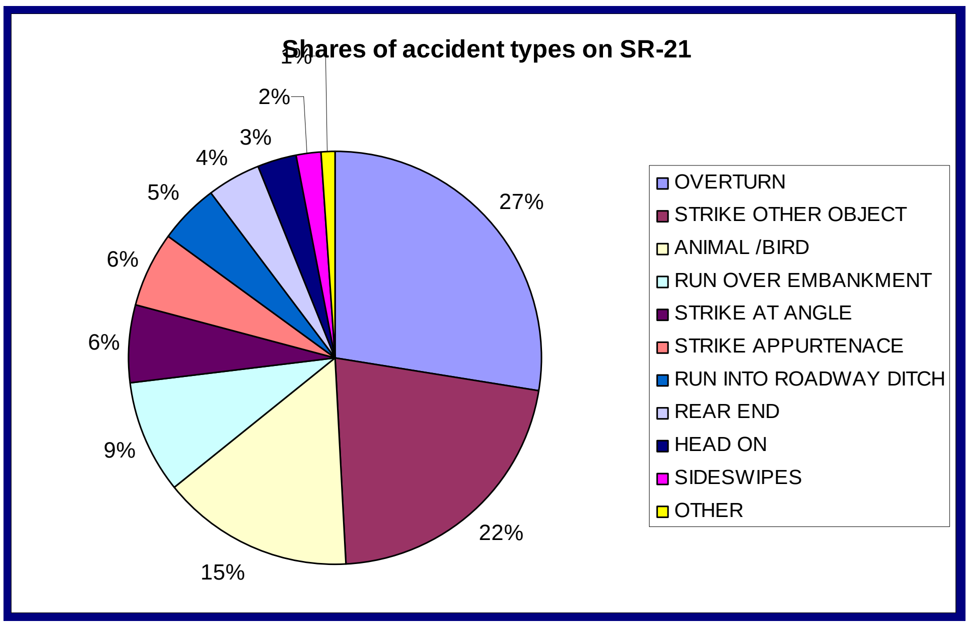

Figure 4-5 Shares of accident types on SR-21

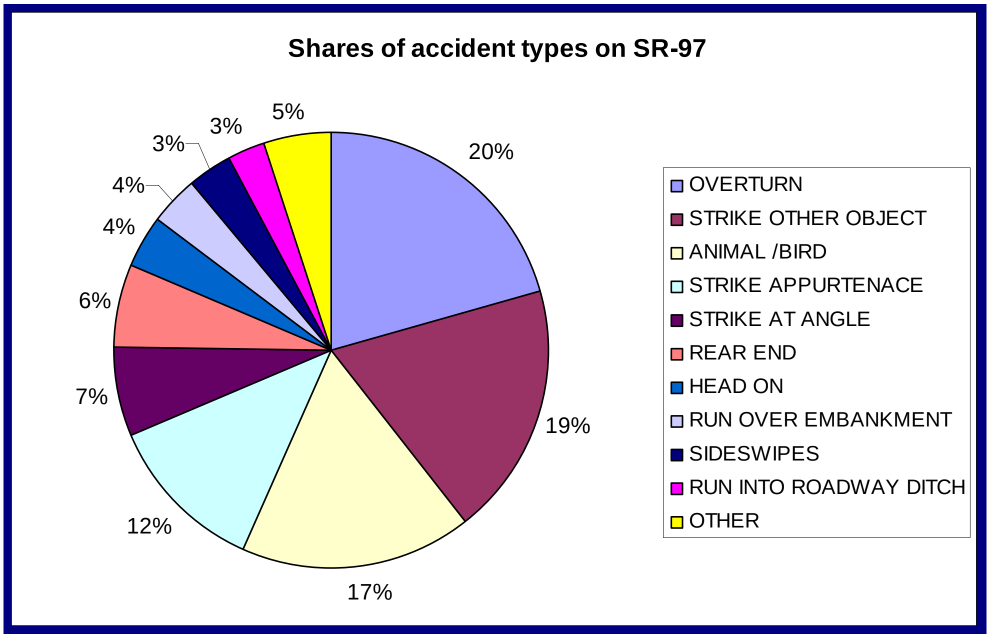

Figure 4-6 Shares of accident types on SR-97

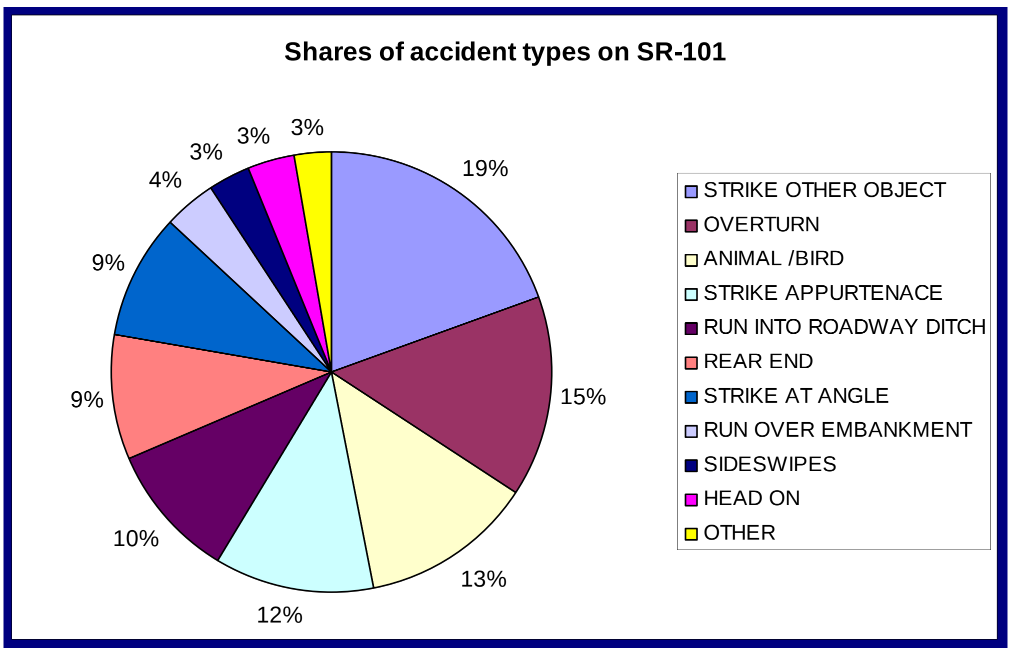

Figure 4-7 Shares of accident types on SR-101

As shown in Figure 4-8, SR-2 has the highest number of accidents per unit length (a mile) among the six study routes whereas SR-21 has the lowest one. Interestingly, the leading type of accident on SR-2 is the rear-end accident and these rear-end accidents could be the reason for higher accident frequency on SR-2. The causal factor of rear-end accidents will be further discussed in Chapter 5.

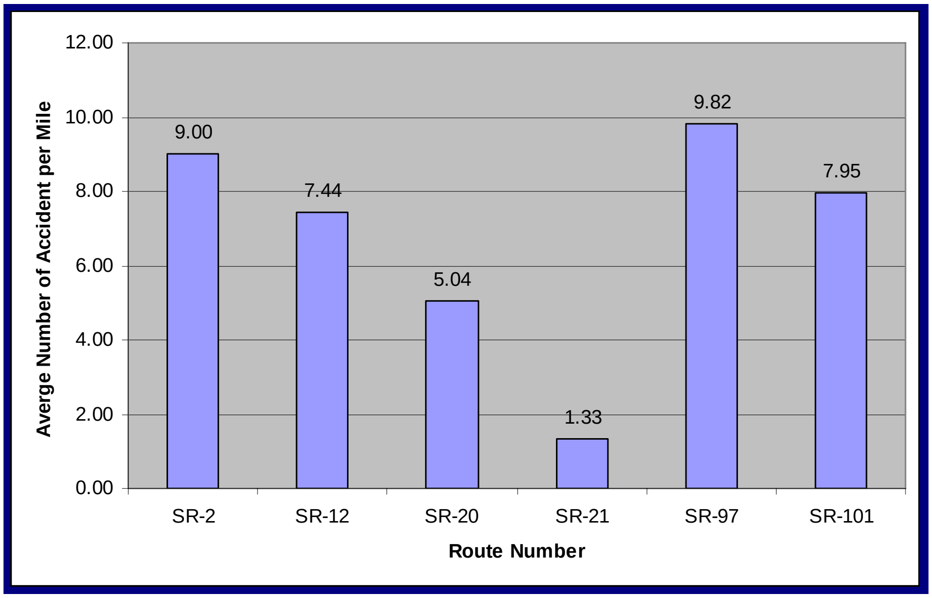

Figure 4-8 Average numbers of accidents per mile by route

As illustrated in Figure 4-9, over 60% of the total accidents occur while there is daylight. The highest commuter traffic volumes are observed during the morning (6-9AM) and afternoon (3-6PM) peak hours. As a result, one may be surprised by the extremely low accident ratio occurring at dawn.

As shown in Figure 4-23, most accidents occurred in clear or cloudy days. More accidents occurred in rainy days (12.96%) than in foggy days (1.75%). In general, when road surface changes from dry to wet, the friction coefficient between a patterned tire and the road surface decreases from 0.7 down to 0.4 (Jones and Childers, 2001). This decrease is worse for worn tires. The friction coefficient between a smooth tire and the road surface drops from 0.9 down to 0.1 as the surface goes from dry to wet (Jones and Childers, 2001). As the friction between the tires and the road decreases, the chance for vehicles to get into accidents increases because the tires can easily lose the cohesiveness with the road surface. However, further analysis with whether information is needed to conclude if rainy days are more dangerous than dry days on the study routes.

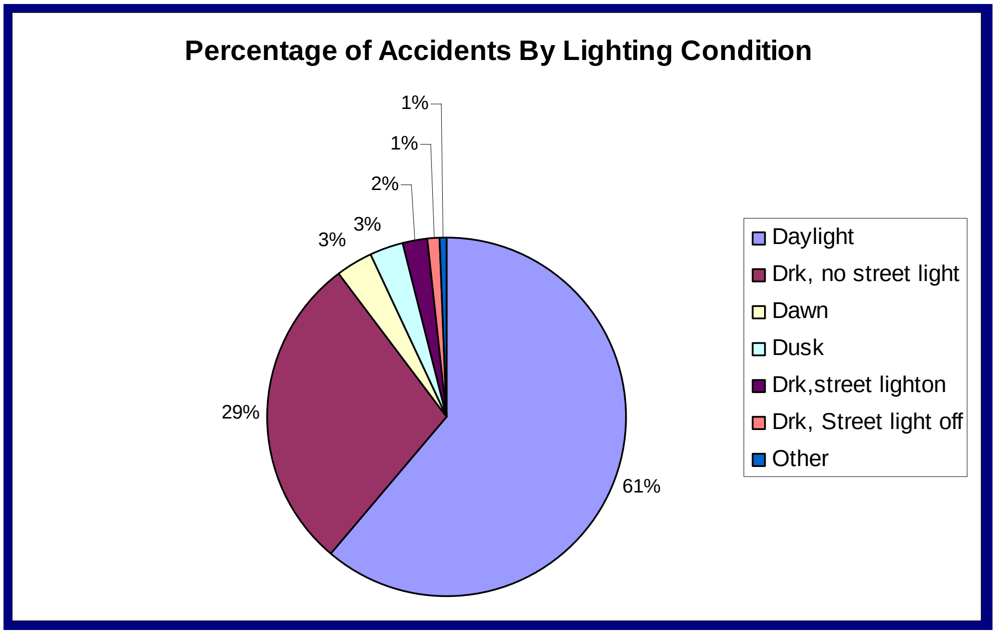

Figure 4-9 Percentage of reported accidents by lighting condition

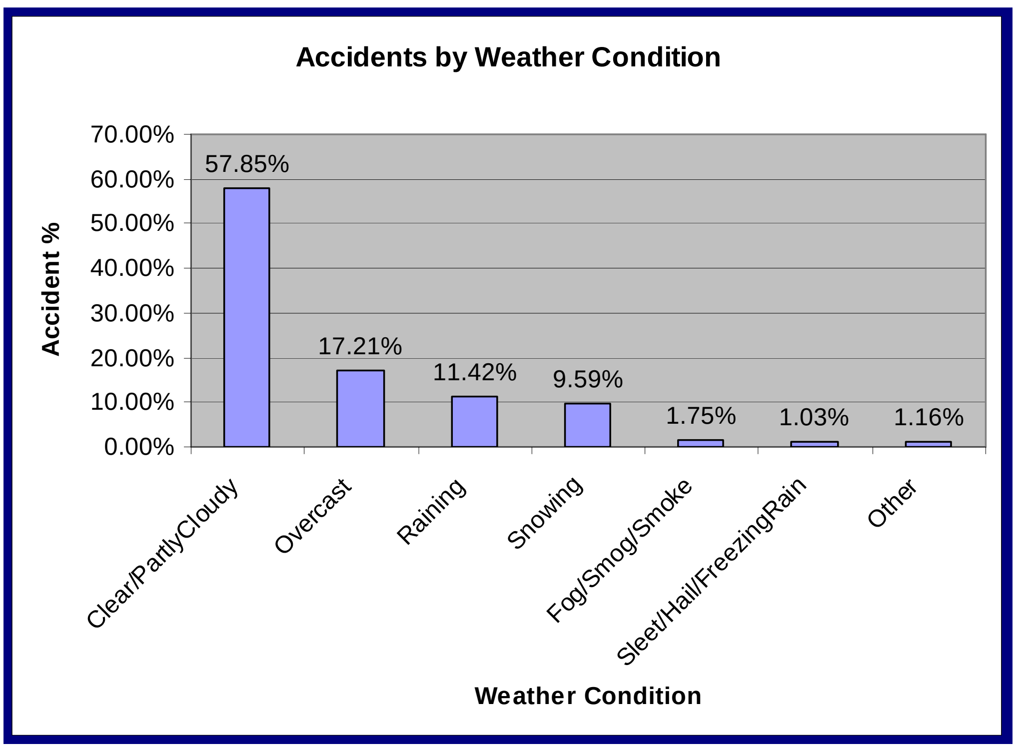

Figure 4-10 Percentage of reported accidents by weather condition

The variation over months seems larger than that over weekdays as can be seen by comparing Figure 4-11 with Figure 4-12. In accordance with the Highway Capacity Manual (TRB, 2000) traffic volume study over weekdays, the largest portion of accidents occurs on Fridays. As one would expect, there are more accidents occuring during the weekend days (Friday through Sunday).

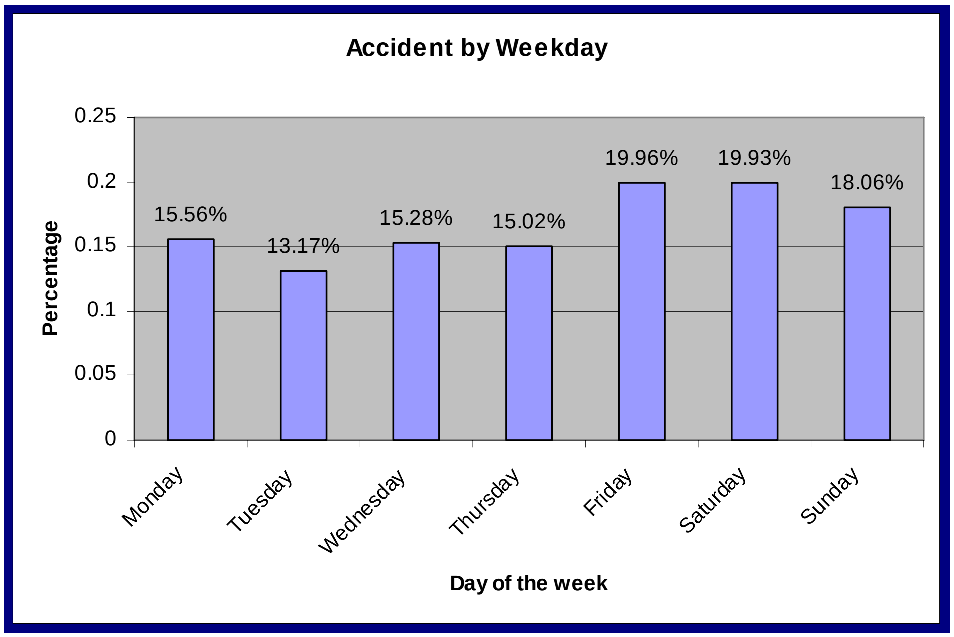

Figure 4-11 Percentage of reported accidents by weekday

Accident data sorted by month are shown in Figure 4-12. It is no surprise that December has the highest number of accidents, followed by January. The month with the fewest number of accidents is April.

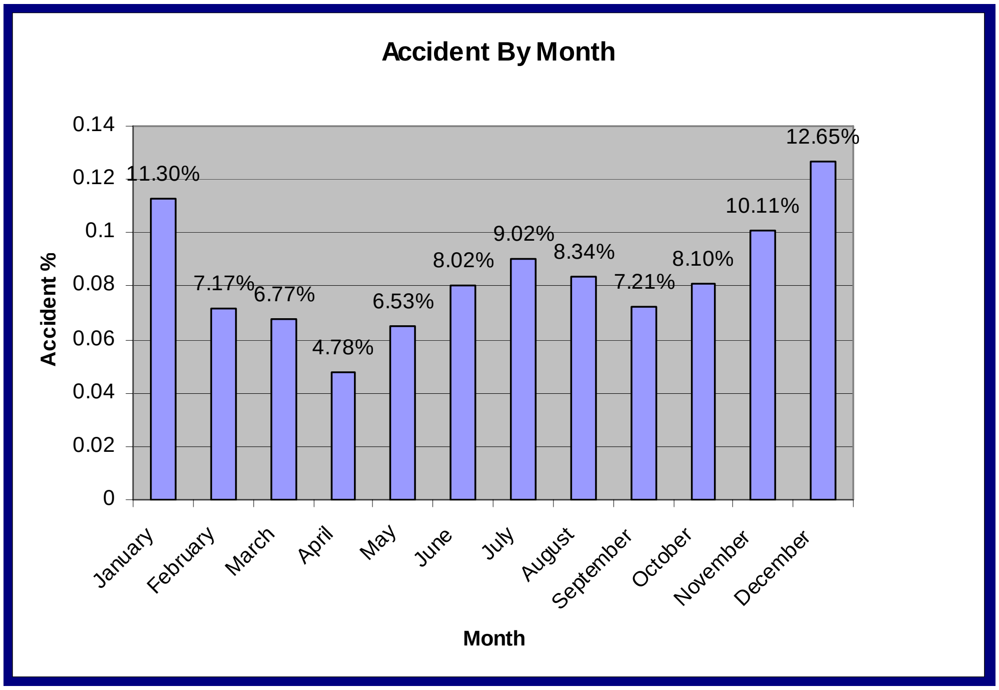

Figure 4-12 Percentage of reported accidents by month

Figure 4-13 illustrates that the numbers of accidents vary between 1999 and 2004. Although the numbers fluctuate over years, it stays around 1300 consistently.

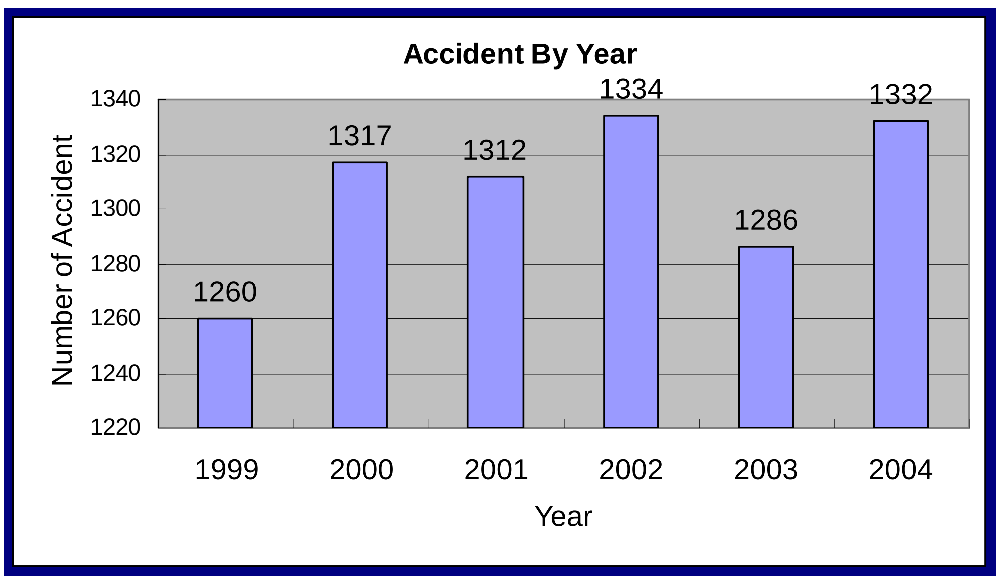

Figure 4-13 Number of reported accidents by year

4.1.2 Intersections

Data used for intersection accident analysis include 3650 accidents which happened at 1881 intersections (or 3762 intersection approaches) of the six study routes: SR-2, SR-12, SR-20, SR-21, SR-97 and SR-101 over a six-year period from 1999 through 2004. Table 4-2 shows numbers of accident by type classified by HSIS for each study route:

Table 4-2 Reported accidents on intersections of the six study routes from 1999 to 2004

Accident Type	SR-2	SR-12	SR-20	SR-21	SR-97	SR-101	total	Rank
Strikes rear end of other vehicle	169	165	285	2	68	134	823	1
Strikes appurtenance	66	73	80	5	47	84	355	2
Strikes or was struck by working object	72	44	63	9	51	54	293	3
All other multi vehicle involvements	45	81	49	5	61	50	291	4
Strikes left side of other vehicle at angle	48	50	54	2	48	72	274	5
Strikes animal or bird	46	58	44	3	49	71	271	6
Was struck on right side at angle by other vehicle	28	46	47	0	38	46	205	7
Was struck on left side at angle by other vehicle	33	48	45	3	29	44	202	8
Strikes right side of other vehicle at angle	26	33	26	1	21	42	149	9
Was struck in rear end by other vehicle	37	28	36	1	24	22	148	10
Strikes front end of other vehicle (not head on)	17	29	40	2	22	34	144	11
Non-collision fire	15	19	41	1	14	44	134	12
Sideswipes left side of other vehicle	11	13	17	0	15	20	76	13
Ran into river, lake, etc.	6	6	12	3	8	8	43	14
Strikes other vehicle head on	7	3	8	0	7	8	33	15
Sideswipes right side of other vehicle	6	5	7	0	6	5	29	16
Was struck in front end by other vehicle (not head on)	3	3	8	1	7	7	29	17
Vehicle overturned	5	5	2	0	7	2	21	18
All other single vehicle involvements	2	5	7	1	3	3	21	19
Ran over embankment - no guardrail present	1	4	4	0	2	7	18	20
Jackknife trailer	0	2	1	1	11	1	16	21
Strikes or was struck by object from other vehicle	3	3	3	0	5	1	15	22
Was sideswiped on left side by other vehicle	2	2	5	0	1	3	13	23
Pedestrian strikes vehicle	1	0	3	0	0	6	10	24
Ran into roadway ditch	0	3	3	0	1	1	8	25
Was sideswiped on right side by other vehicle	3	0	0	0	2	2	7	26
Was struck by other vehicle head on	1	2	1	1	0	1	6	27
Strikes other object	1	1	1	0	3	0	6	28
Pedalcyclist struck by vehicle	1	1	3	0	0	1	6	29
Pedestrian struck by vehicle	1	0	1	0	1	0	3	30
Pedalcyclist strikes vehicle	0	0	0	0	1	0	1	31
Total accident from 1999 - 2004	656	732	896	41	552	773	3650	

Similar to the roadway segment accidents, the 31 accident types are re-classified into 12 main accident types according to the mechanism of accident occurrence. Shares of the 12 accident types are shown in Figure 4-14.

As seen in Figure 4-14, the rear-end accident and strike-at-angle accidents are the top two accident types at intersections. To be more specific, Figure 4-15 to Figure 4-20 show the shares by accident type for each study route over the six-year period. It is worth mentioning that these two types of accident account for more than 50% of total accidents occurred on the study routes.

The two dominating accident types on SR-21 are over-turn and strike-at-angle. Since over-turn accidents often cause injury or death, an in-depth accident risk study is needed to reduce the risk of over-turn accident on SR-21. However, this is beyond the scope of this study.

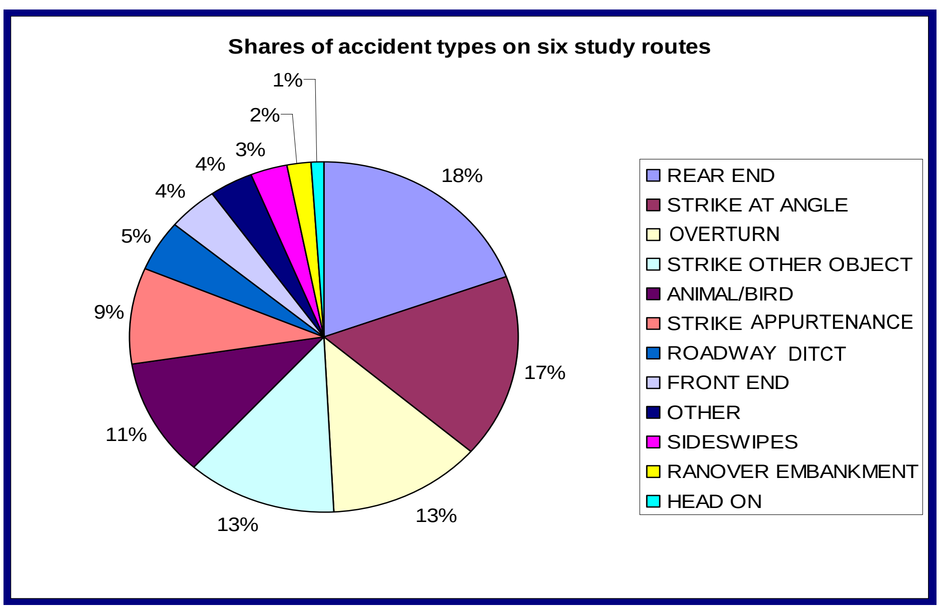

Figure 4-14 Shares of accident types on six study routes

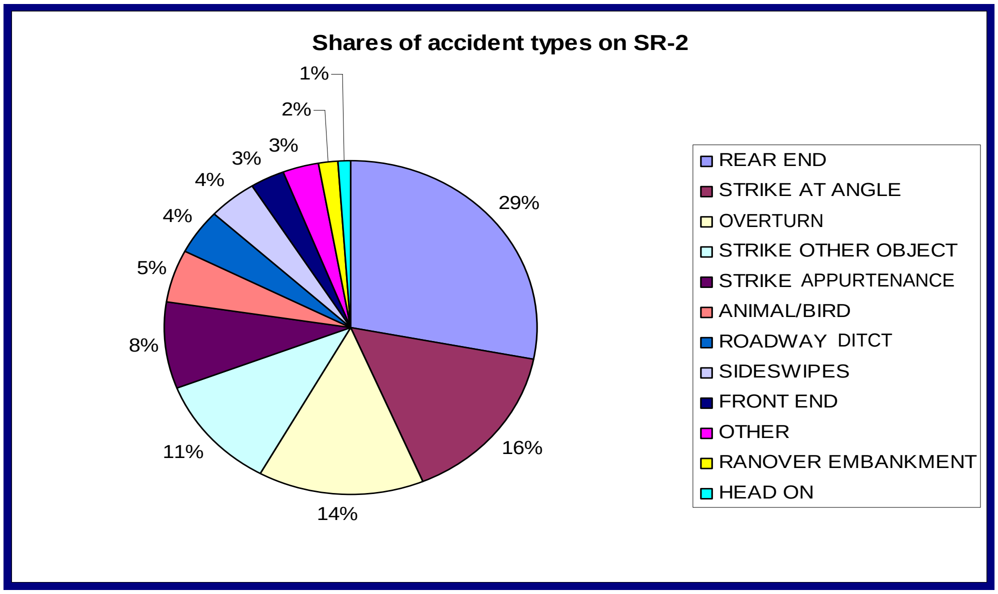

Figure 4-15 Shares of accident types on SR-2

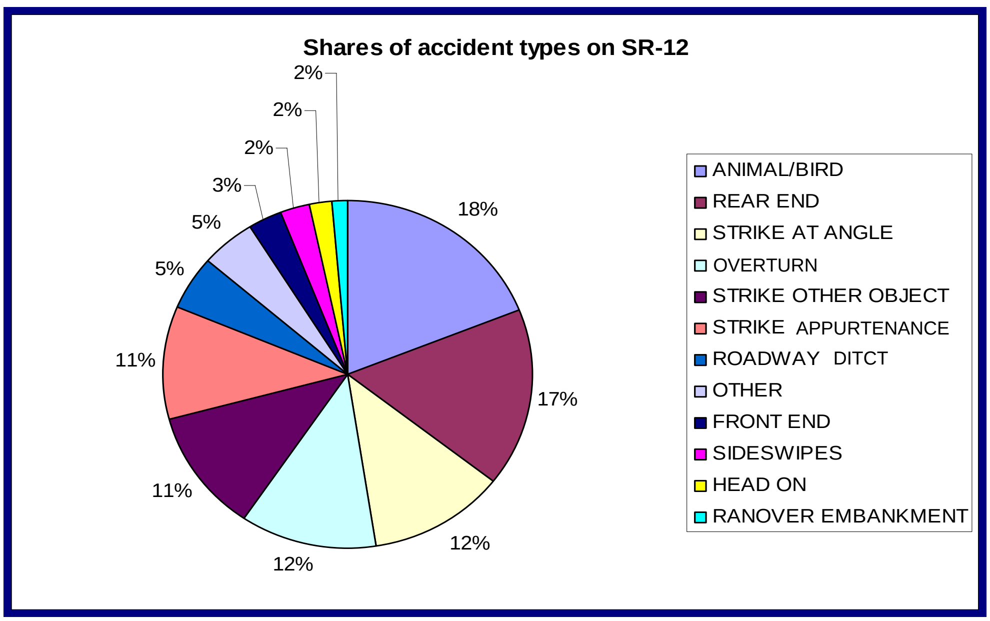

Figure 4-16 Shares of accident types on SR-12

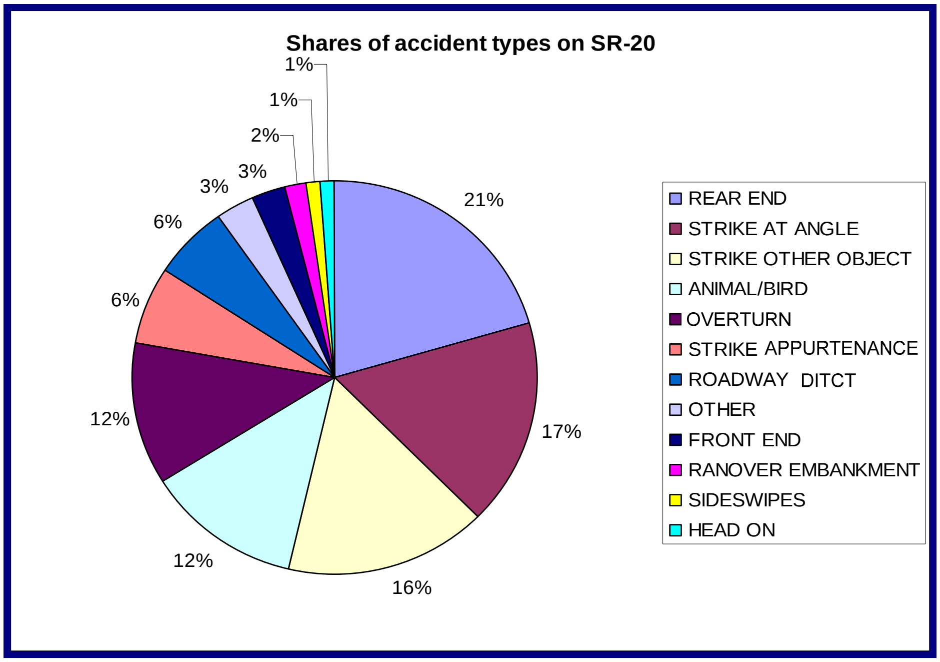

Figure 4-17 Shares of accident types on SR-20

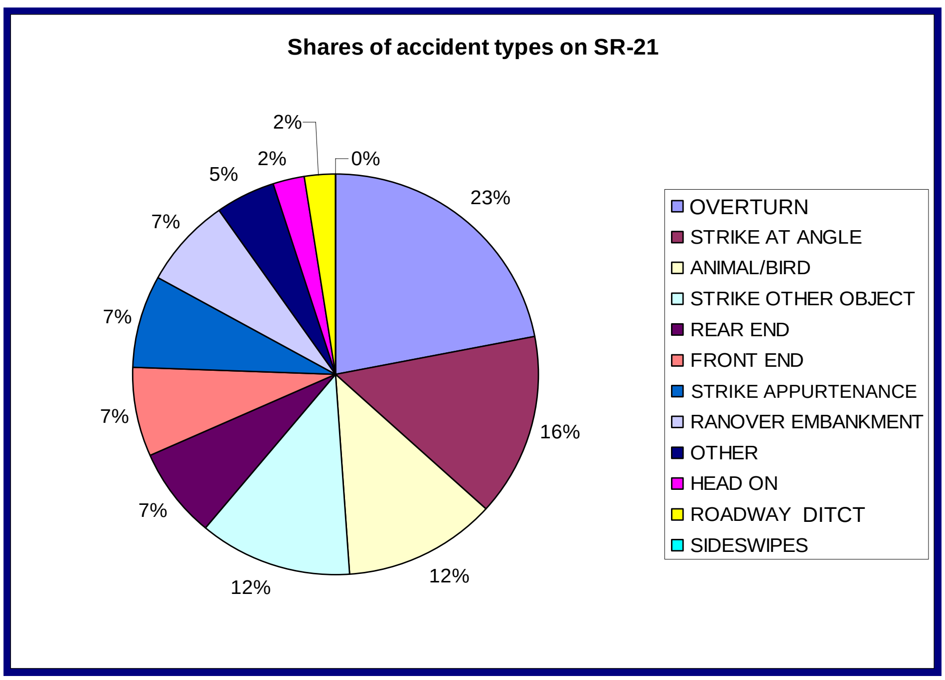

Figure 4-18 Shares of accident types on SR-21

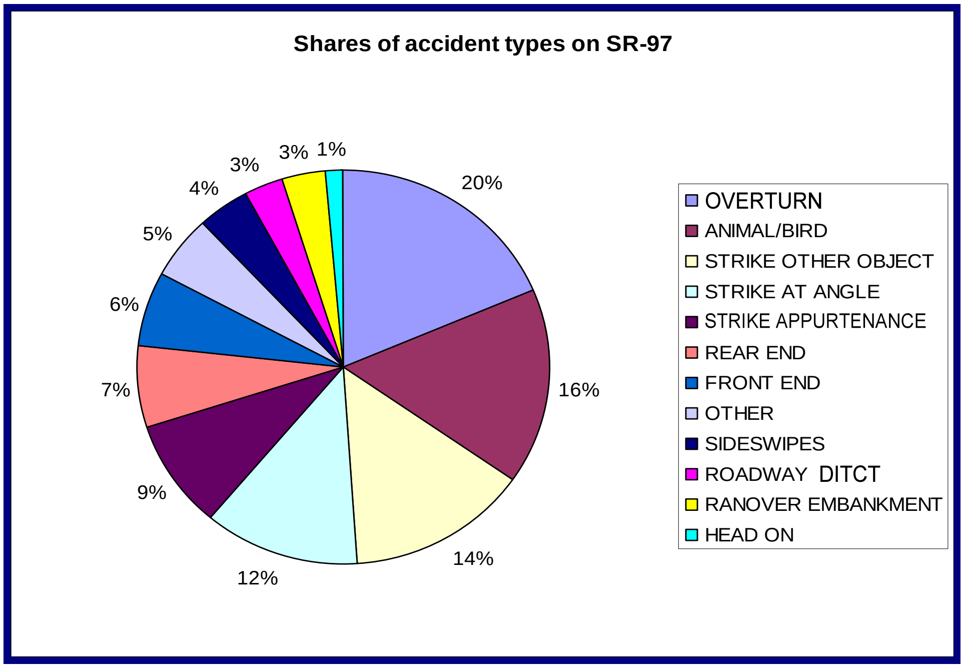

Figure 4-19 Shares of accident types on SR-97

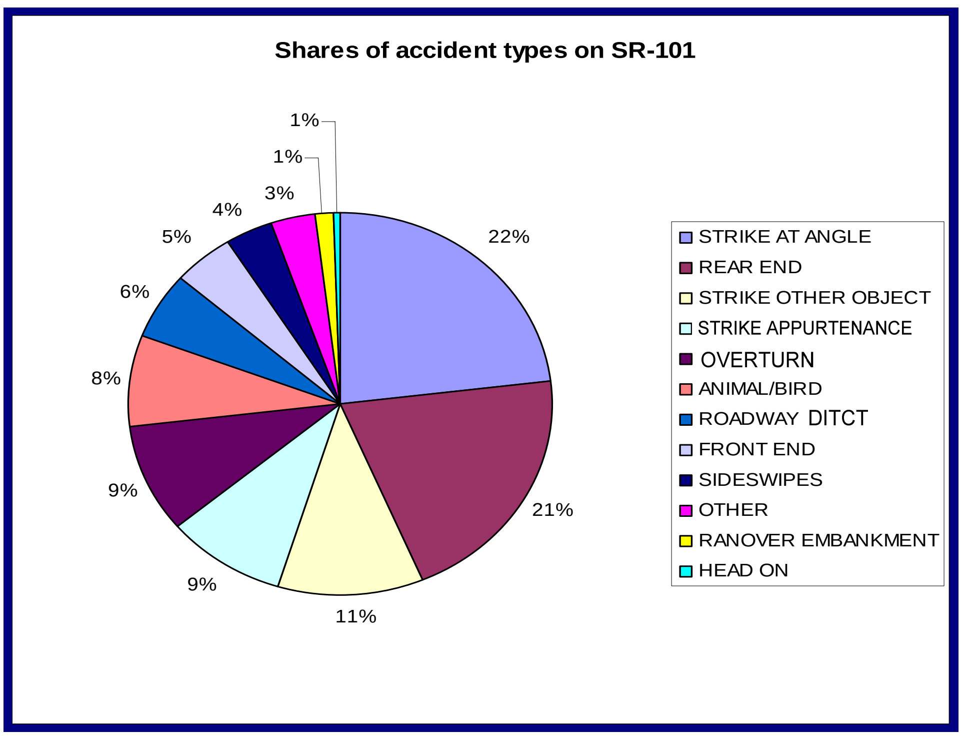

Figure 4-20 Shares of accident types on SR-101

As shown in Figure 4-21, among the six study routes, SR-97 and SR-20 have fairly high accident rate per intersection while SR-21 has the lowest accident rate. However, overturn accidents account for 23% of the total on SR-21 and this type of accident tends to be more severe than many other accident types.

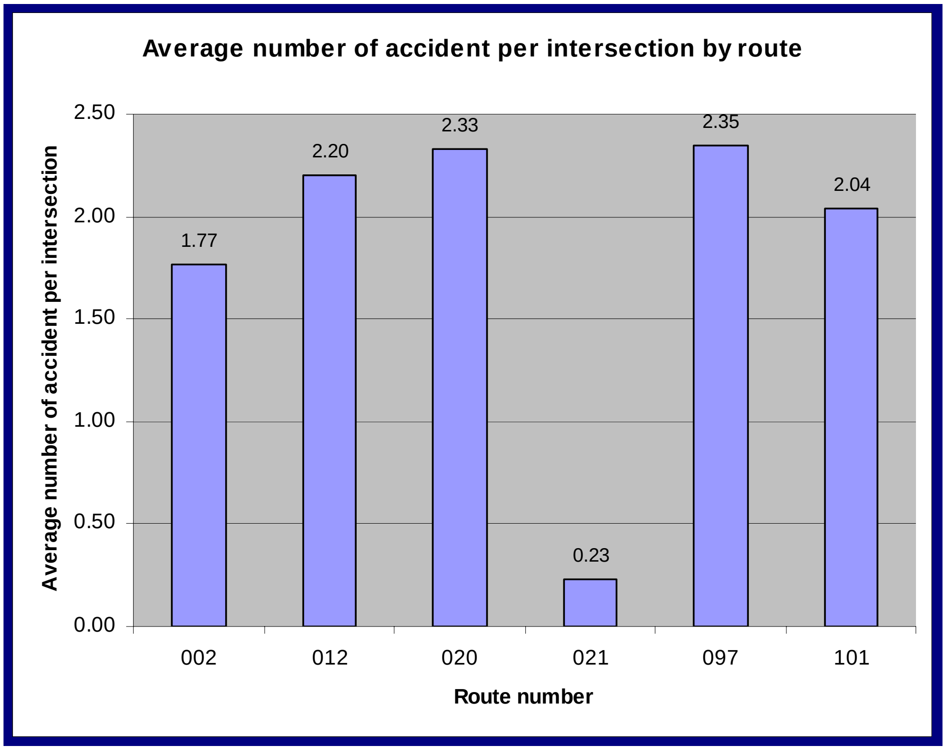

Figure 4-21 Average number of accidents per intersection by route

Figure 4-22 shows that 67.7% of accidents occurred in daylight and almost 19% occurred when it was dark and without streetlights. Only 7.8% of accidents occurred in dark at locations with streetlights on.

As shown in Figure 4-23, most accidents (63.17%) occur in clear or cloudy days. Nearly 13% of accidents occur in rainy days (12.96%), more than the 1.75% in foggy days. As mentioned in Section 4.1.1, when road surface changes from dry to wet, the friction coefficient between tire and the road surface drops significantly. As the friction coefficient decreases, the chance for a vehicle to get involved in an accident becomes higher because longer stopping distance is required. However, further analysis with weather information is needed to conclude if rainy days are more dangerous than dry days on the study routes.

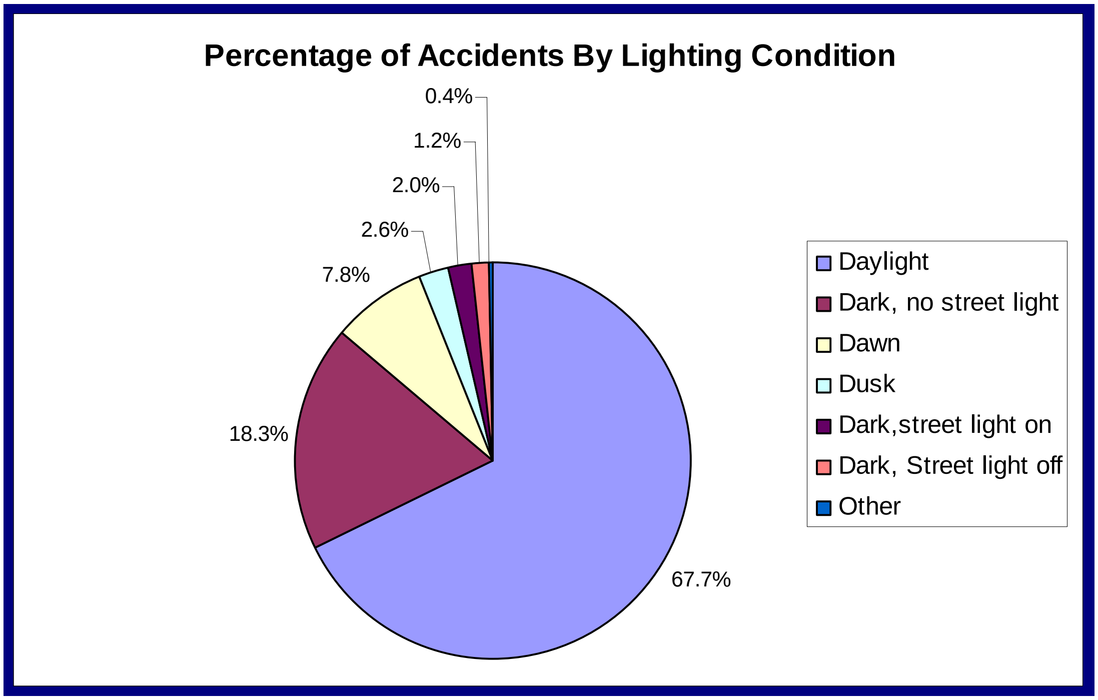

Figure 4-22 Percentage of reported accidents by lighting condition

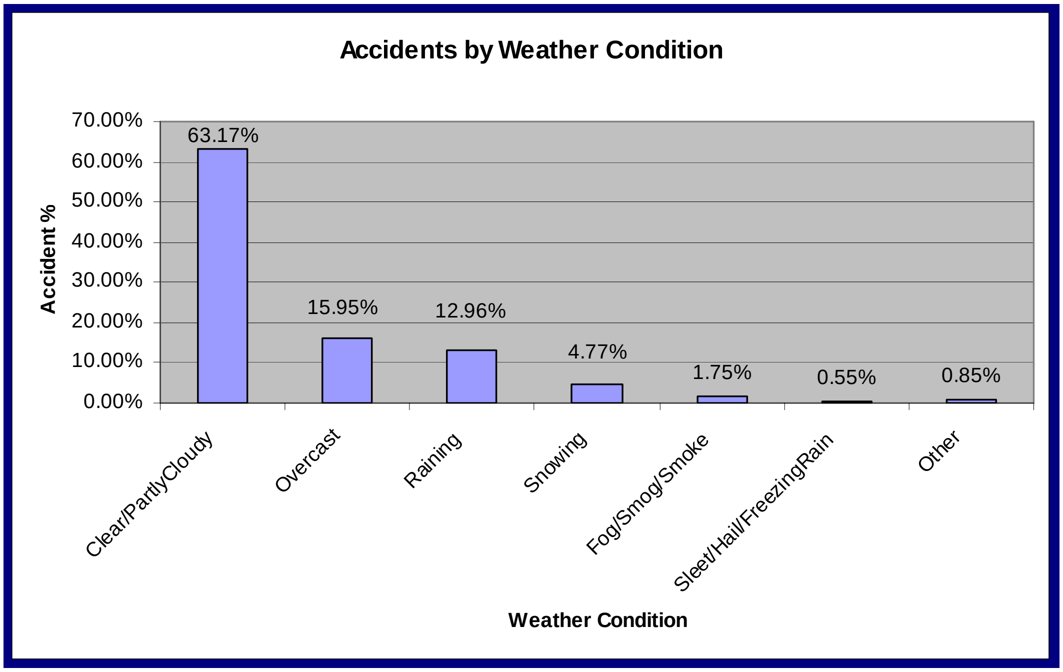

Figure 4-23 Percentage of reported accidents by weather condition

There is not a lot of variation in accident percentage among weekday as seen in Figure 4-24. There are slightly more accidents occuring on the three weekend days (Fridays, Saturdays, and Sundays). Fridays have 16.77% of total accidents, the highest among all days of week. A very noticeable difference between week day traffic and week end traffic is that there are fewer commuters in weekend traffic, assumed more drivers are not familiar with local traffic and roadway conditions. This may partly account for the higher accident rates over Fridays, Saturdays, and Sundays.

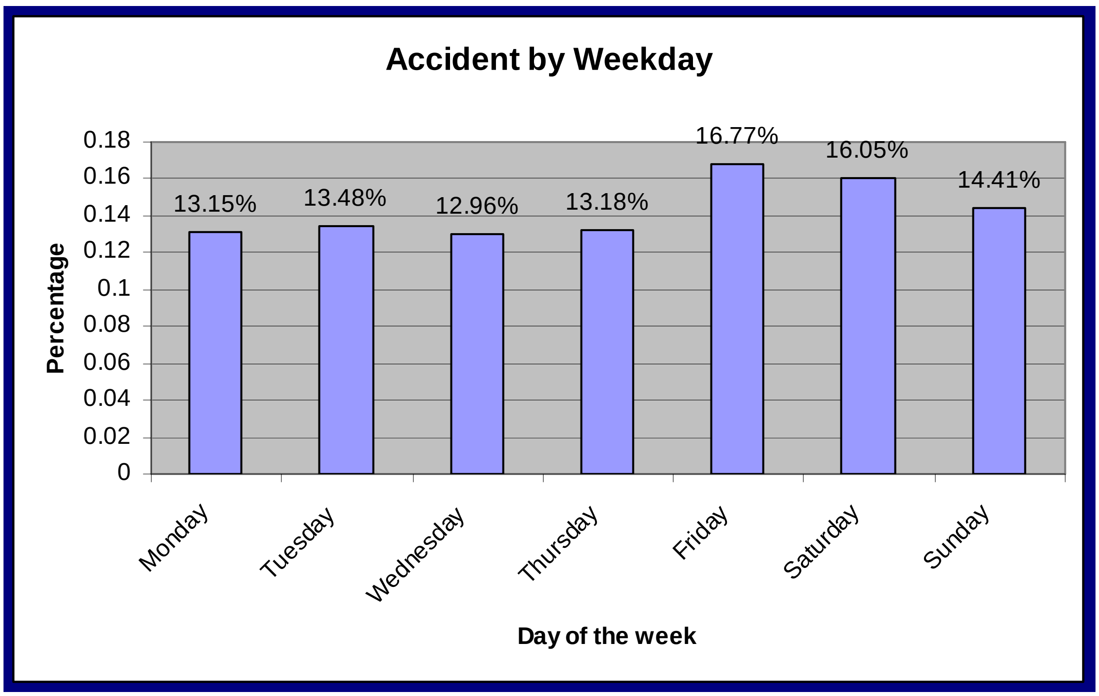

Figure 4-24 Percentage of reported accidents by weekday

Figure 4-25 shows the percentage of accidents for each month. July is the month with the highest percentage of accidents, followed by August. The explanation of the high accident percentages of the two summer months may also be related to the higher volume of site seeing traffic. During summer time, people would like to go out to rural areas more often for hiking, site seeing, etc; therefore there might be an increase in site seeing traffic volume as well as an increase of drivers unfamiliar with local traffic and roadway conditions. March and April are shown to be the two months with low shares of accidents compared to other months.

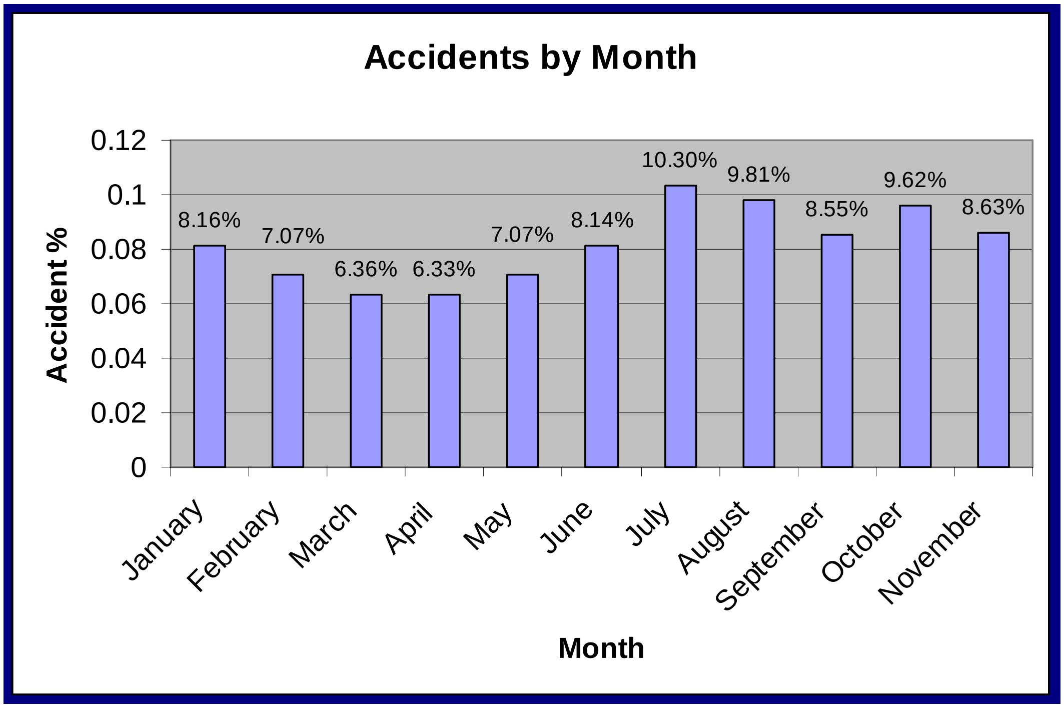

Figure 4-25 Percentage of reported accidents by month

Figure 4-26 shows the variation in accident frequency during the six year period. Number of accident decreased from 1999 to 2001, then went up in 2002 and 2003, and slightly decreased in 2004.

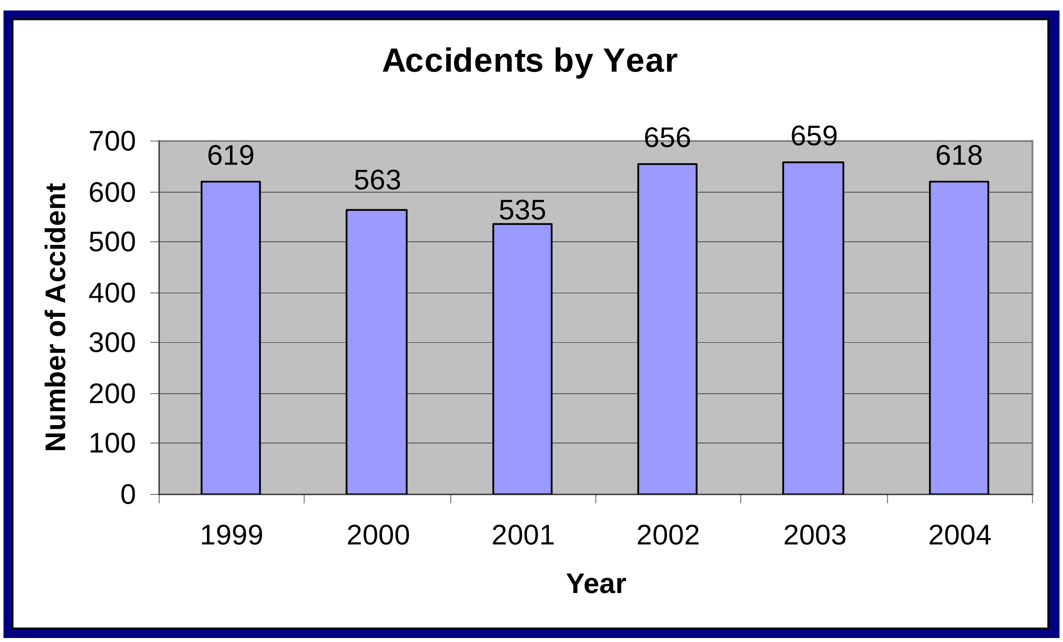

Figure 4-26 Number of reported accidents by year

4.2 Statistical Analysis

4.2.1 Roadway Segments

4.2.1.1 Tested Variables

Table 4-3 includes all the variables and their explanations for roadway segments of this study. Some of these variables are tested by the t-test and F-test to see if they have significant impacts on accident risk. They are also the explanatory variables used for accident risk modeling.

Table 4-3 Tested variables

Independent Variable	Type	Description	Dummy value
Totalt	Numeric	Total number of driveways	
PL	Dummy	Passing lane is present	0 for no; 1 for yes
Splim	Numeric	Speed limit (mph, in ten mph increments)	
curvrad	Numeric	Radius of curvature (feet, in 1000 foot increments)	
degcurv	Numeric	Degree of curvature (degree, in 10 degree increments)	
curvy	Dummy	Degree of curvature is less than 0.25 (radius=2290 ft)	0 for no; 1 for yes
bgrad	Numeric	The grade at a beginning mile post	
egrad	Numeric	The grade at an ending mile post	
mingrad	Numeric	The minimum grade percentage on a given roadway segment	
maxgrad	Numeric	The most extreme grade percentage on a given roadway segment	
mngrdum	Dummy	Minimum grade percentage is greater than 3%	0 for no; 1 for yes
mxgrdum	Dummy	Maximum grade percentage is greater than 6%	0 for no; 1 for yes
AvgGrad	Numeric	Average grade between a beginning and a ending milepost	

Table 4-3 Test variable (Continued)

Independent Variable	Type	Description	Dummy value
blshwd	Numeric	Left shoulder width at beginning mile post (feet, in 10 foot increments)	
elshwd	Numeric	Left shoulder width at ending mile post (feet, in 10 foot increments)	
brshwd	Numeric	Right shoulder width at beginning mile post (feet, in 10 foot increments)	
ershwd	Numeric	Right shoulder width at ending mile post (feet, in 10 foot increments)	
minshwid	Numeric	The minimum shoulder width on a roadway segment (feet, in 10 foot increments)	
bsrfwid	Numeric	Surface width at beginning mile post (feet, in 10 foot increments)	
esrfwid	Numeric	Surface width at ending mile post (feet, in 10 foot increments)	
minsurwd	Numeric	The minimum surface width on a roadway segment (feet, in 10 foot increments)	
SRFWDUM	Dummy	The minimum surface width is greater than 23 feet.	0 for no; 1 for yes
shasp	Dummy	Shoulder type is asphalt	0 for no; 1 for yes
shcurb	Dummy	Shoulder type is curb	0 for no; 1 for yes
shwall	Dummy	Wall at the roadside	0 for no; 1 for yes
shgravel	Dummy	Shoulder type is gravel	0 for no; 1 for yes
walcurb	Dummy	There is a curb or wall present	0 for no; 1 for yes
ShortSec	Dummy	The segment is shorter than 0.1 mile	0 for no; 1 for yes
SR2	Dummy	Roadway segment belong to SR-2	0 for no; 1 for yes
SR12	Dummy	Roadway segment belong to SR-12	0 for no; 1 for yes
SR20	Dummy	Roadway segment belong to SR-20	0 for no; 1 for yes
SR21	Dummy	Roadway segment belong to SR-21	0 for no; 1 for yes
SR97	Dummy	Roadway segment belong to SR-97	0 for no; 1 for yes

4.2.1.2 t-test

Table 4-4 describes the results of t-tests conducted for roadway segment accident rate. Variables statistically significant are marked in bold in the table. Accident rate was calculated for each roadway segment by dividing the number of accidents by the AADT (in thousand of vehicles) and the length of that roadway segment.

As is shown in Table 4-4, accident rate is lower for curvy segments than for straight segments. This may be attributed to more cautious driving and lower speed limit on curvy roadway segments. However, such impacts from multiple factors cannot be separated in t-test. The existence of passing lanes does not have significant impact on accident risk based on the t-test. According to the t-test results of grade dummy variables, *MNGRDUM* and *MXGRDUM*, the higher the grade, the more likely the accidents would occur. These results are highly significant (p=0.001 for MNGRDUM and p=0.004 for MXGRDUM).

Table 4-4 t-test results for roadway segments

Variable	*Groups*	*N*	*Mean Accident Rate*	*t-value*	*p-value*	*Significant at p=0.05*
Curvy	No	2385	4.308	3.275	0.001	**YES**
	Yes	3780	3.210			
PL	No	5726	3.643	0.309	0.758	**NO**
	Yes	439	3.527			
MNGRDUM	Grade less than or equal to 3%	4577	3.328	-3.251	0.001	**YES**
	Grade greater than 3%	1588	4.519			
MXGRDUM	Grade less than or equal to 6%	5491	3.460	-2.880	0.004	**YES**
	Grade greater than 6%	674	5.056			
SHCURB	No	6144	3.602	-1.987	0.061	**FAIRLY**
	Yes	21	13.402			
SHWALL	No	6152	3.633	-0.408	0.691	**NO**
	Yes	13	4.395			
ShortSec	No	3138	3.141	-2.863	0.004	**Yes**
	Yes	3027	4.146			

As for the impact from different types of shoulders, shoulders with curbs seem more dangerous. However, it is not significant at the p=0.05 level. In terms of the segment length, the average segment length is 0.2 mile for all test roadway segments. The t-test of the variable, *ShortSec,* shows the effect of segment length. The threshold to separate short and long segments is 0.1 mile. The t-test result indicates that a short segment tends to have a higher accident risk than a long segment. This may be due to the more frequent steering wheel adjustments required when driving on short segments.

4.2.1.3 ANOVA

Both one-way and two-way ANOVA are applied to several road features provided in the HSIS data, such as minimum curve radius, average curve radius, and average grade percentage. The ANOVA results show that three variables have significant impacts on accident rates; the average speed limit, curvature of roadway segments, and the gradient of roadway segments. As shown in Figure 4-27, roadway segments with a speed limit of 35 mph have significantly lower accident rates. On the other hand, roadway segments with a speed limit of 45 mph have the highest accident rate. The p-value of this one-way ANOVA test for consistent speed limit sections is close to zero, which indicates that the impact from this variable is highly significant.

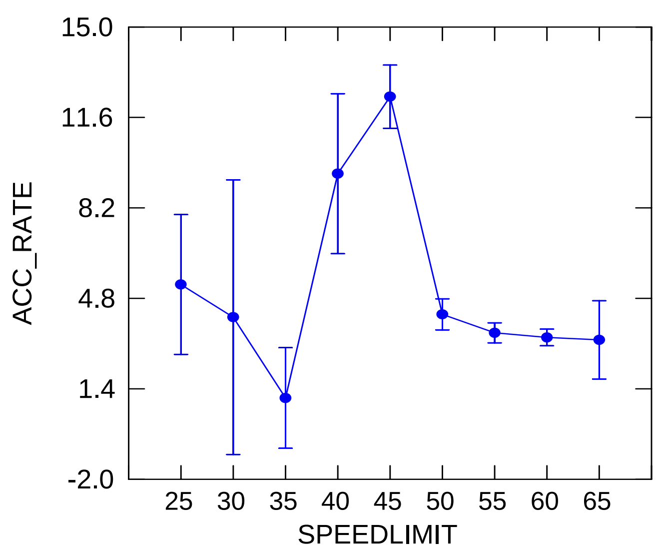

Figure 4-27 ANOVA test for effect of speed limit on accident rate

The curvature is further divided into three groups: curvy (the degree of curvature is greater than 2.5), less curvy (the degree of curvature is between 0~2.5) and straight. As shown in Figure 4-28, the less curvy segments seem to have a higher accident rate but the result is not statistically significant (F-ratio: 0.983, P-value: 0.374). Since road segments with different curvatures are typically associated with different speed limits, it would be interesting to explore the combination impacts of segment curvature and speed limit on accident rate. Therefore, a two-way ANOVA test was conducted to test the affect of the combination of speed limit and curvature on accident rate.

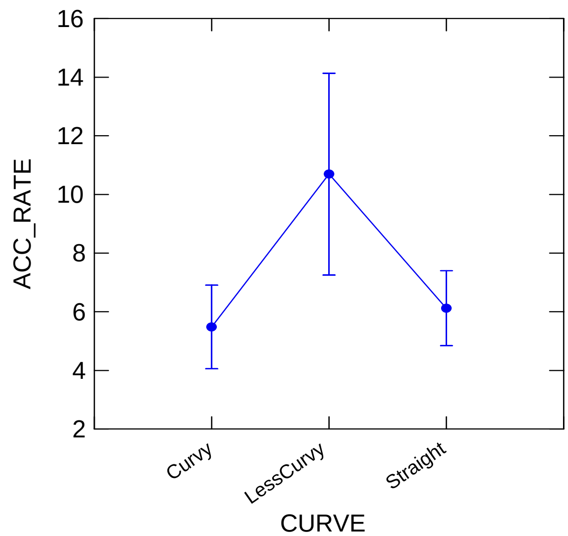

Figure 4-28 Accident rate on segments with different curvy levels

In two-way ANOVA analysis, different speed limits and curvature are compared and shown in Figure 4-29, Figure 4-30, and Figure 4-31. We can see that accident rate is relatively consistent over speed limits for curvy segments in Figure 4-29. For less curvy segments, accident rate is similar to those of curvy segments over most speed levels except that when speed limit is 45 mph. Figure 4-30 shows a peak accident rate at the 45 mph speed limit. The reason on why accident rate is so high at this speed level is unknown and may need further investigation. Similarly, we can see a peak of accident rate for straight segments when speed limit is 40 mph in Figure 4-31. Again, further investigation is needed to understand the reason of this observation. Due to the time constraint of this project, we are not able to address these two issues. The p-value of this test is approximately zero which indicates that the combination effect of these two variables is still highly significant.

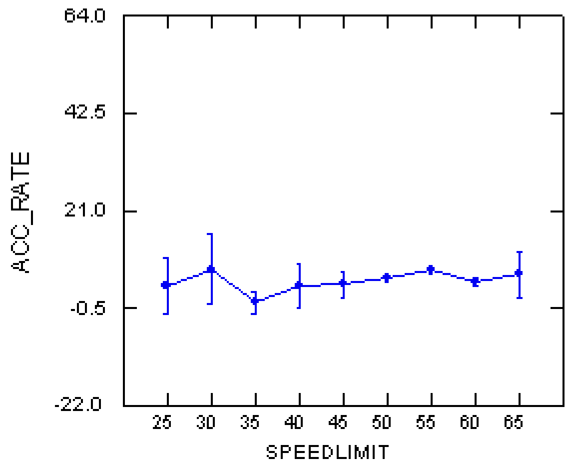

Figure 4-29 Accident rates on curvy segments with different speed limits

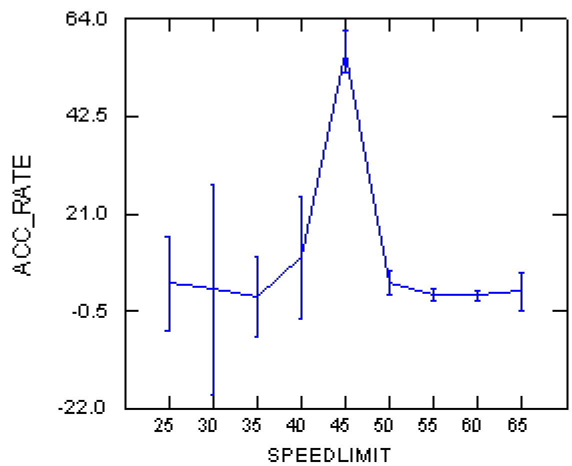

Figure 4-30 Accident rates on less curvy segments with different speed limits

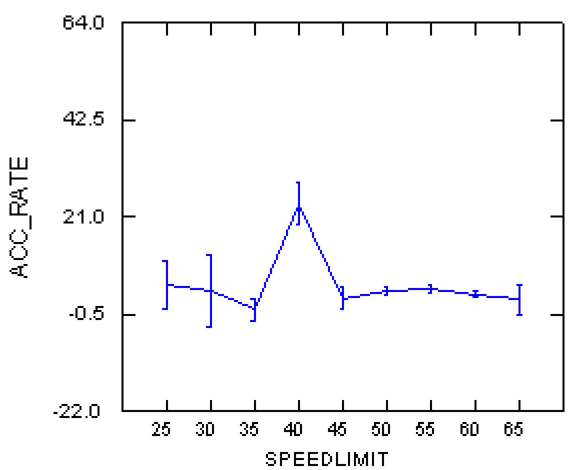

Figure 4-31 Accident rates on straight segments with different speed limits

The effect of speed limit changes on a roadway segment is also worth investigating in this research project. Most drivers are accustomed to driving on a roadway with a consistent speed limit. If speed limit changes over a roadway segment, traffic movement may be disturbed frequently because of the slowing down or speeding up actions. Also, frequent speed limit changes may increase accident potential if drivers miss the speed changing sign. Hence, a one-way ANOVA test for segments with speed limit changes is conducted to further investigate this issue. Any speed limit difference between the beginning and the end of the roadway segment is regarded as a speed limit change. As shown in Figure 4-32, the result shows that speed limit changes on curved sections are associated with more accidents.

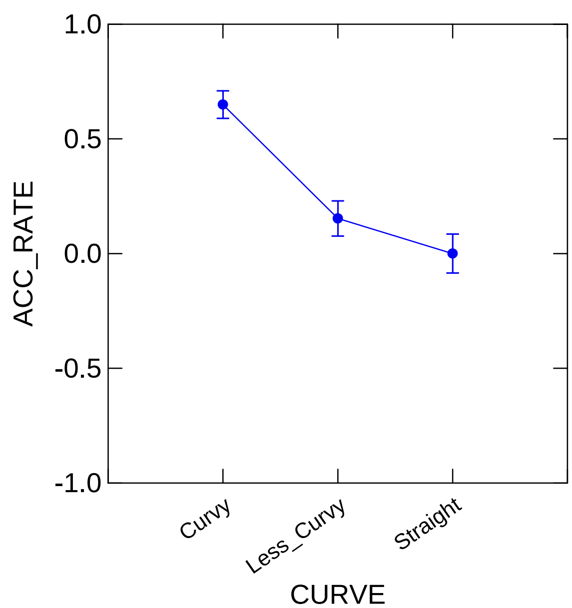

Figure 4-32 ANOVA test for the effect of speed limit changes on curved roadway segments on accident rate

The last ANOVA test for this section is on the effects of different grades of the roadway segments on accident rate. There are four control groups in this test. Group 1 is for grade percentage from 0% to 1%. Group 2 is for grade percentage from 1% to 2%. Group 3 is for grade percentage from 2% to 3%. Group 4 is for grade percentage greater than 3%. The result shows that the steeper the slope of the roadway segment, the higher the accident rate.

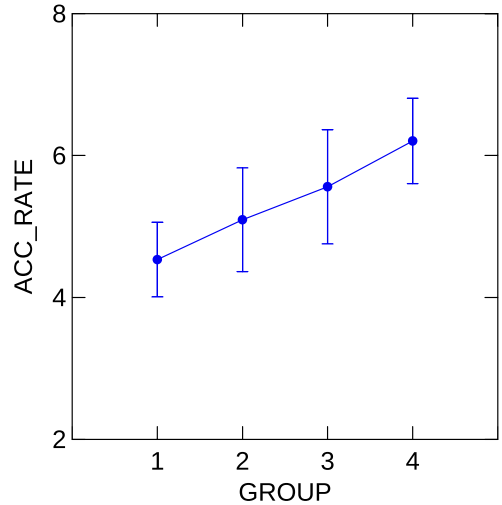

Figure 4-33 ANOVA test result for effect of gradation on accident rate

4.2.2 Intersections

4.2.2.1 Tested Variables

Table 4-5 includes all the variables and their explanations. Some of these variables will be tested by the t-test and ANOVA to see if they have significant impacts on intersection accident risk. They are also the explanatory variables used for intersection accident risk modeling.

Table 4-5 Tested variables

Independent Variable	Type	Description	Dummy value
Control	Dummy	Presence of traffic control	0 for no; 1 for yes
CurvConsist	Dummy	Consistency of directions of curvature	0 for not consistent; 1 for consistent
CurvStraight	Dummy	Curvedness of the intersection section	0 for curvy ;1 for straight
DegCurvA	Numeric	Degree of curvature at the intersection	
DegCurvB	Numeric	Degree of curvature at the beginning of intersection approach	
DegCurvE	Numeric	Degree of curvature at the end of intersection approach	
DiffSW	Dummy	Total absolute value of the difference in shoulder width between the two end of the intersection section and the intersection location	0 for zero value; 1 otherwise
Featillum	Dummy	Presence of artificial illumination at intersection	0 for no; 1 for yes
RadCurvA	Numeric	Radius of curvature at intersection scaled by 0.001	
RadCurvB	Numeric	Radius of curvature at the beginning of intersection approach scaled by 0.001	
RadCurvE	Numeric	Radius of curvature at the end of intersection approach scaled by 0.001	
WallA	Dummy	Presence of wall at end of intersection approach	0 for no; 1 for yes
WallB	Dummy	Presence of wall at the beginning of intersection approach	0 for no; 1 for yes
WallE	Dummy	Presence of wall at the end of intersection approach	0 for no; 1 for yes
CurbA	Dummy	Presence of curb at end of intersection approach	0 for no; 1 for yes
CurbB	Dummy	Presence of curb at the beginning of intersection approach	0 for no; 1 for yes
CurbE	Dummy	Presence of curb at intersection	0 for no; 1 for yes
SlopeChange	Numeric	Total absolute value of the difference in slope between the two ends of the intersection section scaled by 0.1	
SlopeFlat	Dummy	3 parts of the intersections are flat	1 for flat; 0 otherwise

Table 4-5 Tested variables (Continued)

Independent Variable	Type	Description	Dummy value
SlopedA	Dummy	Hilliness at the intersection	0 for slope less than or equal 5%; 1 otherwise
SlopedB	Dummy	Hilliness at the beginning of intersection approach	0 for slope less than or equal 5%; 1 otherwise
SlopedE	Dummy	Hilliness at the end of intersection approach	0 for slope less than or equal 5%; 1 otherwise
Splim	Numeric	Speed limit scaled by 0.1	
SR2	Dummy	Intersection section belong to SR-2	0 for no; 1 for yes
SR12	Dummy	Intersection section belong to SR-12	0 for no; 1 for yes
SR20	Dummy	Intersection section belong to SR-20	0 for no; 1 for yes
SR21	Dummy	Intersection section belong to SR-21	0 for no; 1 for yes
SR97	Dummy	Intersection section belong to SR-97	0 for no; 1 for yes
SWA	Numeric	Shoulder width at the intersection area scaled by 0.1	
SWB	Numeric	Shoulder width at the beginning of intersection approach area scaled by 0.1	
SWE	Numeric	Shoulder width at the intersection area scaled by 0.1	
T4leg	Dummy	Presence of T intersection or Four-leg intersection	0 for Four-leg intersection; 1 for T intersection

4.2.2.2 t-test

Table 4-6 describes the results of t-tests conducted in this study. Variables statistically significant are marked in bold in the table. Accident rates were calculated for each intersection approach by dividing the number of accidents by the AADT (in thousand vehicles) of that intersection approach.

Based on the t-test results, we can conclude that intersections with traffic control devices have higher accident rates than those without. This conclusion does not necessarily infer that traffic control devices make the intersections less safe. It is understandable that intersections with traffic control devices installed are the ones with a lot of human activities which would induce more traffic and human-traffic interactions. Those intersections are considered less safe compared to other intersections.

Table 4-6 t-test results for intersection accidents

Variable	Groups	N	Mean Accident Rate	t-value	p-value	Significant at p=0.05
Control	No	3648	2.140	-4.32	0.000	YES
	Yes	114	6.191			
CurvConsist	Not consistent	1200	2.460	1.865	0.062	FAIRLY
	Consistent	2521	2.160			
CurvStraight	Curvy	1513	2.423	1.862	0.063	FAIRLY
	Straight	2208	2.143			
DiffSW	Zero	3119	2.166	-2.458	0.014	YES
	Greater than zero	643	2.732			
SlopedB	Less than or equal to 5%	390	1.807	-2.067	0.039	YES
	Greater than 5%	3372	2.315			
SlopedE	Less than or equal to 5%	390	1.82	-1.995	0.047	YES
	Greater than 5%	3372	2.314			
SlopeFlat	No	3560	2.321	3.900	0.000	YES
	Yes	202	1.224			
SlopeVaried	No	2848	2.085	-3.322	0.001	YES
	Yes	914	2.817			
SWA	Less than or equal to 6 feet	2302	2.377	2.134	0.033	YES
	Greater than 6 feet	1460	2.082			
SWB	Less than or equal to 6 feet	2303	2.373	2.061	0.039	YES
	Greater than 6 feet	1459	2.088			

The *CurvConsist* variable is a binary dummy variable that describes the consistency of the curvature along the intersection approach. If the curvature for an intersection's inbound approach and outbound approach does not change, *CurvConsist* equals to 1. Otherwise, *CurvConsist* equals to 0. Though the p-value for this variable is 0.062 which is slightly higher than the significance level of 0.05, this variable is still included in Table 4-5 because it is controllable and can be applied for safety improvements. The *t-ratio* shows that accident rate is lower for consistent curvature, i.e. no curvature change from an inbound approach to its corresponding outbound approach (through movements) at intersections.

The *CurvStraight* variable is also a dummy variable. The value of 1 for this variable means that both its inbound approach and its through-movement outbound approach are on the same straight line. It takes the value of 0 otherwise. Though the significance level for this variable (0.063) is slightly higher than 0.05, this variable is also listed in Table 4-5 due to its controllability in practice. The t-test result indicates that accident rate is lower when driving through intersections on straight roadway segments than on curvy roadway segments.

The variable *DiffSW*, again a dummy variable, shows the difference in shoulder width throughout intersection approach. The p-value for this variable shows that it is significant at the 0.05 significance level. The t-test confirms that an intersection approach would be safer if there is no change in shoulder width throughout the intersection approach section.

SlopeB and *SlopeE* are the two dummy variables used to describe the hilliness conditions at the beginning and at the end of an intersection approach section, respectively. If the slope at the beginning or the end of the intersection approach section is less than 5%, *SlopeB* or *SlopeE* takes the value of 0. Otherwise, *SlopeB* or *SlopeE* has the value of 1. The p-values indicate that both variables have significant impacts on accident rate. The t-test results show that an intersection approach section has a lower accident rate if its slope at the beginning or the end is less than 5%.

Based on the p-values, both *SlopeFlat* and *SlopeVaried* have significant impacts on accident rate. *SlopeFlat* has the value of 1 if the slope throughout the intersection approach section is zero; it takes the value of 0 otherwise. The t-ratio shows that if the intersection approach section is flat from the beginning to the end of the section, the accident rate is lower. *SlopeVaried* has the value of 1 when the slope changes sign throughout the intersection approach section, which corresponds to a hilly condition of the road. The *t-ratio* shows that a hilly section has higher accident risk than a non-hilly one.

SWA and *SWB* are both dummy variables used to express if shoulder width at the stop bar or the beginning of an intersection approach is wider than 6 ft. *SWA* = 1 if the shoulder width at the stop bar of an intersection approach is greater than 6 ft. *SWA* = 0 otherwise. Similarly, *SWB* = 1 if the shoulder width at the beginning of an intersection approach is greater than 6 ft. *SWB* = 0 otherwise. The t-test results show that the wider the shoulder width is around the intersection or at the beginning of the intersection approach section, the lower the accident risk is.

4.2.2.3 ANOVA

Table 4-7 describes the basic information of variables included in the ANOVA test.

Table 4-7 Information of the variables used in F-test

Variable	Group 1 (A)	Group 2 (B)	Group 3 (C)	Group 4 (D)	N	DOF
RadCurvA	0-1000 feet	1000-1500 feet	1500-3000 feet	Greater than 3000 feet	3720	3
RadCurvB	0-1000 feet	1000-1500 feet	1500-3000 feet	Greater than 3000 feet	3720	3
RadCurvE	0-1000 feet	1000-1500 feet	1500-3000 feet	Greater than 3000 feet	3720	3
SlopeChange	Less than or equal to 2%	From 2%-4%	Greater than 4%		3762	2
Splim	Less than or equal to 30 mph	From 30-50 mph	Greater than 30 mph		3762	2

The results of ANOVA are summarized in Table 4-8. Only the variables, statistically significant, are included in this table.

Table 4-8 ANOVA results

Variable	Fvalue	F-crit	p-value	Significant at p=0.05
RadCurvA	8.737	2.606	0.000	YES
RadCurvE	4.818	2.606	0.000	YES
SlopeChange	10.067	2.999	0.000	YES
Splim	17.195	2.999	0.000	YES

The ANOVA results show that four variables have significant impacts on accident rate: the radius of curvature at the stop bar of an intersection approach (*RadCurvA*), the radius of curvature at the end of an intersection approach (*RadCurvE*), the change in slope between the beginning and end of an intersection approach section (*SlopeChange*), and the speed limit (*Splim*). As shown in Figure 4-34, the radii of curvature at the stop bar and at the end of intersection approach section have decreasing impacts on accident rate. The larger the radius of the curvature at the stop bar or at the end of intersection approach section, the less dangerous the intersection.

The change in slope from the beginning to the end of intersection approach section has an increasing impact on accident rate. The larger the change, the higher the accident rate. Speed limit also has an increasing impact on the accident rate. The higher the speed limit, the higher the accident rate for the intersection. Further analysis on how much impact these variables have on accident rates can be examined through accident risk models.

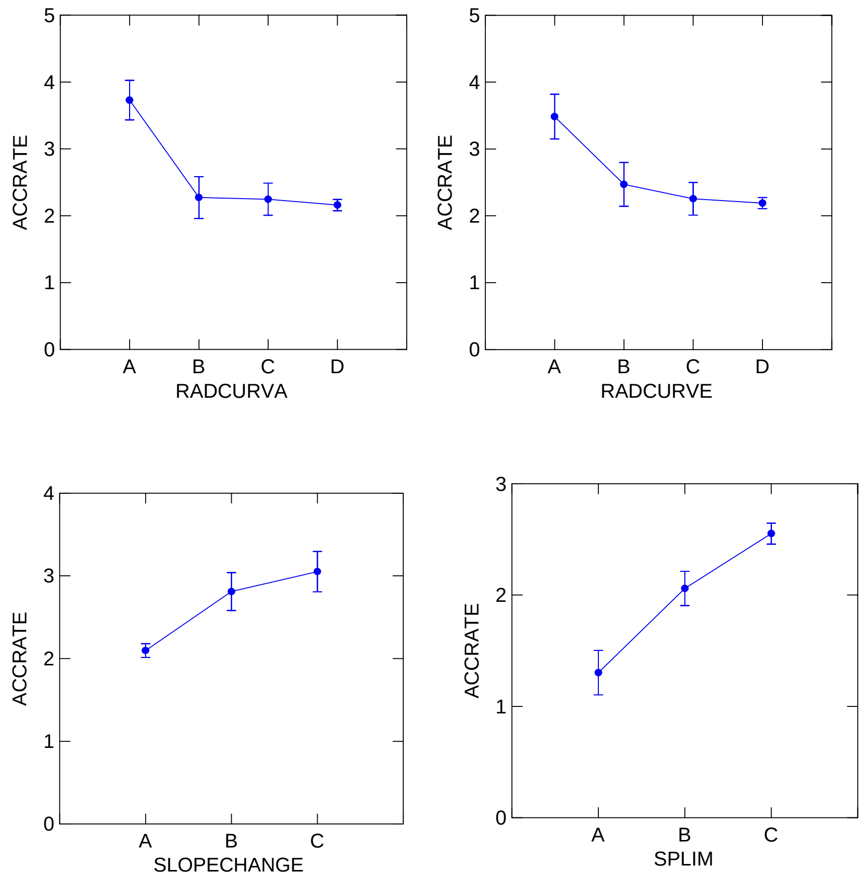

Figure 4-34 Impact of each variable on accident rate in F-test

CHAPTER 5: ACCIDENT RISK MODELING

5.1 INTRODUCTION

The software package used for this modeling process is SYSTAT (version 11). It is a powerful statistical tool that allows users to do analysis on univariate and multivariate data. The fact that SYSTAT allows users to manually define the loss function for MLE of a non-linear model provides a great flexibility for model calibration.

Due to the limitation of linear regression model as discussed in Chapter 1, more appropriate non-linear regression models, such as NB regression and Poisson regression, are investigated for accident risk modeling in this study. Several measures indicating goodness of fit measures are calculated to test how well a model fits the observed data.

5.2 ROADWAY SEGMENTS

5.2.1 Parameter Estimation for the All-Type Accident Risk Model

For roadway segment *i*, the expected number of all-type accidents is modeled in NB regression as

$$\lambda_i = (Vol_i)(L_i)EXP(\overline{\beta}\overline{X}_i + \varepsilon_i) \tag{5-1}$$

Where Vol_i is the six-year traffic volume total at segment *i*, L_i is the roadway segment length, the β-vector stores the coefficients to be estimated, and ε_i is the NB error term as described in Section 3.2.6. If traffic accident risk is defined as accidents per vehicle-mile traveled, then $EXP(\overline{\beta}\overline{X}_i + \varepsilon_i)$ actually models traffic accident risk. Traffic volume, Vol_i, is the total vehicle count over the six year period. It can be calculated from AADT.

The estimated over-dispersion parameter α is 1.26 with a t-statistic of 20.51, which indicates that the α value is significant at p=0.01 level. This implies that accident data used for this study is over-dispersed and the Poisson regression model should not be used.

Before the work proceeds, the appropriateness of NB, ZINB, and ZIP regression models has to be checked. A Vuong statistic value less than -1.96 would reject the ZINB or ZIP regression model and a value greater than 1.96 would recommend the use of the ZINB or ZIP regression model. The Vuong statistics for comparing NB and ZIP, and NB and ZINB result in inconclusive tests. Hence, it is concluded that the NB regression model makes a better fit than the ZINB or ZIP regression model. Consequently, the NB regression model is chosen for modeling all type accident risk for roadway segments in this study.

The NB regression results for all-type accident risk are listed in Table 5-1. The β coefficients and signs (negative or positive) represents how variables impact accident risk. For example, a negative coefficient indicates that the variable has a decreasing impact on accident risk when its value is increased.

Table 5-1 Negative binomial estimation results for roadway segment accident risk (all types)

Variable	Estimated Parameter	Standard error	t-statistic	P-value	Elasticity
Constant	-4.579	0.358	-12.801	0.000	-
SR12	0.267	0.058	4.569	0.000	0.229
SR20	-0.156	0.059	-2.627	0.008	-0.175
SR21	0.198	0.100	1.980	0.047	0.177
SR97	0.285	0.062	4.559	0.000	0.242
PL	-0.174	0.070	-2.495	0.012	-0.194
DEGCURV	0.520	0.040	12.917	0.000	0.157
MINGRAD	0.097	0.012	7.895	0.000	0.164
BRSHWD	-0.032	0.008	-3.807	0.000	-0.139
shcurb	1.208	0.326	3.711	0.000	0.700
shwall	1.131	0.435	2.602	0.009	0.674
BSRFWID	-0.062	0.015	-4.027	0.000	-1.410
ALPHA	1.258	0.061	20.563	0.000	-

The first four variables presented in Table 5-1 are binary dummy variables included to statistically capture the relative differences between the routes included in the study. Originally, a fifth dummy variable representing SR-2 was also included. However, it was not statistically significant and therefore it was removed from the final model estimation. According to the estimated values of the four dummy variables, the overall accident risk goes up for SR-12, SR-21, and SR-97 and drops down for SR-20. Therefore, the associated coefficients of these route-representing dummy variables can be important when prioritizing safety improvement plans amongst all state routes.

Since the degree of curvature for a roadway segment is reversely proportional to its radius, a short curvature radius corresponds to a high degree of curvature. The estimated coefficient for degree of curvature is statistically significant at 1% significance level. The positive value of the estimated parameter indicates that a roadway segment with sharper curves tends to have a higher risk for all type accidents according to the model. Readers may want to pay more attention to this variable since its t-statistic of 12.92 is the highest among all other variables in the model.

With the installation of passing lanes, reduction to all-type accident risk is expected. Another important result is the surface width variable included in the model. The t-statistic of -4.02 (p-value = 0.00) implies that this variable is a key component in the model. Its relatively high elasticity of -1.41 also indicates its importance in changing accident risk. Increasing surface width where possible is definitely an effective solution to safety problems on two-lane rural roads. However, this solution could be expensive economically due to the land acquisition and construction activities involved.

A wall or curb at the roadway shoulder is proven a significant variable at the 5% significance level (t-statistics of 3.71 and 2.60, respectively). Compared to other shoulder types (asphalt and gravel), a wall or curb along the shoulder limits a driver’s maneuverability and hence leads to higher accident risk. This suggests that a certain number of accidents could have been avoided if drivers’ maneuverability had not been constrained by these types of shoulders.

The impacts of increased shoulder width are minimal compared to that of either shoulder types or surface width. The “-” sign indicates that wider shoulders decrease accident risk. The interplay between the two variables representing the space to maneuver is very understandable, but the elasticity value of shoulder width is only a small portion of the elasticity value of surface width.

Last but not least, minimum grade is statistically significant even though the grade directions were not modeled separately. The elasticity value indicates that more extreme slopes account for higher accident risk. The t-statistic of 7.90 (p-value = 0.00) shows its importance in explaining all-type accident risk.

To measure the goodness of fit for this all-type accident model, three measurement coefficients were calculated: G^2=8988, ρ^2 =0.02, and $R_p^{\ 2}$ =0.51. The ρ^2 value is relatively low. This indicates that the model's explanation power is not high. Including more explanatory variables in the model should increase the explanation power. Variables reflecting skid-resistance, wheel path wear, polished aggregates, and even cross-slope are very likely to contribute significantly to the accuracy of the accident risk models. However, collecting such data will be very challenging.

5.2.2 Parameter Estimation for the Rear-End Accident Risk Model

Rear-end accident type has the highest percentage (21.07%) among all accident types on all SRs in the Washington State and is one of the major types on any of the six study routes (10%). As a result, in addition to all types of accidents, the type of rear-end accident is also modeled and estimated to identify accident causal factors. The NB model estimation results for rear-end accident risk are listed in Table 5-2.

Table 5-2 Negative binomial estimation results for rear-end accident risk

Variable	Estimated Parameter	Standard error	t-statistic	P-value	Elasticity
Constant	-3.602	0.844	-4.269	0.000	-
SR2	0.586	0.137	4.284	0.000	0.443
SR12	0.345	0.160	2.155	0.031	0.288
SR97	0.535	0.173	3.092	0.001	0.411
Totalt	0.967	0.251	3.857	0.000	0.036
PL	-0.353	0.199	-1.779	0.075	-0.026
SPDLIM	-0.656	0.073	-9.007	0.000	-3.587
BRSHWD	0.032	0.020	1.608	0.108	0.144
shcurb	1.028	0.634	1.621	0.105	0.639
shwall	2.775	0.882	3.146	0.002	0.936
BSRFWID	-0.077	0.034	-2.268	0.023	-1.757
ALPHA	0.556	0.073	7.590	0.000	-

The over-dispersion parameter α is estimated as 0.56 and statistically significant with a t-statistic of 7.59 (1.96 corresponds to the 95% confidence limit of the two-sided t-test). That indicates the appropriateness of using the negative binomial regression model in comparison with the Poisson regression model for modeling rear-end accident risk. The Vuong statistic of V=-0.02 results in an inconclusive test and therefore it fails to show a statistically better fit for the ZINB regression model to our data. Through a similar process, we also eliminate ZIP to be a good fit of the data. Therefore, the NB regression model is used as the final form.

Similar to those explained in the all-type accident risk model, three route-representing dummy variables are significant. According to the estimated coefficients of these variables, rear-end accident risk increases when SR-2, SR-12, and SR-97 are included in the analysis. Again, the associated coefficients can be important when prioritizing countermeasures against rear-end collisions amongst all the study routes.

The increasing effects of driveway density, *Totalt,* indicates that rear-end accidents are more frequent at or near driveways/intersections. Its elasticity, however, is relatively low (0.04). Nevertheless it is a very significant explanatory variable in the model with one of the relatively high t-statistics of 3.86 (p-value = 0.00).

Speed limit has decreasing effects on rear-end accident risk. This finding may not be surprising since roadway sections with high speed limits are normally associated with good vision, low conflicting movements, and consistent curvature. Although high speed limit also increases the required stopping sight distance which typically leads to higher accident risk, the dual impacts of speed limit cannot be reflected by the current risk model. The decreasing effects on rear-end accident risk reflects the net impact of speed limit. The high t-statistic for speed limit (-9.00) indicates the high significance compared with all other variables in the model.

The passing lane variable is nearly statistically significant at the p=0.05 level with a t-statistic of -1.78 (p-value = 0.08). Since this is a controllable variable for considering safety improvement plans, it is still included in the model. Adding a passing lane on a two-lane rural road is more cost-effective than upgrading to a four-lane road (with two lanes in each direction) and is therefore a viable option.

The impact of increased shoulder width is not as significant as the surface width variable or the passing lane variable. A possible reason for this is that drivers may drive faster than they should with a wide shoulder and therefore increases risk of rear-end collision.
A wall or curb at the roadway shoulder increases rear-end accident risk. The shoulder type of wall variable is statistically significant at the p=0.05. Shoulder type curb was quite close to significance at p=0.10 significance level with a t-statistic of 1.62. Compared to other shoulder types (asphalt and gravel), a wall or curb along the shoulder seems to limit driver's maneuverability in a potential accident situation. This suggests that a certain amount of rear-end accidents could have been avoided if the following vehicles were given some room to maneuver.

When surface width is increased, rear-end accident risk can be decreased. This finding is consistent with that of the all-type accident risk model and is easily acceptable. The elasticity (-1.757) of surface width is significantly higher than that of the shoulder width variable (0.144). The t-statistic of this variable is -2.27, which is significant at the p=0.05 level.

Statistics indicating the goodness of fit for this model are G^2=2443, ρ^2 =0.04, and R_p^2=0.79. Compared to the all-type accident model, this model has a high power of explanation. Nevertheless, there is still plenty of room to include more relevant variables to improve the goodness of fit for this rear-end accident risk model.

5.3 INTERSECTIONS

5.3.1 Parameter Estimation for the All-Type Accident Risk Model

Poisson regression is tried as the first step of this modeling process. Parameters estimated by the Poisson regression are used as the initial values of variables in NB regression. This helps a NB regression process converge sooner.

After running the NB regression for all accident types, the over-dispersion parameter α is found to be 1.27 with a t-statistic of 15.038 which is highly significant compared to the t-ratio of 1.96 at the 95% confidence level in a two-tailed t-test. This indicates accident data used for this modeling process is over-dispersed and NB regression is the right choice. The NB regression model for intersection accidents is expressed in the same form as that for roadway segment accidents shown in Equation (5-1). The statistical result for this NB regression model is presented in Table 5-3.

Table 5-3 Negative binomial modeling results for intersection accident risk (all types)

Variable	Estimated Parameter	Standard error	t-statistic	P-value	Elasticity
Constant	0.6	0.154	3.902	0.000	-
Control	1.018	0.116	8.745	0.000	0.64
SlopeChange	0.33	0.127	2.602	0.005	0.04
Splim	0.378	0.028	13.272	0.000	1.89
SR12	0.133	0.063	2.115	0.035	0.12
SR20	0.192	0.063	3.026	0.003	0.17
SWA	-0.397	0.092	-4.307	0.000	-0.20
DegCurvA	0.367	0.058	6.365	0.000	0.05
T4leg	-0.355	0.059	-5.997	0.000	-0.43
Featillum	0.159	0.062	2.538	0.011	0.15
Alpha	1.267	0.084	15.038	0.000	-

The model estimation results in nine significant explanatory variables. Speed limit (*Splim)* is the most significant variable as the corresponding t-statistic has the highest value (13.272) and also has the highest corresponding elasticity (1.89). The sign of this coefficient or the sign of the associated t-value shows that an intersection with higher speed limit (50 mph or higher) tend to have higher accident risk. This may be because of that vehicles traveling at a higher speed require a longer stopping distance that may not be available under certain conditions.

According to this all-type accident risk model, the *Control* variable also has a big impact on accident risk as indicated by the high value of the associated t-statistic (8.75). The model implies that an intersection with traffic control device is associated with higher accident rate than those without. This finding conflicts with our general thinking of installing traffic control devices to reduce moving conflicts and therefore traffic accident rate at intersections. However, previous studies (e.g. United States, 1995) did find that poorly designed traffic control plans increase accident risk. A closer investigation of the traffic control system at these intersections need to be carried out to find out whether the control systems are defective or malfunctioning. Of course, intersections that warranted signal installations are typically high volume or high risk locations. The fact that signalized intersections showed higher accident risk does not necessarily mean signal control introduces more accidents. To answer this question, a before-and-after analysis for signal installations is needed.

Similar to the *Control* variable, the *Featillum* variable also has a significant impact on accident risk as indicated by the relatively high value of the associated t-statistic (2.538). The positive sign of the estimated coefficient for this variable (0.159) shows that intersection approaches with artificial illumination are associated with higher all-type accident risk. This result does not make a lot of sense for the presence of artificial illumination is supposed to help improve the safety on the road. However, this result might infer that the intersection approaches with artificial illumination usually have more human activities which may result in more disturbances for traffic movements at the intersections.

SlopeChange with the coefficient of 0.33 is the variable that shows the difference in slope between the beginning and the end of an intersection approach section. The estimation result indicates that this variable also has a significant effect on accident risk. Though its elasticity is fairly low (0.04), the t-statistic for this variable is 2.602 indicating a high significance level. The statistical evidence about this variable shows that the higher the difference between the beginning and the end of the intersection approach section, the higher the accident rate. This implies that it is not safe to drive through an intersection approach with a high variance in slope.

The estimated coefficients and high t-statistic values of the two variables *SR-12* and *SR-20* indicate that SR-12 and SR-20 have higher accident rates than the base routes. One other important finding from the model is the significance of the degree of curvature. It is indicated in the model that the higher the degree of curvature at an intersection approach, the higher the accident rate. Though the associated elasticity is relatively low (0.05), the corresponding t-statistic (6.35) and the p-value (0.000) are statistical evidence showing a fairly strong impact of the degree of curvature on accident risk. A high value of the degree of curvature implies a low value of the radius of curvature. The smaller the radius of curvature, the sharper the curve is. The positive parameter of this variable (0.367) indicates that there are more accidents occurring on intersection approaches with sharper curves.

SWA and *T4Leg* are two variables with decreasing impacts on accident risk according to the estimated coefficients of -0.397 and -0.355, respectively. They both have a fairly high t-statistic which indicates a strong influence on accident rate. The *SWA* variable represents shoulder width at the stop bar of an intersection approach. Its t-statistic (-4.307) indicates that the wider the shoulder width is, the safer the intersection is. Though its elasticity (-0.2) is not too high, the significance of this variable needs to be seriously considered because the significance of this variable is at the 1% level. *T4Leg* is a dummy variable indicates whether the intersection is a T-intersection or a four-legged intersection. The estimated parameter of this variable (-0.355) points out that accident risk is lower to drive through a T- intersection than through a four-legged intersection. A relatively high value of the corresponding t-statistic (-5.997) shows that this variable has a strong effect on accident rate. This is a reasonable finding because at a four-legged intersection, traffic flows have more conflicting points than those at a three-legged one. More conflicting points tend to result in more collisions.

In order to check how well the intersection all-type accident risk model fits the observed data, several Goodness of Fit (GOF) statistics are calculated and summarized in Table 5-4:

Table 5-4 Goodness of fit value

Goodness Of Fit	Value
$LL(\beta)$	-4394.61
$LL(0)$	-4547.75
ρ^2	0.03
X^2	306.29
G^2	19260.91

The likelihood ratio test is frequently used to compare two models: the restricted one with all variable coefficients being zeros and the full, non-restricted model. The greater the likelihood ratio test statistic (X^2), the more explanatory power the model has. With the degree of freedom equal to the difference in the number of parameters between the two models which is nine in this case, the likelihood ratio test statistic is χ^2 distributed with the critical value of 16.92. The likelihood ratio test statistic value in this model is 306.29 which is much higher than the critical value; thus, the observed data is explained well by the predicted model. The ρ^2 statistic is another GOF measure. As discussed, the closer the ρ^2 statistic to 1, the more variance the model can explain and thus the better the model fits the observed data. In this case, the ρ^2 statistic is 0.03 which is much less than the value of 1. The sum of deviances, G^2, is the last GOF measure used in this analysis. The closer the G^2 is to zero, the better the model explains the real data. The G^2 value in this case is 19260.91 which is a very high value. These GOF measures indicate that the model does have certain explanatory power on two-lane rural road intersection accidents. Meanwhile, there is still plenty of room to improve the model. Further investigations on this accident risk model are needed. Traffic data from the crossing roads, human activity data, and detailed intersection layout data should be collected to support new research efforts on this model.

5.3.2 Parameter Estimation for the Strike-At-Angle Accident Risk Model

The NB regression for strike-at-angle accidents identifies the over-dispersion parameter α as 0.71 with a t-statistic of 7.929, which is highly significant compared to the critical t-ratio of 1.96 at the 95% confidence level in a two-tailed t-test. This indicates that accident data used for this modeling process is over-dispersed and NB regression is the right choice. The NB regression model takes the same form as that shown in Equation (5-1).

Estimation results for this NB regression model of strike-at-angle accidents are presented in Table 5-5.

Table 5-5 Negative binomial modeling results for intersection strike-at-angle accident risk

Variable	Estimated Parameter	Standard error	t-statistic	P-value	Elasticity
Constant	-0.392	0.256	-1.531	0.126	-
Control	1.135	0.168	6.769	0.000	0.68
Splim	0.331	0.049	6.763	0.000	1.65
SR2	-0.616	0.119	-5.187	0.000	-0.85
SWA	-0.346	0.162	-2.137	0.033	-0.18
T4leg	-0.895	0.098	-9.160	0.000	-1.45
DiffSW	0.176	0.114	1.542	0.123	0.16
Featillum	0.722	0.109	6.606	0.000	0.51
WallB	1.119	0.506	2.213	0.027	0.67
ALPHA	0.71	0.09	7.929	0.000	-

The model contains eight significant explanatory variables. Speed limit (*Splim*) is significant according to its t-statistic of 6.763. It also has the highest corresponding elasticity (1.65). This result implies that the higher the speed limit (50 mph or higher), the more likely a strike-at-angle accident happens.

Another significant variable in this strike-at-angle accident risk model is the *Control* variable. It has a high t-statistic (6.769) and a fairly high elasticity (0.68). The model implies that an intersection with traffic control device is more likely to have strike-at-angle accidents than those without. Similar to the finding in the all-type accident risk model, this finding conflicts with our general understanding of installing traffic control devices to reduce moving conflicts and therefore traffic accident rate at intersections. A close investigation of traffic control systems at these intersections needs to be carried out to find out what has caused the increased accident risk.

The *Featillum* variable is also significant in the strike-at-angle accident risk model. It has a high associated t-statistic (6.606). The positive sign of the estimated coefficient (0.722) shows that the presence of artificial illumination at an intersection approach is associated with a higher accident risk. Again, this result might infer that intersections with artificial illumination usually have more human activities which may result in more disturbances to traffic movements.

SWA and *T4Leg* are two variables with decreasing impacts on accident risk according to the corresponding parameters -0.346 and -0.895, respectively. The *SWA* variable gives the width of approach shoulder at the stop bar. Its t-statistic (-2.317) indicates that the wider the shoulder width is, the less likely the intersection has a strike-at-angle accident. *T4Leg* variable is a dummy variable indicates whether the intersection is a T-intersection or a four-legged intersection. It has the highest t-statistic (-9.160) pointing out that a T-intersection is much less likely to have a strike-at-angle accident than a four-legged intersection.

The variable *DiffSW* is a dummy variable showing whether the shoulder width changes from the inbound to the outbound of an intersection. This variable carries the value of 0 if there is no change and the value of 1 otherwise. Though the t-statistic is not as high as those for other variables, this variable is kept in the model because it is a controllable variable. The strike-at-angle accident rate is higher when there is a difference in shoulder width from the inbound to the outbound of an intersection approach.

WallB is another dummy variable indicating whether there is a wall along the shoulder at the inbound of an intersection approach. This variable is fairly significant as indicated by its t-statistic (2.213). The estimated coefficient (1.119) shows that there is an increase in strike-at-angle accident risk if a roadside wall presents. This result is easy to understand in that a wall by an intersection approach is a sight obstruction for drivers who want to pass through the intersection. They might not be able to see the incoming traffic towards them from other directions in time to avoid collisions. The *SR2* variable and its fairly high t-statistic indicates that SR-2 has much lower strike-at-angle accident risk than other state routes included in the study.

In order to check how well the predicting model fits the observed data, several Goodness of Fit (GOF) statistics are calculated and summarized in Table 5-6

Table 5-6 Goodness of fit value

Goodness Of Fit	Value
$LL(\beta)$	-1769.94
$LL(0)$	-1893.73
ρ^2	0.07
X^2	247.59
G^2	4014.95

The likelihood ratio test statistic value in this model is 247.59 which is much higher than the critical value of 15.51; thus, the observed data is explained well by the predicted model. The ρ^2 statistic of 0.07 is much less than the value of 1. However this ρ^2 statistic for this model is higher than the one in the all-type accident risk model. This risk model for intersection strike-at-angle accidents is well-explained than that for all-type accidents. The G^2 value in this case is 4014.95 which is a fairly high value. However, this value is still smaller than the one for all-type accident risk model, indicating a better explanation power for this model than the all-type accident risk model. As demonstrated by the improved explanation power in the strike-at-angle accident risk model, each specific type of accident has its own occurrence mechanism and therefore is better modeled separately. Further modeling investigations on this and other types of accidents are needed. Traffic data from the crossing roads, human activity data, and detailed intersection layout data should be collected and used to support new research efforts on such accident risk models.

CHAPTER 6: CONCLUSION AND RECOMMENDATION

6.1 CONCLUSIONS

6.1.1 Roadway Segments

The findings of this study provide an important first step to find cost-effective countermeasures against traffic accidents on two-lane rural roads. Through extensive modeling efforts, causal factors to two-lane rural road accidents are identified. The effects of controllable roadway design variables on all-type accident risk (AAR) or rear-end accident risk (RAR) have been quantitatively evaluated. These variables are summarized as follows:

- Passing lane has decreasing effects on both AAR and RAR
- Speed limit has decreasing effects on RAR only
- Degree of curvature has increasing effects on both AAR and RAR
- Grade percentage has increasing effects on AAR only
- Shoulder width has decreasing effects on AAR but increasing effect on RAR
- Roadside curb has increasing effects on both AAR and RAR
- Roadside wall has increasing effects on both AAR and RAR
- Surface width has decreasing effects on both AAR and RAR

Based on the results of modeling and statistical analysis, cost-effective measures that may be applied to reduce roadway segment accident risk are listed below:

- Avoid frequent speed limit changes along the curvy roadway segments.
- Warn drivers before they enter a curved or steep roadway segment since degree of curvature and grade have increasing effects on both AAR and RAR. Warning signs or other pavement-based warning techniques, such as pavement markers and rumble strips, can help reduce the risk.
- Widen the surface width and add an additional passing lane in high accident rate roadway segments.
- Widen shoulder width help reduce AAR but at the cost of increasing RAR.
- Remove roadside curbs and walls.

Furthermore, the elasticity values derived from the modeling results provide information for allocating limited resources to the most important factors in safety improvement projects. The accident risk models developed in this study can also help provide quantitative evaluations on safety improvement plans for two-lane rural roads in Washington State.

6.1.2 Intersections

Impacts from geometric factors, road environment, and traffic operational characteristics on intersection accident risk were investigated using statistical methods i.e. t-test, F-test and accident risk modeling. Accident risk models specific to two-lane rural road intersection collisions were developed for all-type accident frequency and strike-at-angle accident frequency. After exploring several possible regression models, including Poisson, ZIP, NB, and ZINB, NB model was found to be the best choice for modeling the data in this particular study.

Rear-end accidents were found to be the most frequent type of accident for five out of the six study routes. Rear-end accidents usually happen when the leading vehicles slow down or stop due to some disturbances and the following vehicles cannot react in time to avoid collision. A disturbance could be a red signal, a crossing pedestrian, a conflicting vehicle, or a running animal. Intersections are often areas with high rates of disturbance. In order to warn drivers that an intersection is approaching, more signage should be placed in a reasonable distance upstream of each intersection location.

Speed limit, consistency of curvature, curviness of the road, slope of the road, hilliness of the road, shoulder width, and degree of curvature are the factors that have significant impacts on the accident frequency as analyzed through the t-test and F-test. The all-type accident risk model gives similar results. Speed limit, degree of curvature, change in slope between the inbound and the outbound of an intersection approach have increasing impacts on accident risk. On the opposite, shoulder width has a decreasing impact on accident risk.

In the strike-at-angle accident risk model, speed limit, whether shoulder width is consistent through the intersection approach section, and presence of wall at the inbound of an intersection approach have increasing impacts on the strike-at-angle accident risk. Similar to all-type accident risk model, shoulder width has decreasing impact on the accident frequency in the strike-at-angle accident risk model.

Based on the analysis results, cost-effective measures that may be applied to reduce intersection accident risk are listed below in an order from the least expensive to the most expensive:

- Lower speed limit at intersection approaches.
- Put more signs upstream of intersection to make drivers aware of the presence of intersection.
- Remove wall(s) at the inbounds of intersections.
- Increase shoulder width (greater than 6 feet) of intersection approaches.
- Keep shoulder widths consistent along intersection sections.
- Decrease the degree of curvature at intersections.
- Minimize the change in slope between the inbound and outbound of an intersection.

6.2 RECOMMENDATIONS

6.2.1 Roadway Segments

In terms of future work, more samples and more variables, such as driver behavior factors and other roadway design variables, should be included in the modeling process whenever possible. In addition to the variables included in the HSIS data, there is a potential to add more meaningful regressors to the models. Data on skid-resistance, wheel path wear, and polished aggregates are very likely to contribute to the accuracy of the accident risk models. Polished aggregates lead to reduced friction between tires and pavement. Rutting can result in standing water in roadway that may cause potential hydroplaning.

Also, in terms of further studies, it would be interesting to see GIS software incorporated in the field of two-lane rural safety to illustrate high accident risk sites graphically. Segmenting highway sections based on both horizontal and vertical curves may also improve statistical and modeling results.

6.2.2 Intersections

According to the result from the models, intersections with traffic control devices or artificial illumination have more accidents than those without. Although intersections with traffic control or illumination devices are typically associated with higher human activities which is more likely to result in traffic interruptions, this result is still very questionable. Therefore, further studies on these factors using before and after data are desirable.

6.2.3 Modeling Approach

NB regression model fits two-lane accident data better than Poisson, ZIP, and ZINB regression models. It proved to be the correct choice for all the four accident risk models developed in this study and therefore may be considered for future modeling work of two-lane rural road accidents.

REFERENCES

1. Agent, R. and Pigman, J. (2001), "Safety Impacts of Rural Road Construction," Kentucky Transportation Center, College of Engineering, University of Kentucky, Lexington, KY

2. American Association of State Highway and Transportation Officials (AASHTO). (2004). A policy on geometric design of highways and streets: 2004. Washington, D.C.: American Association of State Highway and Transportation Officials.

3. Banihashemi, M. and M. Dimaiuta (2005), "Maximizing Safety Improvement Benefits in Crash Prediction Models with Accident Modification Factors (AMFs)," TRB 2005 Annual Meeting CD-ROM

4. Calvert, James (2004), "Degree of Curvature", University of Denver, Department of mathematics (Available: http://www.du.edu/~jcalvert/railway/degcurv.htm)

5. Cameron, C and P. Trevedi (1990). Regression based tests for overdispersion in the Poisson model. Journal of Econometrics 46, 347-364

6. Chayanan, S., M. Nebergall, V. Shankar, N. Juvva, and Y. Ouyang (2003), "Interaction between the Roadway and Roadside – An Economical Analysis of Design and Environmental Factors affecting Segment Accident Rates," WSDOT, WA-RD 562.1

7. Council, F. M., & William, C. D. (2006). The Highway Safety Information System guidebook for the Washington State data files. Washington, DC: U.S. Dept. of Transportation, Federal Highway Administration.

8. Federal Highway Administration. (2007, February 23). Safety for intersections overviewed by the USDOT Federal Highway Administration. Retrieved July 28, 207 from the World Wide Web: http://safety.fhwa.dot.gov/intersections/index.htm

9. Fitzpatrick, K., Parham, A. H., & Brewer, M. (2002). Treatments for crashes on rural two-lane highways in Texas. College Station: Texas Transportation Institute, Texas A&M University System

10. Garcia-Molina, H., J.D. Ullman, and J. Widom. (2002). Database Systems: The Complete Book. Prentice-Hall, Inc. Upper Saddle River, New Jersey.

11. Gårder, P. (2005), "Segment Characteristics and Severity of Head-on crashes on Two-Lane Rural Highways in Maine," Department of Civil and Environmental Engineering, University of Maine, Accident Analysis and Prevention, v38, 652-61.

12. Geurts, K., Wets, G., Brijs, T, Vanhoof, K. And Karlis, D (2005), "Ranking and selecting dangerous crash locations: Correcting for the Number of Passengers and Bayesian Ranking Plots," Transportation Research Institute, Hasselt University, Agoralaan, Belgium and University of Economics and Business, Department of Statistics, Athens Greece

13. Greene, W (2003), "Econometric Analysis," 5th edition, Prentice Hall, Upper Saddle River, New Jersey

14. Hickey, J (1997). Shoulder Rumble Strip Effectiveness. Transportation Research Record 1573. Transportation Research Board. Washington D.C.

15. Ivan, John N, Garder, Per E., Deng, Zuxuan and Zhang, Chen (2006), "The Effect of Segment Characteristics on the Severity of Head-on-Crashes on Two-Lane Rural Highways," Connecticut Transportation Institute and University of Maine Department of Civil Engineering

16. Jones and Childers (2001), "The Friction of Automobile Tires," Contemporary College Physics, University of South Carolina (Available: http://www.physics.sc.edu/~rjones/phys101/tirefriction.html)

17. Kell, J. H., Fullerton, I. J., & Mills, M. K. (1990). Traffic detector handbook. McLean, Va: Federal Highway Administration.

18. Kim, J., Y. Wang, and G. F. Ulfarsson. (2007). Modeling the Probability of Freeway Rear-End Crash Occurrence. ASCE Journal of the Transportation Engineering, Vol. 133 (1), 11-19.

19. Lamm, R, B. Psarianos, and S. Cafiso (2002). Safety Evaluation Process for Two-Lane Rural Roads: A 10-Year Review. Transportation Research Record 1796. 51-59. Washington D.C.

20. Lee J, & Mannering F. (2002). Impact of roadside features on the frequency and severity of run-off-roadway accidents: an empirical analysis. Accident; Analysis and Prevention. 34(2), 149-61.

21. Levy, A. (2007). Cost-effective safety improvements for two-lane rural roads in Washington State. Thesis (M.S.C.E.)--University of Washington, 2007

22. Miaou SP, Hu P, Wright T, Rathi A, and Davis S (1992), "Relationships Between Truck Accidents and Highway Geometric Design: A Poisson Regression Approach,"

Presented at the Seventy-first Annual Transportation Research Board Meeting, Washington D.C.

23. Miaou, Shaw-Pin and Harry Lum (1993), "Modeling Vehicle Accidents and Highway Geometric Design Relationships," Accident Analysis and Prevention, v25 (6), 689-709

24. National Highway Traffic Safety Administration (2004). Traffic Safety Facts 2003. U.S. Department of Transportation. Washington D.C.

25. National Safety Council (NSC), Estimating the Costs of Unintentional Injuries 2005. Itasca, ILL National Safety Council, Statistics Department, 2006.

26. Ogden KW. (1997). The effects of paved shoulders on accidents on rural highways. Accident; Analysis and Prevention. 29(3), 353-62.

27. Okamoto, H. and Koshi, M. (1989), "A Method to Cope with Random Errors of Observed Accidents Rates in Regression Analysis," Accident Analysis Prevention v21 (4), 317-32

28. Olson, D. and Glad, R. (2004). Cost Effective Safety Improvements for 2-Lane Rural Roads. Research statement for 2005-07 biennium research program at Washington State Department of Transportation. Olympia, Washington.

29. Persaud, B. N., & Griffith, M. S. (2001). Statistical methods in highway safety analysis: a synthesis of highway practice. Washington, D.C.: National Academy Press.

30. Persaud, B. N., Retting, R. A., & Lyon, C. A. (2004). Crash reduction following installation of centerline rumble strips on rural two-lane roads. Accident Analysis and Prevention. 36(6), 1073.

31. Pline, J. L. (1992). Traffic engineering handbook. Englewood Cliffs, N.J.: Prentice-Hall.

32. Poch, M. and Mannering, F. (1996), "Negative Binomial Analysis of Intersection-Accident Frequencies," Journal of Transportation Engineering v122 (2), 105-13

33. Polus, A., & Mattar-Habib, C. (2004). New Consistency Model for Rural Highways and Its Relationship to Safety. Journal of Transportation Engineering. 130(3), 286.

34. Shankar, V., J. Milton, and F. Mannering (1997). Modeling accident frequencies as zero-altered probability processes: an empirical inquiry, Accident Analysis and Prevention, 29(6), 829-837.

35. Shankar, V., Mannering, F., and Barfield, W. (1995), "Effect of Roadway Geometrics and Environmental Factors on Rural Freeway Accident Frequencies," Accident Analysis and Prevention, v27 (3), 371-89

36. Shankar, V., Mannering, F., and Barfield, W. (1995), "Effect of Roadway Geometrics and Environmental Factors on Rural Freeway Accident Frequencies," Accident Analysis and Prevention, v27 (3), 371-89

37. Tsyganov, A., Machemehl, R. and Warrenchuk, N. (2005), "Safety Impact of Edge Lines on Rural Two-Lane Highways," Center for Transportation Research, University of Texas, Austin, TX

38. TRB (Transportation Research Board). (2000).Highway Capacity Manual. TRB, National Research Council, Washington, D.C.,

39. United States. (1995). Improving traffic signal operations: a primer. Washington, D.C.: Institute of Transportation Engineers.

40. Vogt, A., & Bared, J. G. (1998). Accident models for two-lane rural roads segments and intersections. McLean, VA: U.S. Dept. of Transportation, Federal Highway Administration, Research and Development, Turner-Fairbank Highway Research Center.

41. Vuong, Q. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica*, v57, 307-334.

42. Wang, Y (1998). Modeling Vehicle-to-Vehicle Accident Risks Considering the Occurrence Mechanism at Four-Legged Signalized Intersections. Ph.D. Dissertation, the University of Tokyo

43. Wang, Y (1998). Modeling Vehicle-to-Vehicle Accident Risks Considering the Occurrence Mechanism at Four-Legged Signalized Intersections. Ph.D. Dissertation, the University of Tokyo.

44. Wang, Y. and H. Ieda (1999). Using Accident Observations to Evaluate Rear End Accident Risk at Four-Legged Signalized Intersections. World Transport Research, Selected Proceedings of the 8th World Conference on Transport Research, Vol. 2, 123-136.

45. Wang, Y., and N.L. Nihan (2003). Quantitative analysis on angle-accident risk at signalized intersections. World Transport Research, Selected Proceedings of the 9th World Conference on Transport Research. Pergamon Press. Seoul, Korea

46. Wang, Y., Ieda, H. and Mannering, F. (2003), "Estimating Rear-End Accident Probabilities at Signalized Intersections: Occurrence-Mechanism Approach," Journal of Transportation Engineering, v129 (4), 377-84

47. Washington State Department of Transportation (WSDOT) (2004), "2004 Annual State Highway Collision Data Summary," (Available: http://www.wsdot.wa.gov/mapsdata/tdo/)

48. Washington State Department of Transportation (WSDOT) (2005), "2005 Annual State Highway Collision Data Summary," (Available: http://www.wsdot.wa.gov/mapsdata/tdo/)

49. Washington, S. (2005). Validation of accident models for intersections. McLean, Va: Turner-Fairbank Highway Research Center.

50. Washington, S., K. Dixon, D. White, and C.E. Wu (2002). Investigation of Fatal Motor Vehicle Crashes on Two-Lane Rural Highways in Georgia. Research Report FHWA-GA-02-9905 for Geogia Departmet of Transportation. Forest Park. Geogia.

51. Washington, S., Karlaftis, M. G., & Mannering, F. L. (2003). Statistical and econometric methods for transportation data analysis. Boca Raton: Chapman & Hall/CRC.

52. Yuan, F. (2000). Estimating benefits from specific highway safety improvements. Phase II, Safety benefits of intersection approach realignment: final report. Storrs, Conn: Connecticut Transportation Institute, School of Engineering, University of Connecticut.