2009

Study on crash characteristics and injury severity at roadway work zones

Qing Wang
University of South Florida

Study On Crash Characteristics And Injury Severity At Roadway Work Zones

by

Qing Wang

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Science in Civil Engineering
Department of Civil & Environmental Engineering
College of Engineering
University of South Florida

Major Professor: Jian Lu, Ph.D.
Abdul Pinjari, Ph.D.
Yu Zhang, Ph.D.
Zhenyu Wang, Ph.D.

Date of Approval:
March 26, 2009

Keywords: ordered probit regression, descriptive statistics, age groups, traffic safety, marginal effects

ACKNOWLEDGMENTS

It is with great pride that I thank the brilliant minds affiliated with the Department of Civil and Environment Engineering at the University of South Florida. I would like to give special thanks to my major professor, Dr. Jian John Lu, for the guidance he has provided. In addition, I would like to thank committee members Dr. Abdul Pinjari, Dr. Yu Zhang, and Dr. Zhenyu Wang. This thesis would not have been possible without your contributions.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Study on Crash Characteristics and Injury Severity at Roadway Work Zones

Qing Wang

ABSTRACT

In USA, despite recent efforts to improve work zone safety, the number of crashes and fatalities at work zones has increased continuously over several past years. For addressing the existing safety problems, a clear understanding of the characteristics of work zone crashes is necessary. This thesis summarized a research study focusing on work zone traffic crash analysis to investigate the characteristics of work zone crashes and to identify the factors contributing to injury severity at work zones. These factors included roadway design, environmental conditions, traffic conditions and vehicle/driver features. Especially, special population groups, which divided into older, middle Age, and young, were inspected. This study was based on history crash data from the Florida State, which were extracted from the Florida CAR (Crash Analysis Reporting) system. Descriptive statistics method was used to find the characteristics of crashes at work zones. After then, an injury severity predict model, using the ordered probit regression technology, was developed to investigate the impacts of various factors on different the injury severity at work zones. From the model, it can be concluded that some factors,

including the road section with curve, alcohol/drugs involved, a high speed, angle crash and too young or old drivers are more likely to increase the probability of angle crashes. Based on the magnitudes of the variable coefficients, the factor of maximum posted speed have a great impact to injury severity, which shows restriction to driving speed is principle countermeasure for improving work zone safety.

CHAPTER ONE

INTRODUCTION

1.1 Background

In Highway Capacity Manual 2002, the definition of work zone is a segment of highway in which maintenance and construction operations impinge on the number of lanes available to traffic or affect the operational characteristics of traffic flowing through the segment. It should be typically marked by signs, channelizing devices, barriers, pavement marking, and/or work vehicles. It extends from the first warming sign or high-intensity rotating, flashing, oscillating, or strobe lights on a vehicle to the “End Road Work” sigh or the last temporary traffic control device. The Manual on Uniform Traffic Control Devices lists five distinct areas within a work zone. Each of these has a specific purpose and may vary in size and location depending on the specifics of each work zone. The five areas are: advance warning area, transition area, activity area, buffer space, and termination area (Figure 1.1).

The advance warning area is the section of highway where road users are informed about the upcoming work zone or incident area. The transition area is that section of highway where road users are redirected out of their normal path. Transition areas usually involve strategic use of tapers, which because of their importance are discussed

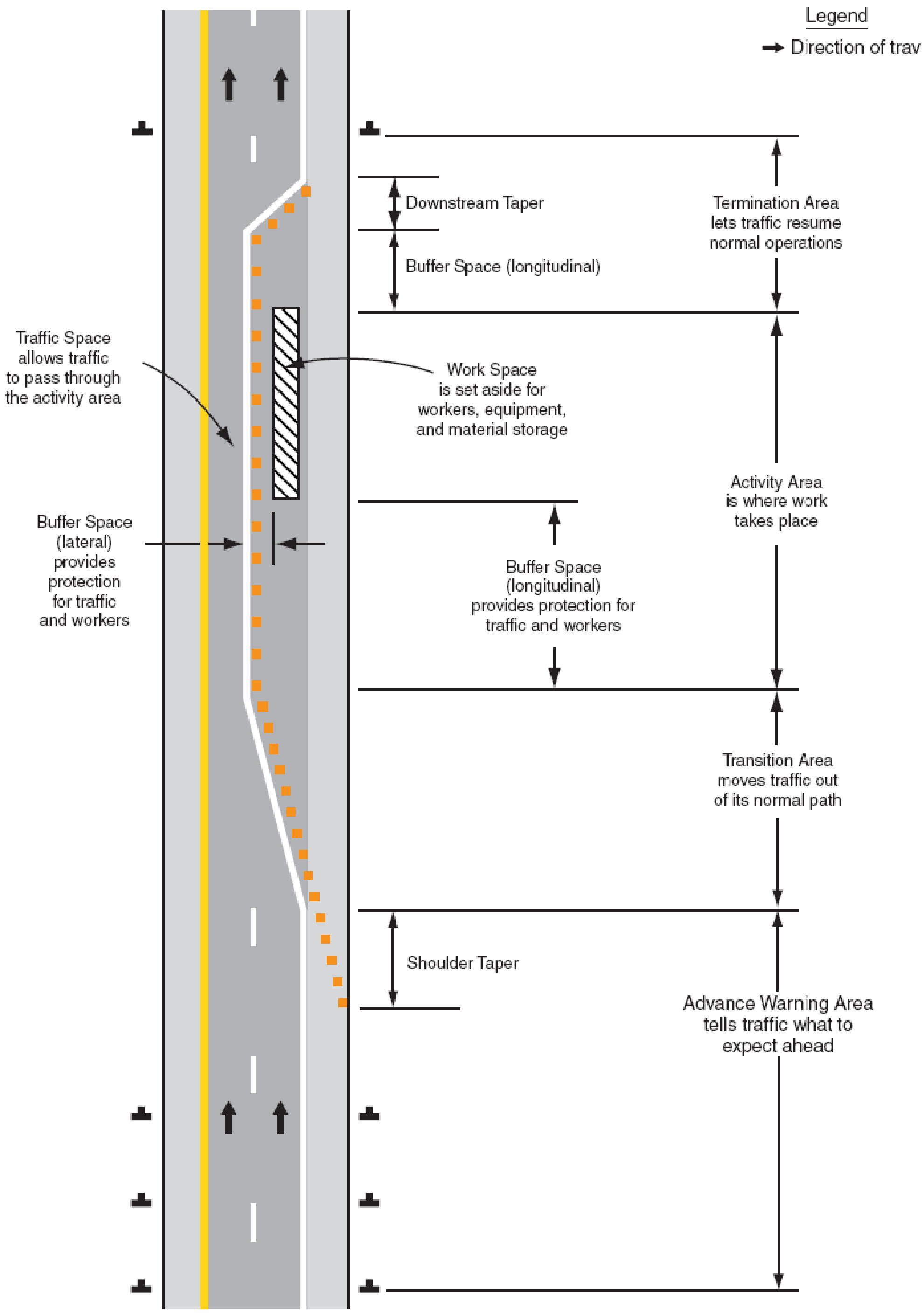

Figure 1.1 Component Parts of a Work Zone

separately in detail. The activity area is the section of the highway where the work activity takes place. It is comprised of the work space, the traffic space, and the buffer space. The work space is that portion of the highway closed to road users and set aside for workers, equipment, and material, and a shadow vehicle if one is used upstream. Work spaces are usually delineated for road users by channelizing devices or, to exclude vehicles and pedestrians, by temporary barriers. Typically, the buffer space is formed as a traffic island and defined by channelizing devices. When a shadow vehicle, arrow panel, or changeable message sign is placed in a closed lane in advance of a work space, only the area upstream of the vehicle, arrow panel, or changeable message sign constitutes the buffer space. The termination is the end area of work zone.

Work zone safety has always been a high priority issue in highway systems but remains unsatisfactory in USA. Based on the statistics from FHWA (Federal Highway Administration), in 2007, there were 835 work zone fatalities, which represent 2.0% of all roadway fatalities for the year. Over four out of every five-work zone fatalities were motorists. In addition, there are over 40,000 injuries at work zones. The total cost of highway work zone injuries calculates to $9.25 billion per year. The highway work zone fatalities per billion dollars spent, are at list 4 times more than in total construction (Maze et al., 2000). Estimating between 1995 and 1997, the direct costs of highway construction zone accidents were as high as $6.2 billion per year, and the average cost is $3687 per accident (Mohan and Gautam, 2002)

To improve work zone safety, four fields need to be approached contemporaneously: engineering, education, enforcement, and coordination with public agencies.

Engineering: This focuses on standardization and evaluation. The standardization part is for traffic control and safety devices in work zone areas. The MUTCD (Manual on Uniform Traffic Control Devices) is the national safety standards to control traffic through work zones, and the NCHRP350 (National Cooperative Highway Research Program Report 350 “Recommended Procedures for the Safety Performance Evaluation of Highway Features”) contains the federal standards and guidelines for all work zone safety devices. The national guidelines regarding planning and implementing work zones is keeping update to address the changing times of more traffic more congestion, greater safety issues, and more work zones.

Education: Public awareness is improved through a variety of activities like clearinghouse website (www.workzonesafety.org); training courses for federal, state, local and tribal highway engineers; conferences, CDs; guidebooks; brochures (for the general public and highway practitioners); bilingual safety public outreach materials; and press events such as National Work Zone Awareness Week.

Enforcement: Engineers in federal highway work closely with state highway to identify appropriate engineering safety countermeasures for high-risk locations new roads. They also work with the enforcement community such as the IACP (International Association of Chiefs of Police). Speed enforcement is a top safety concern in work zones since it has critical relationship with crash severity. In Maryland, Michigan and Virginia, VSL (Variable speed Limits) demonstration projects which determine appropriate speeds for work zones and change them when conditions change were to analyze variations on speed and accompany driver behavior.

Association: Working with emergency medical services, police and fire organizations can ensure that public safety is maintained at high levels and access for emergency vehicles is possible during work zone operations. AASHTO (American Association of State Highway and Transportation Officials), ATSSA (American Traffic Safety Services Association) and FHWA found the National Work Zone Awareness Week in April every year to bring national attention to motorist and worker safety and mobility issues in work zones. Beside this, lots of other publications like Basic Traffic Control for Utility Operations manual and Strategic Highway Safety Plan are the productions by more than one partner or sponsor.

Researching the characteristics of crashes is the very first step of learning the deficiencies of work zone safety and countermeasures. In addition, studying the characteristic differences between each crash injury severity level may cause the discovery of factors influencing injury severity change, which could benefit the development of traffic controls for reducing the proportion of high-severity crashes in total crashes.

1.2 Research Objectives and Approaches

The main objectives of this study are to investigate the characteristics of accidents in work zones, to identify the factors contributing to injury severity levels, and to study how these factors influence injury levels. For more specifically, this study follows these steps:

(1) Review the previous researches in the field of work zone crash characteristics and injury severity models.

(2) Determine the most promising model for model development part by comparing various models in literature review part.

(3) Investigate the differences of characteristics such as crash severity, environmental conditions, crash types and contributing factors among three driver age groups.

(4) Develop a crash severity model for the identification of the most significant factors contributing to the injury severity levels.

1.3 Organization

This thesis consists of five chapters. Chapter 1 provides a brief introduction of the research, including the background of the research, research objective and approaches. Chapter 2 discusses the past studies in both work zone crash characteristics and crash injury severity models, and chooses the most appropriate model to develop the work zone injury severity model for this study. Chapter 3 compares the descriptive characteristics of work zone crashes in three age groups, including the crash severity, environmental conditions and some other contributing factors. A crash injury severity model is produced and interpreted; the factors that influence crash severity levels are found are given in chapter 4. Finally, chapter 6 provides a summary and the conclusion of this research.

CHAPTER TWO

LITERATURE REVIEW

2.1 Previous Studies on Work Zone Crashes

Many studies have been performed on accident experience within work area in the United States. Most of them focus on the crash characteristic in diverse work zone types, crash severity levels, and different locations within work zone.

Ullman et al. (2005) presented an analysis of the safety effects of night work activity upon crashes at two types of construction projects in Texas. The first project type involved both day and night work, whereas the other project type involved pavement resurfacing activities performed only at night. They found that crashes increased more significantly during periods of work activity than during periods when the work zone was inactive. Overall, the increase during work activity was somewhat higher at night than during the day. Researchers also found that crashes increased more at night than during the day at the hybrid projects even when the work zone was inactive, presumably reflecting a disproportionate influence of the temporary geometrics and traffic control upon nighttime travel at these sites.

77 fatal work zone crash sites throughout Texas from Feb. 2003 to Apr. 2004 were analyzed by Schrick (2004). Based on these investigations, researchers concluded that

only 8 percent of the investigated crashes had a direct influence from the work zone, whereas 39 percent of the investigated crashes had an indirect influence from the work zone. Researchers also concluded that 45 percent of the investigated crashes appeared to have no influence from the work zone (included in this subset are the 16 percent of the investigated crashes which occurred in work zones that were work zones in name only, such as work zones that consisted only of project limit signing).

The characteristics of highway work zone collisions and their detailed locations within work zones were studied by Garber and Zhao (2002) to enhance the selection of effective countermeasures. The objective was to determine the distribution and characteristics of crashes in specific areas within a work zone and to compare selected characteristics of work zone crashes with those of non-work zone crashes. In their study, the different locations in the work zone were referred to as the advance warning area, transition area (taper), longitudinal buffer area, activity area, and termination area. Based on the crash percentages regarding location, severity, and collision type, the researchers concluded several major findings. First, the activity area had the highest number of crashes and the highest number of fatal crashes while the termination area was the safest area in terms of numbers of crashes. Second, property-damage-only (PDO) crashes were the predominant severity type, followed by the injury crashes. Third, rear-end crashes were predominant for all areas and all road types except for the termination area, where all crashes were angle crashes. Fourth, as traffic moved from the transition area to the work area, the proportions of rear-end and same-direction sideswipe crashes decreased and the proportions of fixed-object, off-road, and angle crashes increased, although rear-end crashes were still predominant. Last, most nighttime work zone crashes were in the activity area and the severities of nighttime and daytime work zone crashes were not significantly different.

In 2000, Daniel et al. performed a study which was expanded further to examine the difference between fatal crash activity within work zones compared with fatal crashes in non-work-zone locations. Using data from three work zone locations in Georgia, fatal crash activity within work zones also was compared with nonfatal crashes within work zones. Finally, fatal crash activity was examined to determine the influence of the work zone activity on the frequency of fatal crashes. The overall findings of the study indicate that the work zone influences the manner of collision, light conditions, truck involvement, and roadway functional classification under which fatal crashes occur. The study also indicates that fatal crashes in work zones are more likely to involve another vehicle than non-work-zone fatal crashes, and fatal crashes in work zones are less influenced by horizontal and vertical alignment than are non-work-zone crashes.

Khattak et al. (2002) created a unique dataset of California freeway work zones that included crash data (crash frequency and injury severity), road inventory data (average daily traffic and urban/rural character), and work zone related data (duration, length, and location). Crash rates and crash frequencies were investigated in the pre-work zone and during-work zone periods. For the freeway work zones investigated in this study, the total crash rate in the during-work zone period was 21.5% higher (0.79 crashes per million vehicle km) than the pre-work zone period (0.65 crashes per million vehicle km). Compared to the pre-work zone period, the increase in non-injury and injury crash rates in the during-work zone period was 23.8% and 17.3%, respectively. Next, crash frequencies were investigated using negative binomial models, which showed that frequencies increased with increasing work zone duration, length, and average daily traffic.

Wang et al. (1996) discussed the primary questions that safety researches are attempting to answer. The results were presented of an investigation to (a) determined what is known about the magnitude of highway work zone crashes, (b) examined characteristics of highway work zone crashes using the Highway Safety Information System, (c) investigated how work zone accidents are reported on police accident report forms and within state accident report systems, (d) identified critical voids in the knowledge of the relative safety of work zones, and (e) examined possible ways to address unfulfilled information needs related to work zone safety.

2.2 Previous Studies on Crash Severity Model

Researchers have employed many statistical techniques to analyze crash severity level. Among these techniques were log-linear, logit, and probit models.

2.2.1 Log-linear Model

Using 1994 and 1995 crash data from Florida, Abdel-Aty et al. (1998) used log-linear technique to examine relationships between driver age and crash characteristics. The three injury severities in their study were no injury, injury and fatality, and their results suggest that injury severity is positively associated with age; they also concluded that middle-age drivers are more likely to be involved in some crashes, but older drivers are more likely to be involved in fatal crashes. Kim et al. (1995) used log-linear models to predict automobile crash and injury severity. The results suggested that alcohol or drug use and lack of seat belt use increase the odds of more severe crashes and injuries.

2.2.2 Logit Model

Logistic regression models were developed by Donnell and Mason (2004) using both an ordinal and a nominal response. The results indicateed that modeling crash severity as an ordinal response provided appropriate results for cross-median crashes, whereas a nominal response was more appropriate for median barrier crashes. Explanatory variables such as pavement surface conditions, use of drugs or alcohol, presence of an interchange entrance ramp, horizontal alignment, crash type, and average daily traffic volumes affect crash severity. The analysis results might be used by practitioners to understand the trade-off between geometric design decisions and median-related crash severity. Approximately 0.7% median barrier crashes on the

Interstate system resulted in a fatality, whereas 43% were property-damage-only crashes and about 56% were injury crashes. More than 17% of cross-median collisions were fatal, and 67% involved injury.

Modeling severity as a discrete outcome involves estimating the probability that a vehicular crash has a certain severity by determining the likelihood of outcomes given that a crash has occurred. Lee and Chang (2002) estimated the severity of run-off-road crashes in the state of Washington, again by using the nested logit model. Temporal, environmental, driver, roadway, and roadside characteristics were used to estimate property damage and possible injury probabilities for rural run-off-road crashes conditioned on no evident injury. The findings indicated that wet pavement surfaces resulted in possible injury, drivers younger than 25 were more likely to be involved in injury crashes, alcohol-impaired drivers were more likely to be involved in injury crashes, and crashes in the presence of a horizontal curve were more likely to involve an injury.

Dissanayake and Lu (2002) used binary logistic regression model takes the following form. Factors that prove most influential in predicting severity in young driver crashes included influence of alcohol or drugs, ejection in the crash, point of impact, crash location, existence of horizontal curve or vertical grades at the crash site, speed of the vehicle, and restraint device usage.

Krull, Khattak, and Council (2000) used logit models to analyze driver injury severity involved in a single-vehicle crash. Three-year crash data from Michigan and

Illinois were analyzed to explore the effect of rollover, while controlling for roadway, vehicle, and driver factors. Results showed that driver injury severity increases with: (a) failure to use a seatbelt, (b) passenger cars as opposed to pick-up trucks, (c) alcohol use, (d) daylight, (e) rural roads as opposed to urban, (f) posted speed limit, and (g) dry pavement as opposed to slippery pavement.

Chang and Mannering (1999) estimated a nested logit model to study the occupancy crash injury severity relationship. Crash data of principle arterials, state highways, and interstates in Seattle, Washington, during 1994 were used in the analysis. The dependent variable was the crash severity, which represents the most severe level of injury sustained by any vehicle occupant involved in the crash. The occupancy can be significant because vehicles with large occupancies have an increased likelihood of having someone seriously injured. Separate models were estimated for non-truck-involved crashes and for non-truck-involved crashes. Results showed that increased severity was more likely for truck-involved crashes, high speed limits, crashes occurring when a vehicle is making a right or left turn, and rear-end types of collisions.

Shankar, Mannering, and Barfield (1996) estimated a nested logit model to analyze crash severity of single-vehicle crashes on rural freeways. All possible nesting structures (which examine possible correlation among the choices) were considered and statistically tested by the likelihood ratio test. The authors found that a nested-logit model, which treated property damage only (no injury) and possible shared characteristics of injury crashes, fits the data best.

Shankar and Mannering (1996) used a multinomial logit specification for estimating motorcycle rider crash severity likelihood conditioned on the occurrence of a crash. Five levels of severity are considered: property damage only, possible injury, evident injury, severe injury, and fatality. Crash data were 5-year statewide data on single-vehicle motorcycle crash from the state of Washington. Results showed that the multinomial logit formulation is a promising approach to evaluate the determinants of motorcycle crash severity.

Nassar, Saccomanno, and Shortreed (1994) estimated a nested logit model to predict crash severity. Three separate models were calibrated for three crash situations: single-vehicle, two-vehicle, and multi-vehicle crashes. Factors that affect the level of damage experienced by individuals involved in traffic crashes include a crash dynamic term, seating position, seat belt use, vehicle condition, vehicle mass, driver condition, and driver action. Road surface condition was insignificant in the models. Bad weather conditions may prompt drivers to slow down and keep a safe distance from other vehicles.

2.2.3 Probit Model

Abdel-Aty and Keller (2005) produced ordered probit models for crash severity level and used the tree-based regression to explore the factors which affect injury level. The results of this research showed that when attempting to forecast the number of expected crashes of different severity levels, it is imperative that models are developed for each level of collision instead of aggregating crash types to predict the overall severity level. While the ordered probit model approach had been adopted, as did many previous researchers, using the tree-based regression for each severity level improved our understanding of the specific factors and their importance for each severity level. Furthermore, the results showed that crashes reported on short-forms are important and should therefore be retained and included in crash databases. Ignoring this data could lead to biasing the results by under reporting crashes of certain severity or type that could be related to specific explanatory factors. Other crash types or severities might appear to have higher percentages, and therefore, their effect could be artificially exaggerated.

Khattak and Targa (2004), Khattak et al. (2002, 2003) used ordered probit models to predict the injury level for crashes occurring at construction zones and involving trucks, to predict injury severity for single-vehicle truck rollovers, and to determine vehicle, roadway, driver, crash, and environmental characteristics that influence the severity level of older drivers involved in crashes, respectively.

Abdel-Aty (2003) applied the ordered probit models to predict crash injury severity on roadway sections, signalized intersections and toll plazas. Models explained a driver's violation was significant in the case of signalized intersections. Alcohol, lighting conditions, and the existence of a horizontal curve affected the likelihood of injuries in the roadway sections' model. A variable specific to toll plazas, vehicles equipped with Electronic Toll Collection (ETC), had a positive effect on the probability of higher injury severity at toll plazas. Other variables that entered into some of the models were weather condition, area type, and some interaction factors. This study illustrates the similarities and the differences in the factors that affect injury severity between different locations.

Kockelman and Kweon (2002) described the use of ordered probit models to examine the risk of different injury levels sustained under all crash types, two-vehicle crashes, and single-vehicle crashes. The results suggested that pickups and sport utility vehicles are less safe than passenger cars under single-vehicle crash conditions. In two-vehicle crashes, however, these vehicle types were associated with less severe injuries for their drivers – and more severe injuries for occupants of their collision partners. Other conclusions also were presented; for example, the results indicated that males and younger drivers in newer vehicles at lower speeds sustain less severe injuries.

Toshiyuki and Shankar (2002) used a bivariate ordered-response probit model to study driver and most severely injured passenger severity in collision with fixed objects in Washington State. Results showed that icy roadway surface and rain decrease the probability of more severe driver injury. The type of fixed objects significantly affects driver's injury severity. Guardrails have different effects on driver's injury whether the collisions are with its face or with its leading end. Proper use of a restraint system significantly decreases the probability of more severe driver injury. Male and younger drivers have a lower probability of more severe injury, probably because of their physical strength. Also, driver's unconsciousness causes more severe driver injury.

Duncan, Khattak, and Council (1999) used ordered probit modeling to examine the occupant characteristics and roadway and environmental conditions that influence injury severity in rear-end crashes involving truck-passenger car collisions. Two models were developed, one with the basic variables and the other including interactions among the independent variables. Results revealed that an increased severity risk exists for higher speed crashes, those occurring at night, for women, when alcohol is involved, and for crashes when a passenger car rear-ends a truck at a large differential speed between the two vehicles.

Khattak (1999) applied the ordered probit model to examine the effect of information (accuracy of information conveyed by brake and turning lights) and other factors on rear-end crash propagation and the propensity of driver injury in such crashes. Results on injury severity showed that in a two-vehicle crash, the leading driver is more likely to be injured, whereas, in a three-vehicle crash, the driver in the middle is likely to be more severely injured. Furthermore, as rear-end crashes propagate from two-vehicles to three-vehicles the last driver is relatively less severely injured.

Klop (1998) examined the impacts of physical and environmental factors on the severity of injury to bicyclists in North Carolina. Using the ordered probit model, the effect of a set of roadway, environmental, and crash variables on injury severity was explored. Separate models were estimated for rural and urban locations. Results indicated that straight grades, curved grades, darkness, and fog significantly increase injury severity.

Renski, Khattak, and Council (1998) estimated ordered probit models to explore the effects of policy variables on injury severity. Results showed that highway segments where speed limits were raised by 10 mph resulted in a higher probability of increased severity than those raised by only 5 mph. No significant changes in injury severity were found for the highway segments where speed limits were raised from 65 to 70 mph.

In assessing the probabilities of four levels of injury severity as a function of driver attributes, O'Donnell and Connor (1996) compared ordered logit and ordered probit specifications. Their results suggest that injury severity rises with speed, vehicle age, occupant age (squared), female gender, blood alcohol levels over 0.08 percent, non-use of a seatbelt, manner of collision (e.g., head-on crashes), and travel in a light-duty truck. And, according to their comparison of effects, seating position of crash victims was most relevant (e.g., the left-rear seat of the vehicle was found to be most dangerous) and gender least relevant. Many of their results are echoed in the models presented here; the key distinction is that here collision partners and crash-type are examined and emphasized.

Hutchinson (1986) developed an ordered probit model to study occupants' injury severity when involved in traffic crashes. British crash data for 1962–1972 had been processed to give a cross-tabulation of the severity of injury to the driver and to the front seat passenger in four types of single-vehicle crashes (overturning and non-overturning, each in rural and urban areas). Results showed that passengers tend to be more seriously injured than drivers in nonoverturning, but that there is no difference in overturning crashes.

CHAPTER THREE

DESCRIPTIVE STATISTICS ANALYSIS

3.1 The Trend of Crashes

The trend of work zone crashes and fatal crashes are ascending continuously from 2002 to 2006 in Florida (see Figure 3.1). The average annual increase rate of work zone crashes is 18.8%, and the fatal crashes in 2006 are 64.4% more than one in 2002. This trend indicates that the work zone safety in Florida remained a serious concern.

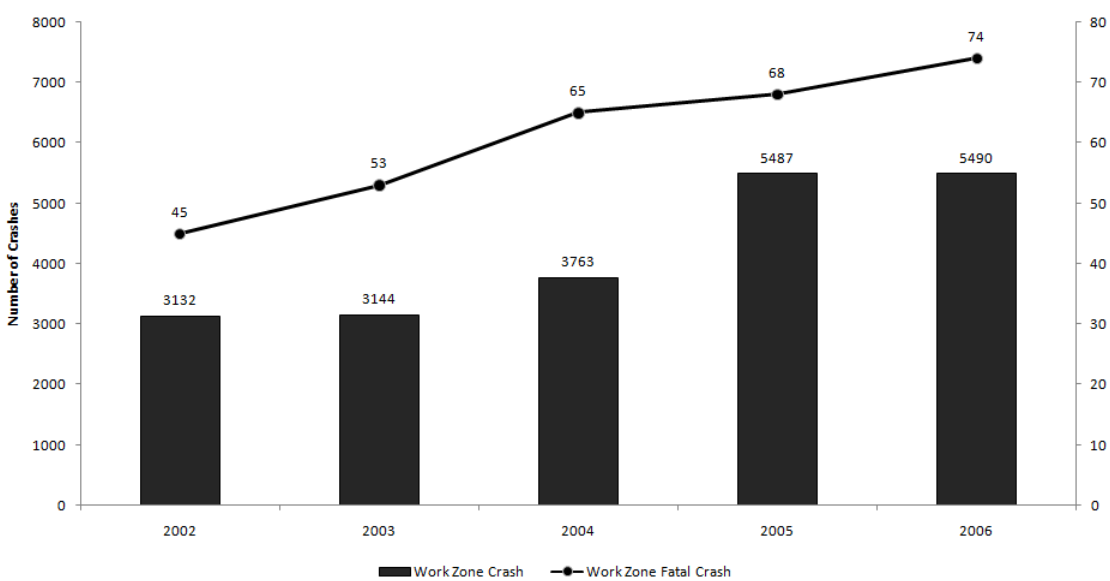

Figure 3.1 Work Zone Crashes and Work Zone Fatal Crashes in Florida

3.2 Distribution of Crashes by Drivers' Age

Figure 3.2 shows the age distribution of the at-fault drivers for work zone and non-work zone crashes. The drivers are divided into three age groups: Young Age (less than 25), Middle Age (25 – 64) and Elderly Age (greater than 65). In work zone area, the middle age drivers cause the highest proportion (67%) of crashes, while the elderly drivers are only responsible for 9% of the crashes. The driver group having the second highest crash rate (24%) is the young age drivers. Compared to work zone crashes, middle age drivers in non-work zone area have a lower possibility of occurring crashes (63%).

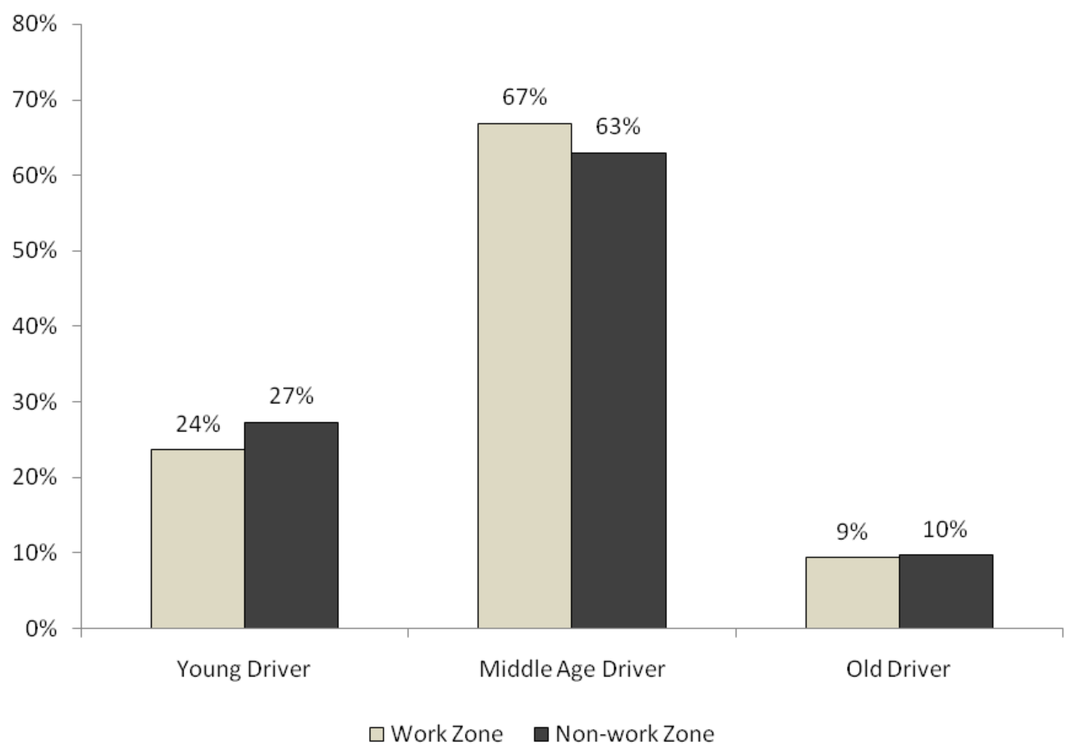

Figure 3.2 Distribution of Work Zone and Non-work Zone Crashes by Age Group

3.3 Distribution of Crashes by Crash Severity

The distribution of work zone crashes by crash severity is shown in Figure 3.3, which indicates that the middle age drivers involved the highest percentage in the no injury crashes which is 49%, and always has the lowest percentage in other severity levels. While in the more severe level crashes, elderly drivers contribute more than the other two age groups (Incapacitating Injury: Old Drivers 9% and Fatal Injury: Old Drivers 2%).

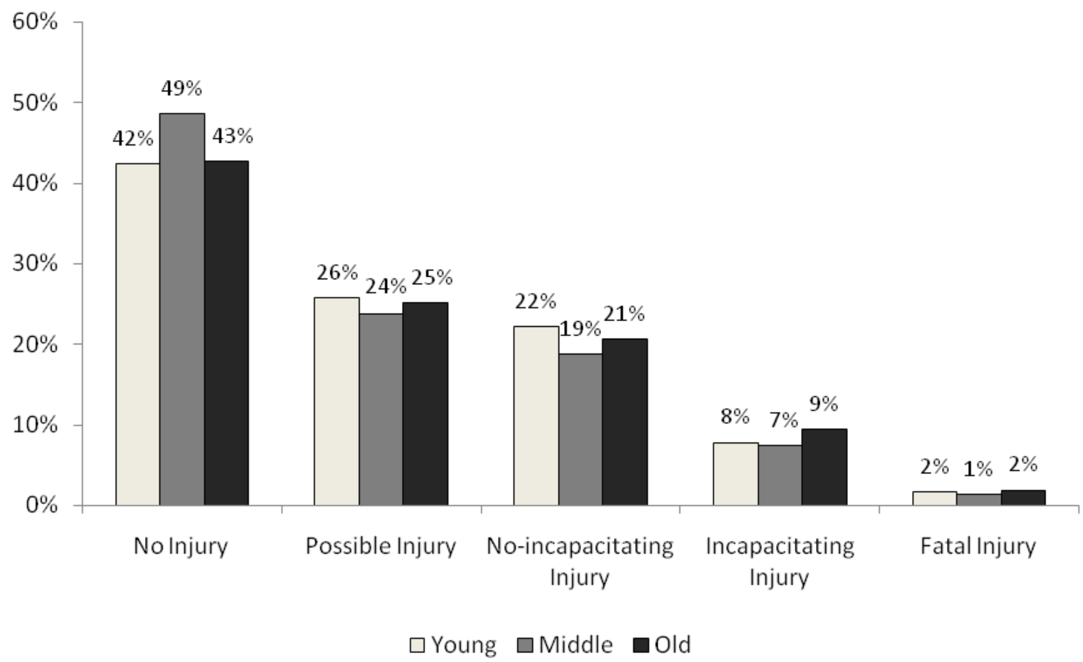

Figure 3.3 Distribution of Work Zone Crashes by Crash Severity

3.4 Distribution of Crashes by Climatic Environmental Conditions

Climatic Environmental conditions include lighting conditions, weather conditions, and road surface conditions. Figure 3.4 summarizes the distribution of crashes by lighting conditions. Most crashes occur when lighting condition is good. Elderly drivers is most likely to having crashes under good lighting condition (daylight), and only has 18% crashes under non-daylight condition including dawn, dusk and dark conditions. In contrast, the difference of crash rate between these two lighting conditions in young drivers is not remarkable.

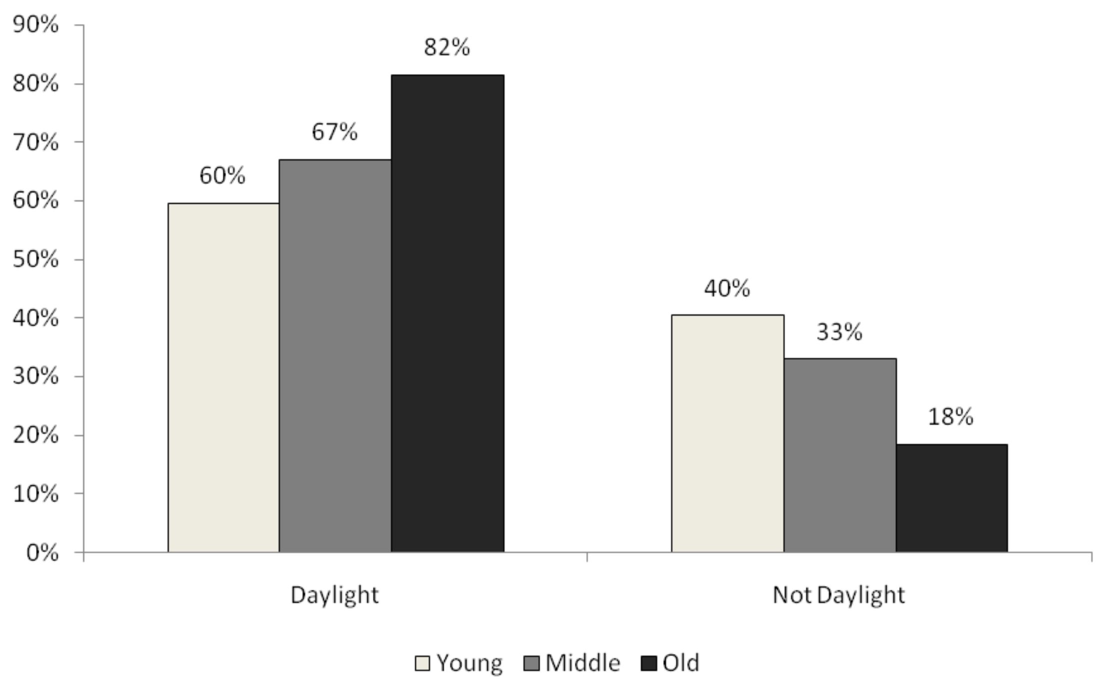

Figure 3.4 Distribution of Work Zone Crashes by Lighting Conditions

The results of analysis of the distribution of work zone crashes by weather and road surface conditions are shown in Figures 3.5 and 3.6 respectively. The results indicate that in all three age groups only a small proportion of work zone crashes occur in bad weather or bad road surface conditions. In contrast to the common sense, the adverse weather and road conditions do not have significant influence on the work zone fatal crashes.

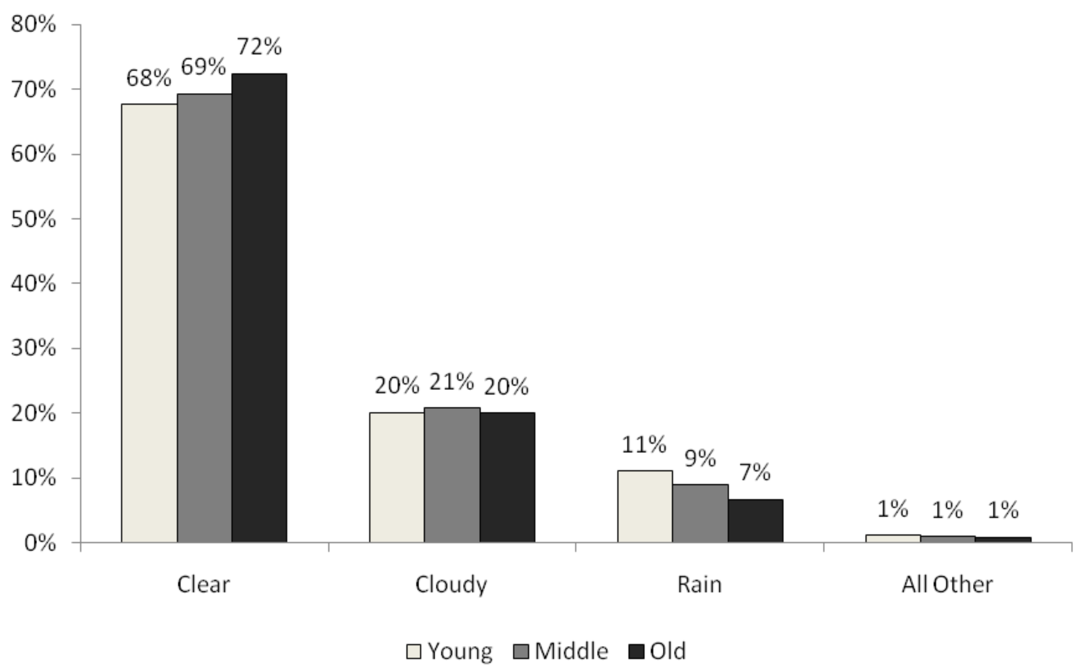

Figure 3.5 Distribution of Work Zone Crashes by Weather Conditions

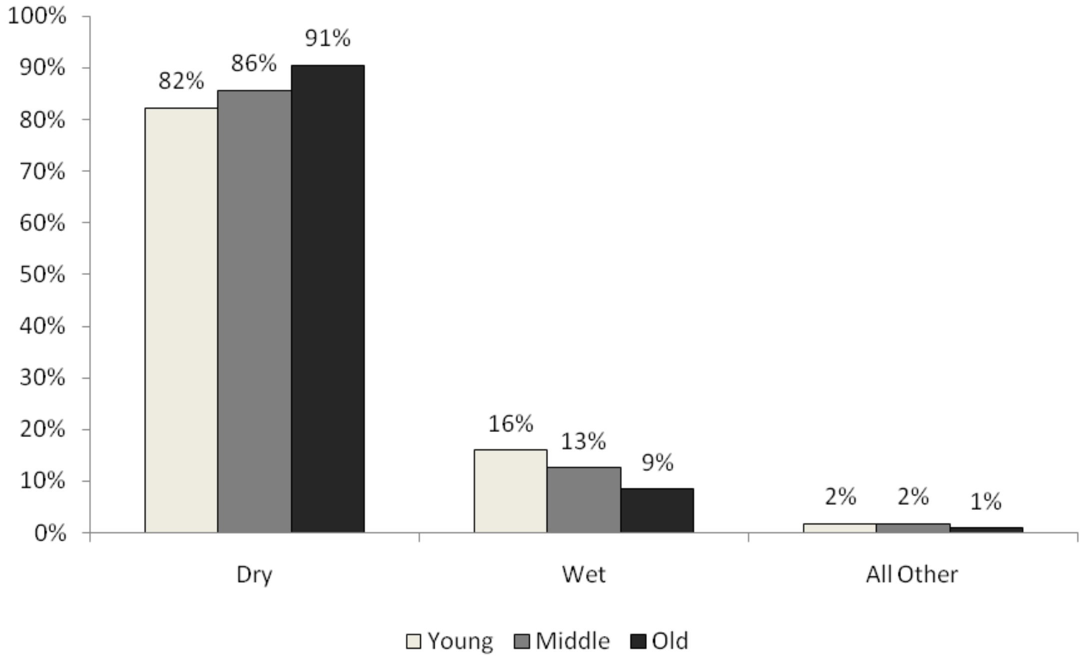

Figure 3.6 Distribution of Work Zone Crashes by Road Surface Conditions

3.5 Distribution of Crashes by Crash Types

As illustrated in Figure 3.7, the top three work zone crash types in all age groups are the same. There are rear-end, angle and sideswipe which are defined as the principle crash types in this study. In young and middle age groups, the percentage of rear-end crashes is obviously higher than angle and sideswipe crashes. Elderly age group shows higher rate in angle crashes than others. Compared work zone and non-work zone crash types in Figure 3.8, read-end and sideswipe crashes are more likely to be occurred in work zone area.

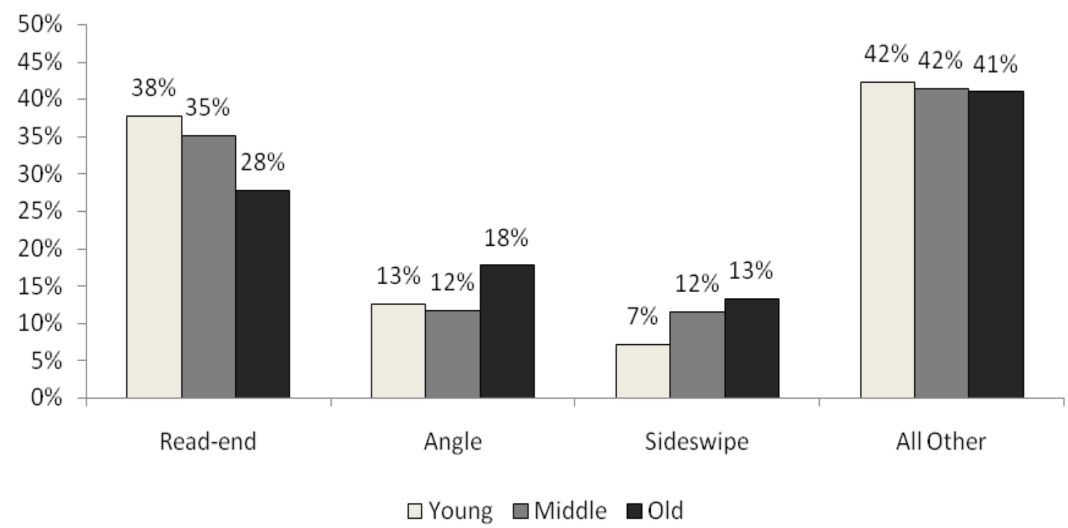

Figure 3.7 Distribution of Work Zone Crashes by Crash Types

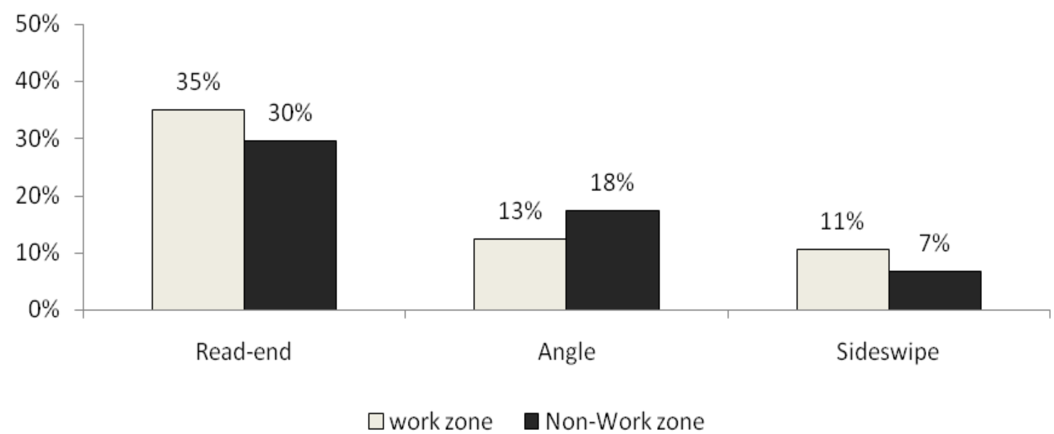

Figure 3.8 Distribution of Work Zone and Non-work Zone Crashes by Crash Types

3.6 Distribution of Crashes by Contributing Factors

Figure 3.9 represents the distribution of contributing factors by all drivers and each age group. Among all drivers, careless driving, the most predominant contributing factor, is responsible for 43% of total crashes. Another predominant factor is failed to yield right of way (11%) followed by no improper driving action (10%) and improper lane change (7%) respectively. In young and middle age group, the distributions are basically same as which of all drivers, except that young drivers show slightly higher rate in careless driving (48%), and the second and third factors which are not variant too much in rate. But in elderly age group, the rate of first factor is just 34% and second one is more than 10% higher than other two age groups.

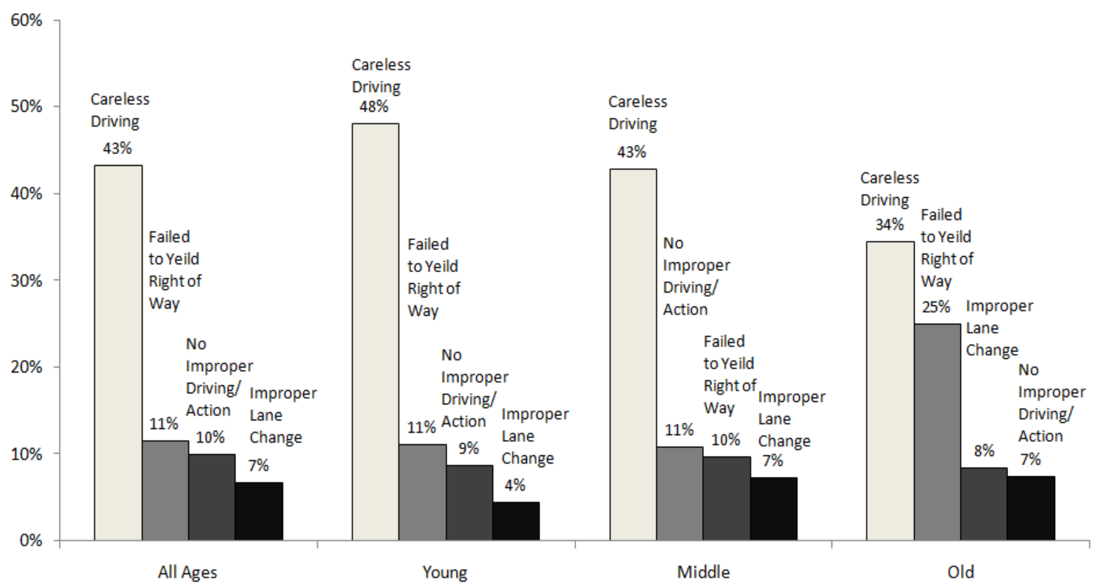

Figure 3.9 Distribution of Work Zone Crashes by Contributing Factors

Figure 3.10 to 3.12 express the distribution of predominant contributing factors over the principal crash types. The most predominant contributing factor for rear-end crashes is careless driving (average 74% in all three age groups). A difference between elderly age group and the other two age groups is that improper lane change is not a predominant contributing factor for older age drivers but it is for young age drivers and middle age drivers.

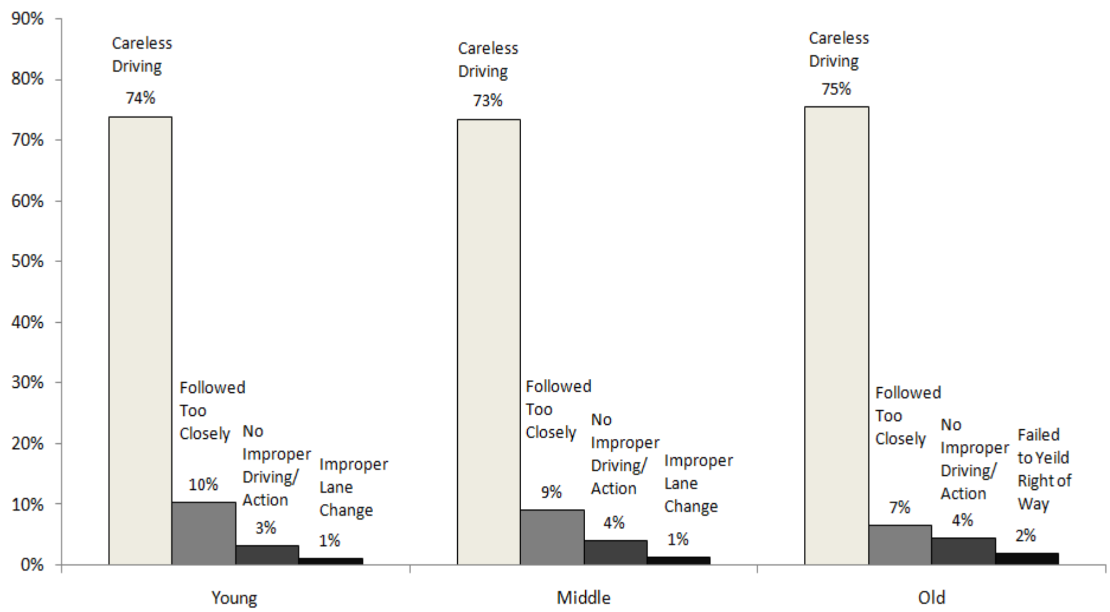

Figure 3.10 Distribution of Work Zone Rear-end Crashes by Contributing Factors

Failed to yield right of way is the most predominant contributing factors for angle crashes. In elderly age group, the rate of this crash type is significantly higher than young and middle age groups; otherwise the rate of careless driving is less than others.

For sideswipe crashes, the improper lane change is the most frequent contributing factor in middle (36%) and elderly (40%) age group, and second most one is careless driving (19% for both groups). However, for young drivers, the top two factors have no much difference (27% for improper lane change and 30% for careless driving).

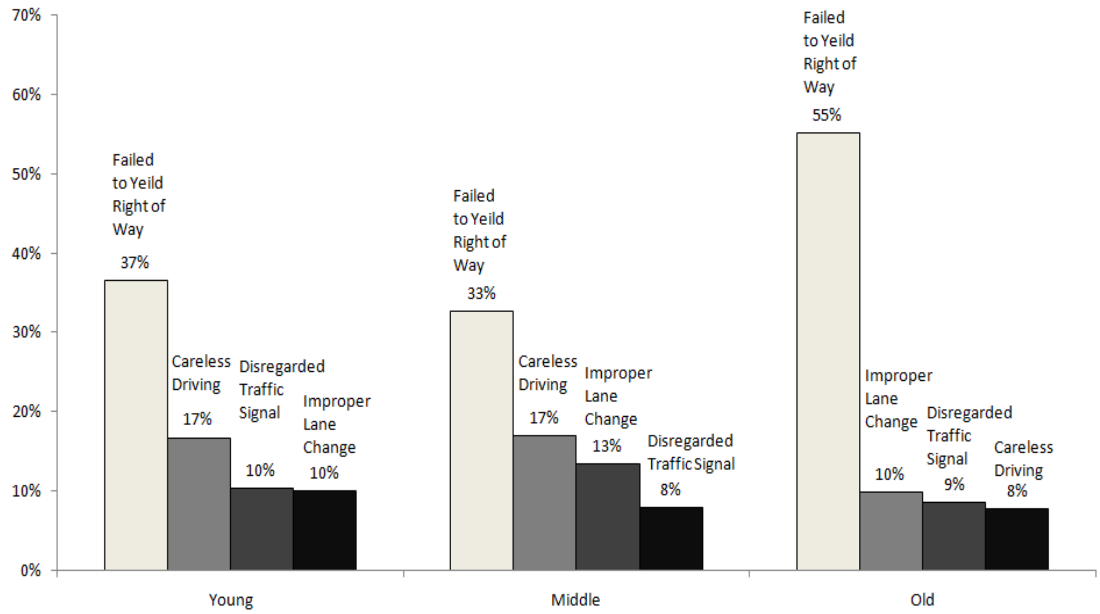

Figure 3.11 Distribution of Work Zone Angle Crashes by Contributing Factors

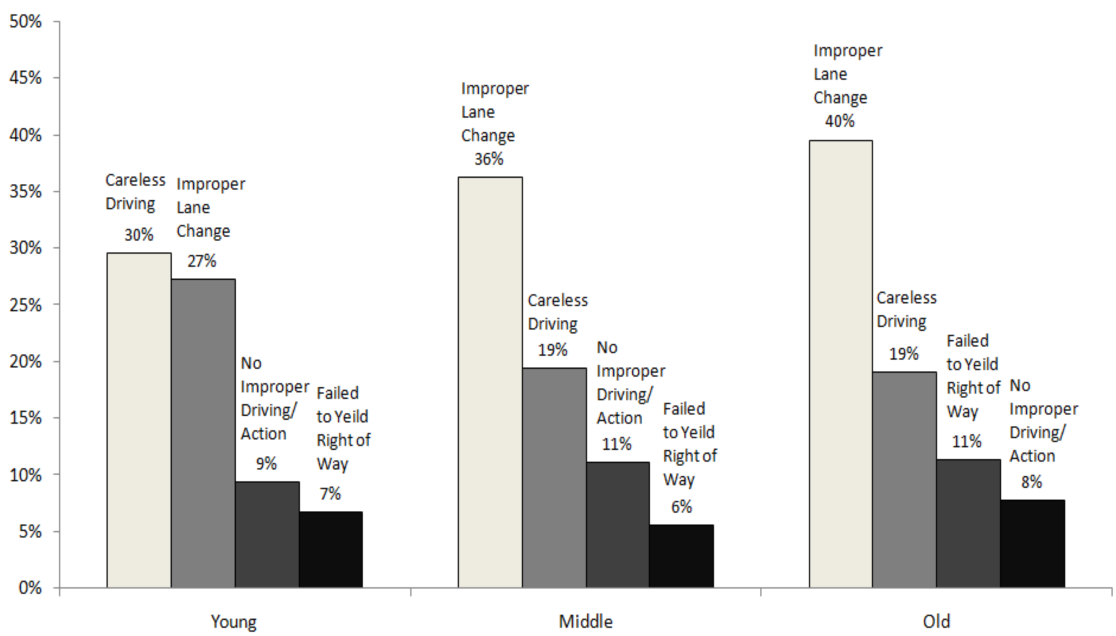

Figure 3.12 Distribution of Work Zone Sideswipe Crashes by Contributing Factors

3.7 Predominant Factors for Other Variables

The distributions of alcohol/drug involved and heavy vehicle (heavy truck and truck tractor) involved are given in Figure 3.13, 3.14 and 3.15. Old drivers are seldom influenced by alcohol/drug (only 1% involved), and most work zone crashes for young age group is not included by heavy vehicle. But heavy vehicle is more easily related to work zone crashes (14%) than non-work zone crashes (7%).

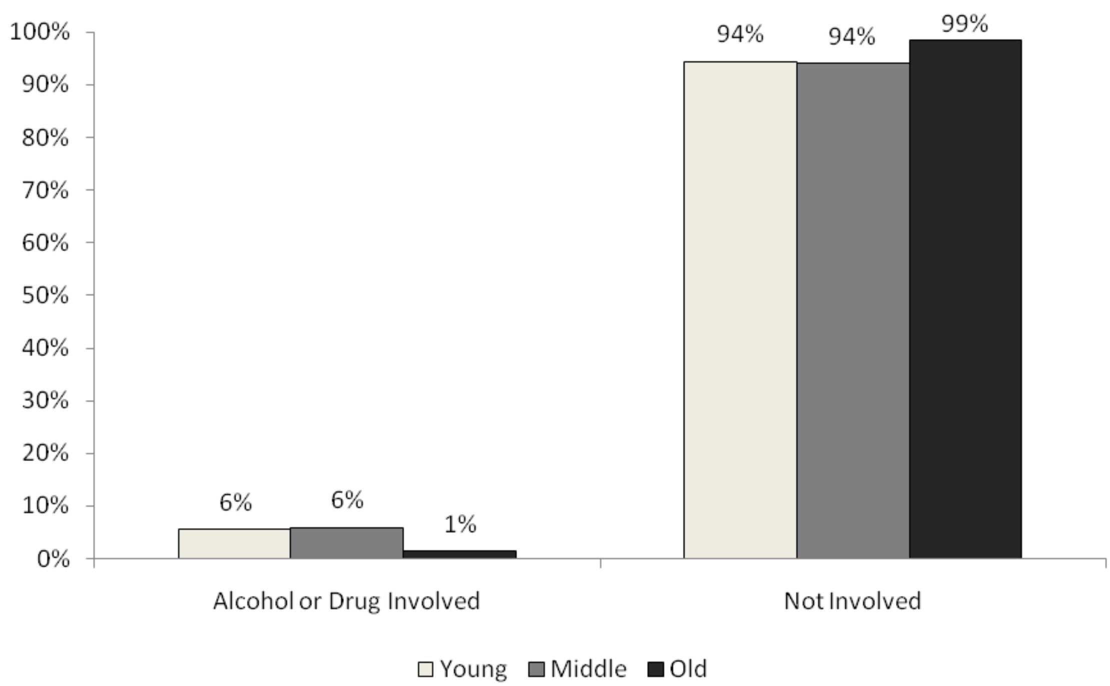

Figure 3.13 Distribution of Work Zone Crashes by Alcohol/Drug Involved

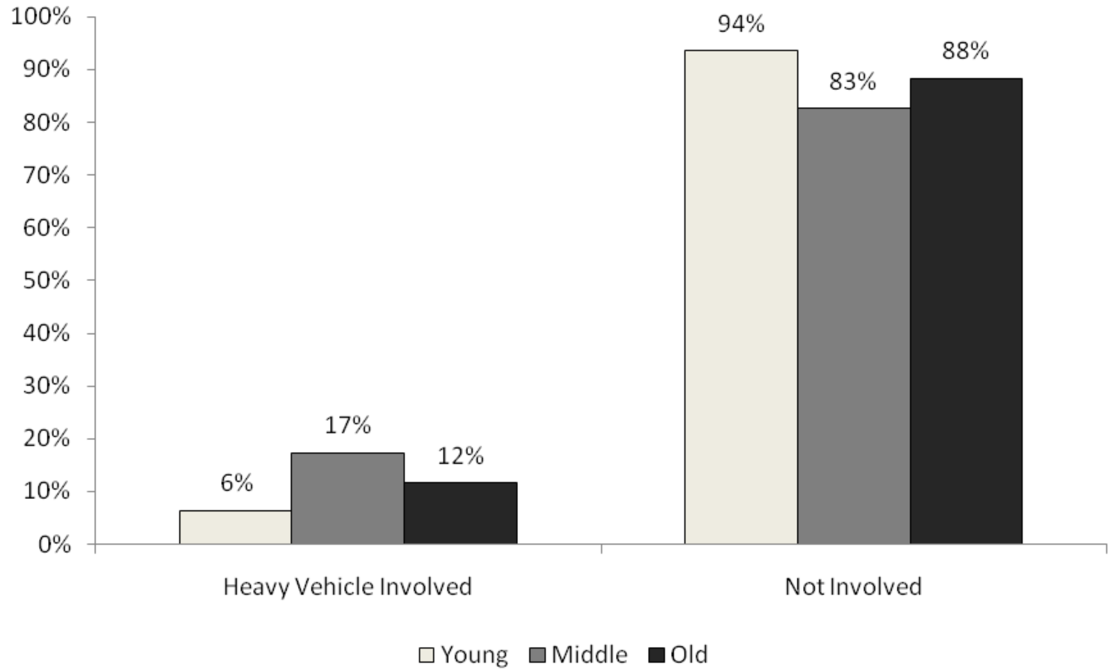

Figure 3.14 Distribution of Work Zone Crashes by Heavy Vehicle Involved

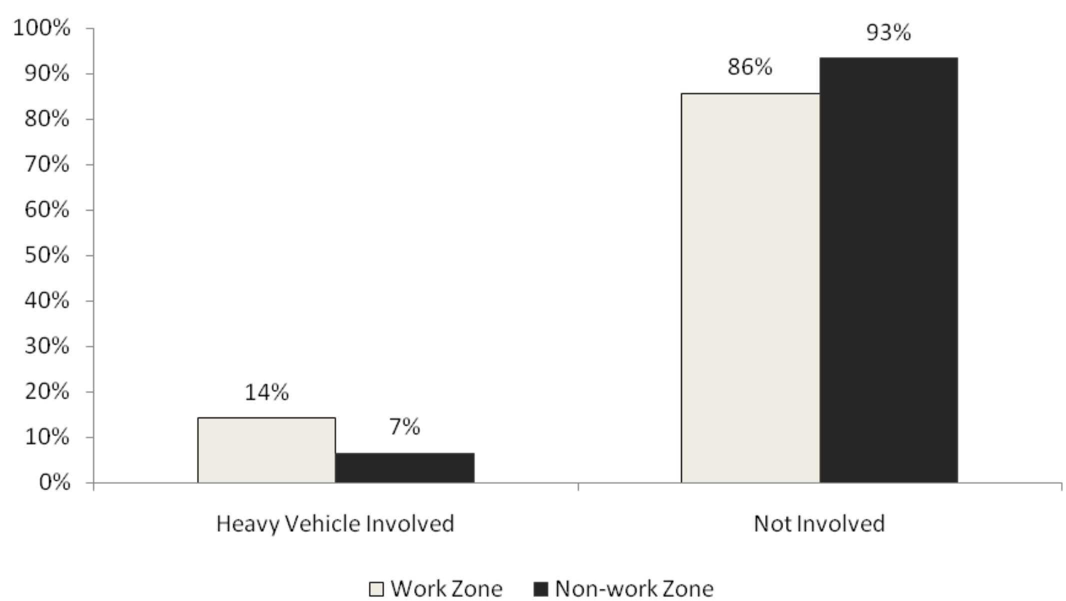

Figure 3.15 Distribution of Work Zone and Non-work Zone Crashes by Heavy Vehicle Involved

CHAPTER FOUR

CRASH SEVERITY MODEL

4.1 Methodology

As stated in previous papers, In contrast to the multinomial models which neglect the data's ordinarily and require more parameters estimated and nested logit models that produce better results but have complexity in identifying the nesting structure, the ordered probit models with a relatively simple approach recognize the indexed nature of various response variables. They are recommended to analyze the crash severity levels.

4.1.1 Crash Severity Models

The crash severity model in this study was developed to investigate the factors that affect crash severity in work zone area. The dependent variable in the model is injury severity level, and the independent variables are the factors which have significant influence on the crash severity. The crash injury severity is a typical ordinal variable which could be categorized at five levels from the least severe level to the most severe level (shown in Table 4.1).

Table 4.1 Definition and Description of Crash Severity Level

Level	Definition	Description
1	No Injury	there is no reason to believe any person received bodily harm from the crash
2	Possible Injury	No visible signs of injury but complaint of pain or momentary unconsciousness
3	Non-incapacitating Injury	Visible injuries from the such as bruises, abrasions, limping, etc.
4	Incapacitating Injury	Any visible signs of injury from the crash and person(s) had to be carried from the scene.
5	Fatal Injury	an injury sustained in a motor vehicle crash that results in death within 90 days

4.1.2 Ordered Probit Regression

The ordered probit model is as followed:

$$y_i^* = \alpha + x_i\beta + \varepsilon_i \tag{4.1}$$

where y_i^* is the latent and continuous measure of crash injury severity; *i* is the number of crashes faced by this severity level; x_i is a vector of parameters to be estimated; ε_i is a random error term which assumed to follow a normal distribution with mean 0 and variance 1. The pdf (Probability Density Function) is

$$\phi(\varepsilon) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{\varepsilon^2}{2}\right) \tag{4.2}$$

and the cdf (Cumulative distribution Function) is

$$\Phi(\varepsilon) = \int_{-\infty}^{\varepsilon}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{\varepsilon^2}{2}\right)dt \tag{4.3}$$

The observed and coded discrete crash injury severity variable y is determined from the model as follows:

$$y_i = \begin{cases} 1 & \text{If } -\infty \le y_i^* < \tau_1 \\ 2 & \text{If } \tau_1 \le y_i^* < \tau_2 \\ 3 & \text{If } \tau_2 \le y_i^* < \tau_3 \\ \vdots & \\ n-1 & \text{If } \tau_{n-2} \le y_i^* < \tau_{n-1} \\ n & \text{If } \tau_{n-1} \le y_i^* < \infty \end{cases} \quad (4.4)$$

This mapping from the latent variable to the observed crash injury severity class is illustrated in Figure 4.1.

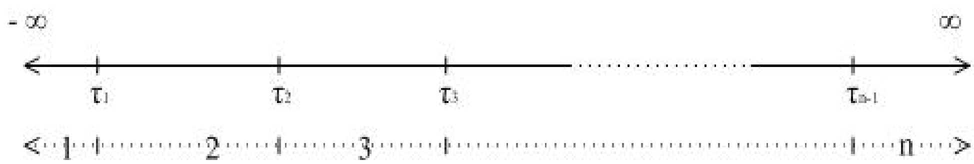

Figure 4.1 Latent Variables to the Observed Categories

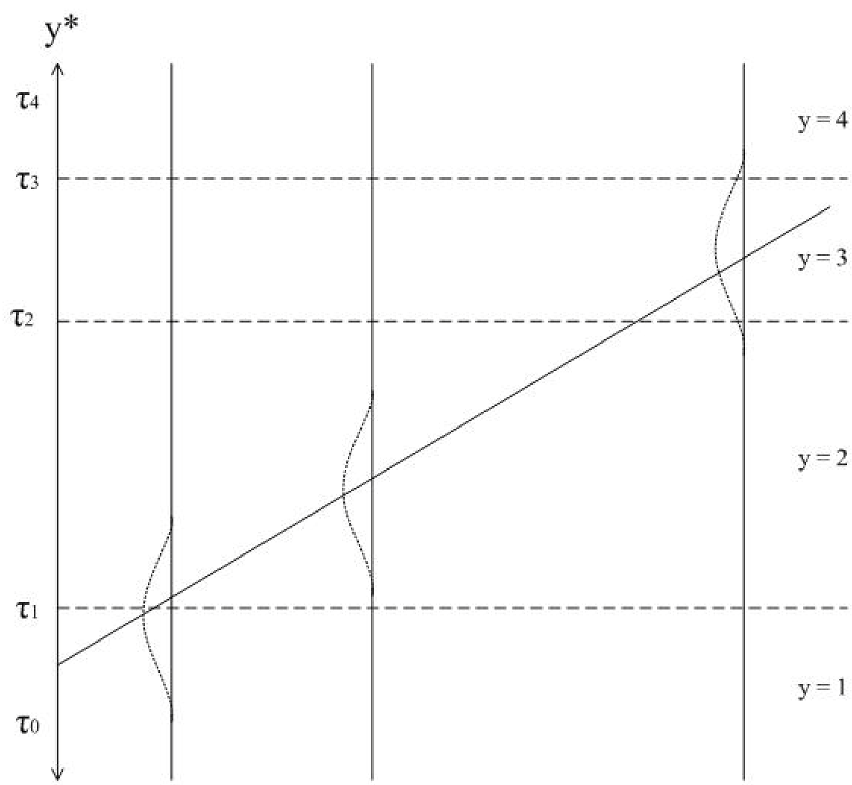

Figure 4.2 Distribution of y^* Given x for the Ordered Regression Model

Consider Figure 4.2 which shows the distribution of y^* for four values of *x*. The errors are distributed normally around the regression line $E(y^*|x) = \alpha + \beta x$. The Probability of outcome *m* corresponds to the area of the error distribution between the cutpoints τ_{m-1} and τ_m. This area is computed as follows.

First consider the formula for the probability that *y* = 1. We observe *y* = 1 when y^* falls between $\tau_0 = -\propto$ and τ_1. This implies that

$$\Pr(y_i = 1|x_i) = \Pr(\tau_0 \le y_i^* < \tau_1 | x_i) \tag{4.5}$$

Substituting $y_i^* = \alpha + x_i\beta + \varepsilon_i$,

$$\Pr(y_i = 1|x_i) = \Pr(\tau_0 \le \alpha + x_i\beta + \varepsilon < \tau_1 | x_i) \tag{4.6}$$

Then, subtracting $x\beta$ within the inequality,

$$\Pr(y_i = 1|x_i) = \Pr(\tau_0 - \alpha - x_i\beta \le \varepsilon < \tau_1 - \alpha - x_i\beta | x_i) \tag{4.7}$$

The probability that a random variable is between two values is the difference between the cdf evaluated at these values. Therefore,

$$\begin{aligned}\Pr(y_i = 1|x_i) &= \Pr(\varepsilon < \tau_1 - \alpha - x_i\beta | x_i) - \Pr(\varepsilon < \tau_0 - \alpha - x_i\beta | x_i) \\ &= \mathrm{F}(\tau_1 - \alpha - x_i\beta) - \mathrm{F}(\tau_0 - \alpha - x_i\beta)\end{aligned} \tag{4.8}$$

These steps can be generalized to compute the probability of any observed outcome *y* = *m* given *x*:

$$\Pr(y_i = m|x_i) = \mathrm{F}(\tau_m - \alpha - x_i\beta) - \mathrm{F}(\tau_{m-1} - \alpha - x_i\beta) \tag{4.9}$$

When computing $\Pr(y = 1|x)$, the second term on the right-hand side drops out since $\mathrm{F}(\tau_0 - x\beta) = \mathrm{F}(-\propto - x\beta) = 0$; when computing $\Pr(y = J|x)$, the first term equal 1 since

$F(\tau_J - x\beta) = F(\propto - x\beta) = 1$. Thus, for a model with four observed outcomes, such as shown in Figure 4.2, the formulas for the ordered probit model are

$$\Pr(y_i = 1|x_i) = \Phi(\tau_1 - \alpha - \beta x_i)$$
$$\Pr(y_i = 2|x_i) = \Phi(\tau_2 - \alpha - \beta x_i) - \Phi(\tau_1 - \alpha - \beta x_i)$$
$$\Pr(y_i = 3|x_i) = \Phi(\tau_3 - \alpha - \beta x_i) - \Phi(\tau_2 - \alpha - \beta x_i) \quad (4.10)$$
$$\Pr(y_i = n-1|x_i) = \Phi(\tau_n - \alpha - \beta x_i) - \Phi(\tau_{n-1} - \alpha - \beta x_i)$$
$$\Pr(y_i = n|x_i) = 1 - \Phi(\tau_{n-1} - \alpha - \beta x_i)$$

where *i* is an individual; 1, 2, 3…*n*-1, *n* are response alternatives; $\Phi(\cdot)$ is the standard normal cumulative distribution function.

Since y^* is latent, its mean and variance cannot be estimated. The variance is identified by using that $\mathrm{Var}(\varepsilon|x) = 1$. While these assumptions identify the variance, the mean of y^* is still unidentified. The consequences of this can be seen by considering the model $y^* = \alpha + \beta x + \varepsilon$ with cutpoints τ_m. Think of α and the τ 's as the “true” parameters in the sense that they were used to generate the observed data. Define an alternative set of parameters:

$$\alpha^* = \alpha - \delta \quad \text{and} \quad \tau_m^* = \tau_m - \delta \quad (4.11)$$

where τ is an arbitrary constant. The probability that $y = m$ is identical, whether the true or alternative parameters are used:

$$\Pr(y = m|x) = \mathrm{F}(\tau_m - \alpha - x\beta) - \mathrm{F}(\tau_{m-1} - \alpha - x\beta)$$

$$= \mathrm{F}([\tau_m - \delta] - [\alpha - \delta] - \beta x) - \mathrm{F}([\tau_{m-1} - \delta] - [\alpha - \delta] - \beta x)$$

$$= \mathrm{F}(\tau^*_{m-1} - \alpha^* - x_i\beta) - \mathrm{F}(\tau^*_{m-1} - \alpha^* - x_i\beta) \qquad (4.12)$$

Since both sets of parameters generate the same value for the probability of an observed outcome, there is no way to choose between the two sets of parameters using the observed data: a change in the intercept in the structural model can always be compensated for by a corresponding change in the thresholds. That is to say, the model is unidentified.

While there are an infinite number of assumptions that could be made to identify the model, only two are commonly used:

(1). Assume that $\tau_1 = 0$. This involves setting $\delta = \tau_1$ in Equation 4.11.

(2). Assume that $\alpha = 0$. This involves setting $\delta = \alpha$ in Equation 4.11.

Both assumptions identify the model by imposing a constraint on one of the parameters. The different identifying assumptions lead to what are known as different parameterizations of the model. The choice of which parameterization to use is arbitrary and does not affect the β's (except for β_0) or associated significance tests. Further, as known by Equation 4.12, the probabilities are not affected by the identifying assumption. However, understanding the different parameterizations is important since different software uses different parameterizations. Programs such as LIMDEP uses the first assumption, while programs such as Markov, SAS's LOGISTIC, and Stata use the second one. The choice of parameterization does not affect estimates of the slopes, but does affect the estimates of β_0 and the τ 's.

4.1.3 Criteria for Ordered Probit Models

4.1.3.1. z - Test

z - Test is used to test the statistical significance of individual estimated coefficient in ordered porbit models. Maximum likelihood estimators possess a number of desirable properties when certain general conditions apply. Independent and identically distributed observations, and independence of the x_i and the model errors (the ε_i) are all that is required. With these conditions satisfied, the maximum likelihood estimator is asymptotically unbiased (consistent), is normally distributed, and has the smallest variance among all consistent and asymptotically normal estimators. The t – ratios for the null hypothesis H_0 that $\beta_i = 0$, and the test statistic is

$$z = \frac{\hat{\beta}_i}{\hat{\sigma}_i / \sqrt{k}} \tag{4.11}$$

where $\hat{\beta}_i$ is the estimator of β_i; and β_i is the *i*th coefficient of the model; $\hat{\sigma}_i$ is the estimator of standard deviation of the coefficient β_i; *i* is number of observations. If H_0 is true, the coefficient β_i of the model is not statistically significant. If H_0 is rejected at a confidence level (usually is 0.05), the coefficient β_i is significant to the response.

4.1.3.2. Pseudo - R^2

A Pseudo - R^2 is often used as a goodness-of-fit measure in non-linear models. They look like R^2 in the sense that they are on a similar scale, ranging from 0 to 1, but they cannot be interpreted as one would interpret an ordinary least squares (OLS) R^2 and different Pseudo - R^2 can arrive at very different values.

Here, the Pseudo - R^2 is provided as

$$R^2 = 1 - \frac{\ln \hat{L}(M_{full})}{\ln \hat{L}(M_{intercept})} \tag{4.12}$$

where M_{full} is the model with predictors; $M_{intercept}$ is the model without predictors; $\hat{L}$ is the estimated likelihood.

A likelihood falls between 0 and 1, so the log of a likelihood is less than or equal to zero. If a model has a very low likelihood, then the log of the likelihood will have a larger magnitude than the log of a more likely model. Thus, a small ratio of log likelihoods indicates that the full model is a far better fit than the intercept model.

4.1.3.3. Likelihood Ration (LR) Test

The likelihood ratio test is a statistical test of the goodness-of-fit between two models. It relies on a test statistic computed by taking the ratio of the maximum value of the likelihood function under the constraint of the null hypothesis to the maximum with that constraint relaxed. The null hypothesis is $H_0 : \beta = 0$, where β is the intercept. This statistic is given as

$$G^2 = -2[\ln L(M_{constrained}) - \ln L(M_{unconstrained})] \tag{4.13}$$

where $L(M_{constrained})$ is the likelihood of the constrained model; $L(M_{unconstrained})$ is the likelihood of the unconstrained model.

This LRT statistic approximately follows a chi-square distribution. The degree of freedom is equal to the number of additional parameters in the unconstrained model. If the null hypothesis is rejected (the confidence level is usually 0.05), it can be concluded that at least one independent variable has significant influence for the dependent variable.

4.1.4 Interpretation of Model Coefficients

4.1.4.1. The Partial Change in y^*

In the ordered regression model,

$$y^* = x\beta + \varepsilon \tag{4.16}$$

and the partial change in y^* with respect to x_k is

$$\frac{\partial y^*}{\partial x_k} = \beta_k \tag{4.17}$$

Since the model is linear in y^*, the partial change can be interpreted as: for a unit increase in x_k, y^* is expected to change by β_k units, holding all other variables constant. Because the variance of y^* cannot be estimate from the observed data, the meaning of a change of β_k units in y^* is unclear. Interpretations should be based on y^*-standardized coefficients.

If σ_{y^*} is the unconditional standard deviation of the latent y^*, then the y^*-standardized coefficient for x_k is

$$\beta_k^{Sy^*} = \frac{\beta_k}{\sigma_{y^*}} \tag{4.18}$$

which can be interpreted as: for a unit increase in x_k, y^* in expected to increase by $\beta_k^{Sy^*}$ standard deviations, holding all other variables constant.

y^*-standardized coefficients indicate the effect of an independent variable in its original unit of measurement. This is sometimes preferable for substantive reasons and is necessary for binary independent variables.

The variance of y^* can be estimated by the quadratic form:

$$\hat{\sigma}_{y^*}^2 = \hat{\beta}' \, \hat{\mathrm{V}}\mathrm{ar}(x)\hat{\beta} + \mathrm{Var}(\varepsilon) \tag{4.19}$$

where $\hat{\mathrm{V}}\mathrm{ar}(x)$ is the covariance matrix for the x's computed from the observed data; $\hat{\beta}$ contains Maximum Likelihood (ML) estimates; and $\mathrm{Var}(\varepsilon) = 1$ in the ordered probit model.

4.1.4.2. Partial Change in Predicted Probabilities

The predicted probability that $y = m$ given x is

$$\Pr(y = m|x) = \mathrm{F}(\tau_m - x\beta) - \mathrm{F}(\tau_m - x\beta) \tag{4.20}$$

Taking the partial derivative with respect to x_k,

$$\frac{\partial \Pr(y=m|x)}{\partial x_k} = \frac{\partial \mathrm{F}(\tau_m - x\beta)}{\partial x_k} - \frac{\partial \mathrm{F}(\tau_{m-1} - x\beta)}{\partial x_k}$$

$$= \beta_k f(\tau_{m-1} - x\beta) - \beta_k f(\tau_m - x\beta)$$

$$= \beta_k \left[f(\tau_{m-1} - x\beta) - f(\tau_m - x\beta) \right] \tag{4.21}$$

The partial change or marginal effect is the slope of the curve relating x_k to $\Pr(y=m|x)$, holding all other variables constant. The sign of the marginal effect is not necessarily the same as the sign of β, since $f(\tau_{m-1} - x\beta) - f(\tau_m - x\beta)$ can be negative. Indeed, it is possible for the marginal effect of x_k to change signs as x_k changes.

In general, the marginal effect does not indicate the change in the probability that would be observed for a unit changes in x_k. However, if an independent variable varies over a region of the probability curve that is nearly linear, the marginal effect

can be used to summarize the effect of a unit change in the variable on the probability of an outcome.

4.2 Data Collection

4.2.1 Data Base

The dataset used to fit the ordered probit model was extracted from the Florida Crash Analysis Reporting (CAR) system. CAR system is a relational database for State System crashes consisting of nine tables which contain different data relevant to a certain facet of a traffic crash (Table 4.2). It maintains electronic crash records based on crashes reported on the long-form crash report. That the variable “FIRST ROAD CONDITION

CRASH COD" is equal to 04 (road under repair/construction) is used as the indicator of work zone crashes. In this study, the work zone crash dataset contained all the work zone crashes from 2002 to 2006.

Some variables in the database were selected for modeling. They may include ordinal variables, nominal variables, or continuous variables. In order to get better result performance all categorical variables should be purposely converted to binary ones (dummy variable). The continuous variables need to be normalized (by dividing by each maximum value) to have values which lie between 0 and 1. The reason for this is that the dummy variables have means between 0 and 1, and ordered multiple choice models are almost never estimable if the variables are of very different magnitudes (Greene 1993). All the missing values are deleted from database. Appendix A lists the description of every original variable in this work zone crashes database.

Table 4.2 Tables from Florida Traffic Crash Records Database

File Name	Description
Events	Contains information about the crash event (i.e. date, time, harmful events, etc.). This is the "parent file" of the database.
Drivers	Contains information about each driver involved in the crash demographic and causal).
Passengers	Contains information about each passenger involved in the crash (demographic and causal).
Pedestrians	Contains information about each pedestrian involved in the crash (demographic and causal).
Property	Contains information about property (other than vehicles) damaged in the crash.
Vehicles	Contains information about each vehicle involved in the crash.

Table 4.2 (Continued)

File Name	Description
Violations	Contains information about citations issued to drivers or pedestrians involved in crashes (limited to the first eight citations issues per party).
ComVeh	The newest table, contains information about commercial vehicles and carriers involved in crashes.
DOT	Contains Department of Transportation location and road data.

4.2.2 Data Description

For developing the work zone crash injury severity model, 10 variables (Table 4.3) are selected. The dependent variable is the crash injury severity which has 5 levels from no injury to fatal injury at an ascending order. The other independent variables can be categorized as 4 classes: environmental condition, roadway condition, driver’s condition, and crash-related information.

Table 4.3 Description of Selected Variables for Model Development

Variable	Description	Type	Value	Definition
ACCISEV	Crash Severity Level	Ordinal	1	No Injury
			2	Possible Injury
			3	Non-incapacitating injury
			4	Incapacitating Injury
			5	Fatal Injury
Environmental Condition				
LGHTCOND	If the crash occurred under the good lighting condition (daylight condition)	Binary	0	No
			1	Yes
Roadway Condition				
CURVE	If there is a curve at the crash location	Binary	0	No
			1	Yes

Table 4.3 (Continued)

URBAN	If the crash occurred in a urban area	Binary	0	No
			1	Yes
MAXSPEED	Maximum Posted Speed Limit	Continuous		
SECTADT	Section average annual daily traffic	Continuous		
Driver's Condition				
AGE_AT_FAULT	At fault driver's age	Categorical	1	Young (15-24)
			2	Middle (25-64)
			3	Old (≥ 65)
ALDGUSE_AT_FAULT	If at fault driver was under influence of alcohol or drugs	Binary	0	No
			1	Yes
Crash-Related				
VEHTYPE	If heavy vehicle (heavy truck and truck tractor)was involved	Binary	0	No
			1	Yes
HARMEVN	Crash Type	Categorical	1	Rear-end
			2	Angle
			3	Sideswipe
			4	Other Types

Table 4.4 describes the minimum value, maximum value, range, mean, and standard deviation of the two continuous variables. The minimums, maximums, ranges, means, and standard deviations of the original unnormalized variables can be obtained easily by multiplying the values in Table 4.4 by the appropriate scaling factors (the original maximum values in each variable). The range of AADT in work zone area is very large from $0.0045 \times 289{,}000 = 1{,}300$ vehicles per day to $1 \times 289{,}000 = 289{,}000$ vehicles per day. The minimum speed limit is $0.2143 \times 70 = 15$ miles per hour, the maximum one is $1 \times 70 = 70$ miles per hour, and mean value is $0.7455 \times 70 = 52$ miles per hour.

Table 4.4 Description Statistic of Continuous Variables

Varibale	N	Minimum	Maximum	Range	Mean	Std. Deviation	Scaling Factors
SECADT	14217	0.0045	1	0.9955	0.2205	0.1774	289000
MAXSPEED	14217	0.2143	1	0.7857	0.7455	0.15984	70

Table 4.5 illustrates the discrete variables' frequency statistic. When the crash injury severity increases, the frequency of crashes decreases. The total percentage of slight injury crashes (ACCISEV = 1, 2, and 3) in work zone area is 90.41%. Incapacitating injury crash only holds 7.94%, and the fatal crash has the least proportion which is 1.66%. More than one third of work zone crashes (34.17%) occur under the not good lighting condition (non-daylight), and 85.84% of them in the urban area. Only 8.10% of locations where work zone crash happen has curve, 14.62% work zone crashes occur with heavy vehicle involvement, and 5.15% drivers are influenced by drugs or alcohol.

The top three crash types here are rear-end (37.15%), angle (12.04%), and sideswipe (11.26%). The distribution of at-fault driver's age group is 23.61% young age drivers, 66.81% middle age drivers, and 9.59% old age drivers.

Table 4.5 Frequencies of Discrete Variables

Variable	Value	Frequency	Percent
Sample Size 14217			
ACCISEV	1	6477	45.56
	2	3555	25.01
	3	2820	19.83
	4	1129	7.94
	5	236	1.66
LGHTCOND	0	4858	34.17
	1	9359	65.83
CURVE	0	13065	91.90
	1	1152	8.10
URBAN	0	2013	14.16
	1	12204	85.84
AGE_AT_FAULT	1	3356	23.60
	2	9498	66.81
	3	1363	9.59
ALDGUSE_AT_FAULT	0	13485	94.85
	1	732	5.15
VEHTYPE	0	12139	85.38
	1	2078	14.62
HARMEVN	1	5282	37.15
	2	1712	12.04
	3	1601	11.26
	4	5622	39.55

4.3 Work Zone Crash Injury Severity Model

4.3.1 Estimation Procedure

This section presents the estimation results of the work zone crash severity model for all work zone crashes. At first, cross tabulation analysis is performed to check the distribution of explanatory variables across injury severity levels and ensure enough observations in each cell. And AGE_AT_FAULT variable was transformed to three dummy variables: YOUNG_AGE (AGE_AT_FAULT = 1), MIDDLE_AGE (AGE_AT_FAULT = 2), and OLD_AGE (AGE_AT_FAULT = 3). Be similar, another categorical variable HARMEVN was converted to four dummy variables: REAR-END (HARMEVN =1), ANGLE (HARMEVN = 2), SIDESWIPE (HARMEVN = 3), and OTHERS (HARMEVN = 4). After then, the ordinal probit regression model was developed using the OPROBIT procedure in the STATA software package. In the procedure, the stepwise option was added for selecting independent variables for which the significant level is greater than 95%. The theory of variable selection is: at first, there was no variable in this ordered probit model, then the variables whose p-value is less or equal to 0.05 were added into the model one by one.

4.3.2 Cross Tabulations between Explanatory Variables and Crash Severity

In order to obtain a better understanding about the selected explanatory variables, cross tabulations of binary or categorical variables with crash severity were developed and given in Tables 4.6.

Table 4.6 Cross Tabulation between explanatory Variables and Crash Severity

Frequency Row %	Value	Crash severity					Total
		1	2	3	4	5	
LGHTCOND	0	2131	1118	1013	452	144	4858
		43.9%	23.0%	20.9%	9.3%	3.0%	100%
	1	4346	2437	1807	677	92	9359
		46.4%	26.0%	19.3%	7.2%	1.0%	100%
	Total	6477	3555	2820	1129	236	14217
		45.6%	25.0%	19.8%	7.9%	1.7%	100%
CURVE	0	5950	3316	2574	1016	209	13065
		45.5%	25.4%	19.7%	7.8%	1.6%	100%
	1	527	239	246	113	27	1152
		45.7%	20.7%	21.4%	9.8%	2.3%	100%
	Total	6477	3555	2820	1129	236	14217
		45.6%	25.0%	19.8%	7.9%	1.7%	100%
URBAN	0	787	408	491	252	75	2013
		39.1%	20.3%	24.4%	12.5%	3.7%	100%
	1	5690	3147	2329	877	161	12204
		46.6%	25.8%	19.1%	7.2%	1.3%	100%
	Total	6477	3555	2820	1129	236	14217
		45.6%	25.0%	19.8%	7.9%	1.7%	100%
AGE_AT_FAULT	1	1391	891	748	267	59	3356
		41.4%	26.5%	22.3%	8.0%	1.8%	100%
	2	4496	2330	1795	730	147	9498
		47.3%	24.5%	18.9%	7.7%	1.5%	100%
	3	590	334	277	132	30	1363
		43.3%	24.5%	20.3%	9.7%	2.2%	100%
	Total	6477	3555	2820	1129	236	14217
		45.6%	25.0%	19.8%	7.9%	1.7%	100%

Table 4.6 (Continued)

Frequency Row %	Value	Crash severity 1	2	3	4	5	Total
HARMEVN	1	2110	1782	1043	314	33	5282
		39.9%	33.7%	19.7%	5.9%	0.6%	100%
	2	708	410	388	172	34	1712
		41.4%	23.9%	22.7%	10.0%	2.0%	100%
	3	1183	226	144	43	5	1601
		73.9%	14.1%	9.0%	2.7%	0.3%	100%
	4	2476	1137	1245	600	164	5622
		44.0%	20.2%	22.1%	10.7%	2.9%	100%
	Total	6477	3555	2820	1129	236	14217
		45.6%	25.0%	19.8%	7.9%	1.7%	100%
ALDGUSE_AT_FAULT	0	6140	3445	2696	1050	154	13485
		45.5%	25.5%	20.0%	7.8%	1.1%	100%
	1	337	110	124	79	82	732
		46.0%	15.0%	16.9%	10.8%	11.2%	100%
	Total	6477	3555	2820	1129	270	14251
		45.4%	24.9%	19.8%	7.9%	1.9%	100%
VEHTYPE	0	5141	3223	2553	1030	192	12139
		42.4%	26.6%	21.0%	8.5%	1.6%	100%
	1	1336	332	267	99	44	2078
		64.3%	16.0%	12.8%	4.8%	2.1%	100%
	Total	6477	3555	2820	1129	236	14217
		45.6%	25.0%	19.8%	7.9%	1.7%	100%

4.3.3 Estimation Results

The estimation of results of the ordinal probit regression is given in Table 4.7. The sample size is 14,217 observations, and the Likelihood Ratio (LR) test statistic falls into the rejection area ($p - value = 0 < 0.05$). That means the overall explanatory variables of the model have significant influence on the responses (crash severity levels) at a statistical significance level 95%. Except for ANGLE, all slope coefficients are significant at a confidence level 0.05. Although the *p - value* of ANGLE is little greater than 0.05, the variable was still included in the model since angle crash was an important crash type and more variables increase the explanation ability of the model.

Table 4.7 Estimation of Ordered Probit Regression for Work Zone Crash Severity Model

Ordered probit regression					Number of observation = 14217	
					LR chi2(12) = 1094.6	
					Prob > chi2 = 0.000	
Log likelihood = -17861.331					Pseudo R^2 = 0.0297	
ACCISEV	Coef.	Std. Err.	z	P>z	[95% Conf.Interval]	
LGHTCOND	-0.0981	0.0206	-4.77	0.000	-0.1384	-0.0578
CURVE	0.0818	0.0344	2.38	0.018	0.01432	0.1494
URBAN	-0.1768	0.0308	-5.74	0.000	-0.2372	-0.1164
VEHTYPE	-0.3846	0.0295	-13.02	0.000	-0.4425	-0.3267
ALDGUSE_AT_FAULT	0.2096	0.0430	4.87	0.000	0.1252	0.2939
YOUNG_AGE	0.0506	0.0224	2.26	0.024	0.0067	0.0945
OLD_AGE	0.1229	0.0326	3.77	0.000	0.0590	0.1867
REAR-END	-0.0752	0.0217	-3.47	0.001	-0.1178	-0.0327
ANGLE	0.0569	0.0305	1.87	0.062	-0.0028	0.1166
SIDESWIPE	-0.7253	0.0363	-19.98	0.000	-0.7964	-0.6541
SECADT	-0.3851	0.0656	-5.87	0.000	-0.5136	-0.2565
MAXSPEED	0.7702	0.0742	10.38	0.000	0.6248	0.9156
/cutpoint1	0.0434	0.0677			-0.0892	0.1761
/cutpoint2	0.7261	0.0679			0.5931	0.8591
/cutpoint3	1.5236	0.0686			1.3892	1.6579
/cutpoint4	2.3867	0.0722			2.2452	2.5281

Based on the estimated results in Table 4.7, the probability models for five crash injury severity levels are given as:

$$\Pr(y_i = 1 | x_i) = \Phi(\tau_1 - \beta x_i)$$

$$\Pr(y_i = 2 | x_i) = \Phi(\tau_2 - \beta x_i) - \Phi(\tau_1 - \beta x_i)$$

$$\Pr(y_i = 3 | x_i) = \Phi(\tau_3 - \beta x_i) - \Phi(\tau_2 - \beta x_i) \quad (4.22)$$

$$\Pr(y_i = 4 | x_i) = \Phi(\tau_4 - \beta x_i) - \Phi(\tau_3 - \beta x_i)$$

$$\Pr(y_i = 5 | x_i) = 1 - \Phi(\tau_4 - \beta x_i)$$

where τ is the cutpoint, and β is the coefficient of the corresponding variable.

4.3.4 Interpretation

The crash severity model estimated by the ordinal probit regression has the same slope coefficients across all severity levels. For example, the coefficient for LGHTCOND is -0.0981 and the standardized coefficient for it is -0.0931, which means that the presence of day light (LGHTCOND = 1) tends to reduce the injury severity of work zone crashes, and when driving in daylight condition, the probability of having a higher injury severity crash is 0.0931 standard deviations lower than in non-daylight condition, holding all other variables constant. Table 4.8 and 4.9 shows the estimated results of the partial changes in y^* and in predicted probabilities for this ordered model respectively.

Table 4.8 Partial Change in y^*

ACCISEV	Coef.	z	P>z	y standardized coef.
LGHTCOND	-0.0981	-4.773	0.000	-0.0931
CURVE	0.0818	2.38	0.018	0.0777
URBAN	-0.1768	-5.74	0.000	-0.1678
VEHTYPE	-0.3846	-13.02	0.000	-0.3650
ALDGUSE_AT_FAULT	0.2096	4.87	0.000	0.1989
YOUNG_AGE	0.0506	2.26	0.024	0.0480

Table 4.8 (Continued)

OLD_AGE	0.1229	3.77	0.000	0.1166
REAR-END	-0.0752	-3.47	0.001	-0.0714
ANGLE	0.0569	1.87	0.062	0.0541
SIDESWIPE	-0.7253	-19.98	0.000	-0.6882
SECADT	-0.3851	-5.87	0.000	-0.3654
MAXSPEED	0.7702	10.38	0.000	0.7308

Table 4.9 Partial Change in Predicted Probabilities

	No Injury	Possible Injury	Non-incapacitating Injury	Incapacitating Injury	Fatal Injury
LGHTCOND	0.0388	-0.0053	-0.0179	-0.0123	-0.0034
P>z	0.000	0.000	0.000	0.000	0.000
CURVE	-0.0323	0.0040	0.0149	0.0105	0.0029
P>z	0.017	0.005	0.017	0.022	0.029
URBAN	0.0694	-0.0074	-0.0318	-0.0234	-0.0068
P>z	0.000	0.000	0.000	0.000	0.000
VEHTYPE	0.1524	-0.0333	-0.0693	-0.0403	-0.0096
P>z	0.000	0.000	0.000	0.000	0.000
ALDGUSE_AT_FAULT	-0.0817	0.0072	0.0374	0.0286	0.0086
P>z	0.000	0.000	0.000	0.000	0.000
YOUNG_AGE	-0.0200	0.0027	0.0092	0.0063	0.0017
P>z	0.024	0.017	0.024	0.026	0.030
OLD_AGE	-0.0483	0.0056	0.0222	0.0160	0.0045
P>z	0.000	0.000	0.000	0.000	0.001
REAR-END	0.0299	-0.0045	-0.0137	-0.0092	-0.0024
P>z	0.001	0.001	0.001	0.000	0.000
ANGLE	-0.0225	0.0030	0.0104	0.0072	0.0020
P>z	0.061	0.039	0.061	0.068	0.076
SIDESWIPE	0.2793	-0.0792	-0.1231	-0.0632	-0.0138
P>z	0.000	0.000	0.000	0.000	0.000
SECADT	0.1527	-0.0222	-0.0703	-0.0475	-0.0127
P>z	0.000	0.000	0.000	0.000	0.000
MAXSPEED	-0.3054	0.0443	0.1406	0.0950	0.0255
P>z	0.000	0.000	0.000	0.000	0.000

4.3.4.1. Signs

In the Tables 4.8, the variables recording daylight condition, urban area, heavy vehicle involved, rear-end crash type, sideswipe crash type and average annual daily traffic have negative coefficients, that means when the value of these variables increase, the crash injury severity is more likely to be slight. In contrast, the increase of other variables with positive coefficients tends to make a higher probability of more severe injury crashes. The summary is in Table 4.10

Table 4.10 Analysis of the Coefficient Signs

Independent Variable	Sign	Influence for Crash Severity Level
LGHTCOND	-	← Decrease
CURVE	+	→ Increase
URBAN	-	← Decrease
VEHTYPE	-	← Decrease
ALDGUSE_AT_FAULT	+	→ Increase
YOUNG_AGE	+	→ Increase
OLD_AGE	+	→ Increase
REAR-END	-	← Decrease
ANGLE	+	→ Increase
SIDESWIPE	-	← Decrease
SECADT	-	← Decrease
MAXSPEED	+	→ Increase

4.3.4.2. Magnitude of Coefficients

The injury severity level y^* is specified as a linear function of the independent variables, the relative magnitudes of estimated variable coefficients are, in most cases, a measure of the relative impacts of these variables on the average severity level of injury severity (O'Donnell and Connor, 1996). For example, the increase in injury severity of an old driver is about 2.43 times higher than the increase in injury severity of a young driver, all other things being equal, because the estimated coefficient of the variable OLD_AGE ($\hat{\beta} = 0.1229$) is about 2.43 larger than the estimate of the coefficient of the variable YOUNG_AGE ($\hat{\beta} = 0.0506$). Then, the estimated variable coefficients can be compared in this way and the influences of different variables on average injury severity level can be ranked (see Table 4.11).

Table 4.11 Ranked Magnitudes of Coefficients

Rank	Independent Variable	Coefficient (Positive)	Independent Variable	Coefficient (Negative)
1	MAXSPEED	0.7702	SIDESWIPE	-0.7253
2	ALDGUSE_AT_ FAULT	0.2096	SECADT	-0.3851
3	OLD_AGE	0.1229	VEHTYPE	-0.3846
4	CURVE	0.0818	URBAN	-0.1768
5	ANGLE	0.0569	LGHTCOND	-0.0981
6	YOUNG_AGE	0.0506	REAR-END	-0.0752

4.3.4.3. Detailed Interpretations

(1) Under good lighting conditions (such as daylight), the work zone crash severity is more likely to decrease.

(2) A curved design at the work zone sections, which means the driving condition turns to be difficult, is easily to result in a severe sever crash.

(3) In urban work zone area, the level of crash injury tends to decrease. It may because of the lower driving speed.

(4) Heavy vehicle involved can induce to less sever crashes. This is not the same as we think usually. The reason might be that the most people drive carefully when there is a truck around them.

(5) Alcohol and drugs tend to increase the crash injury severity level.

(6) In two special age groups, young age drivers who are more aggressive and have less experience and old age drivers whose physical, visual, and cognitive abilities may deteriorate are easily involved into severe crashes. But the influence of the old age is more than which of the young age.

(7) Two major crash types in work zone area, read-end and sideswipe may not contribute directly hurt to drivers, so if these two types of crashes happen, the probability of having injury would decrease. The condition of angle crash type occurring is totally contrary. The impact of the sideswipe crashes is much more than the impact of the rear-end crashes (0.7253 / 0.0752 = 9.64).

(8) The increase of maximum speed limit tends to increase the crash severity level and the condition is totally contrary to the variable AADT.

(9) According to the different magnitudes of estimated variable coefficients, the increase of maximum posted speed ($\hat{\beta} = 0.7702$) has the highest impact to increase the crash severity level, which is the 3.67 times higher than the second ranked variable ALDGUSE_AT_FAULT ($\hat{\beta} = 0.2096$). In contrast, the sideswipe crash type ($\hat{\beta} = -0.7253$) has the highest impact to reduce the crash severity level, which is the 1.88

times higher than the second ranked variables SECADT ($\hat{\beta} = -0.3851$), and VEHTYPE ($\hat{\beta} = -0.3846$).

4.3.5 Possible Countermeasures to Improve Work Zone Safety

Since the explanatory variables are the factors which have significant influence on the crash severity, the countermeasures can be suggested based on the variables in the models.

(1) Driving in daylight can reduce crash severity level, so a good lighting condition is important for work zone safety, especially during the nighttime periods. When nighttime work is being performed, floodlights should be used to illuminate in work zones, but the disabling glare condition for approaching road users which might be produced should be noticed.

(2) Be careful the work zone transition beginning in existing horizontal curve. We can keep continuous curve radii on work zone transitions which can help drivers from overestimating the appropriate speed, resulting in fewer runoff-the-road crashes, or move transition upstream so that it does not start in an existing horizontal curve instead.

(3) Speed limit is to keep drivers at a constant safe speed in work zones. Several other signs besides regular speed limit sign such as speed feedback signs and changeable message signs with radar (CMR) can be used.

Speed feedback signs usually measure using radar and display an individual vehicle's speed. These signs can only display speed, but several have the capability of displaying other text, such as "Slow Down."

CMR displays warning messages when a vehicle is traveling at an unsafe speed. The standard message on the CMS unit changes when a vehicle is traveling faster than the programmed speed, typically 3 mph above the speed limit. The messages used might included: "YOU ARE SPEEDING, SLOW DOWN," "HIGH SPEED, SLOW DOWN," "REDUCE SPEED IN WORK ZONE," and "EXCESSIVE SPEED, SLOW DOWN."

CHAPTER FIVE

SUMMARY

5.1 Summary

The main objectives of this study are to investigate the characteristics of accidents in work zones, to identify the factors contributing to injury severity levels, and to study how the factor influence injury levels. To achieve this purpose, two different statistics are processed. One is descriptive statistics and the other ordered regression modeling.

Descriptive statistic analysis was used to get the distribution of work zone crashes over three age groups for various factors which were paid attentions by researchers. In this part, crash severity level, environmental conditions, crash types, contributing factors, heavy vehicle involvement, and alcohol/drugs involvement were discussed over age groups, in some characteristics even the distribution between work zone and non-work zone were compared. The main results are:

(1) In work zone area, the middle age drivers cause the highest proportion (67%) of crashes, while in non-work zone area they have a lower possibility of occurring crashes (63%).

(2) Middle age drivers involved the highest percentage in the no injury crashes which is 49%, and always has the lowest one in other crashes. While in the more severe level crashes, elderly drivers contribute more than the other two age groups

(3) Rear-end, angle and sideswipe are the principle crash types in all three age groups. In young and middle age groups, the percentage of rear-end crashes is obviously higher than angle and sideswipe crashes, and elderly age group shows higher rate in angle crashes than others. Read-end and sideswipe crashes are more likely to be occurred in work zone area.

(4) The most predominant factor for work zone crashes is careless driving, and others are failed to yield right of way, no improper driving action and improper lane change in all age groups. But in elderly age group, the distribution (proportion and rank) has slight difference. In the distribution of predominant contributing factors over the principal crash types, careless driving, failed to yield right of way, and improper lane change are three most predominant contributing factor for rear-end, angle, and sideswipe crashes respectively.

(5) Heavy vehicle is more easily related to work zone crashes (14%) than non-work zone crashes (7%). Most driver especially old driver is not influenced by alcohol/drugs.

Crash severity is an important criterion reflecting the cost of work zone crashes in social and economy, and affected by various factors including driver's characteristics, vehicle characteristics, environmental factors, and roadway features. A full understanding of the impacts of the factors on the crash severity is beneficial to select proper countermeasure for reducing the crash severity at work zones and decrease the loss of construction/maintenance on roadway. A probit regression for ordinal output was used to estimate the crash severity models for overall work zone crashes. Based on the results of crash severity modeling and analysis, some conclusions can be obtained:

(1) According to the ordered probit model for work zone crash severity, lighting condition, road section with curves, urban or rural area, heavy vehicle involved, alcohol/drug involvement, young and old age group, three predominant crash types, AADT and maximum posted speed have the main influence to work zone crash severity.

(2) The factors of daylight condition, urban, rear-end crash type, sideswipe crash type and high average annual daily traffic are more likely to reduce the severity of work zone crashes.

(3) In contrast to the common sense, heavy vehicle involved could induce work zone crash severity. That's maybe because of driving carefully when there is a truck or tractor around.

(4) Based on the magnitudes of the variable coefficients, the variables of maximum posted speed and the sideswipe crash have the major impact to crash severity level. That shows restriction to driving speed is principle factor for work zone safety.

Based on these statistical analyses for work zone crashes, several countermeasures can be given:

(1) Floodlights needs to be used to illuminate in work zones in the nighttime in order to build a good lighting condition.

(2) Discourage traffic control plan designs that include transition areas for the work

zone on an existing horizontal curve, and encourage that the transition be accomplished on a tangent section instead.

(3) Speed limit signs are very important for work zone safety. Some dynamic signs like changeable message signs with radar and speed feedback signs have better effectiveness to reduce driver speed.

REFERENCES

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

Abdel-aty, M., Chen, C., and Schott, J. R. An assessment of the effect of driver age on traffic accident involvement using log-linear models. Accident Analysis and Prevention. 30(6), 851–861., 1998.

Abdel-Aty, M., Keller, J. Exploring the overall and specific crash severity levels at signalized intersections. Accident; analysis and prevention. 37(3), 417-25., 2005.

Chang, L., & Mannering, F. Analysis of Injury Severity and Vehicle Occupancy in Truck and Non-truck-Involved Accidents. Accident Analysis and Prevention. 31, 579–592., 1999.

Daniel, J., Dixon, J., and Jared, D. Analysis of fatal crashes in Georgia work zones. Transportation Research Record 1715. Transportation Research Board, Washington, D.C., pp. 18–23., 2000.

Dissanayake, S., and Lu, J., Analysis of Severity of Young Driver Crashes: Sequential Binary Logistic Regression Modeling. In Transportation Research Record: Journal of the Transportation Research Board, No. 1784, TRB, National Research Council, Washington, D.C., pp. 108–114., 2002.

Donnell, E., and Mason, J. Predicting the Severity of Median-Related Crashes in Pennsylvania by Using Logistic Regression. Transportation Research Record, No. 1897, TRB, National Research Council, Washington, D.C., pp. 55-63., 2004.

Duncan, C., Khattak, A., & Council, F. Applying the Ordered Probit Model to Injury Severity in Truck-Passenger Car Rear-End Collisions. Transportation Research Record, vol. 1635, pp. 63– 71., 1999.

Garber, N. J. and Zhao, M. Distribution and Characteristics of Crashes at Different Locations within Work Zones in Virginia. Research Report VTRC 02-R12, Virginia Transportation Research Council, Charlottesville, Virginia, 2002.

Hutchinson, T. Statistical Modeling of Injury Severity,With Special reference to Driver and Front Seat Passenger in Single-Vehicle Crashes. Accident Analysis and Prevention, 18(2), 157– 167., 1986.

Khattak, A. Effect of Information and Other Factors on Multivehicle Rear-end Crashes: Crash Propagation and Injury Severity. Presented at the 78th Annual Meeting of the Transportation Research Board, Washington, D.C., 1999.

Khattak, A.J., Khattak, A.J., and Council, F.M. Effects of work zone presence on injury and non-injury crashes, Accident Analysis and Prevention. Vol. 34(1), pp. 19-29., 2002.

Khattak, A., Pawlovich, M., Souleyrette, R., Hallmark, S. Factors related to more severe older driver traffic crash injuries. J. Transportation Eng. 128 (3), 243–249., 2002.

Khattak, A., Schneider, R., Targa, F., Risk factors in large truck rollovers and injury severity: analysis of single-vehicle collisions. In: Presented at the 82nd Annual Meeting of the Transportation Research Board, Washington, DC., 2003.

Khattak, A., Targa, F., Injury severity and total harm in truck-involved work zone crashes. In: Presented at the 83rd Annual Meeting of the Transportation Research Board, Washington, DC., 2004.

Kim, K., L. Nitz, J. Richardson, and L. Li. Personal and Behavioral Predictors of Automobile Crash and Injury Severity. Crash Analysis and Prevention. Vol. 27, No. 4, pp. 469–481., 1995.

Klop, J. Factors Influencing Bicycle Crash Severity on Two-Lane Undivided Roadways in North Carolina. Presented at the 78th Annual Meeting of the Transportation Research Board, Washington, D.C., 1998.

Kockelman, K., Kweon, Y. Driver injury severity: an application of ordered probit models. Accident Analysis and Prevention. 34 (3), 313-321(9)., 2002.

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

Maze, T. Kamyab, A. and Schrock, S. Evaluation of Work Zone Speed Reduction Measures. Final Contract Report. Iowa Department of Transportation, IA, 2000.

McFadden, J., Yang, W. T., and Durrans, S. R. Application of Artificial Neural Networks to Predict Speeds on Two-Lane Rural Highways. Transportation Research Record. No. 1751, Transportation Research Board, National Research Council, Washington D.C., 2005.

Mohan, S. B. and Gautam, P. Cost of Highway Work Zone Injuries. Practical Periodical on Structural Design and Construction. Vol. 7, No.2, 2002, pp68 – 73., 2002.

Nassar, S., Saccomanno, F., and Shortreed, J. Road Accident Severity Analysis: A Micro Level Approach. Canadian Journal of Civil Engineering, 21, 847– 855., 1994.

O'Donnell, C., and Connor, D. Predicting the Severity of Motor Vehicle Accident Injuries Using Models of Ordered Multiple Choices. Accident Analysis and Prevention, 28(6), 739– 753., 1996.

Renski, H., Khattak, A., and Council, F. Impact of Speed Limit Increases on Crash Injury Severity: Analysis of Single-Vehicle Crashes on North Carolina Interstate Highways. Presented at the 78th Annual Meeting of the Transportation Research Board, Washington, D.C., 1998.

Schrick, S., et al. An analysis of Fatal Work Zone Crashes in Texas. Texas Transportation Institute. No. 0-4028-1., 2004.

Shankar, V., and Mannering, F. An Exploratory Multinomial Logit Analysis of Single-Vehicle Motorcycle Accident Severity. Journal of Safety Research, 27(3), 183–194., 1996.

Shankar, V., Mannering, F., and Barfield, W. Statistical Analysis of Accident Severity on Rural Freeways. Accident Analysis and Prevention. 28(3), 391–401., 1996.

Toshiyuki, Y., and Shankar, V. Bivariate Ordered-Response Probit Model of Driver's and Passenger's Injury Severities in Collision with Fixed Objects. Presented at the 81th Annual Meeting of the Transportation Research Board, Washington, D.C., 2002.

Ullman, G., et al. Analysis of Crashes at Active Night Work Zones in Texas Presented at the Annual Meeting of the Transportation Research Board, Washington, D.C., 2005.

Wang, J., Hughes, W.E., Council, F.M., and Paniati, J.F. Investigation of highway work zone crashes: what we know and what we don't know, Transportation Research Record 1529, pp. 54-62., 1996.

BIBLIOGRAPHY

FHWA. Manual on Uniform Traffic Control Devices for Streets and Highways. 2003 Edition, Federal Highway Administration. U.S. Department of Transportation, 2003.

Greene H. Econometric Analysis. Second Edition, 1993.

Long. J. Regression Models for Categorical and Limited Dependent Dependent Variables, 1997.

Long, J. and Freese, J. Regression Models for Categorical Dependent Variables Using Stata. Second Edition, 2006.

Powers, D. and Xie, Y. Statistical Methods for Categorical Data Analysis, 2000

Transportation Research Board, Highway Capacity Manual, 4th edition, National Research Council, Washington, D.C., 2000.

Washington, S., Karlaftis, M., and Mannering, F. Statistical and Econometric Method for Transportation Data Analysis, 2003

APPENDICES

Appendix A: Variables and Codes of Work Zone Crash

Table A-1 Variable of Work Zone Crashes

Variable	Description	Type
YEAR	The year of work zone fatal crash	Nominal
TIME	The time of work zone fatal crash	Nominal
AGE	The age of driver at fault	Ordinal
VEHMOVEMENT	The movement of vehicle at fault before accident	Nominal
CRASHTYPE	The type of crash	Nominal
VEHICLETYPE	Heavy vehicle involved?	Nominal
FUNCLASS	The function of roads	Nominal
TRWAYCHR	Road Characteristics (level / curve?)	Nominal
MAXSPEED	The speed limit	Continue
SECTADT	The AADT of the section of work zones	Continue
TYPESUR	The type of road surface	Nominal
SITELOCA	Site Location	Nominal
LIGHTCONDITION	Light condition	Nominal
WEATHERCONDITION	Weather condition	Nominal
ROADSURFACE	Road surface condition	Nominal
VISION	Vision Obstructed	Nominal
RDACCESS	Access control type	Nominal
SURWIDTH	The width of roads	Continue
CONTRIBUTINGFACTORS	The contributing factors	Nominal
TRAFCONT	Traffic Control	Nominal

Table A-2 Codes for TIME

Codes	Description
1	6:00-10:00
2	10:00-16:00
3	16:00-20:00
4	20:00-6:00

Appendix A (Continued)

Table A-3 Codes for AGE

Codes	Description
1	<19
2	20-24
3	25-34
4	35-44
5	45-54
6	55-64
7	>65

Table A-4 Codes for VEHMOVEMENT

Codes	Description
01	STRAIGHT AHEAD
02	SLOWING/STOPPED/STALLED
03	MAKING LEFT TURN
04	BACKING
05	MAKING RIGHT TURN
06	CHANGING LANES
07	ENTERING/LEAVING PARKING SPACE
08	PROPERLY PARKED
09	IMPROPERLY PARKED
10	MAKING U-TURN
11	PASSING
12	DRIVERLESS OR RUNAWAY VEH.
77	ALL OTHERS
88	UNKNOWN

Appendix A (Continued)

Table A-5 Codes for CRASHTYPE

Codes	Description
01	COLL. W/MV IN TRANS. REAR-END
02	COLL. W/MV IN TRANS. HEAD-ON
03	COLL. W/MV IN TRANS. ANGLE
04	COLL. W/MV IN TRANS. LFT-TURN
05	COLL. W/MV IN TRANS. RGT-TURN
06	COLL. W/MV IN TRANS. SIDESWIP
07	COLL. W/MV IN TRANS. BAKD INTO
08	COLL. W/PARKED CAR
09	COLLISION WITH MV ON ROADWAY
10	COLL. W/ PEDESTRIAN
11	COLL. W/ BICYCLE
12	COLL. W/ BICYCLE (BIKE LANE)
13	COLL. W/ MOPED
14	COLL. W/ TRAIN
15	COLL. W/ ANIMAL
16	MV HIT SIGN/SIGN POST
17	MV HIT UTILITY POLE/LIGHT POLE
18	MV HIT GUARDRAIL
19	MV HIT FENCE
20	MV HIT CONCRETE BARRIER WALL
21	MV HIT BRDGE/PIER/ABUTMNT/RAIL
22	MV HIT TREE/SHRUBBERY
23	COLL. W/CONSTRCTN BARRICDE/SGN
24	COLL. W/TRAFFIC GATE
25	COLL. W/CRASH ATTENUATORS
26	COLL. W/FIXED OBJCT ABOVE ROAD
27	MV HIT OTHER FIXED OBJECT
28	COLL. W/MOVEABLE OBJCT ON ROAD
29	MV RAN INTO DITCH/CULVERT
30	RAN OFF ROAD INTO WATER
31	OVERTURNED
32	OCCUPANT FELL FROM VEHICLE
33	TRACTOR/TRAILER JACKNIFED
34	FIRE
35	EXPLOSION
36	DOWNHILL RUNAWAY
37	CARGO LOSS OR SHIFT

Appendix A (Continued)

Table A-5 (Continued)

38	SEPARATION OF UNITS
39	MEDIAN CROSSOVER
77	ALL OTHER (EXPLAIN)

Table A-6 Codes for VEHICLETYPE

Codes	Description
00	UNKNOWN/NOT CODED
01	AUTOMOBILE
02	PASSENGER VAN
03	PICKUP/LIGHT TRUCK (2 REAR TIR)
04	MEDIUM TRUCK (4 REAR TIRES)
05	HEAVY TRUCK (2 OR MORE REAR AX)
06	TRUCK TRACTOR (CAB)
07	MOTOR HOME (RV)
08	BUS (DRIVER + 9 - 15 PASS)
09	BUS (DRIVER + > 15 PASS)
10	BICYCLE
11	MOTORCYCLE
12	MOPED
13	ALL TERRAIN VEHICLE
14	TRAIN
15	LOW SPEED VEHICLE
77	OTHER
88	PEDESTRIAN NO VEHICLE

Table A-7 Codes for TRWAYCHR

Codes	Description
1	STRAIGHT-LEVEL
2	STRAIGHT-UPGRADE/DOWNGRADE
3	CURVE-LEVEL
4	CURVE-UPGRADE/DOWNGRADE

Appendix A (Continued)

Table A-8 Codes for TYPESUR

Codes	Description
01	SLAG/GRAVEL/STONE
02	BLACKTOP
03	BRICK/BLOCK
04	CONCRETE
05	DIRT
77	ALL OTHER

Table A-9 Codes for SITELOCA

Codes	Description
01	NOT AT INTERSECTION/RRX/BRIDGE
02	AT INTERSECTION
03	INFLUENCED BY INTERSECTION
04	DRIVEWAY ACCESS
05	RAILROAD CROSSING
06	BRIDGE
07	ENTRANCE RAMP
08	EXIT RAMP
09	PARKING LOT/TRAFFIC WAY
10	PARKING LOT AISLE OR STALL
11	PRIVATE PROPERTY
12	TOLL BOOTH
13	PUBLIC BUS STOP ZONE
77	ALL OTHER

Table A-10 Codes for LIGHTCONDITION

Codes	Description
01	DAYLIGHT
02	DUSK
03	DAWN
04	DARK (STREET LIGHT)
05	DARK (NO STREET LIGHT)
88	UNKNOWN

Appendix A (Continued)

Table A-11 Codes for WEATHERCONDITION

Codes	Description
01	CLEAR
02	CLOUDY
03	RAIN
04	FOG
77	ALL OTHER
88	UNKNOWN

Table A-12 Codes for ROADSURFACE

Codes	Description
01	DRY
02	WET
03	SLIPPERY
04	ICY
77	ALL OTHER
88	UNKNOWN

Table A-13 Codes for VISION

Codes	Description
01	VISION NOT OBSCURED
02	INCLEMENT WEATHER
03	PARKED/STOPPED VEHICLE
04	TREES/CROPS/BUSHES
05	LOAD ON VEHICLE
06	BUILDING/FIXED OBJECT
07	SIGNS/BILLBOARDS
08	FOG
09	SMOKE
10	GLARE
77	ALL OTHER (EXPLAIN)

Appendix A (Continued)

Table A-14 Codes for RDACCESS

Codes	Description
1	FULL
2	PARTIAL
3	NONE

Table A-15 Codes for CONTRIBUTINGFACTORS

Codes	Description
01	NO IMPROPER DRIVING/ACTION
02	CARELESS DRIVING
03	FAILED TO YEILD RIGHT OF WAY
04	IMPROPER BACKING
05	IMPROPER LANE CHANGE
06	IMPROPER TURN
07	ALCOHOL-UNDER INFLUENCE
08	DRUGS-UNDER INFLUENCE
09	ALCOHOL DRUGS-UNDER INFLUENCE
10	FOLLOWED TOO CLOSELY
11	DISREGARDED TRAFFIC SIGNAL
12	EXCEEDED SAFE SPEED LIMIT
13	DISREGARDED STOP SIGN
14	FAILED TO MAINTAIN EQUIP/VEHIC
15	IMPROPER PASSING
16	DROVE LEFT OF CENTER
17	EXCEEDED STATED SPEED LIMIT
18	OBSTRUCTING TRAFFIC
19	IMPROPER LOAD
20	DISREGARDED OTHER TRAFFIC CONT
21	DRIVING WRONG SIDE/WAY
22	FLEEING POLICE
23	VEHICLE MODIFIED
24	DRIVER DISTRACTION
77	ALL OTHER (EXPLAIN)

Appendix A (Continued)

Table A-16 Codes for TRAFCONT

Codes	Description
01	NO CONTROL
02	SPECIAL SPEED ZONE
03	SPEED CONTROL SIGN
04	SCHOOL ZONE
05	TRAFFIC SIGNAL
06	STOP SIGN
07	YIELD SIGN
08	FLASHING LIGHT
09	RAILROAD SIGNAL
10	OFFICER/GUARD/FLAGMAN
11	POSTED NO U-TURN
12	NO PASSING ZONE
77	ALL OTHER