5-2010

Statistical modeling and analysis of injury severity sustained by occupants of passenger vehicles involved in crashes with large trucks

Ahmad Koupaenejad
University of Nevada, Las Vegas

STATISTICAL MODELING AND ANALYSIS OF INJURY SEVERITY SUSTAINED BY OCCUPANTS OF PASSENGER VEHICLES INVOLVED IN CRASHES WITH LARGE TRUCKS

by

Ahmad Kouhpaenejad

Bachelor of Science in Economics
ATU College of Economics, Tehran
1977

Bachelor of Science in Industrial Technology
California State University at Long Beach, Long Beach
1981

Master of Science in Industrial and System Engineering
University of Southern California, Los Angeles
1982

A dissertation submitted in partial fulfillment
of the requirements for the

Doctor of Philosophy in Engineering
Department of Civil and Environmental Engineering
Howard R. Hughes College of Engineering

Graduate College
University of Nevada, Las Vegas
May 2010

THE GRADUATE COLLEGE

We recommend the dissertation prepared under our supervision by

Ahmad Kouhpaenejad

entitled

Statistical Modeling and Analysis of Injury Severity Sustained by Occupants of Passenger Vehicle Involved in Crashes with Large Trucks

be accepted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Engineering
Civil and Environmental Engineering

Mohamed S. Kaseko, Committee Chair

Moses Karakouzian, Committee Member

Nader Ghafoori, Committee Member

Alexander Paz, Committee Member

Ashok Singh, Graduate Faculty Representative

Ronald Smith, Ph. D., Vice President for Research and Graduate Studies
and Dean of the Graduate College

May 2010

ABSTRACT

Statistical Modeling and Analysis of Injury Severity Sustained by Occupants of Passenger Vehicles Involved in Crashes with Large Trucks

by

Ahmad Kouhpaenejad

Mohamed Kaseko, Ph.D., Examination Committee Chair
Associate Professor of Civil and Environmental Engineering
University of Nevada, Las Vegas

Crashes are the result of complex interactions between several factors representing driver, roadway, vehicle, and environmental characteristics. Understanding to what degree each factor contributes to the severity of a crash is not a simple task. The outcomes of crashes in the US have been an average of 42,000 deaths and 3 million injuries per year. To better understand the role of significant contributors to crashes, three sets of models using multinomial logit and one using ordered probit were calibrated. To calibrate these models, forty two independent predictor variables including driver/occupants characteristics, crash environment at the crash location, crash characteristics, and vehicle characteristics were used. In total, twenty variables were found to be statistically significant.

The crash data used for this study was from the state of North Carolina. The obtained data included all the crash records for the year 2003. Vehicle dimensions were incorporated into the final database.

The contributions of this study were twofold: First, the evaluation of the impact of passenger-vehicle dimensions on the injury severity. The following is a condensed summary of the findings:

An increase in the vehicle front overhang was more likely to decrease the risk of suffering an evident injury for two groups: drivers age 66 and older, and the male drivers. In addition, an increase in the vehicle rear overhang was more likely to reduce the risk associated to fatal injury for three groups: female drivers, drivers age <= 25, and drivers age 66 and older. Further, an increase in the vehicle width was more likely to increase the risk of sustaining injury for drivers age 66 and older.

Second, although the findings of this research were consistent with other researches, some differences identified as discussed below.

An increase in vehicle weight increased the risk of sustaining a fatal injury for two groups: drivers age<= 25 and female drivers. Furthermore, an increase in number of occupants did not pose an extra risk of fatal injury for two groups: drivers age 46-65 and female drivers. Moreover, dark roads with no lighting posed an extra risk of sustaining a fatal injury for drivers age <=25, but posed the lowest risk of injuries for female drivers. Further, head on crashes imposed a higher risk of sustaining a fatal injury for two groups: drivers age<=25 and female drivers. Finally, roads with no divided medians posed a higher risk of injury for two groups: drivers age <= 25 and male drivers.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to take this opportunity to thank my advisor, mentor, and committee chair, Dr. Mohamed Kaseko for his continuous support, guidance, and motivation throughout my doctoral study. Let it be known that the extent of his analytical skills, knowledge, teaching style, integrity, and the patience never ceased to amaze me. No thank you would ever convey the magnitude of my deepest gratitude and true appreciation.

I would like to express my appreciation to Dr. Ashok Singh for his invaluable guidance and time that he devoted to this research. His vast knowledge in statistics and valuable suggestions in the modeling portion of this study were instrumental. I gratefully acknowledge the help of my other committee members Dr. Moses Karakouzian for his valuable questions and concerns, Dr. Nader Ghafoori for showing me the importance of focal points of the dissertation, and Dr. Alexander Paz for his numerous inputs.

I will always remember my friends and colleagues, including but not limited to Mr. Xuecai Xu, Dr. Saidi Suihi, Mr. Timur Mauga, and the students at the TRC which without their friendship this would have been a much harder road.

I further acknowledge the financial support of the state of Nevada through the UNLV department of Civil and Environmental Engineering for providing me with the funding through teaching and research.

I would like to extend my deepest thanks to Mr. Yusuf Mohamedshah and Mr. Behrang Hejazi of Highway Safety Information System in the Washington DC for providing me with the data for this study.

DEDICATION

I dedicate this dissertation to my wife and children, Ashkaan and Bita. Their continuous support and understanding throughout this journey has been my guiding light. Especially my children's cooperation and patience in the last few months have conveyed to me that their futures are brighter than the sun.

CHAPTER 1

INTRODUCTION

1.1 Background

Road safety is a major concern because of the economic and social impacts of traffic crashes. From 1990-2006, an average of 42,000 people per year have died from vehicle crashes in the USA (Bureau of Transportation Statistics (BTS), 2008) (Figure 1-1). As shown in the figure, since 1998 there has been an upward trend in the total number of fatalities. During the same period, an average of 3 million people were injured in over 6.3 million crashes (BTS, 2008). According to Blincoe et al. (2002), motor vehicle crashes cost $ 230.6 billion in the year 2000 in the US.

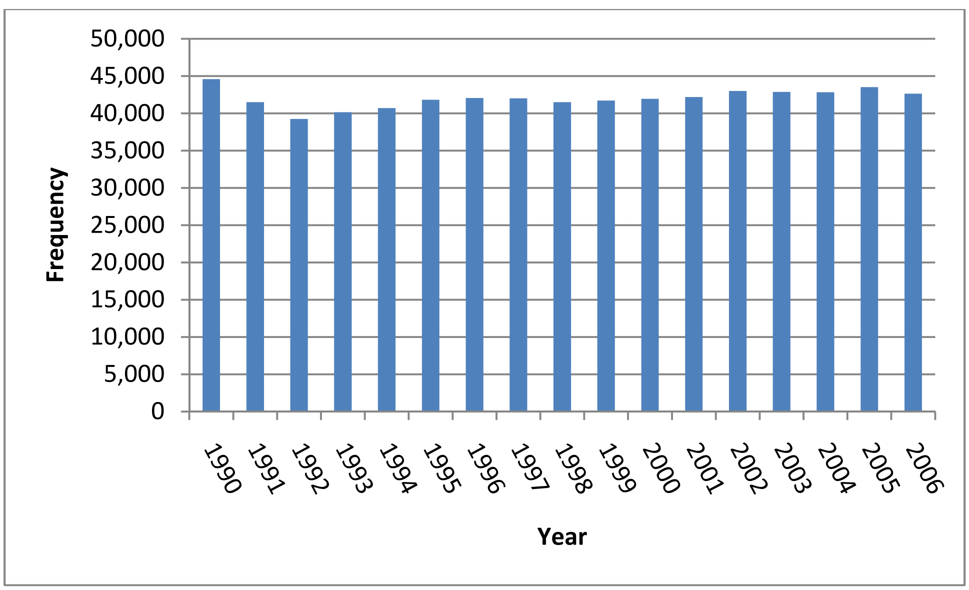

Figure 1-1 Crash fatality distribution in US

During the same period, an average of 3,800 fatalities and 127,000 injuries per year were due to crashes between large trucks (LT), trucks with gross vehicle weight greater than 10,000 lbs., and Passenger Vehicles (PV). Figure 1-2 illustrates the fatality distribution of PV occupants involved in crashes with LT. This study attempts to further the knowledge in this area.

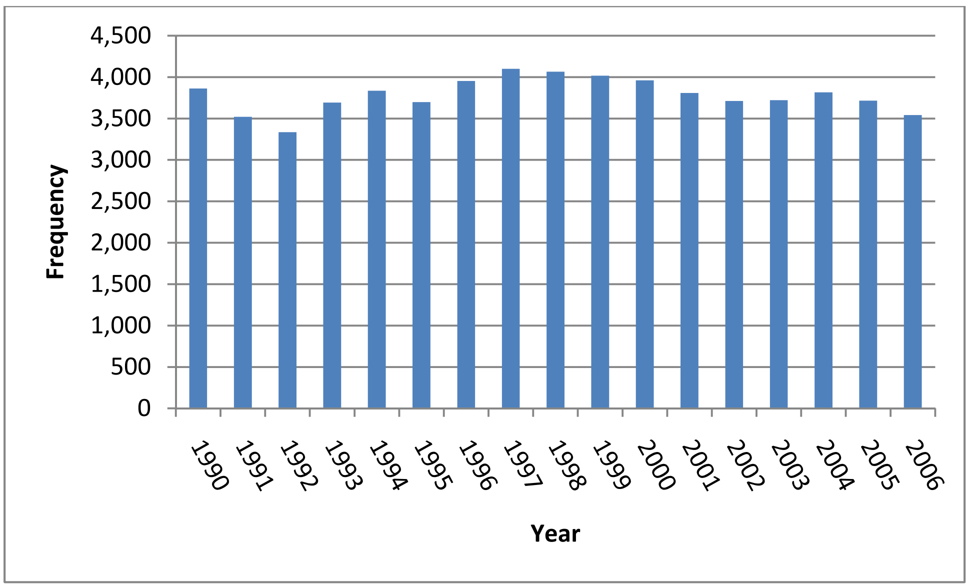

Figure 1-2 Passenger vehicle fatality distribution.

Large trucks are a very important mode of transportation for goods movement throughout the USA. According to BTS (2002), commercial truck traffic increased 75% over the past three decades, and this trend is likely to continue over the next 10 years.

Zaloshnja et al. (2004) stated that each crash costs $59,153 (in year 2000 dollars) and $88,483 for LTs and multiple combination LTs respectively. Their cost components include medical costs, emergency service costs, property damage, loss of productivity, monetized value of pain and suffering, and loss of quality of life due to injury or death.

Understanding the factors that contribute to injury as a result of crashes will help policy makers and road designers implement countermeasures, which could reduce crash injury severity, and cost. These alarming facts and trends convey the importance of further studies in road safety.

Crashes are the result of complex interactions between several factors representing driver, roadway, vehicle, and environmental characteristics. Understanding to what degree each factor contributes to the severity of a crash is not a simple task. Researchers have utilized mathematical and statistical modeling-schemes to solve this complex road-safety problem. In order to further knowledge in the field of road safety, this study will develop a statistical model for estimating the probabilities of an injury severity level to occupants of PVs involved in crashes with LTs. For this study PVs include passenger cars, pickup trucks, vans, and SUVs.

1.2 Research Objectives

The objectives of this study were to determine the impact of driver factors, roadway conditions, vehicle and environmental characteristics in injury severity sustained by occupants of PVs involved in crashes with LTs. To that extent, this research has attempted to:

(1) Identify explanatory variables including vehicle dimensions that were significant predictors of various injury severity levels using statistical models.

(2) Estimate the probability of injury severity level sustained by the most severely injured occupants of the PV, given a crash has occurred.

(3) Estimate the odds ratios and marginal impacts of each significant explanatory variable on the predicted probability of each injury severity level.

(4) Test an unprecedented number of explanatory variables in explaining occupants injury severity level.

The results of this study will provide invaluable direction to policy makers, traffic engineers, vehicle manufacturers, vehicle buyers, and insurance companies.

1.3 Dissertation Organization

This dissertation is divided into six chapters. Chapter 1 introduces the background of the problem and the research objectives. Chapter 2 presents a general review of the most relevant literature used in previous injury severity research. Chapter 3 presents the methodology, including a thorough description of the modeling concept. Chapter 4 describes data, data sources, and descriptive statistics. Chapter 5 presents and discusses model results. Chapter 6 presents conclusions and recommendations.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

This review was centered on the studies, where the injury severity was a target variable. Furthermore, two relevant issues were the focal points; first, models used to estimate the contribution of each significant variable to injury severity, second, the variables which were identified as significant determinants of injury severity.

A sufficient number of studies with injury severity as dependent variable were selected and reviewed.

The following sections are organized based on the modeling approaches used by the researchers.

2.2 Ordered Probability Models

Ordered Regression Models (ORM) include Ordered Probit (OP) and Ordered Logit (OL). In this modeling technique, the dependent variable is assumed to be categorical and ordinal.

O'Donnell and Connor (1996) were first researchers to apply ORM in road safety. They developed two models using OP and OL modeling methodologies. The study used the data from 1991 police reported crashes in New South Wales, Australia. Models could estimate probabilities of four injury severity levels for the vehicle occupants, given a crash has occurred. Among the selected explanatory variables, age of the occupant and vehicle speed led to slight increase in the probability of serious injury and death. The remaining independent variables in the model such as, blood alcohol level, vehicle type, vehicle make and type of collision had various influence on the probabilities of injury severity levels.

Duncan et al. (1998) applied OP modeling methodology to estimate injury severity sustained by passenger cars' occupants involved in rear-end crashes with trucks on divided roadways. The study used data from North Carolina crashes during 1993-1995. The following attributes were used as ordinary and dummy variables: passenger car rear impact, impact speed differential, impact speed differential & rear impact, station wagon & rear impact, speed limit, congestion (AADT/lane), AADT/lane not reported, grade, grade & wet road, darkness, lighted darkness, icy or snowy road surface, wet road surface, age under sixteen, child restraint, drinking driver, female, station wagon, defective brakes, and car rollover. Among these variables darkness, high speed differentials, high speed limits, grades, wet grades, being a female, and driving while drunk increased occupant injury severity. Furthermore, variables that decrease injury severity include snowy or icy road, young children being in a child restraint, and being in a station wagon struck.

Renski et al. (1999) researched the effect of speed limit increases on injury severity due to single vehicle crashes. They applied OP methodology, odds ratio tests, and quasi-experimental research design. The two years of data (1995 and 1997) from North Carolina was used in this study. It is worth noting that speed increase went into effect in 1996. In their design, they used the data from one year before the policy change and one year after. The results showed that increasing the speed from 55 mph to either 60 mph or 65 mph did affect the likelihood of evident and complaint of pain injuries. In the case of raising the speed limit from 65 mph to 70 mph no significant changes in the level of injuries were identified.

Khattak (2000) investigated the injury severity sustained by the drivers of two and three car rear-end crashes. He included the role of the information and vehicle technology in an OP modeling scheme. The data was obtained from the Highway Safety Information System (1994-1995) for North Carolina. Three separate OP models were estimated. Each model was conditioned on injury severity sustained by the driver of car 1, 2, or 3 (car 1 was the leading vehicle). From the data, in a two-vehicle crash, the leading vehicle-driver was likely to be injured more severely. In regard to three-vehicle crashes, the driver of the second car was likely to be injured more severely. The variable (technology) was statistically significant across three models.

Khattak et al. (2002) identified significant variables contributing to injury severity sustained by drivers 65 years of age and older. They applied OP modeling technique for estimating significant variables. They used the police reported crash data (1990-1999) from the State of Iowa. The following variables were statistically significant for this age group: drivers age, drivers gender, alcohol usage, level terrain, speed limit, farm vehicles, and crashes in the dark with no lighting.

Kockelman and Kweon (2002) calibrated six OP regression models to estimate various parameters corresponding to a set of preselected independent variables. Three sets of data from the 1998 National Automotive Sampling System (NASS) were extracted. The injury severities sustained by vehicle drivers under all crash types, two-vehicle crashes, and single vehicle crashes were analyzed. The results of the analysis indicated that the following variables played an important role: manner of collision, number of vehicles, gender, vehicle type, and driver alcohol use. Head-on collisions and rollovers were major contributors. Females sustained more severe injuries than males. In two vehicle crashes, light-duty trucks protect the drivers better. Moreover, pick-ups and SUVs cause more severe injuries on the colliding partner and they are more prone to rollover.

Abdel-Aty (2003) calibrated three models for driver injury severity levels for roadway sections, signalized intersections, and toll plazas in central Florida. 1996-1997 crash data from the central Florida area were used for both road way section and signalized intersection. Crash data from 1999-2000 was used for toll plazas. He applied OP modeling methodology and the results for the three models indicated that drivers age, gender, seatbelt use, point of impact, speed, and vehicle type were statistically significant on the drivers injury severity level. There were other statistically significant variables for each specific case. For signalized intersections drivers violation and for the case of toll plazas vehicles with electronic toll collection apparatus had higher impact on the probability of driver injury severity. In the roadway section model, lighting condition, alcohol, and the presence of horizontal curve affected the likelihood of injuries.

Donnell and Mason (2004) estimated median-related crash severity using OL and unordered regression models (ordered logistic and multinomial logit). Response variable (injury severity) was divided into three categories: no injury, injury, and fatality. The data for this study was collected from the field and provided by Pennsylvania Department of Transportation. The results showed that, explanatory variables such as daily traffic, crash cause, pavement surface, crash type, horizontal alignment, and presence of interchange entrance ramps affect the crash severity.

Ma and Kockelman (2004) accessed data from six freeways in the Orange County areas of southern California (1998) to estimate injury severity levels. They fitted OP into the data. Their findings were consistent with other researchers; namely, females and older persons are more at risk than others and dense traffic flow reduces the likelihood of sustaining severe injury.

Abdel-Aty and Keller (2005) explored explanatory predictors contributing to injuries occurred at signalized intersections. They used the data from central Florida areas (2000-2001). Several models were estimated using OP and tree-based regression methodology. The results of the study conveyed that left turn, pedestrians, and bicyclist crashes had the highest probability of more severe injury. Moreover, the speed limit, median, crash type, and, intersection characteristics on the minor road were significant in their final ordered probit model.

Wang and Kockelman (2005) used a Heteroscedastic Ordered Logit (HOL) and an OL for estimating injury severity sustained by vehicle occupants in single and two vehicle crashes. The major difference between the HOL and OL is that the HOL allows the error-term's variance to vary. They obtained (1998-2001) data from National Automotive Sampling System's Crashworthiness Data System (NASS CDS). The results of the model indicated that vehicle weight, seating position, seatbelts, crash type, posted speed limit, weather condition, roadway medians, and drivers age were significant determinants of injury severity.

Rifaat and Chor (2005) analyzed injury severity due to single-vehicle crashes in Singapore. They used police-reported accidents data from the year 1992-2001. The methodological approach in the study was the calibration of an OP model. The results of their study indicated that the most important determinants of the crash injury were the crash type, vehicle type, roadside objects, trees, expressways, night time, young and male drivers. Among the statistically significant variables, colliding with trees had the highest probability of fatal injury.

Wang and Abdel-Aty (2008) calibrated three partial proportional odds models (generalization of ordered regression) to estimate injury severities due to left-turn at signalized intersections. Six years of data was used from the central Florida area. The models showed that traffic entering the intersection, traffic of opposing approach, left turning traffic, left-turn lane offset, alcohol or drug use, drivers age (very young and very old), and point of impacts of both vehicles affect the injury severity.

Wang et al. (2009) calibrated OP and partial proportional odds models to identify variables contributing to injury severity at freeway diverge-areas in the state of Florida. They tested the parallel regression assumption of OP using Wald test and since the assumption did not hold, they used proportional odds model (relaxing the assumption). The significant variables were: crash type, surface condition, average daily traffic, number of lanes, length of deceleration lanes, light condition, weather condition, and alcohol/drug involvement.

Xie et al. (2009) investigated injury severity sustained by drivers of passenger cars, SUVs, and vans. They calibrated a Bayesian Ordered Probit (BOP) and an OP modeling scheme using data from the 2003 National Automotive Sampling System General Estimates System (NASSGES). The results from both the BOP and OP methodologies indicated that, the following factors were statistically significant: age, gender, alcohol, vehicle type, vehicle age, crash type, point of impact on the vehicle, crash location type, road surface, and lighting condition. To evaluate the impact of sample size, the authors reduced the sample size from 76,994 records to 100 and calibrated new models using both methodologies. The new results indicated that for small samples the BOP could be a better estimator than OP. One of the shortcomings of the BOP modeling methodology is that it requires a prior distribution assumption, which is subjective and difficult to determine. Variable coefficients lack t-values and corresponding confidence intervals.

2.3 Unordered Probability Models

In the case where the dependent variable was assumed to be categorical and unordered Multinomial Logit and its extensions were used.

Chang and Mannering (1999) estimated two separate nested logit models. The focus of their study was to investigate the relationship between injury severity and occupancy for truck-involved and non-truck-involved crashes. They used data from police reported crashes on principal arterials, state highways, and interstate highways in King County of Washington state during 1994. For truck-involved crashes high speed limits, rear end, and vehicles making right or left turns significantly increased injury severity. Moreover, effects of trucks on injury severity of PV occupants are more significant for multi-occupants vehicles than single occupant vehicles.

Abdel –Aty and Abdelwahab (2002) calibrated three models to investigate the role of Light Truck Vehicles (LTV) in four rear-end crash configurations. They used multinomial logit (MNL), heteroscedastic extreme value (HEV), and bivariate probit (BVP) modeling schemes. The latter two are considered as the extensions to MNL (relaxing the irrelevant alternative assumption). The data from the General Estimates System (GES 2000) was used to estimate the above mentioned models. In the case of MNL, the significant variables were: driver age, gender, distraction, and lighting condition. The significant variables in the calibrated HEV were the same as MNL with an addition of traffic signals. The significant variables for BVP were consistent with MNL and HEV. The models results indicated that when a passenger car is behind an LTV, the driver of a passenger car would experience sight distance and discomfort problems. Moreover, the probability of rear end crashes increases when the driver of the following vehicle is distracted. Furthermore, young drivers and old drivers are more likely to be involved in rear end crashes.

Ulfarsson and Mannering (2004) estimated the differences in injury severity sustained by male and female drivers involved in single and two-vehicle crashes. To estimate the predicted probability for four levels of injury severity, they estimated fourteen multinomial logit models conditioned on the drivers gender, number, and type of vehicles involved. They obtained data from the Washington State Department of Transportation. The dataset contained reported accidents from January 1993 to July 1996. Due to the vast number of models involving multinomial logit schemes, some of the most significant findings are listed here. For both genders, drivers who did not use seat belts experienced more severe injuries. The predicted probabilities of higher injury severity were increased for drivers age 25 or younger and for 65 and older. Defective tires increased the predicted probability of possible or evident injury for female drivers. Wet, icy or snowy roads reduced the severity of injuries for both genders. Female drivers striking a barrier or guard rails increased the probability of fatal/disabling injury.

Khorashadi et al. (2005) investigated the differences in passenger vehicle and large truck drivers injury severity in rural and urban areas. They used four years (1997-2000) of data provided by the California Department of Transportation (Caltrans). Two multinomial logit models were estimated, one for rural and the second one for urban areas. Some of their major findings were; crashes involving large trucks at intersections in rural areas result in a 725% increase in the likelihood of severe/fatal injuries, whereas, in urban area the same kind of crash would result in 10% decrease in the likelihood of the same injury. The most significant variable across the models was influence of alcohol or drug. Furthermore, roads with median barriers in rural areas reduced the likelihood of severe/fatal injury by 69%.

Holdridge et al. (2005) estimated multivariate nested logit models to determine the impact of fixed roadside objects that affect crash severities. The data was from state of Washington from January 1993 to July 1996. From the database all single-vehicle crashes which involved roadside objects in an urban setting were selected. The results of their research showed that utility poles, trees, leading ends of guardrails, traffic poles, overhead poles, sign boxes, and bridge rails would increase the probability of fatal injury. Moreover, the following factors increased the predicted probability of fatal injury: driving above the posted speed limit and alcohol usage.

Milton et al. (2007) calibrated a mixed logit model (random parameter logit) to estimate accident severities on Washington highways. The data for 274 roadway segments from 1990-1994 was obtained from Washington State Department of Transportation. The model was estimated by incorporating injury severity proportions for individual roadway segments. In a mixed logit scheme, parameters can vary randomly across roadway segments. The average daily traffic per lane, average daily truck traffic, truck percentage, interchange per mile, and snow falls were random parameters. The number of horizontal curves, number of grade breaks per mile, and pavement friction were fixed parameters.

Malyshkina and Mannering (2008) estimated the influence of posted speed limit on the injury severities in the State of Indiana. They used the data for rural interstate and non-interstate routes from 2004 and 2006. They estimated the parameters using MNL models. The findings of the models indicated that increasing the posted speed limit on interstate highways did not have statistically significant impact on increasing the severity of the crashes in 2006. With regard to non-interstate highway crashes, increasing the speed limits were statistically significant.

Angel and Hickman (2009) used ten years of the crash records from state of Utah to estimate injury severity sustained by occupants in two-vehicle crashes. They calibrated two models using MNL and linear models. The statistically significant variables were: age, alcohol usage, population density, crash type, seating position, gender, use of seatbelts, and vehicle curb weight. Furthermore, the data shows that children are least likely to get injured while the older people are more likely to sustain more severe injury.

Schneider et al. (2009) estimated drivers injury severity due to single-vehicle crashes on horizontal curves on rural two-lane highways in the state of Texas. Separate multivariate MNL models were developed for small, medium, and large radius curves. To estimate these models, five years of crash data (1997-2001) for rural two-lane normal cures were obtained. The findings of the models indicated that female drivers were more likely to sustain more severe injury than male drivers. Moreover, older drivers were more likely to sustain more severe injuries. The following variables were found to increase the probability of severe injury: not wearing safety belts, fatigue, and drug or alcohol use.

2.4 Logistic Regression Models

Researchers used Logistic regression models when there were two levels of injury severities.

Farmer et al. (1997) investigated the relationship between crash characteristics and injury severity on two-vehicle side impact crashes. The data (1988 -1992) was obtained from National Accident Sampling System Crashworthiness Data System (NASS/CDS). Logistic regression (binary) technique was employed to estimate the various effects of the predictors. The results of the model showed that light trucks were 14 times as likely as cars to roll when struck. The occupants of the cars seating closer to the point of impact were more likely to suffer more severe injury than the occupants in a light truck. The occupants of the heavier vehicle were less likely to suffer high injury severity. Elderly were particularly at high risk of injury in side-impact crashes, specifically those 65 and older.

Al-Ghamdi (2002) applied a logistic regression methodology to estimate two levels of injury severities. A sample of 560 serious crashes from city of Riyadh in Saudi Arabia during August 1997 to November 1998 was extracted. The sample was divided into two categories; namely, fatal and non-fatal crashes. Nine independent variables were selected. These variables were: location, collision type, accident type, accident cause, age of driver at fault, nationality, vehicle type, and license status. Two variables (location and cause) were found to be statistically significant at 5% level. Furthermore, the models results conveyed that the probabilities of fatal crashes occurring at intersections are lower. The author mentioned that the data in the study suffered a great deal of validity because of the unskilled police processing crash reports. Variables such as: crash type (collision type), vehicle type, accident type, and age should have been statistically significant.

Hill and Boyle (2006) estimated the relative risk of severe injury for older female occupants involved in car crashes. They calibrated a logistic regression model using the General Estimates System (GES) data from the year 2000. Female occupants between 55 years and 74 years of age were at high risk. Seat belts, front seat, head on, side impact were significant variables. The risk of serious injury decreases after age of 75 for females, while it increases for male occupants. Occupants seated in the front seats are in a higher risk compared to the ones in the rear seats. Among these seating positions, seating in right front seat increased the risk.

2.5 Other Models

Kim et al. (1995) studied causal relationship among driver characteristics and behaviors and injury severity. The study was a part of Hawaii CODES project (Crash Outcome Data Evaluation System). All the police-reported crashes from 1990 in Hawaii were analyzed. To achieve a structural model that could illustrate the causal links among driver characteristics, crash severity, and injury severity, they used a log linear model. The study's categorical casual relationship of injury severity with respect to, driver behavior, driver age, sex, alcohol or drug use, driver error, and crash type were established. They constructed a structural model which could estimate the odds multiplier for each factor. Odds multiplier is defined as, how much each factor increases or decreases the odds of injury severity. The results indicated that drug or alcohol use and no seat belt greatly increased the odds of more severe injury. Furthermore, driver error had small impact while drivers age and sex were insignificant variables in determining causal relationship of injury severity.

Abdelwahab and Abdel-Aty (2002) applied Artificial Neural Networks (ANN) theory to predict injury severity level sustained by drivers involved in vehicle crashes on highways, signalized intersections, and toll plazas. Two databases from central Florida (1996-1997) and (1999-2000) were used. Explanatory variables in the study were alcohol, age, gender, violation, seat belt use, point of impact, speed ratio (running speed/posted speed), vehicle type, time of crash, area type, day of the week, pavement type, traffic condition, road alignment, type of toll plaza (epay or manual), and weather condition. Different models yielded different results. The following factors were significant in all three models: age, gender, seat belt, point of impact, and vehicle type. Day of the week and traffic condition were insignificant across all models.

Kweon and Kockelman (2003) estimated the risk of injury to different drivers across various vehicle types. They used multinomial probability models to estimate the probability of injury severity sustained by different groups of drivers. The data in this study was from the 1995 Nationwide Personal Transportation Survey (NPTS). Furthermore, to estimate vehicle miles driven by various drivers, they incorporated the results from NPTS to that of General Estimates System (GES). They estimated a series of crash exposure rates, such as: crash rates for different crash types, young drivers, middle aged drivers, old drivers, and vehicle types. The results suggest that women driving light duty trucks are in a higher risk groups comparing to men and older drivers are in low risk group. Moreover, SUVs and PUs are more often involved in rollovers comparing to other passenger vehicles.

Delen et al. (2006) utilized a series of ANN models to prioritize importance of crash-related variables as they apply to various levels of crash severity sustained by the driver. Data for the study were acquired from General Estimating System (GES). The GES data is a nationally representative sample of all police reported crashes in the US. The dataset contained 30,358 crashes from 1995-2000. The independent variables were as follows: age, sex, alcohol or drugs, vehicle age, body type, restraint system, highway type (interstate highway versus other), light conditions, road surface conditions, crash-type, time of day, and day of week. Rollover was important predictor in all models except one. The importance of drivers gender diminished as the level of injury severity increased. Age was an important predictor for injury severity level. Striking and struck variables had reverse impact on the severity of injury; importance of striking decreases with increased injury severity. Body type had various impacts across different models and could not have been explained singularly. Weather conditions or time of crash had no impact on severity of the crash.

Chang and Wang (2006) developed a Classification And Regression Tree (CART) model which established the relationship between injury severity and twenty explanatory variables. They used National Traffic Accident Investigation Reports crash data from Taipei area during 2001. The data were divided into two subsets; one was used for learning and the other for testing. Model prediction accuracy for individual level of severity was over 94% for both learning and testing data. In the case of fatality, model failed in both learning and testing data (0% prediction). Model identified that pedestrian, motorcycle and bicycle riders are the most vulnerable groups. Crash type, contributing circumstances, and driver actions were found to be important factors in determining the injury severity level.

2.6 Summary

There exists an abundant body of literature in the area of crash severity research. As mentioned before, this review was centered on two topics, the methodological approaches and variables used by previous researchers. This thorough literature review yielded many methodological approaches and variables which were applied by the researchers. In all the reviewed literature, injury severity was treated as a discrete outcome variable.

Methodological approaches

The majority of the researchers have extensively applied the OP and OL methodologies when the outcome variable was considered categorical and ordinal. In the case of the failure of OP and OL model assumptions, research requirements, odds ratio, or some other reasons, the outcome variable was considered unordered and categorical. In that case MNL and its extensions were used.

Other researchers limited the injury severity into two (binary) levels; in that case, logistic regression techniques were applied. Log linear models were applied if the focus of the research was on the association patterns among the categorical predictors. The ANN and multinomial probability models were calibrated but not as often as other models. Therefore, due to extensive use of ORM and MNL, in this study these two methodologies were considered.

Variables

The following variables were found to be statistically significant by other researches: drivers age, vehicle speed, crash type, vehicle weight, elderly, dark with no lighting, speed limit, gender, grades, number of occupants, alcohol usage, farm vehicles, level train, seat belts, seating position, vehicle type, drivers age less than or equal to 25, drivers age greater or equal to 65, median, daily traffic, daily traffic per lane, utility poles, trees, truck percentage, interchange per lane, snowfalls, road type, point of impact, weather condition, population density, and fatigue.

Therefore, in this study the majority of the above mentioned variables as well as other variables were used.

Although an extensive body of knowledge exists in this area, the literature still suffers from significant limitations. This section provided a relevant summary of the studies conducted in this field. The reviewed body of knowledge had the following characteristics in common:

1. The lack of a comprehensive study which could address the injury severity sustained by the occupants of passenger vehicles involved in large truck crashes.
2. Passenger-vehicle dimensions were not incorporated in the models for estimating injury severity.

In other words, vehicles were treated as dimensionless objects. Each vehicle has specific dimensions such as: wheelbase, rear overhang, front overhang, and etc. One would expect that the occupants of a vehicle with smaller front overhang sustain higher injury severities, given a crash has occurred. This study has attempted to address these concerns while using an unprecedented number of variables.

CHAPTER 3

RESEARCH METHODOLOGY

3.1 Introduction

In this study, PVs include passenger cars, pickup trucks, mini vans, and SUVs. Injury severity levels have been categorized using "KABCO" scale as defined in a report by National Safety Council (1990). Within the system there are five categories of injury severity levels denoted as follows: "K" for fatal injury, "A" for incapacitating injury, "B" for non- incapacitating injury, "C" for no visible injury but complaint of pain, and "O" for no injury. In this study the five levels of injury severity were assigned numbers ranging from 1 to 5, i.e., K=5, A=4, B=3, C=2, and O=1.

As stated in chapter 1, one of the objectives of this study was to develop a model for estimating the probabilities of different injury severity levels to PVs' occupants involved in crashes with LTs. The mathematical formulation for the estimated conditional probability of crash ***i*** resulting in injury severity level ***m*** is given as;

$$\Pr (y_i = m | \mathbf{x_i}) = F (x_i) \tag{3-1}$$

Where $\mathbf{x}_i$ is a vector of explanatory variables representing crash features, F (x_i) is a function of a set of predetermined explanatory variables.

In order to estimate equation 3-1, the discrete response variable(y) was treated as ordered-categorical and unordered-nominal. In the case of ordered-categorical response variable, two of the most commonly used ORM were considered. In the case of unordered-nominal scheme, MNL model was considered.

3.2 Ordered Regression Models

Ordered regression models include OP and OL. For the last fifty years ORM methodology has been applied in many fields; namely, biology, econometrics, education, naval studies, road safety, and etc. Ordered regression model was the outcome of Aitchison and Silvey's (1957) work on organism's tolerance to various exposures. Their models were of univariate nature. McKelvey and Zavonia (1975) extended the work of Aitchison and Silvey by incorporating multiple independent variables. They studied votes for the 1965 Medicare bill. Winship and Mare (1984) estimated educational attainment using four categories of educational levels. Marcus and Greene (1985) studied three levels of work skills for navy recruits job assignments. Hedström (1994) analyzed organizational ranks in Sweden using four categorical levels of dependent variable. Hartog et al. (1994) analyzed the job levels that desired by workers. Meng and Miller (1995) estimated the occupational attainment coefficients in China using three ordered categorical levels.

O'Donnell and Connor (1996) from Australia were the first researchers applying ORMs in the field of road safety. Following O'Donnell and Connor, Duncan et al. (1998) published an inspiring paper applying OP in the US.

Ordered probit and OL are suitable for application "when a variable is ordinal and its categories can be ranked from low to high, but the distances between adjacent categories are unknown" (Long, 1997). As an example, in this study difference between injury severity level 1 and 2 is not the same as 4 and 5. Other words, numbers have no cardinal significance. For further information on OP and OL refer to chapter five (Long 1997).

Table 3-1 presents the specific features of OP and OL. As shown in the table the characteristics of OL are similar to OP. The differences between these two modeling methodologies are in their random error distribution and mathematical formulation. In the OL methodology, the random error has a logistic distribution with a mean of zero and a variance of $\pi^2/3$ and the random error associated with OP is assumed to follow standard normal distribution. Furthermore, there exist differences between the equations presenting their Probability Density Functions (PDF) and Cumulative Density Functions (CDF). These differences do not result in major variation in the estimation of the coefficients. Generally speaking there could be a 1.7 ratio between OL and OP coefficients. Therefore, due to minor differences in these two modeling methodologies and common use of OP in this area of research, it was decided to estimate the coefficients by calibrating an OP model. The end results of OP are summarized as follows:

1. Coefficients: Estimated coefficients of the model represent the directional association between the explanatory and dependent variable. As an example, if the estimated coefficient is positive, one can conclude that the variable in the discussion increases the probability of the risk associated with the injury sustained by the occupants.

2. Significant contributors: Identifies statistically significant contributors to the dependent variable.

3. Dependent Variable: Interaction among a set of explanatory variables and their impact on the dependent variable.

4. Predicted probability: The impact of a single or a combination of independent variable(s) can be examined while holding the rest of the variables constant at their means.

Table 3-1 Characteristics of Ordered Regression Models

Variables	Ordered Probit	Ordered Logit
Response outcome	Discrete	Discrete
Ranked and Ordinal Dependent Variable	Yes	Yes
Relation of (DV) to a latent variable	Yes	Yes
DV ordered and increasing sequentially	Yes	Yes
Equal distance among observed (DV)	No	No
Error Distribution	Normal	Logistic
Distribution of the error's mean	0	0
Distribution of the error's standard deviation	1	$\pi / \sqrt{3}$
Error's probability density function (PDF)	$\frac{1}{\sqrt{2\pi}} \exp(-\frac{x^2}{2})$	$\frac{\exp(x)}{[1+\exp(x)]^2}$
Error's cumulative density function (CDF)	$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp(-\frac{x^2}{2}) dx$	$\frac{\exp(x)}{1+\exp(x)}$
Maximum Likelihood Estimation (MLE)	Yes	Yes
Factor of coefficient difference	1	#1.7
Parallel regression assumption	Yes	Yes
Proportional odds	No	Yes
AIC	YES	YES
Pseudo ρ^2	YES	YES
LR (Likelihood Ratio)	YES	YES

As an example, the effect of the number of occupants on the probability of injury severity can be calculated while holding other variables constant. This procedure can be extended to the calculation of probabilities for each injury category for a range of conditions.

5. Marginal effects: The impact of one unit change of an explanatory variable on the predicted probability of each injury severity level.

3.3 Ordered Probit Model

3.3.1 Model Specification

The OP has the following specification:

$$Y_n^* = \boldsymbol{\beta} x_n + \epsilon_n \tag{3-2}$$

where Y_n^* = latent and continuous estimate of injury sustained by occupant n in a crash,

x_n = a vector of independent variables,

$\boldsymbol{\beta}$ = a vector of parameters to be estimated, and

ϵ_n = random error term assumed to have N(0,1).

Figure 3-1 illustrates the mapping between the latent continuous injury variable, Y_n^*, and the observed injury severity level, Y_n.

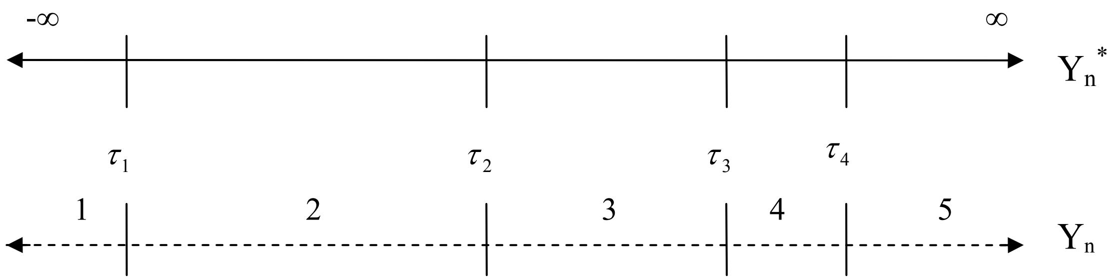

Figure 3-1 Mapping of latent and coded injury severity variable

Figure 3-1 represents the following characteristics about OP:

1. The distances between the thresholds are unknown and may not be equal.
2. Domain for category five (fatal) is from τ_4 to ∞.
3. Domain for category one (no injury) is from τ_1 to $-\infty$.
4. The thresholds for the other categories are bounded within two thresholds of categories one and five.
5. The latent continuous variable ranging from $-\infty$ to ∞ is drawn to an observed discrete random variable, Y_n.

In this study the observed discrete random variable, Y_n, represents the injury severity levels defined below:

$$Y_n = \begin{cases} 5 \implies & \text{(Fatal)} \\ 4 \implies & \text{(Incapacitating)} \\ 3 \implies & \text{(Non-incapacitating)} \\ 2 \implies & \text{(No visible injury but complain of pain)} \\ 1 \implies & \text{(No injury)} \end{cases}$$

The observed and coded, Y_n, is determined from the model as follows:

$$
Y_n = \begin{cases} 1 \implies & \text{if } -\infty \leq Y_n^* < \tau_1 \\ 2 \implies & \text{if } \tau_1 \leq Y_n^* < \tau_2 \\ 3 \implies & \text{if } \tau_2 \leq Y_n^* < \tau_3 \\ 4 \implies & \text{if } \tau_3 \leq Y_n^* < \tau_4 \\ 5 \implies & \text{if } \tau_4 \leq Y_n^* < \infty \end{cases}
$$

Where τ_is are thresholds or cut points.

3.3.2 Model Estimation

Maximum Likelihood Estimation (MLE) is one of several methods used in obtaining parameter estimates **ß** and **τ** (Powers et al. 2000). According to Ben-Akiva and Lerman (1985), "A maximum likelihood estimator is the value of the parameters for which the observed sample is most likely to have occurred".

Maximum Likelihood estimators are distributed asymptotically normally. Other words, if H_o is true then z is distributed normally with a mean of zero and variance of one.

In this study, "STATA" was used for estimating the unknown parameters. STATA utilizes the MLE method. The log likelihood equation used in OP is:

$$
\text{Log L } (\boldsymbol{\beta}, \tau \mid \mathbf{y}, \mathbf{X}) = \sum_{j=1}^{J} \sum_{yi=j} \ln [F(\tau_j - x_i \boldsymbol{\beta}) - F(\tau_{j-1} - x_i \boldsymbol{\beta})] \tag{3-3}
$$

Everything else being equal, models with a larger value of log likelihood are preferred (Akiva & Lerman, 1985).

3.3.3 Scalar Measures of Model Fit and Model Test

1. Likelihood Ratio Test:

Likelihood Ratio (LR) provides a comparative test between the coefficients of the initial model (All the variables included) and the coefficients of the final model (model with an imposed restriction).

Likelihood Ratio is computed by comparing the log likelihood of the initial model with a restricted model. The equation for this test statistic is:

$$LR= -2(Ł^{R}- Ł^{U}) \tag{3-4}$$

where $Ł^R$ and $Ł^U$ are the values of the log likelihood function at its maximum for restricted and unrestricted models, respectively. The LR is distributed as χ^2 (chi-square) with r degrees of freedom (Ben-Akiva et al. (1985)). To determine the significance of a model for a given set of explanatory variables, one can use equation 3-4 to calculate the LR. The calculated LR can be compared to $\chi2$ value obtained from the table of critical values of χ^2 at a predetermined confidence level. The larger the calculated LR value comparing to critical value, the better the model fit.

In this study, LR test of multiple coefficients were used (α=0.10). As an example, comparing the final model to the initial one, there will be n fewer variables which results in n fewer coefficients. The hypothesis that the effects of the removed variables are simultaneously equal zero was tested by:

H_o: $\beta_1 = \beta_2 = \beta_3 = \ldots\ldots\ldots \beta_n$.

2. Pseudo ρ^2

Pseudo ρ^2: Provides a convenient way to compare log likelihoods across different models (Long 1997). ρ^2 equation is as follows:

$$\rho^2 = 1 - \frac{\log L^R}{\log L^U} \tag{3-5}$$

The ρ^2 might aid in comparing competing models and ultimately, in choosing the final model (Long (1997)). Based on the reviewed literature in chapter II, the reported ρ^2 varied in value from .057 to 0.116; namely, Renski et al. (1991, ρ^2= 0.116), Khattak (2000, ρ^2= 0.066), Khattak et al. (2002, ρ^2=057), Abdelwahab and Abdel-Aty (2002, ρ^2=0.10), Wang and Abdel-Aty (2008, ρ^2=0.086), and Angel and Hickman (2009, ρ^2=0.0848). It appears in this area of research ρ^2 is small in magnitude. Therefore, it was decided that for this study, final models with ρ^2 of 0.10 or greater to be considered.

3. Hypothesis testing of the estimated coefficients

Z statistic is used to test a single hypothesis about the parameters in the OP model. The equation for this statistic is:

$$Z \equiv \hat{\beta}_j / se(\hat{\beta}_j) \tag{3-6}$$

where $\hat{\beta}_j$ is unbiased estimator of β_j and $se(\hat{\beta}_j)$ is standard error of β_j. Having the calculated values of Z statistic, one can test the following hypothesis:

$$H_o\text{: } \beta_j = 0 \tag{3-7}$$

$$H_A\text{: } \beta_j \neq 0 \tag{3-8}$$

where β_j is unknown population parameter, H_o and H_A are null and alternative hypotheses respectively. If the calculated values are far enough from zero, then H_o is rejected. In this study, the significance level or the probability (p) of rejecting H_o is 10%. Others words it is possible to reject H_o when it is true 10% of the time.

4. Akaike Information Criterion

Akaike (1973) Information Criterion (AIC) is an index used in choosing the best fitted model. All else being equal (Long 1997) the model with smallest AIC is considered better. The equation for the index is as follows:

$$\text{AIC}=\frac{\{-2\ln\hat{L}(M_k)+2P_k\}}{N} \tag{3-9}$$

where $\hat{L}(M_k)$ is the model's likelihood and P_k is the number of parameters contained in the model.

5. Count R^2

The count R^2 is the proportion of correctly predicted levels for the entire model.

$$R^2_{Count}=\frac{1}{N}\sum_j n_{jj} \tag{3-10}$$

where n_{jj} is the total number of correctly predicted for discrete level j.

3.3.4 Model Assumption Test

By changing the intercept on an S shaped probability curve would result in parallel shifting of the curve to the right or to the left. Consequently, the slope at a given value does not change due to this parallel shifting. The parallel regression assumption is based on this phenomenon and it implies that $ß_1=ß_2=\cdots=ß_{k-1}$, if the assumption holds, then the estimated coefficients $\hat{\beta}_1=\hat{\beta}_2=\cdots=\hat{\beta}_{k-1}$ or very close in magnitude. To test the validity of the assumption, a Wald test proposed by Brandt (1990) using STATA software package was employed. This statistic test compares the log likelihood from OP with that of (J-1) binary probit models. The null hypothesis assumes all the estimated coefficients are equal and the alternative assumes they are not. If parallel regression assumption doesn't hold, one can utilize the MNL model (Scott Long and Jeremy Freese (2006)). According to Scott Long (1997) "a researcher might prefer to treat an outcome as nominal, even though it is ordered".

3.3.5 Analysis of Results

Predicted probability for each outcome

As previously mentioned, the assumption of OP model is that ε is distributed normally with mean zero and variance one (Standard normal distribution). By substituting these values in the PDF and CDF of normal distribution, one can show:

PDF for error distribution is ==> $$f = \phi(\varepsilon) = \frac{1}{\sqrt{2\pi}} \exp\left(\frac{e^2}{2}\right) \quad (3\text{-}11)$$

CDF for error distribution is ==> $$F = \Phi(\varepsilon) = \int_{-\infty}^{\varepsilon} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) dt \quad (3\text{-}12)$$

Probability that a random variable is between two values is the difference between the Fs evaluated at these values, thus

$$\Pr(y_i = m \mid \mathbf{x_i}) = F(\tau_m - x_i \boldsymbol{\beta}) - F(\tau_{m-1} - x_i \boldsymbol{\beta}) \quad (3\text{-}13)$$

Therefore, probabilities associated with injury severity levels are formulated as follows:

$$\Pr(y_i = 1 \mid x_i) = \Phi(\tau_1 - \beta_o - \beta x_i) \quad (3\text{-}14)$$

$$\Pr(y_i = 2 \mid x_i) = \Phi(\tau_2 - \beta_o - \beta x_i) - \Phi(\tau_1 - \beta_o - \beta x_i) \quad (3\text{-}15)$$

$$\Pr(y_i = 3 \mid x_i) = \Phi(\tau_3 - \beta_o - \beta x_i) - \Phi(\tau_2 - \beta_o - \beta x_i) \quad (3\text{-}16)$$

$$\Pr(y_i = 4 \mid x_i) = \Phi(\tau_4 - \beta_o - \beta x_i) - \Phi(\tau_3 - \beta_o - \beta x_i) \quad (3\text{-}17)$$

$$\Pr(y_i = 5 \mid x_i) = 1 - \Phi(\tau_4 - \beta_o - \beta x_i) \quad (3\text{-}18)$$

where y_i is the observed injury severity level for the probability of outcome m. These probabilities are positive if the threshold parameters satisfy the restriction $\tau_1 < \tau_2 < \tau_3 < \tau_4$. Predicted probabilities for each severity level can be estimated using equation 3-19. X is vector of values based on the observation or theoretical values for any hypothetical case. In this study X is the average value of the significant variables.

$$\Pr(y = m \mid X) = F\left(\hat{\tau}_m - X\hat{\beta}\right) - F\left(\hat{\tau}_{m-1} - X\hat{\beta}\right) \quad (3\text{-}19)$$

Predicted probability for each observation

The percentage of correctly predicted probability for each level versus the observed level of injury can be calculated. The objective of this analysis is to find out how well the model predicted each outcome. Upon estimating the predicted probabilities for each observation, a table presenting the results was tabulated.

Marginal effect

The effect of increasing a continuous variable x_k by one unit on the predicted probability of an outcome is called marginal effect. The change in the probability is the slope of the curve relating x_k to Pr(y=m|X) and it can be calculated using equation 3-20.

$$\frac{\delta \Pr(y = m \mid X)}{\delta x_k} = \frac{\delta F(\tau_m - X\beta)}{\delta x_k} - \frac{\delta F(\tau_{m-1} - X\beta)}{\delta x_k} \quad (3\text{-}20)$$

As an example, using equation 3-20 one could calculate the impact of increasing the drivers age by one year on the predicted probability of an injury severity level.

The effect of a dummy variable can be thought of a discrete change. Discrete changes can be calculated using equation 3-21.

$$\frac{\Delta \Pr(y=m \mid X)}{\Delta x_k} = \Pr(y=m|X, x_k = x_E) - \Pr(y=m|X, x_k = x_S) \quad (3\text{-}21)$$

Other words, when x_k changes from x_s to x_E, the predicted probability changes by Δ Pr(y=m|x)/ Δ x_k. As an example, the effect of a male driver versus a female driver on the predicted probability of an injury severity can be calculated using equation 3-21.

3.3.6 Modeling Procedure

In summary, Figure 3-2 presents a general procedural approach for estimating and analyzing the OP model.

An initial model with all the preselected explanatory variables was calibrated. In order to achieve the final model, explanatory variables with p greater than 10% were removed one at a time.

The final model was tested for the final fit by using scalar measures of fit. The model assumption was tested using a Brandt test.

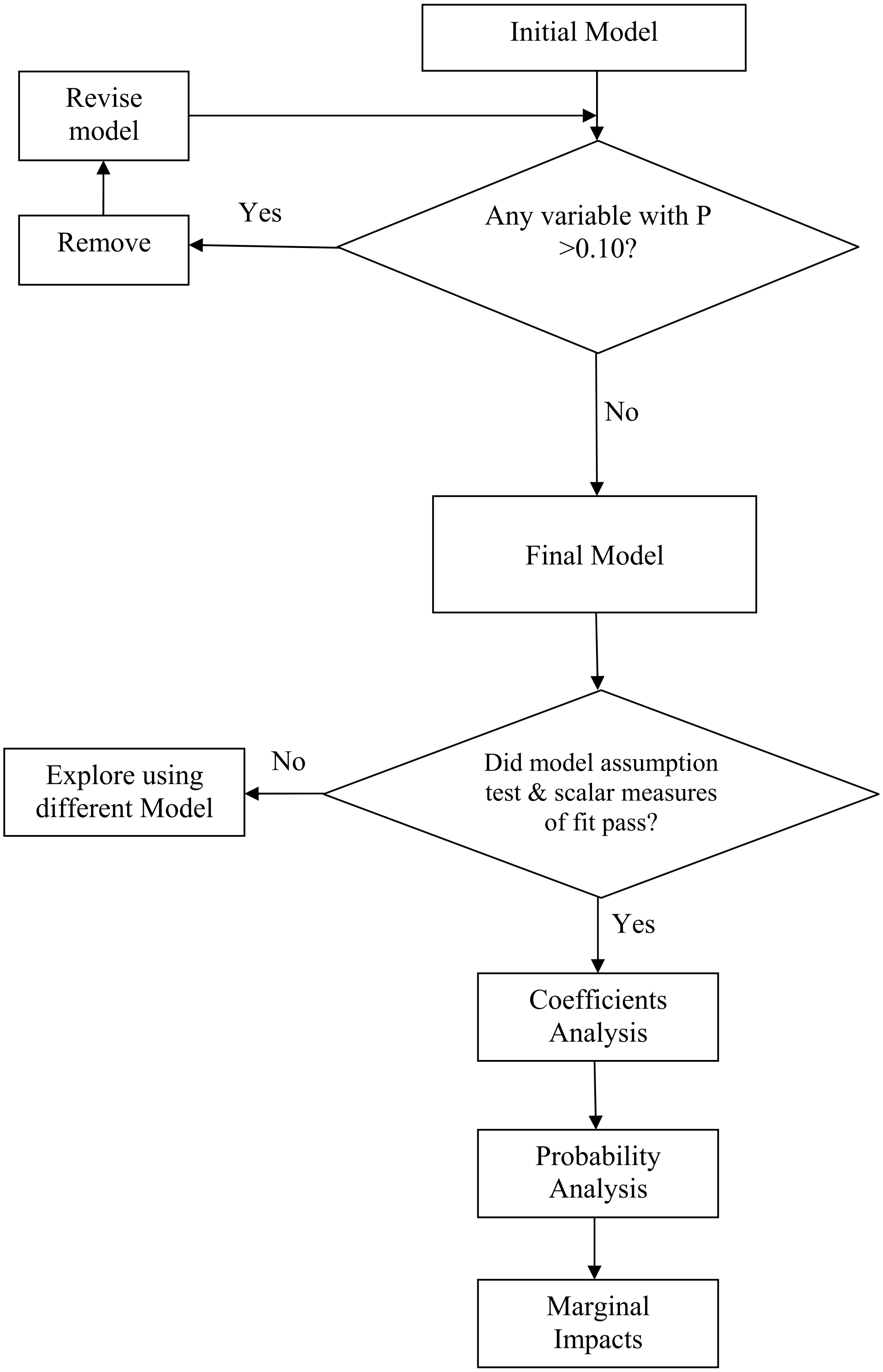

Figure 3-2 Modeling procedure for Ordered Probit Model

3.4 Multinomial Logit Model

Multinomial logit model is one of the most commonly used nominal regression models (refer to chapters 3 and 6 of Scott Long 1997 for detailed discussion of the model). As shown in chapter two, MNL has been employed in the area of injury severity research almost as frequently as OP.

In the MNL, crash outcomes can be grouped like OP, however outcome categories are not ordered.

3.4.1 Model Specification

Multinomial Logit could be thought of as an extension of the binary logit model. The MNL is a nonlinear probability model and a linear model in the log of the odds. In this study MNL is applied as odds and probability model.

The MNL as an Odds model

An important characteristic of MNL modeling scheme is the calculated odds of various events for a given vector of X. The odd of an event is denoted as:

$$\Omega = \left[\frac{\Pr(y=1 \mid X)}{1-\Pr(y=1 \mid X)} \right] \quad (3\text{-}22)$$

where Ω stands for the odd, Pr(y=1|X) is the probability of an event taking place, and [1-Pr(y=1|X)] is the probability of the same event not taking place given X. By taking the log of equation 3-22, the outcome is known as logit and denoted by $\ln\Omega$. Conversely, the exponent of the logit is the odd.

$$\ln\Omega = \ln\left[\frac{\Pr(y=1\mid X)}{1-\Pr(y=1\mid X)}\right] = \text{logit} \tag{3-23}$$

In MNL model for each pair of outcome one logit model is being estimated.

The general equation for MNL is written as:

$$\ln\Omega_{m|b}(x) = \ln\left[\frac{\Pr(y=m\mid X)}{\Pr(y=b\mid X)}\right] = \beta X \qquad \text{for m=1 to j} \tag{3-24}$$

where b is formally known as base or comparison group. As an example, for a set of data with three nominal outcomes (e.g., A, B, C), we will be simultaneously estimating logit coefficients (logits) for A versus C and B versus C. In this example C is set as base.

There is a necessary relationship among the logits as shown in equation 3-25, such that;

$$\ln\left[\frac{\Pr(A\mid x)}{\Pr(C\mid x}\right] - \ln\left[\frac{\Pr(B\mid x)}{\Pr(C\mid x}\right] = \ln\left[\frac{\Pr(A\mid x)}{\Pr(B\mid x}\right] \tag{3-25}$$

Due to this necessary relationship, the following equality must hold too.

$$\beta_{1,\,A|C}{}^{x} - \beta_{1,\,B|C}{}^{x} = \beta_{1,\,A|B}{}^{x} \tag{3-26}$$

Therefore, in the example mentioned above, the logits for A|B is the difference between the logits of A|C and B|C, as shown in equation 3-26.

In this study, Level 1 (no injury) is set as base (comparison group) and every injury level is compared to the base, such that;

$$\ln\Omega_{Level5|Level1}(x) = \ln\left[\frac{\Pr(y=5\mid X)}{\Pr(y=1\mid X)}\right] = \beta X \quad (3\text{-}27)$$

$$\ln\Omega_{Level4|Level1}(x) = \ln\left[\frac{\Pr(y=4\mid X)}{\Pr(y=1\mid X)}\right] = \beta X \quad (3\text{-}28)$$

$$\ln\Omega_{Level3|Level1}(x) = \ln\left[\frac{\Pr(y=3\mid X)}{\Pr(y=1\mid X)}\right] = \beta X, \text{ and} \quad (3\text{-}29)$$

$$\ln\Omega_{Level2|Level1}(x) = \ln\left[\frac{\Pr(y=2\mid X)}{\Pr(y=1\mid X)}\right] = \beta X \quad (3\text{-}30)$$

The last equation (Level1|Level1) is not shown, because the log of [Pr(y=1|x) ÷ Pr(y=1|x)] is zero and that makes the effect of all explanatory variables as zero.

The MNL as a Probability model

From equation 3-24 one can derive the predicted probability for y=m|x as follows:

$$\Pr(y=m|x_i) = \frac{\exp(X\beta_{m|b})}{\sum_{j=1}^{J}\exp(X\beta_{j|b})} \quad (3\text{-}31)$$

3.4.2 Model Estimation

Maximum likelihood estimation used in obtaining parameter estimates of ß. The logit coefficients for different levels comparing to the base are estimated simultaneously. The likelihood equation is as follows (Long 1997):

$$L(\beta_2,\ldots,\beta_J|y,X)=\prod_{m=1}^{J}\prod_{y_i=m}\frac{\exp(x_i\beta_m)}{\sum_{j=1}^{J}\exp(x_i\beta_j)} \quad (3\text{-}32)$$

3.4.3 Scalar Measures of Model Fit and Model Test

1. Likelihood Ratio Test

Likelihood Ratio is computed by comparing the log likelihood of the initial model with a restricted model. The equation for this test statistic is:

$$LR=-2(Ł^R - Ł^U) \quad (3\text{-}33)$$

where $Ł^R$ and $Ł^U$ are the values of the log likelihood function at its maximum for restricted and unrestricted models, respectively.

2. Hypothesis testing of the estimated coefficients

Z statistic is used to test hypothesis about each estimated coefficient. The following hypothesis is set for each logit coefficients:

$$H_o: \beta_j = 0 \quad (3\text{-}34)$$

$$H_A: \beta_j \neq 0 \quad (3\text{-}35)$$

where β_j is unknown population parameter. In this modeling scheme, the significance level or the probability of rejecting H_o was 10%.

3. Pseudo ρ^2

Pseudo ρ^2 has the same definition and application as in OP. ρ^2 equation is as follows:

$$\rho^2 = 1 - (\log Ł^R / \log Ł^U) \tag{3-36}$$

4. Akaike Information Criterion

Akaike Information Criterion (AIC) has the same definition and application as in OP. The equation for the index is as follows:

$$\text{AIC} = \frac{\{-2\ln \hat{L}(M_k) + 2P_k\}}{N} \tag{3-37}$$

5. Count R^2

In MNL, count R^2 is applied in the same manner as in OP. The formulation for the count R^2 is as follows:

$$R^2_{Count} = \frac{1}{N}\sum_j n_{jj} \tag{3-38}$$

where n_{jj} is the total number of correctly predicted for the outcome j.

3.4.4 Model Assumption Test

The Independence of Irrelevant Alternatives (IIA) in MNL implies that, the alternatives are irrelevant, other words deleting or adding an alternative does not impact the odds between other alternatives. To test this important assumption, the estimated coefficients ($\hat{\beta}$) of the full model are compared to that of the restricted model with at least one less alternative. The assumption of IIA is rejected if the test statistic is significant, other words the use of MNL should not be considered. In this study, Small and Hsiao (SH) test is employed. In order to calculate SH, sample was randomly divided into two equal sizes.

$\hat{\beta}_u^{S_1}$ and $\hat{\beta}_u^{S_2}$ are the estimated coefficients for the first and second subsamples of the unrestricted (full) models. The weighted average of theses coefficients are formulated as follows:

$$\hat{\beta}_u^{S_1 S_2} = \left(\frac{1}{\sqrt{2}}\right) \hat{\beta}_u^{S_1} + \left\{1 - \left(\frac{1}{\sqrt{2}}\right)\right\} \hat{\beta}_u^{S_2} \tag{3-39}$$

Using a restricted model for the second subsample will result $\hat{\beta}_r^{S_2}$. The SH is distributed as chi-squared and calculated using formula (3-40)

$$\text{SH} = -2\left\{ \text{L}(\hat{\beta}_u^{S_1 S_2}) - \text{L}(\hat{\beta}_r^{S_2}) \right\} \tag{3-40}$$

3.4.5 Analysis of Results

Predicted probability for each outcome

Predicted probabilities for each outcome (severity level) can be estimated using equation 3-31.

Predicted probability for each observation

The percentage of correctly predicted probability for each level versus the observed level of injury was calculated. The objective of this test was to find out how well the model predicted each outcome. Upon estimating the predicted probabilities for each observation, a table containing these values was tabulated.

Marginal effect

The impact of one unit change in the continuous explanatory variable on the predicted probability of each level is calculated using equation 3-41.

$$\frac{\delta \Pr(y=m \mid X)}{\delta x_k} = \Pr(y=m \mid X)\left\{\beta_{k,m|J} - \sum_{j=1}^{J} \beta_{k,j|J} \Pr(y=j \mid X)\right\} \tag{3-41}$$

The impact of the one unit change in the dummy explanatory variable on the predicted probability of each level is calculated using equation 3-42.

$$\frac{\Delta \Pr(y=m \mid X)}{\Delta x_k} = \Pr(y=m \mid X, x_k = xE) - \Pr(y=m \mid X, x_k = x_s) \tag{3-42}$$

Odds ratio

Odds ratio or factor change coefficient can be estimated while holding other variables constant. The odds of outcome m versus n as x_k increases by δ can be calculated by equation 3-43:

$$e^{\beta_{k,m|n}} = \frac{\Omega_{m|n}(X, x_k + \delta)}{\Omega_{m|n}(X, x_k)} \tag{3-43}$$

For δ =1, the odds of m versus n will change by a factor of $e^{\beta_{k,m|n}}$, holding all explanatory variables constant.

3.4.6 Modeling Procedure

In summary, Figure 3-3 presents a general procedural approach for estimating and analyzing the MNL model. The following steps were undertaken:

- No injury (level 1) was set as the base for the comparison,
- An initial model with all the preselected explanatory variables was calibrated,
- Explanatory variables with P value greater than 10% in all its corresponding injury levels were removed one at a time,
- Tested the model assumption,

- Assessed scalar measures of fit,
- Estimated the Odds ratios,
- Estimated the predicted probability of each injury severity level,
- Estimated the marginal impact of each significant variable on the predicted probability of each injury severity level.

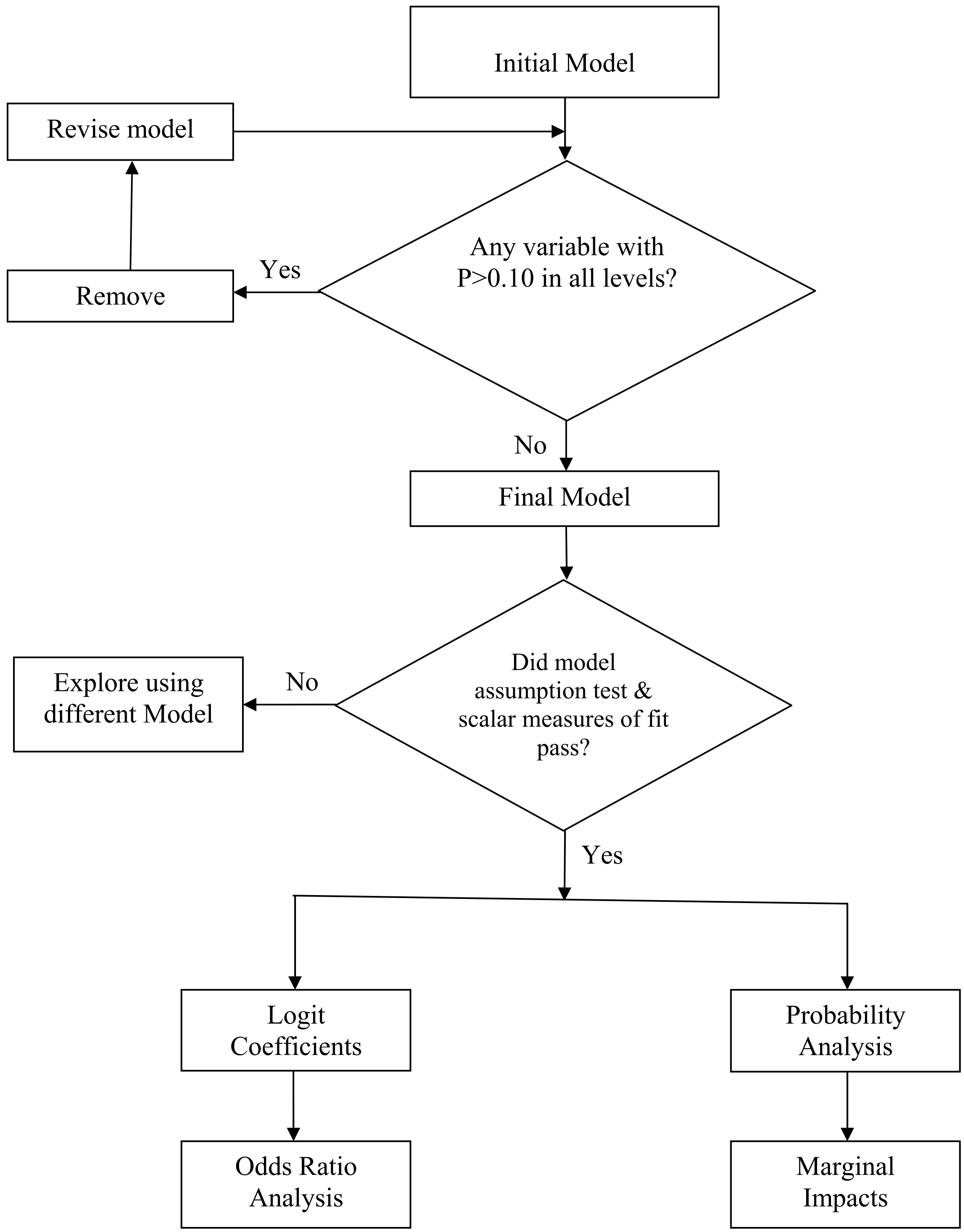

Figure 3-3 Modeling procedure for Multinomial Logit model

CHAPTER 4

DATA ANALYSIS

4.1 Introduction

North Carolina crash data for 2003 were obtained from Highway Safety Information System (HSIS). The obtained database contained a great deal of crash information namely, number of vehicles, drivers age, number of occupants, vehicle makes, time of the crash, road surface condition, and numerous other data. Two vehicle crashes involving a PV and an LT were selected.

4.2 Explanatory Variables

For this study, variables contributing to crashes are categorized into four groups: driver/occupant characteristics, crash environment characteristics, crash characteristics, and vehicle characteristics. A complete list of these explanatory variables is presented in Table 4-1. Table 4-1 represents twenty-nine selected explanatory variables and their descriptions.

In addition to the variables from the reviewed literature, there are variables, which distinctly include vehicle specifications and were not specified in the reviewed literature. These specifications are; weight of (LT), front-overhang (PV), rear-overhang (PV), wheelbase (PV), height (PV), and width (PV). Figure 4-1 depicts wheelbase (A), front overhang (B), and rear overhang (C). Wheelbase is the distance from the center of the front wheel to the center of the rear wheel. Moreover, front overhang is measured from the center of the front wheel to the outer tip of the front bumper and rear overhang is measured from the center of the rear tire to the outer tip of the rear bumper.

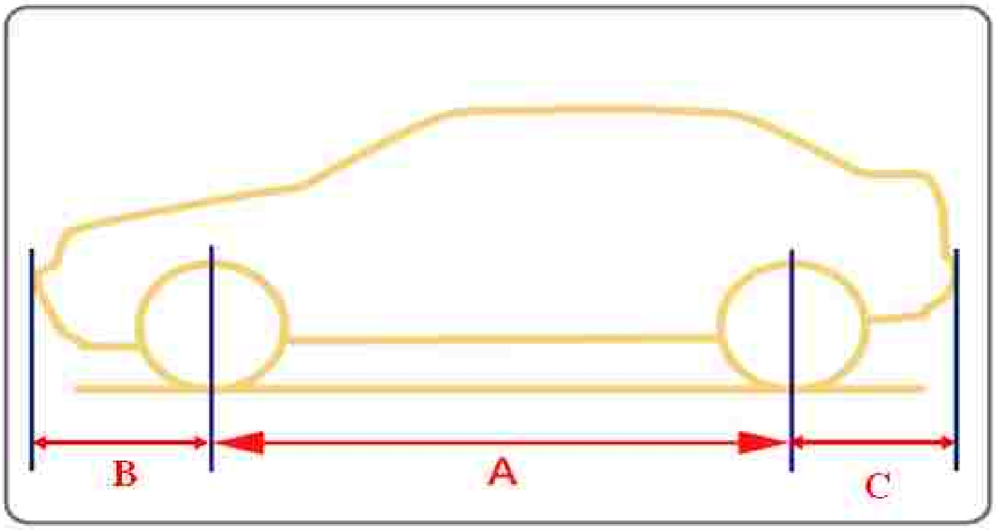

Figure 4-1 Wheelbase (A), Front overhang (B), and Rear overhang (C)

4.3 Data Source

The Federal Highway Administration (FHWA), a branch of U.S. Department of Transportation, has created a highway safety database called "Highway Safety Information System (HSIS)". Under contract with FHWA, the University of North Carolina Highway Safety Research Center (HSRC) and LENDIS Corporation jointly operate HSIS. The HSIS uses data already being collected by nine states. The participating states are California, Illinois, Maine, Michigan, Minnesota, North Carolina, Ohio, Utah, and Washington.

Table 4-1 Explanatory variables

Categories	VARIABLES	DESCRIPTION
Driver/ Occupants Characteristics	Drivers age (PV)	Age of the PV's driver
	Drivers gender(PV)	Gender of the PV's driver
	# Of occupants (PV)	Number of occupants in PV (Includes driver)
	Drivers age (LT)	Drivers age (Large truck=LT)
	Drivers gender (LT)	Drivers gender (LT)
	Alcohol	Alcohol flagged
Crash Environment at the Crash Location	AADT	Average Annual Daily Traffic
	Number of lanes	Total number of lanes
	MVMT	Million vehicle miles travelled
	% of trucks	Percent of trucks at crash location
	Speed limit	Posted speed limit
	Median type	Median type (e.g. Divided, Undivided)
	Light condition	Light condition (e.g. Light, dark)
	Road surface condition	Road Surface condition (e.g. Dry , wet)
	Pavement type	Pavement type (e.g. Asphalt, concrete)
	Weather condition	Weather condition (e.g. Dry, rain)
	Roadway class	Roadway classification (e.g. Freeway, 2-lane)
	Road geometry	Road Geometry (e.g. Straight, Curve)
Crash Characteristics	Impact speed (PV)	PV's estimated speed at impact
	Travel speed (PV)	PV's estimated travel speed before impact
	Impact speed (LT)	LT's estimated speed at impact
	Travel speed(LT)	LT's estimated travel speed before impact
	Crash type	Accident type (e.g. Read end, Sideswipe)
Vehicle Characteristics	Weight(LT)	Gross commercial vehicle weight (LT)
	Weight(PV)	Curb weight=Total weight of a vehicle with standard equipment and all the necessary fluids (PV).
	Front-overhang (PV)	Distance between center of front wheel to the front of the vehicle (PV)
	Rear-overhang (PV)	Distance between center of rear wheel to the rear of the Vehicle (PV)
	Wheelbase (PV)	Distance between center of front wheel to the center of rear wheel (PV)
	Height (PV)	Overall height of the vehicle (PV)
	Width (PV)	Overall width of the vehicle (PV)

The HSIS maintains a large number of crash, roadway, and traffic variables from each participating state. The North Carolina crash database includes accident characteristics, vehicles involved, occupants in the vehicles, and roadway inventory (HSIS- 2006). Ninety variables from the database were selected and obtained (Appendix B). Three sub-files containing crash data for 2003 were analyzed. Based on the crash data, there were 145,769 crashes, involving 253,490 vehicles with a total number of 371,639 occupants. Of these crashes, 4068 are crashes involving a PV and a commercial truck.

4.4 Data Cleaning

Crash records with significant omitted data were excluded. Moreover, records which lacked the required variables listed in Table 4-1 were removed. The final data contained the observations with complete data points. The observations with the following characteristics were removed from the database:

1. Unrecorded impact speeds
2. Unrecorded LT's weight
3. LT's gross weight less than 10,000 lbs
4. Unknown injury severity level of the most severely injured occupant
5. Zero impact speeds for both vehicles
6. Unknown drivers age
7. Unknown drivers gender
8. Unknown or posted speed limit less than five miles per hour
9. Injury severity value outside of the defined range
10. Total number of occupants greater than seven

11. AADT equal to zeros
12. Vehicle Identification numbers (VIN) with less than or greater than seventeen characters
13. Both travel speeds were zero before crash
14. For the same vehicle, if travel speed was zero and impact speed was greater zero
15. Unknown roadway class
16. Unknown roadway geometry,
17. Unknown or ambiguous VIN

The resulting database contained 1,804 two vehicles crashes, with one being a PV and the other an LT.

4.5 Data Processing

The need for vehicle specifications led this research to access other databases. One of these commercially available databases is Auto Check (AC). The AC provides information with regard to vehicle make, year, model, history, and etc. All the VINs which are available in the NC crash database were decoded by accessing AC's database. The next step was to obtain vehicle specifications for each PV based on the vehicle make, model, and year. Vehicle specifications extracted from Visual Statement Inc's database (VS). The VS has developed a comprehensive database containing vehicle specifications for various vehicles. The following vehicle specifications for PVs were extracted; curb weight, overall height, overall length, wheelbase, front overhang, and rear overhang. Although VS had most of the vehicle specifications for the majority of the PVs, however, there were instances where they were missing specifications for some vehicles. In these cases, other sources were consulted; such as, Car Direct (CD). The CD had more of vehicle specifications for less vehicle years.

4.6 Dummy Variables

Dummy variables are used to represent categorical variables. A dummy variable is generally a binary variable taking values of 0 or 1. In this study the following variables were converted to dummy variables; namely, gender, crash type, road geometry, median type, roadway class, road surface condition, weather condition, pavement type, and light condition. A variable with k categories is normally represented by k-1 binary values. As an example, gender with two categories (male and female) was represented by one dummy variable.

4.6.1 Drivers Gender

Table 4-2 represents codes used by NC and this research for the drivers gender. Variable denoted as DRV_SEX in NC's database was converted to drivers gender. As shown in the table NC uses one for male and two for female. In this study male is equal to one (Base) and female as zero. LT drivers genders have same codes as PVs.

Table 4-2 Drivers gender

North Carolina	North Carolina Code	In this study
Variable : DRV_SEX		Variable: Drivers Gender
Male	1	Male=1
Female	2	Female=0

4.6.2 Crash Type

Table 4-3 represents variable name, classification, codes, abbreviation, and new name used by this study. Variable denoted as acctype in NC's database stands for accident type. There are twenty two accident types starting with ran-off road and ending with angle. The numerical values used by NC vary from 1 to 30. As shown in the table, twenty-two categories of crashes were reduced to six. The base for the dummy variable is "Sideswipe".

4.6.3 Road Geometry

In table 4-4 variable denoted as RD_CHAR is used by NC for road characteristics. For this study RD_CHAR converted to road geometry. Eight categories of road geometry reduced to two. As shown in the table, a straight road coded as one and curves as zero.

4.6.4 Median Type

There are eight coded median types in NC database. As shown in table 4-5, median types vary from undivided to curb. Variable denoted as Med_type in NC's database converted to Median type. Eight categories of road median types reduced to two. As shown in the table, divided roads coded as one and undivided as zero.

Table 4-3 Crash type

North Carolina Variable: acctype	North Carolina CODE	In this study Variable: Crash type
RAN OFF ROAD-RIGHT	1	Others
RAN OFF ROAD-LEFT	2	
RAN OFF ROAD-STRAIGHT	3	
JACKKNIFE	4	
OVERTURN/ROLLOVER	5	
OTHER NON-COLLISION	13	
ANIMAL	17	
PARKED MOTOR VEHICLE	20	
BACKING UP	31	
OTHER COLLISION	32	
MOVABLE OBJECT	18	
FIXED OBJECT	19	
REAR END,SLOW OR STOP	21	Rear end
REAR END,TURN	22	
LEFT TURN , SAME ROADWAYS	23	Turns
LEFT TURN , DIFFERENT ROADWAYS	24	
RIGHT TURN, SAME ROADWAY	25	
RIGHT TURN, DIFFERENT ROADWAY	26	
HEAD ON	27	Head on
SIDESWIPE, SAME DIRECTION	28	Side swipe-BASE
SIDESWIPE, OPPOSITE DIRECTION	29	
ANGLE	30	Angle

Table 4-4 Road geometry

North Carolina	North Carolina	In this study
Variable: RD_CHAR	CODE	Variable:Roadway Geometry
STRAIGHT-LEVEL	1	Straight=1
STRAIGHT-HILLCREST	2	
STRAIGHT-GRADE	3	
STRAIGHT-BOTTOM	4	
CURVE-LEVEL	5	Curve=0
CURVE- HILLCREST	6	
CURVE-GRADE	7	
CURVE- BOTTOM	8	

Table 4-5 Median type

North Carolina	North Carolina	In this study
Variable: Med_type	CODE	Variable: Median type
UNDIVIDED	1	Undivided=0
CONTINUOUS TURN LANE	2	Divided=1
POSITIVE BARRIER	6	
GRASS	5	
PAVED MOUNTABLE	3	
PARKLAND, BUSINESS	7	
COUPLET	8	
CURB	4	

4.6.5 Roadway Class

In the first column of table 4-6, variable denoted as rodwycls stands for roadway class. Ten categories of roadway class ranging from urban 2-lane to rural multilane-divided reduced to four categories. As shown in the table, the base for the dummy variable is “Freeways” and the remaining three (2-lane, multilane divided, and multilane undivided) were categorized as non-base.

Table 4-6 Roadway class

North Carolina	North Carolina CODE	In this study
Variables: rodwycls		Variable: Roadway class
URBAN 2-LANE ROADS	3	2-Lane
RURAL 2-LANE ROADS	8	
URBAN FREEWAYS	1	Freeways-BASE
URBAN FREEWAYS< 4 LANES	2	
RURAL FREEWAYS	6	
RURAL FREEWAYS < 4 LANES	7	
URBAN MULTILANE DIVIDED	4	Multilane divided
RURAL MULTILANE DIVIDED	9	
URBAN MULTILANE UNDIVIDED	5	Multilane undivided
RURAL MULTILANE DIVIDED	10	

4.6.6 Road Surface Condition

The variable denoted as RDSURF in NC's database converted to surface condition. As presented in table 4-7, there are eight different surface conditions from unknown to surface covered with ice. Eight categories of road surface condition reduced to three. As shown in the table, the base for the dummy variable is "Dry" and the remaining two (others and wet) will take values of zero and one where applicable.

Table 4-7 Road surface condition

North Carolina	North Carolina	In this study
Variable: RDSURF	CODE	Variable: Surface condition
UNKNOWN	10	
SAND, MUD, DIRT, GRAVEL	7	Others
SNOW	5	
DRY	1	Dry-BASE
WET	2	
WATER STANDING	3	Wet
SLUSH	6	
ICE	4	

4.6.7 Weather Condition

Table 4-8 represents a summary of variable names, classifications, codes, and new name used by NC database and this study. Variable denoted as WEATHER in NC's database converted to weather condition. Seven categories of weather conditions reduced to three. The base for the dummy variable is "clear" and the remaining two (others and rain) are referred to as non-base.

Table 4-8 Weather condition

North Carolina	North Carolina	In this study
Variable: WEATHER	CODE	Variable: Weather condition
SNOW	4	Others
FOG, SMOG,SMOKE	5	
SLEET, HAIL, FREEZING RAIN/ DRIZZLE	6	
OTHER	9	
CLOUDY	2	
CLEAR	1	Clear-BASE
RAIN	3	Rain

4.6.8 Pavement Type

Table 4-9 contains information for variable denoted as RD_PAVE in NC's database. As shown in the table RD_PAVE converted to pavement type. The eight categories of pavement types reduced to three. The base for the dummy variable is "Asphalt" and the remaining two (others and concrete) are referred to as non-base.

4.6.9 Light Condition

In table 4-10, there are seven categories for light condition. Variable denoted as LIGHT in NC's database covers a wide range of light condition from dark unknown to dark road-way not lighted. Seven categories of light conditions reduced to four. As shown in the table, the base for the dummy variable is "day light" and the remaining three (others, dark-lighted, and dark-not lighted) are referred to as non-base.

Table 4-9 Pavement type

North Carolina	North Carolina CODE	In this study
Variable: RD_PAVE		Variable: Pavement type
GRAVEL	5	Others
SAND	6	
SOIL	7	
OTHER	8	
SMOOTH ASPHALT	3	Asphalt-BASE
COARSE ASPHALT	4	
CONCRETE	1	Concrete
GROOVED CONCRETE	2	

4.6.10 Alcohol Flagged

Table 4-11 represents codes used by NC and this study for whether or not there were any alcohol usage detected by the police officer at the scene of the crash. Variable denoted as ALCFLAG in NC's database was converted to alcohol flagged. As shown in the table, NC uses one for yes and zero for no. In this study "yes" was converted to one (Base) and "no" to zero.

Table 4-10 Light condition

North Carolina	North Carolina CODE	In this study
Variable: LIGHT		Variable: Light condition
DARK-UNKNOWN LIGHTING	6	Others
UNKNOWN	8	
DUSK	2	
DAWN	3	
DAYLIGHT	1	Day Light-BASE
DARK-LIGHTED ROADWAY	4	Dark-Lighted
DARK- ROADWAY NOT LIGHTED	5	Dark- Not Lighted

Table 4-11 Alcohol flagged

North Carolina	North Carolina Code	In this study
Variable : ALCFLAG		Variable: Alcohol flagged
N	0	No=0
Y	1	Yes=1

In summary, table 4-12 represents a list of all the explanatory variables that were used in this study. Row one through five consists of binary variables with two distinct categories and ten possible outcomes. Row four through eleven contains six dummy variables with three or more categories and twenty four possible outcomes. Furthermore, for each categorical variable, there is a predetermined base in the last column. The remaining nineteen explanatory variables are called “Ordinary Independent Variables”.

Table 4-12 Binary, dummy, and ordinary explanatory variables

Binary Variables		Corresponding Categories	Base	
1	PV drivers gender	Male-Female	Male	
2	LT drivers gender	Male-Female	Male	
3	Road geometry	Straight-Curve	Straight	
4	Median type	Undivided-Divided	Divided	
5	Alcohol Flagged	Yes-No	Yes	
Dummy Variables			Base	
6	Crash type	Rear end, Turn, Head on, Others, Side swipe, and Angle	Side swipe	
7	Roadway class	2-lane, Freeway, Multilane divided, and Multilane undivided	Freeway	
8	Surface condition	Others, Dry, and Wet	Dry	
9	Weather condition	Others, clear, and rain	Clear	
10	Pavement type	Asphalt, Concrete, and Others	Asphalt	
11	Light condition	Others, Dark lighted ,Dark not-lighted, and Day light	Day light	
Ordinary Independent Variables				
PV drivers age	PV total # of occupants	PV travel speed	PV impact speed	Speed limit
PV overall height	PV front overhang	PV overall width	PV wheel base	PV weight
PV rear overhang	AADT	MVMT	LT impact speed	No of lanes
LT drivers age	LT travel speed	LT weight	LT percentage	

4.7 Variable Analysis

This section will present a thorough graphical and tabular representation for both dependent variable and explanatory variables.

4.7.1 Dependent Variable

Dependent variable is a discrete random variable, representing five levels of injury severities. In NC database, numerical value ranging from 1 to 5 (killed to no injury) is assigned to each observed level of injury severity sustained by the most severely injured occupant in a crash. Incorporating these five numerical values and re-coding them into KABCO categorization would result, K (Killed) =5, A (Incapacitated) =4, B (Non-incapacitated) =3, C (Complaint of pain) =2, and O (No injury) = 1.

Table 4-13 represents frequency distribution for the dependent variable. As shown in the table, no injury (Level 1) has the highest and fatal injury (Level 5) has the lowest frequency respectively.

Table 4-13 Dependent variable distribution

Injury Severity	No. of crashes	Percent of total
1=No Injury (O)	1,179	65.35
2=Possible Injury (C)	381	21.12
3=Non Incapacitating , Evident Injury(B)	158	8.76
4=Incapacitating Injury (A)	46	2.55
5= Fatal Injury (K)	40	2.22
Total	1,804	100

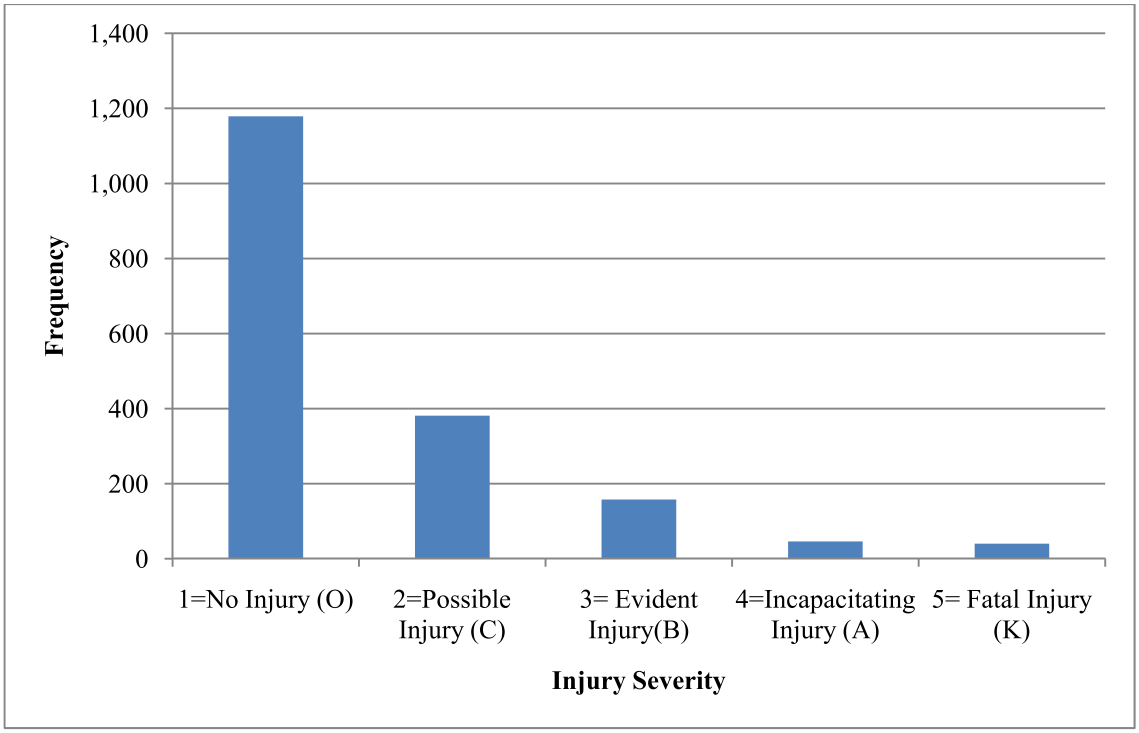

Figure 4-2 Injury severity distribution

Figure 4-2 represents five levels of injury severities sustained by the most severely injured occupant in the NC's sample data.

A graphical test was conducted to see whether or not this sample represents the population (US) trend. The nation's estimates are the calculated averages for KABCO categorization from year 2000 to 2006. Table 4-13 was obtained from Department of Transportation, General Estimates System (GES) through HSIS. Figure 4-3 exhibits the KABCO trends for NC's and the nation's estimates. By dividing the nation's estimates by 100 and comparing them to the NC's sample data, they both seem to have similar trends.

Based on the literature review (chapter 2) drivers age, drivers gender, number of occupants, and crash type were statistically significant across different modeling schemes; therefore, in the following sections these variables and their relationship with the dependent variable are reviewed in detail.

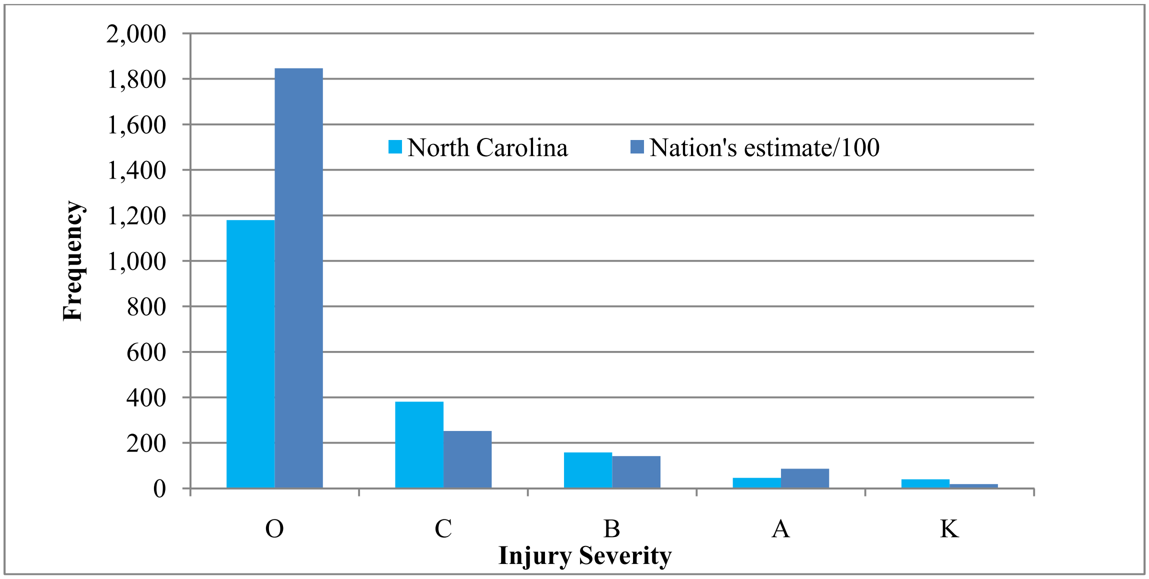

Figure 4-3 KABCO trend for the nation and North Carolina

Table 4-14 Nation's KABCO estimates

Injury Severity	2000	2001	2002	2003	2004	2005	2006	2007
No Injury (O)	187,045	192,331	182,203	193,981	179,826	192,038	171,596	178,402
Possible Injury (C)	33,470	27,096	25,922	24,287	23,776	21,143	23,933	22,276
Evident Injury (B)	14,370	14,318	14,449	15,304	13,761	14,232	14,654	12,589
Incapacitating Injury (A)	9,632	9,740	9,465	7,389	9,010	9,392	7,706	6,654
Fatal Injury (K)	2,451	1,708	1,699	2,211	1,379	1,468	1,937	2,057

4.7.2 Descriptive Statistics

In this section, the interaction between the dependent variable and categorical explanatory variables; namely, drivers gender and crash type are analyzed.

1. Injury severity distribution by drivers gender

Figure 4-4 illustrates injury severity distribution by drivers gender. 57% of the injured were male and 43% female. Other words, there is a ratio of 1.33 of the number of male to female. By referring to Table 4-15 and comparing this average ratio to ratio for each level, one can see that for level 4 and level 5 the ratios change to 1.7 and 2.3 respectively. The data suggest that there were 2.3 times more male drivers involved in fatal crashes than female drivers. In the case of level 4 injury severity, male drivers have been involved 1.7 times more than female drivers.

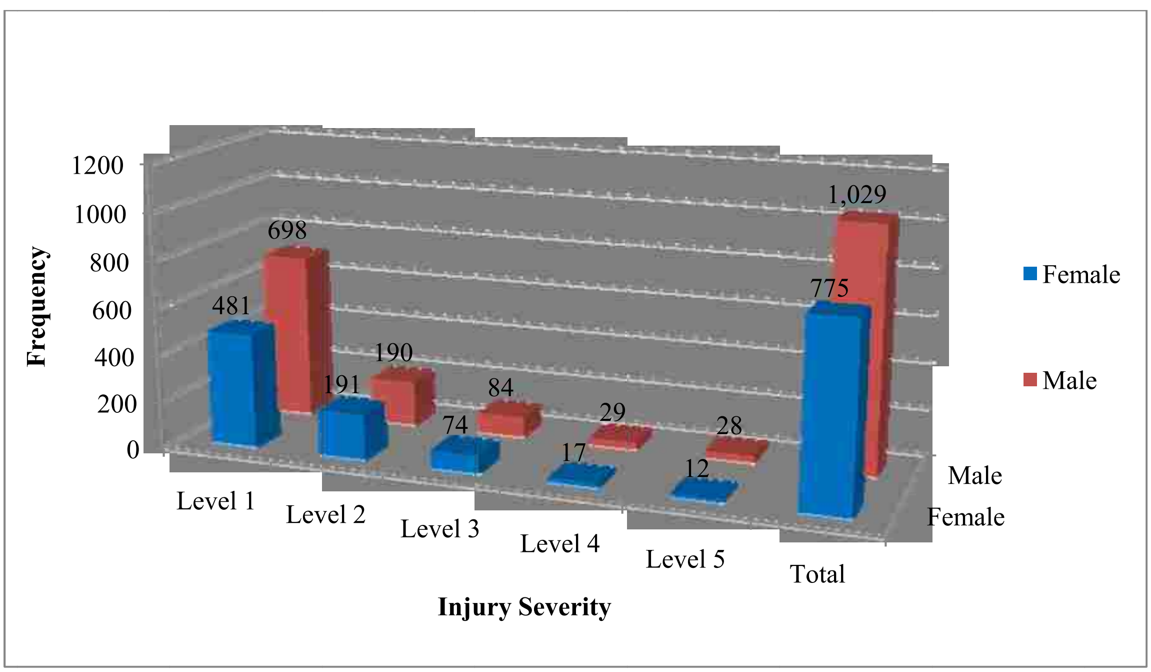

Figure 4-4 Injury severity by drivers gender

Table 4-15 Injury severity ratio by drivers gender

Injury Severity	Female	Male	Male%	Female%	Ratio Male/Female	Total
1	481	698	59.2	40.8	1.5	1,179
2	191	190	49.9	50.1	1.0	381
3	74	84	53.2	46.8	1.1	158
4	17	29	63.0	37.0	1.7	46
5	12	28	70.0	30.0	2.3	40
Total	775	1,029	57.0	43.0	1.3	1,804

2. Injury severity distribution by crash type

In this study, crash type was divided into six categories namely; rear end, angle, sideswipe, head on, turn, and others. As show in Figure 4-5 sideswipe had the highest frequency of the occurrence for level 1 and 2 among all the crash types. On the other hand rear end crashes had the highest frequency for level 2. Table 4-16 represents percentages for each injury severity level and its corresponding crash type. Head on crash type had an increasing trend in terms of injury severity, other words, probability of fatality is much higher given a head on crash occured. On the other hand, in the case of sideswipe, injury severity trend had an opposite direction, which that translates to, given a sideswipe crash occurred, the probability of having a fatal injury was the lowest among the five categories.

Table 4-16 Injury severity by crash type

Injury Severity	Other	Head On	Side Swipe	Angle	Rear End	Turn	Total
1	13%	1%	45%	10%	17%	15%	100%
2	11%	2%	27%	18%	28%	15%	100%
3	9%	3%	21%	22%	17%	28%	100%
4	9%	9%	22%	30%	15%	15%	100%
5	5%	15%	20%	30%	18%	13%	100%
Total	12%	2%	38%	13%	19%	16%	100%

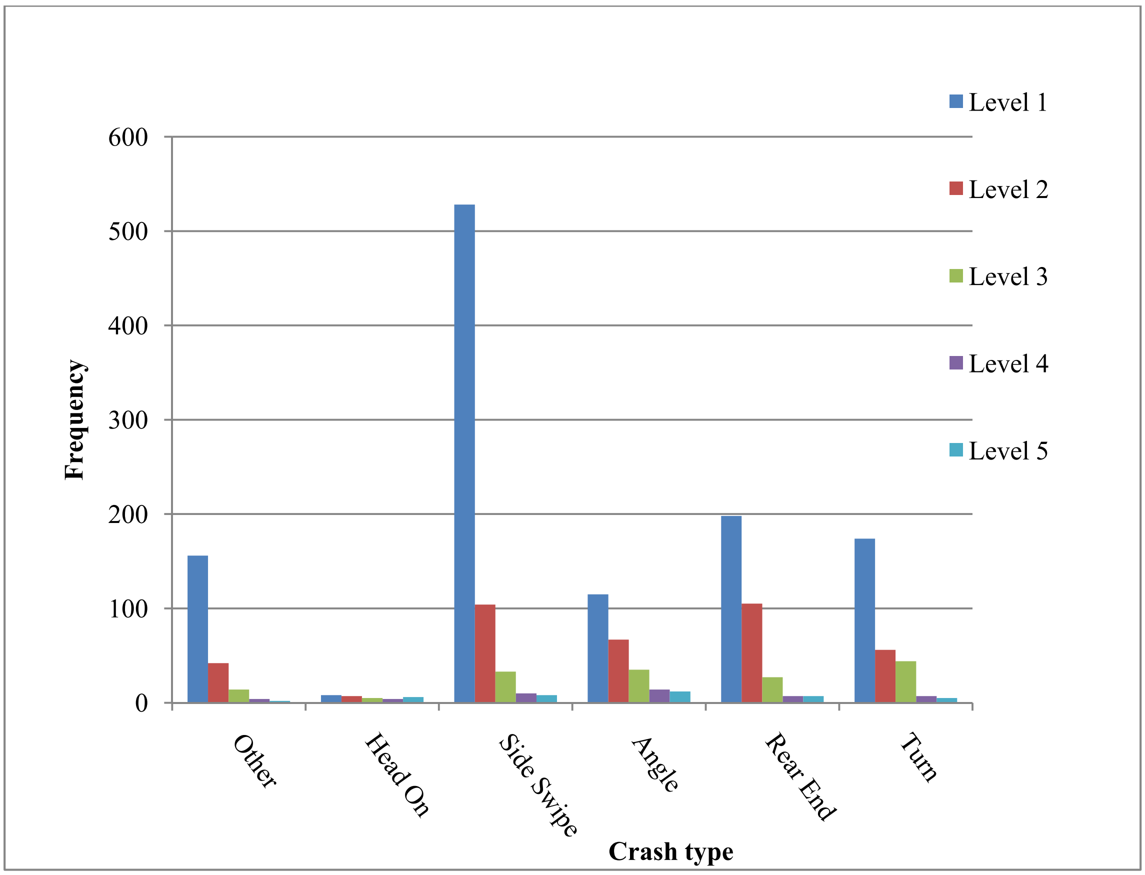

Figure 4-5 Injury severity by crash type

4.7.3 Dependent and Continuous Explanatory Variables

In this section, the interaction between the dependent variable and continuous explanatory variables; namely, number of occupants and drivers age are analyzed.

1. Injury severity distribution by number of occupants

As described in chapter 3, crashes with number of occupants greater than seven were removed from the database. Table 4-17 represents the frequency for different number of occupants and their corresponding injury level. As shown in the table PVs with one occupant (driver) have the highest frequency (1195). Figure 4-6 illustrates a decreasing trend for frequency of occurrence as the number of occupant increases. The data indicate that the highest frequency of fatal and incapacitating crashes occurred when the number of occupant was one or two. This phenomenon is expected, because that is almost ninety percent of the sample size.

Table 4-17 Injury severity by number of occupants

	Number of occupant							Total
	1	2	3	4	5	6	7	
Level 1	822	230	80	29	13	4	1	1179
Level 2	218	103	29	24	5	2	0	381
Level 3	99	40	17	1	1	0	0	158
Level 4	33	8	2	1	2	0	0	46
Level 5	23	11	1	3	1	1	0	40
Total	1195	392	129	58	22	7	1	1804

2. Injury severity distribution by drivers age

To gain more insight in the drivers age distribution, data was divided into four age groups: drivers age<=25, drivers age 26-45, drivers age 46-65, and drivers age 66 and older. Figure 4-7 illustrates that the age group 26-45 has the highest frequency (697) in this dataset.

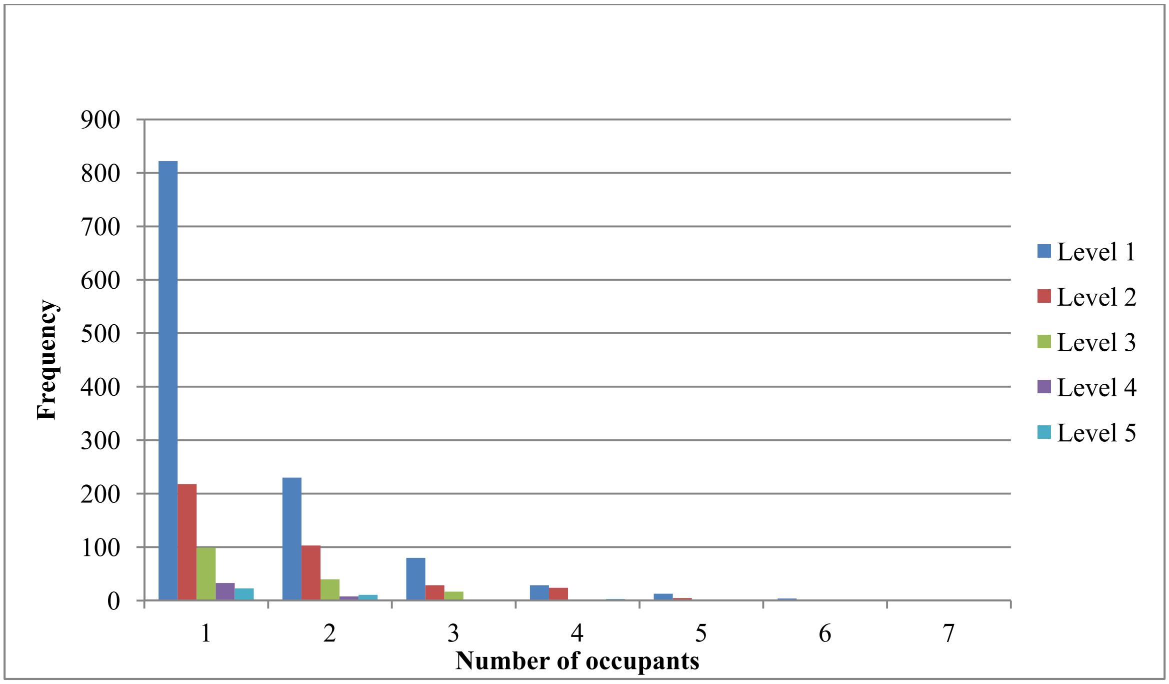

Figure 4-6 Injury severity by number of occupants

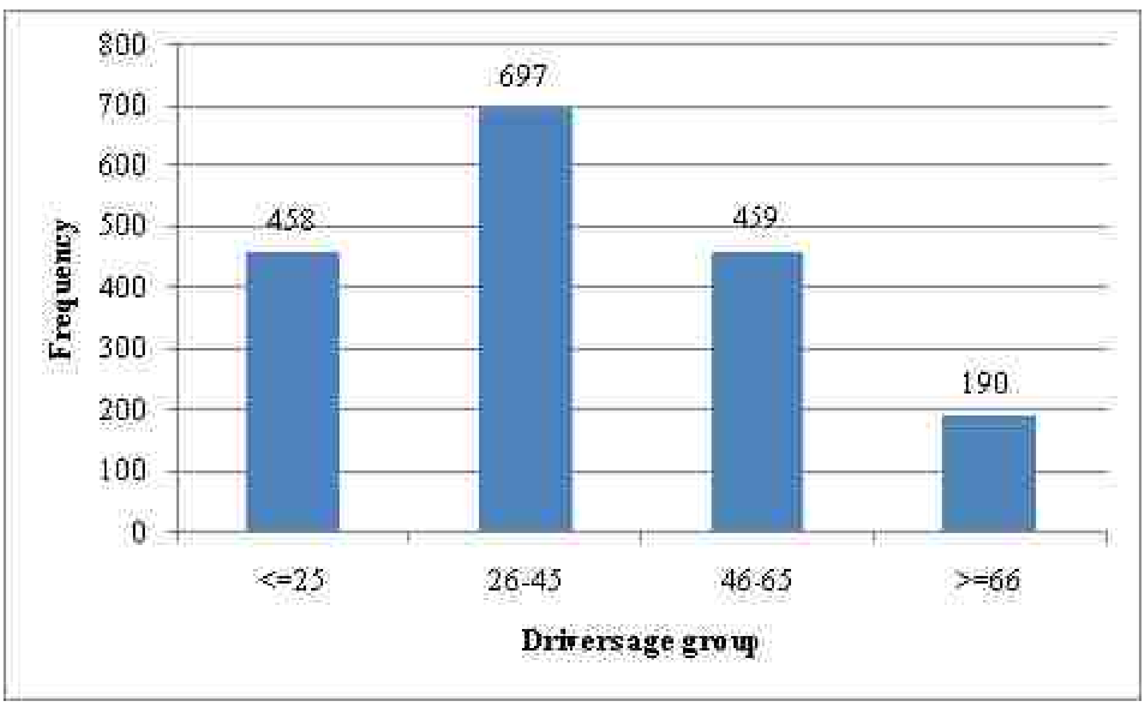

Figure 4-7 Drivers age-group distribution

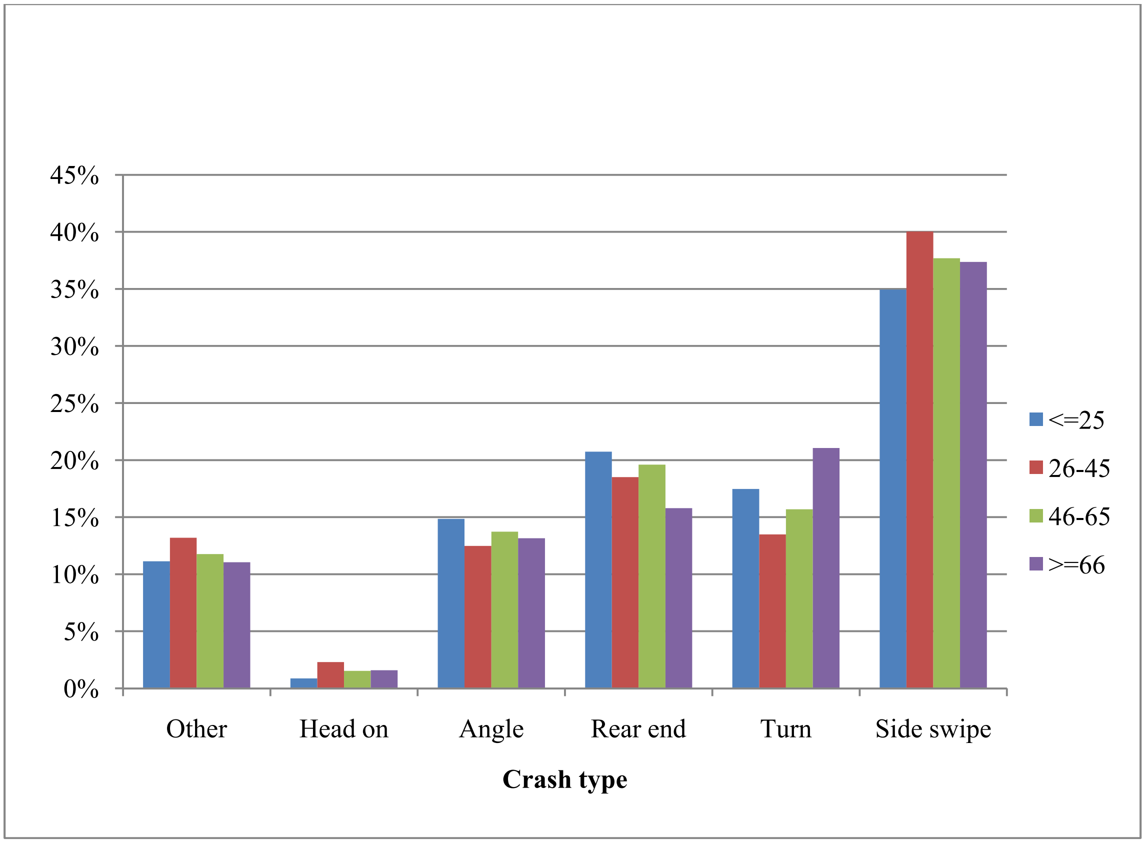

Figure 4-8 Drivers age-group by crash type

Figure 4-8 illustrates crash type distribution by age group. Although age group 26-45 had the highest frequency among other groups, it had the lowest involvement in angle and turn crashes. On the other hand, age group 66 and older with the lowest frequency (190) had the highest involvement in turn crashes and same percentage of involvement in head on crashes as (46-65) age group. Age group <=25 with third frequency had the highest involvement in angle and rear end crashes.

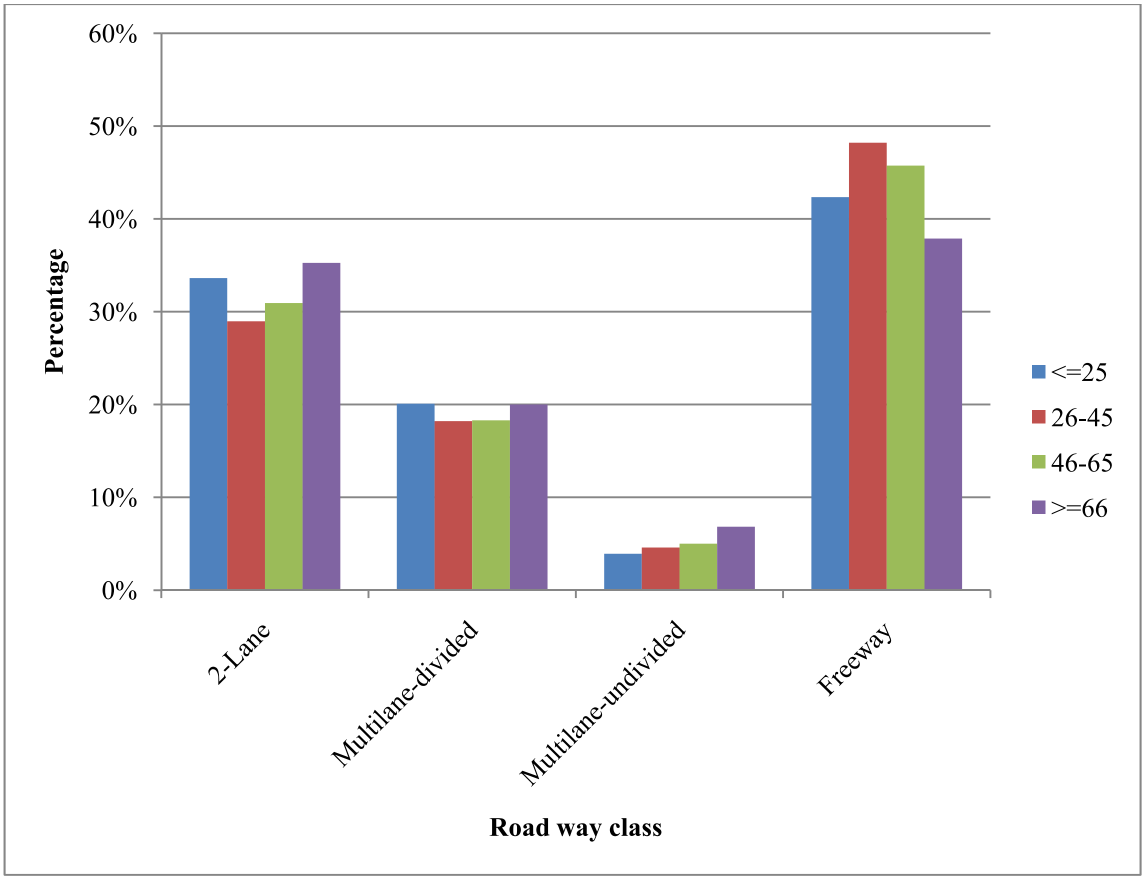

Figure 4-9 Drivers age-group by roadway class

Figure 4-9 illustrates roadway class distribution by age group. Age group with the lowest frequency (66 and older) had the highest crash involvement on multilane undivided and 2-lane facilities. On the other hand, age group with the highest frequency (26-45) had the lowest crash involvement on 2-lane roadways. Furthermore, the youngest age group and the oldest age group had the highest crash involvement on multilane divided facilities.

Figure 4-10 represents injury severity level by drivers age-group. As shown in the graph the youngest and the oldest group had almost the same involvement in level 5 injury crashes. Age group 46-65 had the highest involvement in the most severe crash injuries.

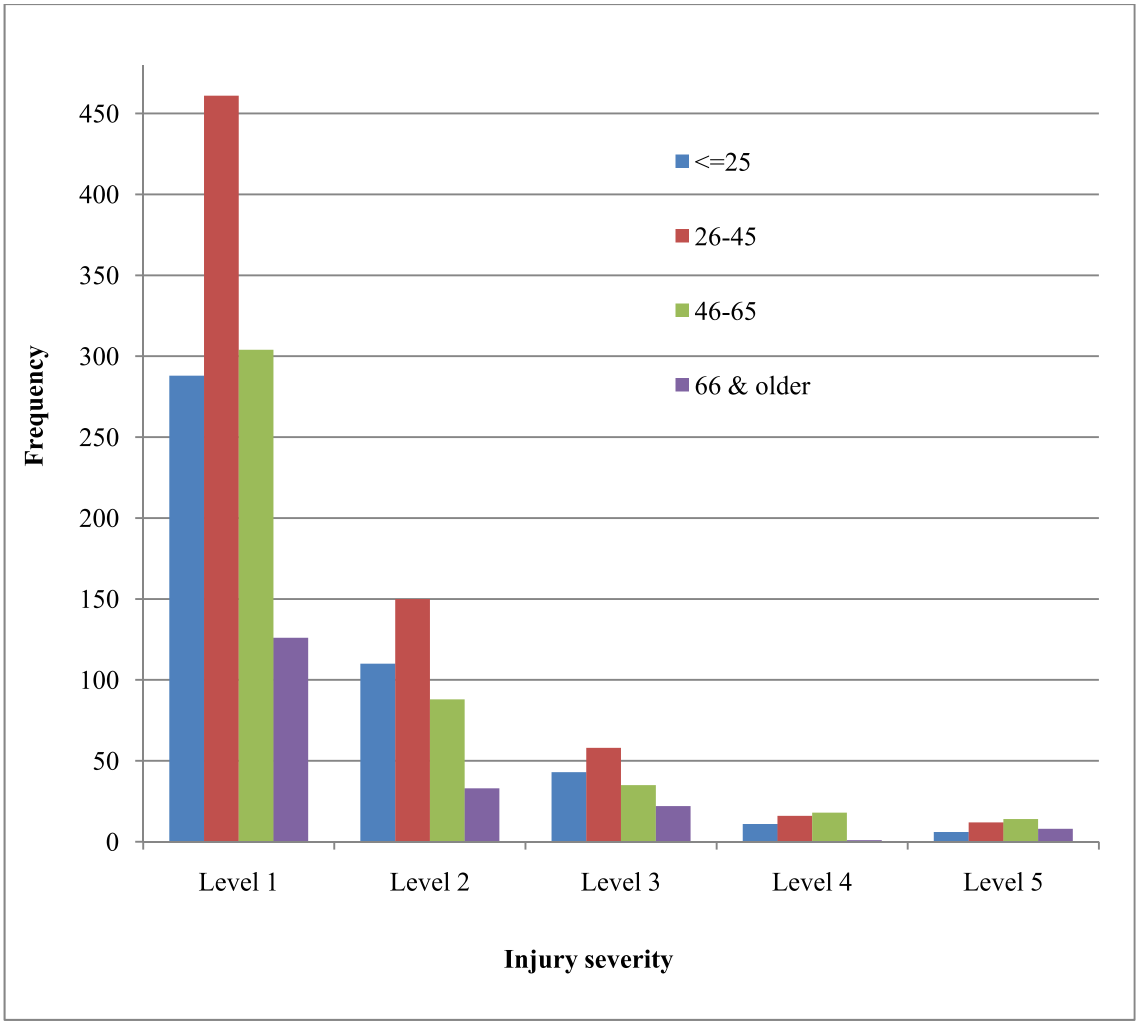

Figure 4-10 Injury severity by drivers age

Table 4-18 and 4-19 illustrate a summary of descriptive statistics for four categories; namely, occupant characteristics, crash environment, crash characteristics, and vehicle characteristics.

Out of 1804 PV drivers, 57% were male and 43% female. The age of these drivers ranged from 15 to 92 with an average of 40.50 years. For the same sample, the LT drivers had 96% male and 4% female with the average age three years older than PV drivers. The percentage of LT drivers suggests that, there were twenty four times more male drivers than female drivers in that line of business.

The number of occupants varied from 1 to 7. A vehicle with one occupant translates into driver-only or an observation with occupant number equal one. Over 66% of the data is one-occupant observation.

In this sample, 99.96% of the observations lacked any alcohol involvement or at least not flagged. In the cases that they were flagged, no injury severity above level 2 was reported.

The Average Annual Daily Traffic (AADT), Million Vehicle Miles Traveled (MVMT), number of lanes, % of trucks and speed limit (Posted speed limit) showed a wide range of data fluctuation. This range was expected because the data covered a wide range of roadway class; namely, Freeway, 2-lane, multilane divided, and multilane undivided.

The median type had a binary value of zero and one. One was assigned to roads divided by median and zero otherwise. 66% of the roadway network had medians and 34% had no medians.

Around 78% of the total crashes took place during daylight, and this was expected since the majority portion of the AADT was taken place during the day light.

Road surface condition had three distinct categories; namely, dry, wet, and other. 78% of the crashes happened on dry and 20% on the wet roads. This lower percentage was expected due to two reasons, first; there were more dry days than wet days, and second, on wet days people would drive slower to avoid crashes.

Table 4-18 Descriptive statistics for driver/occupants and crash environment characteristics

Categories	Explanatory Variable	Mean	SD	Min	Max
Driver/ Occupants Characteristics	Drivers age (PV)	40.50	17.47	15	92
	Drivers gender(PV)	.57	.5	0	1
	# Of occupants (PV)	1.53	.92	1	7
	Drivers age (LT)	43.52	11.76	18	103
	Drivers gender (LT)	.96	.19	0	1
	Alcohol	.00	.04	0	1
Crash Environment at the Crash Location	AADT	40,906	39,762	10	160,000
	Number of lanes	3.85	1.71	2	8
	MVMT	18.53	24.89	.001	203.84
	% of trucks	12.35	6.40	1	46
	Speed limit	53.43	10.05	10	70
	Median type	.66	.47	0	1
	Light condition	(Proportion)			
	1. Other	.04	.19	0	1
	2. Day light	.78	.42	0	1
	3. Dark lighted	.05	.22	0	1
	4. Dark not-lighted	.13	.34	0	1
	Road surface condition	(Proportion)			
	1. Other	.02	.13	0	1
	2. Dry	.78	.42	0	1
	3. wet	.20	.40	0	1
	Pavement type	(Proportion)			
	1. Other	.01	.09	0	1
	2. Asphalt	.94	.23	0	1
	3. Concrete	.05	.21	0	1
	Roadway class	(Proportion)			
	1. Freeway	.45	.50	0	1
	2. 2-lane	.31	.46	0	1
	3. Multilane divided	.19	.39	0	1
	4. Multilane undivided	.05	.21	0	1
	Road geometry	(Proportion)			
	Horizontal tangent	.86	.35	0	1

Table 4-19 Descriptive statistics for crash and vehicle characteristics

Crash Characteristics	Impact speed (PV)	36.20	20.55	1	90
	Travel speed (PV)	42.85	19.54	1	95
	Impact speed (LT)	35.69	20.78	0	73
	Travel speed(LT)	41.30	20.36	0	80
	Crash type	(Proportion)			
	1. Head on	0.02	0.13	0	1
	2. Sideswipe	0.38	0.49	0	1
	3. Angle	0.13	0.34	0	1
	4. Rear end	0.19	0.39	0	1
	5. Turn	0.16	0.37	0	1
	6. Other	0.12	0.33	0	1
Vehicle Characteristics	Weight(LT)	79,432	48,123	10,000	845,000
	Weight(PV)	3,371	825	1,651	7,189
	Front-overhang (PV)	36.95	4.37	24	56
	Rear-overhang (PV)	42.87	5.25	21	67
	Wheelbase (PV)	111.64	14.75	87	172
	Height (PV)	60.59	8.11	47	104
	Width (PV)	71.27	4.57	57	96

The data indicates that 94% of pavement surfaces were asphalt and 6% were concrete and other pavement materials.

Around 45% of the crashes happened on the freeways and 31% on the 2-lane roads. Multilane divided and multilane undivided had the lower percentage of all the crashes.

Travel speeds and impact speeds for both vehicles had a large range with averages ranging from 36 mph to 43 mph.

In the category of crash type, sideswipe had the highest percentage (38%) and head on with the lowest (2%).

For the purposes of this study LTs with minimum 10,000 lbs were selected from the database. The average weight of LTs and PVs were 79,500 lbs to 3,371 lbs respectively. On the average, LTs were twenty four times heavier than PVs. The weights of PVs ranged from 1651 lbs (1993 Suzuki Swift) to 7189 lbs (2000 Ford Excursion).

Front overhangs of PVs ranged from 24 inches (2000 Jeep Wrangler) to 56 inches (1999 Lincoln Town car). Another variable of interest was vehicle height. The vehicle height ranged from 47 to 104 inches with an average height of 60.60.

4.7.4 PV-Drivers Gender

In the following sections descriptive statistics for PV-drivers gender are presented. Table 4-20 and 4-21 illustrate a summary of descriptive statistics for male and female PV drivers.

The study of vehicle characteristics (dimensions) conveyed what the literature in this area of research has been suffering from for decades and that is, researchers have been categorizing the vehicle based on their type; namely, passenger cars, pickup trucks, vans, and SUVs. A thorough analysis of the data showed that a passenger car can be heavier, bigger, and taller than a pickup truck or a van or an SUV or vice a versa. Therefore, it is suggested to introduce additional classification for vehicle type by dimensions.

Table 4-22, 4-23, 4-24, and 4-25 illustrate descriptive statistics for four drivers age groups.

Table 4-20 Comparative descriptive statistics for PV- drivers gender

Categories	Explanatory Variable	Male (N=1029)		Female (775)	
		Mean	SD	Mean	SD
Driver/ Occupants Characteristics	Drivers age (PV)	40.41	17.46	40.61	17.49
	# Of occupants (PV)	1.50	.89	1.56	0.95
	Drivers age (LT)	43.44	11.73	43.61	11.79
	Drivers gender (LT)	.96	.20	0.97	0.18
	Alcohol	.001	.04	0.00	0.04
Crash Environment at the Crash Location	AADT	40468	38171	41488	41800
	Number of lanes	3.85	1.71	3.85	1.71
	MVMT	18.81	24.43	18.16	25.50
	% of trucks	12.80	6.50	11.75	6.27
	Speed limit	53.85	10	52.87	10.10
	Median type	.67	.47	0.65	0.48
	Light condition				
	1. Other	.041	.20	0.03	0.18
	2. Day light	.76	.43	0.80	0.40
	3. Dark lighted	.052	.22	0.05	0.22
	4. Dark not-lighted	.15	.35	0.11	0.31
	Road surface condition				
	1. Other	.024	.15	0.01	0.10
	2. Dry	.76	.43	0.80	0.40
	3. wet	.21	.41	0.19	0.39
	Pavement type				
	1. Other	.009	.098	0.01	0.07
	2. Asphalt	.94	.24	0.96	0.20
	3. Concrete	.05	.22	0.04	0.19
	Roadway class				
	1. Freeway	.46	.50	0.43	0.50
	2. 2-lane	.31	.46	0.32	0.47
	3. Multilane divided	.18	.38	0.20	0.40
	4. Multilane undivided	.048	.21	0.05	0.21
	Road geometry	.87	.34	0.85	0.36

Table 4-21 Comparative descriptive statistics for PV- drivers gender

Categories	Explanatory Variable	Male (N=1029)		Female (N=775)	
		Mean	SD	Mean	SD
Crash Characteristics	Impact speed (PV)	37.97	20.42	33.86	20.49
	Travel speed (PV)	44.63	18.96	40.48	20.05
	Impact speed (LT)	36.56	20.78	34.53	20.73
	Travel speed(LT)	41.97	20.18	40.40	20.58
	Crash type				
	1. Head on	.017	.13	0.02	0.12
	2. Sideswipe	.37	.48	0.38	0.49
	3. Angle	.12	.33	0.15	0.35
	4. Rear end	.20	.40	0.17	0.38
	5. Turn	.16	.36	0.15	0.36
	6. Other	.11	.32	0.13	0.33
Vehicle Characteristics	Weight(LT)	79798	46115	78947	50692
	Weight(PV)	3489	867	3214	737
	Front-overhang (PV)	36.60	4.43	37.40	4.25
	Rear-overhang (PV)	43.40	5.49	42.16	4.82
	Wheelbase (PV)	114.18	16.77	108.28	10.65
	Height (PV)	61.81	8.70	58.95	6.93
	Width (PV)	71.81	4.93	70.55	3.91

Table 4-22 Descriptive statistics drivers age groups of <=25 and 26-45

Categories	Explanatory Variable	Drivers age<=25 (N=458)		Drivers age 26-45 (N=697)	
		Mean	SD	Mean	SD
Driver/ Occupants Characteristics	Drivers gender(PV)	0.59	0.49	0.57	0.50
	# Of occupants (PV)	1.63	0.99	1.54	0.97
	Drivers age (LT)	43.75	11.48	43.76	11.82
	Drivers gender (LT)	0.97	0.16	0.96	0.20
	Alcohol	0.00	0.07	0.00	0.00
Crash Environment at the Crash Location	AADT	37,621	37,343	44,504	42,162
	Number of lanes	3.79	1.72	3.97	1.76
	MVMT	17.42	25.04	19.95	25.34
	% of trucks	11.79	6.13	12.54	6.47
	Speed limit	53.96	9.20	53.76	10.00
	Median type	0.65	0.48	0.69	0.46
	Light condition				
	1. Other	0.04	0.19	0.05	0.22
	2. Day light	0.76	0.43	0.76	0.43
	3. Dark lighted	0.06	0.24	0.05	0.22
	4. Dark not-lighted	0.14	0.35	0.14	0.35
	Road surface condition				
	1. Other	0.02	0.13	0.02	0.15
	2. Dry	0.75	0.43	0.76	0.43
	3. wet	0.23	0.42	0.22	0.41
	Pavement type				
	1. Other	0.01	0.11	0.01	0.08
	2. Asphalt	0.94	0.24	0.94	0.24
	3. Concrete	0.05	0.21	0.05	0.23
	Roadway class				
	1. Freeway	0.42	0.49	0.48	0.50
	2. 2-lane	0.34	0.47	0.29	0.45
	3. Multilane divided	0.20	0.40	0.18	0.39
	4. Multilane undivided	0.04	0.19	0.05	0.21
	Road geometry	0.86	0.35	0.85	0.35

Table 4-23 Descriptive statistics drivers age groups of <=25 and 26-45

Categories	Explanatory Variable	Drivers age<=25 (N=458)		Drivers age 26-45 (N=697)	
		Mean	SD	Mean	SD
Crash Characteristics	Impact speed (PV)	36.80	20.78	37.59	19.83
	Travel speed (PV)	44.36	19.41	44.52	18.43
	Impact speed (LT)	35.07	20.83	36.72	20.69
	Travel speed(LT)	40.34	20.22	42.37	20.09
	Crash type				
	1. Head on	0.01	0.09	0.02	0.15
	2. Sideswipe	0.35	0.48	0.40	0.49
	3. Angle	0.15	0.36	0.12	0.33
	4. Rear end	0.21	0.41	0.19	0.39
	5. Turn	0.17	0.38	0.13	0.34
	6. Other	0.11	0.31	0.13	0.34
Vehicle Characteristics	Weight(LT)	76841	31150	83276	62732
	Weight(PV)	3154	800	3449	884
	Front-overhang (PV)	36.41	3.86	36.54	4.18
	Rear-overhang (PV)	41.77	4.68	42.73	5.21
	Wheelbase (PV)	109.23	14.00	112.54	15.58
	Height (PV)	59.22	7.87	61.52	8.53
	Width (PV)	70.19	4.35	71.46	4.71

Table 4-24 Descriptive statistics drivers age groups of 46-65 and 66 and older

Categories	Explanatory Variable	Drivers age 46-65 (N=459)		Drivers age 66 &older (N=190)	
		Mean	SD	Mean	SD
Driver/ Occupants Characteristics	Drivers gender(PV)	0.56	0.50	0.57	0.50
	# Of occupants (PV)	1.43	0.81	1.47	0.76
	Drivers age (LT)	43.18	12.35	42.88	10.70
	Drivers gender (LT)	0.96	0.20	0.96	0.20
	Alcohol	0.00	0.05	0.00	0.00
Crash Environment at the Crash Location	AADT	42,672	40,156	31,357	32,921
	Number of lanes	3.88	1.70	3.50	1.48
	MVMT	18.20	23.84	16.80	25.28
	% of trucks	12.65	6.50	12.25	6.51
	Speed limit	52.55	10.75	53.08	10.38
	Median type	0.67	0.47	0.58	0.49
	Light condition				
	1. Other	0.03	0.17	0.02	0.12
	2. Day light	0.79	0.41	0.86	0.34
	3. Dark lighted	0.06	0.23	0.03	0.16
	4. Dark not-lighted	0.12	0.33	0.09	0.29
	Road surface condition				
	1. Other	0.01	0.11	0.02	0.12
	2. Dry	0.80	0.40	0.86	0.34
	3. wet	0.19	0.39	0.12	0.33
	Pavement type				
	1. Other	0.01	0.08	0.01	0.07
	2. Asphalt	0.94	0.23	0.98	0.14
	3. Concrete	0.05	0.22	0.02	0.12
	Roadway class				
	1. Freeway	0.46	0.50	0.38	0.49
	2. 2-lane	0.31	0.46	0.35	0.48
	3. Multilane divided	0.18	0.39	0.20	0.40
	4.Multilane undivided	0.05	0.22	0.07	0.25
	Road geometry	0.87	0.34	0.86	0.35

Table 4-25 Descriptive statistics drivers age groups of 45-65 and 66 and older

Categories	Explanatory Variable	Drivers age 45-65 (N=459)		Drivers age 66 &older (N=190)	
		Mean	SD	Mean	SD
Crash Characteristics	Impact speed (PV)	35.17	20.84	32.13	21.33
	Travel speed (PV)	41.04	20.20	37.47	21.00
	Impact speed (LT)	34.24	21.07	36.87	20.14
	Travel speed(LT)	40.36	21.08	41.95	19.81
	Crash type				
	1. Head on	0.02	0.12	0.02	0.12
	2. Sideswipe	0.38	0.49	0.37	0.49
	3. Angle	0.14	0.34	0.13	0.34
	4. Rear end	0.20	0.40	0.16	0.37
	5. Turn	0.16	0.36	0.21	0.41
	6. Other	0.12	0.32	0.11	0.31
Vehicle Characteristics	Weight(LT)	77255	41586	76839	31207
	Weight(PV)	3454	795	3405	616
	Front-overhang (PV)	37.13	4.58	39.28	4.89
	Rear-overhang (PV)	43.36	5.41	44.84	5.58
	Wheelbase (PV)	113.08	15.71	110.72	9.29
	Height (PV)	61.01	8.05	59.46	6.56
	Width (PV)	71.71	4.58	72.11	4.05

CHAPTER 5

MODEL RESULTS AND ANALYSIS

5.1 Introduction

The contributions of significant variables from driver/occupants characteristics, crash environment at the crash location, crash characteristics, and vehicle characteristics to injury severity were modeled using an OP and MNL models. In the case of OP a general model was estimated applying the methodology set forth in chapter three. Three sets of models were estimated using MNL technique.

In the case of MNL, odds ratio analysis and the marginal impact of the significant variables on the predicted probabilities are presented. All the models were estimated using the STATA statistical software package.

Furthermore, it was decided to reduce the number of injury levels from five to three levels. To achieve the new injury categories level 2 and level 3 were combined together and so as level 4 and level 5. Table 5-1 represents the injury severity level categorization.

Table 5-1 Injury severity distribution

Injury Severity	No. of crashes	% of total
Level 1= No injury	1,179	65.35
Level 3= (level 2 + level 3)= Evident injury	539	29.88
Level 5= (level 4 + level 5)=Fatal/incapacitating injury	86	4.77
Total	1,804	100.00

5.2 Ordered Probit Model

1. Final model

Table 5-2 presents the final model using OP technique. Out of forty two variables, fifteen were statically significant.

2. Scalar measures of model fit

Table 5-3 presents a summary of scalar of measures of fit. Final model has an LR value of 293 with P=0.000 which is an indication of a highly significant model. Furthermore, additional scalars are: AIC (1.424), pseudo ρ^2 (.104) and count R^2 of 67%.

Table 5-2 Final model using an Ordered Probit Model

Variables		Coefficient	P-value
PV Drivers Gender		-0.1239	0.047
PV Total # of Occupants		0.1226	0.000
Posted Speed Limit		0.0173	0.000
PV Travel Speed		0.0103	0.000
Crash Type Head On		1.2848	0.000
Crash Type Angle		0.8237	0.000
Crash Type Rear		0.4590	0.000
Crash Type Turn		0.3165	0.001
PV Curb weight		-0.0002	0.007
PV overall Width		0.0205	0.086
AADT/1000/Lane		-0.0123	0.058
Road Class 2 Lane		0.7449	0.000
Road Class Multilane Divided		0.4334	0.001
Road Class Multilane Undivided		0.4505	0.022
Lighting Condition Dark- No Light		0.2822	0.001
Thresholds	τ_1	3.2956	
	τ_2	4.7520	

Table 5-3 Scalar measures of fit for Full and Final models using Ordered Probit.

SCALARS	Final Model	Initial Model
Likelihood Ratio (Degrees of Freedom) , P=0.000	293.57 (15)	327.08 (42)
Log Likelihood (Intercept Only)	-1414.346	-1414.34
Log Likelihood (Full Model)	-1267.561	-1250.80
Akaike Information Criteria (AIC)	1.424	1.435
Count R^2 (% of Correctly Predicted)	67%	68%
Pseudo ρ^2	0.104	0.116
Number of Observations	1804	1804

3. Model assumption test

The validity of the assumption was tested by a Wald test proposed by Brandt (1990). An approximate LR was estimated. The LR test compares the log likelihood from OP with that of (J-1) binary probit models. The null hypothesis assumes all the estimated coefficients are equal.

$$H_o:\ \hat{\beta}_1 = \hat{\beta}_2 = \cdots\cdots = \hat{\beta}_{j-1}$$

The results of the assumption test rejected the null with the calculated chi^2 (15) =29.46. As previously mentioned, if parallel regression assumption doesn't hold, one can utilize the MNL model (Scott Long and Jeremy Freese (2006)).

In this study a systematic approach for calibrating each model was employed. It was decided to show the final model using OP for the comparisons with other models using MNL.

5.3 Multinomial Logit Model

Three sets of multivariate MNL models of injury severity were estimated. The first one was a general model which consists of all data points combined into one model. The second set was gender models which separates the data by the gender of the driver of the PV. The last set was age models which separate the data into four groups based on the age of the drivers of PVs.

Each set was analyzed individually by interpreting its Odds Ratio (OR) and illustrating their marginal impacts on the predicted probability of injury severity. The OR ranges from zero to positive infinity (Washington et al. 2003). When a continuous variable increases by one unit and the corresponding OR ranges between zero and one that translates into a reducing impact on the injury severity. In the same example, if the OR is greater than one, it means increasing impact on the injury severity. For this study to show these impacts graphically, it was decided to subtract a one from the OR values and call it Differential Odds Ratio (DOR). As an example, OR=0.30 for a continuous variable means that, increasing that variable by one unit reduces the odds of the corresponding injury severity comparing to base by 70% (1-0.30), therefore, by subtracting a one from 0.30 would result a DOR of -0.70. The DOR represents the direction (-) and the magnitude (0.70). Same approach was used for all ORs regardless of the values. As an example, OR=2, translates into DOR of 100% increase (2-1=100%).

Unlike OP, explanatory variables in a MNL model may have opposing impacts on the predicted probabilities of injury severity. As an example in an OP model, a variable may have either a positive or a negative impact on the predicted probability of all the injury severity levels, whereas in a MNL model a variable may have a positive impact on one injury severity level and a negative impact on another level. Therefore, for each model there are graphical representations of their impacts for each level. At the end, variables of interest are analyzed across all the models.

5.3.1 General Model

1. Final model

Table 5-4 presents the final general model using MNL technique. Out of forty two variables, seventeen were statically significant and are presented in the table. In this table, each predictor is illustrated with its corresponding logit coefficients, ORs, and P-values. By referring to the table, level 3/ level 1, translates into evident injury compared to no injury. As mentioned in chapter 3, the comparison group is level 1 and each level was compared to the comparison group.

2. Scalar measures of model fit

Table 5-5 shows the scalar of the measures of fit. As shown in the table, final model has an LR value of 349 with P=0.000 which is an indication of a highly significant model. Furthermore, additional scalars are: AIC (1.415), pseudo ρ^2 (.123), and the count R^2 of 68%.

Table 5-4 Final general model using multinomial logit methodology

VARIABLES	LEVEL 3/ LEVEL 1			LEVEL 5/ LEVEL 1		
	Logit Coeffcient	Odds ratio	P-value	Logit Coeffcient	Odds ratio	P-value
Drivers age	*NA*	*NA*	*NA*	0.0187	1.02	0.008
Drivers gender	-0.4288	0.65	0.000	*NA*	*NA*	*NA*
# Of occupants	0.2219	1.25	0.000	0.3202	1.38	0.008
2-Lane	0.6756	1.97	0.000	2.1766	8.82	0.000
Multilane-divided	*NA*	*NA*	*NA*	1.3590	3.89	0.008
Dark-not lighted	0.5016	1.65	0.002	0.6078	1.84	0.065
AADT/1000/lane	-0.0292	0.97	0.008	-0.0678	0.93	0.053
Speed limit (Posted)	*NA*	*NA*	*NA*	0.0572	1.06	0.007
Head on	1.5216	4.58	0.002	3.1462	23.25	0.000
Rear-end	0.8826	2.42	0.000	0.6704	1.96	0.065
Angle	1.2717	3.57	0.000	2.1282	8.40	0.000
Turn	0.6489	1.91	0.000	*NA*	*NA*	*NA*
Travel speed	0.0148	1.01	0.000	0.0344	1.04	0.000
Weight	-0.5659	0.57	0.000	*NA*	*NA*	*NA*
Front-overhang	-0.3423	0.71	0.057	*NA*	*NA*	*NA*
Rear-overhang	0.4390	1.55	0.007	-0.9366	0.39	0.007
Width	0.7144	2.04	0.016	*NA*	*NA*	*NA*
Constant	-5.5017	*NA*	0.000	-9.2271	*NA*	0.002

Table 5-5 Scalar measures of fit for the final general model

Scalars	Full model	Final model
Likelihood Ratio (Degrees of Freedom, P=0.000)	403.645 (84)	348.926(34)
Log Likelihood (Intercept Only)	-1414.346	-1414.346
Log Likelihood (Full Model)	-1212.524	-1239.883
Akaike Information Criteria(AIC)	1.440	1.415
Pseudo ρ^2	0.143	.123
Count R^2 (% of Correctly Predicted)	69%	68%
Number of Observations	1804	1804

3. Model assumption test

The Independence of Irrelevant Alternatives (IIA) in MNL implies that, the alternatives are irrelevant, other words deleting or adding an alternative does not impact the odds between other alternatives. As mentioned in the methodology section, in this study, Small-Hsiao (SH) test was employed. In order to calculate SH, the sample was randomly divided into two equal sizes. $\hat{\beta}_u^{S_1}$ and $\hat{\beta}_u^{S_2}$ were estimated for the unrestricted models. Using a restricted model for the second subsample will result $\hat{\beta}_r^{S_2}$. The SH is distributed as chi-squared and calculated using equation 3-32. The hypotheses for the SH test was set as follows.

H_o: Odds (Outcome-J versus outcome- K) are independent of other alternatives

H_A: Odds (Outcome-J versus outcome- K) are not independent of other alternatives

Table 5-6 represents the results of SH test. The null hypothesis was accepted, other words, the ORs for (level 3) / (level 1) do not depend on (level 5) / (level 1).

Table 5-6 Model assumption test for the general model

Injury level	lnL(full)	lnL(omit)	df	evidence
Evident	-134.392	-127.729	18	for H_o
Fatal/Incapacitating	-517.518	-509.744	18	for H_o

Analysis of results

1. Predicted probability for each injury level

The results of using equation 3-31 and calculating the Pr(y=m|X) are tabulated in Table 5-7. In this study, X is the average value of the independent predictor. As shown in the table, the predicted probability for level 1 has the highest and level 5 has the lowest value (as expected). In the following sections, the impacts of significant variables on the predicated probability for each outcome are analyzed.

Table 5-7 Predicted probability for the general model

Pr (y=m|X)			
	Pr(y=1|X)	Pr(y=3|X)	Pr(y=5|X)
Predicted probability	0.6866	0.2910	0.0224

2. Predicted probability for each observation

Table 5-8 summarizes the predicted versus observed values of each injury severity level. This table is the result of counting all the predicted probabilities for each injury level and comparing them with the observed level. As shown in this table, count R^2 for each level is as follows: level 1 (91%), level 3 (25%), and level 5 (16%). Figure 5-1 is a graphical representation of the observed versus predicted injury severity levels (missed prediction). As shown in the graph, the decreasing trends among various levels were maintained.

Table 5-8 Predicted correctly for each injury level

Injury Severity		Observed levels			Total	Missed predictions
		Level 1	Level 3	Level 5		
Predicted levels	Level 1	1077	97	5	1179	9%
	Level 3	400	134	5	539	75%
	Level 5	51	21	14	86	84%
Total		1528	252	24	1804	32%
Count R^2 =% correctly predicted		91%	25%	16%	68%	

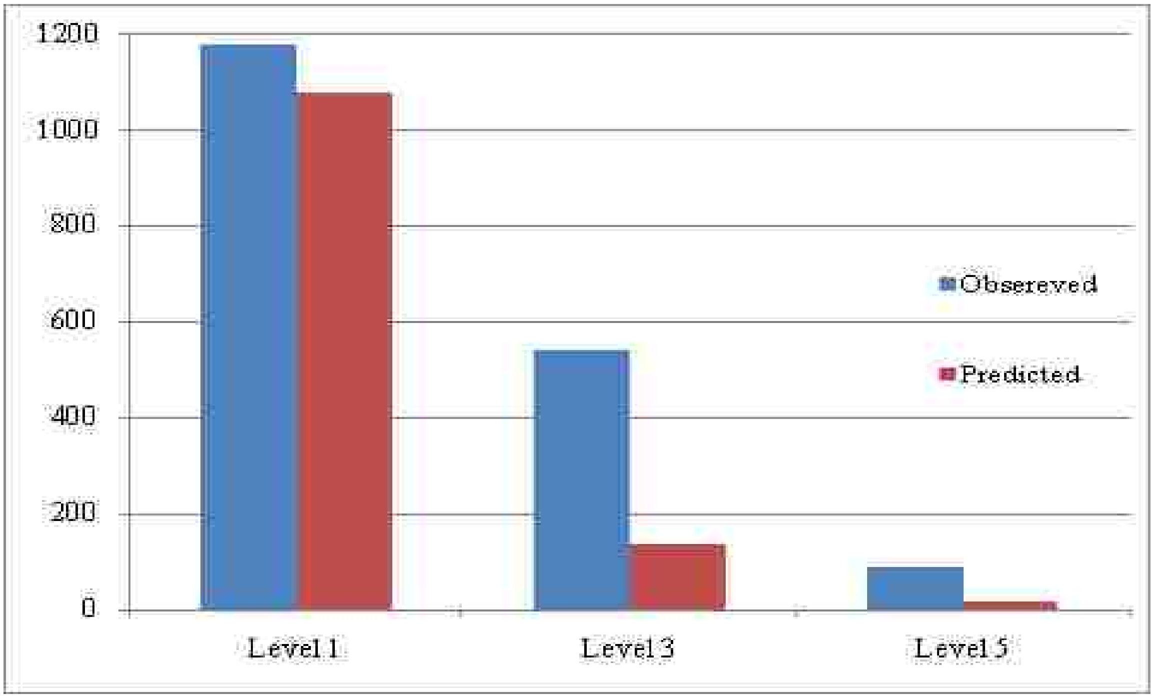

Figure 5-1 Model's performance

3. Odds Ratio and marginal impact

The marginal impacts of each variable on the predicted probabilities are presented in Table 5-9. There are two types of variables in the table; namely, continuous and dummy variables. Dummy variables are denoted by 0->1. As an example, head on crashes compared to sideswipes increase the predicted probability of an evident injury by 24% and fatal injury by almost 18%. Similarly, by increasing the continuous variable by one unit while holding the remaining variables constant would indicate the marginal impact of that specific variable on the predicted probability of each level. As an example, by increasing the rear overhang by one foot, the predicted probability of a fatal injury would decrease by almost 2% while holding other variables constant.

Table 5-9 Marginal effects for the general

Variables	Discrete effect	Level 1	Level 3	Level 5
Drivers age		0.0002	-0.0006	0.0004
Drivers gender	0->1	0.0834	-0.0908	0.0074
# of occupants		-0.0493	0.0437	0.0056
2-Lane	0->1	-0.186	0.1182	0.0678
Multilane-divided	0->1	-0.0911	0.0489	0.0422
Dark-not lighted	0->1	-0.1162	0.105	0.0112
AADT/1000/lane		0.0069	-0.0056	-0.0013
Speed limit		-0.0032	0.002	0.0012
Head on	0->1	-0.4156	0.2378	0.1778
Rear-end	0->1	-0.2007	0.1922	0.0085
Angle	0->1	-0.3205	0.2553	0.0652
Turn	0->1	-0.1457	0.1406	0.0051
Travel speed		-0.0035	0.0028	0.0007
Weight		0.1133	-0.1167	0.0034
Front-overhang		0.069	-0.0704	0.0014
Rear-overhang		-0.0733	0.0967	-0.0234
Width		-0.149	0.145	0.004

In the following sections the OR and marginal impacts on predicted probabilities of injury levels are evaluated for each significant variable by category.

1. Driver/ Occupant characteristics

PV-drivers age

By referring to Table 5-4, while increase on the drivers age by one year has no statistical significance on increasing the likelihood of sustaining a level 3 injury, however, the odds of level 5 increase by 2% and the predicted probability increases by (+0.0004). This finding is consistent with the finding of other researcher (O'Donnell and Connor 1996; Farmer et al. 1997; Khattak et al. 2002; Ulfarsson and Mannering 2004). As an example, comparing two drivers with 20 years of age difference would result in an OR of 1.45. Meaning that the older driver has a 45% higher risk of sustaining a fatal injury compared to the younger driver while holding other variables constant, given a crash occurred. Perhaps this is an indicative of physiological reality of the older drivers population. As drivers get older their perception reaction time increases which that could increase the likelihood of sustaining a level 5 injury.

PV-drivers gender

Gender was identified as a significant variable in this study. This finding is consistent with the finding of other researchers (Duncan et al. 1998; Khattak et al. 2002; Abdel –Aty and Abdelwahab 2002; Kockelman and Kweon 2002; Angel and Hickman 2009). According to the results, having a male driver reduces the odds of sustaining a level 3 injury by 35% with a decrease in the predicted probability by (-0.0908), while gender difference does not play a role in affecting the likelihood of a fatal crash.

Number of occupants

The model indicates, as the number of occupant increases, the injury severity increases (OR: 1.25 to 1.38) with increases on both predicted probabilities for evident and fatal/incapacitating injuries by (+0.0437 and +0.0056) respectively. With each addition of an occupant, the odds of evident and fatal injuries increase by 25% and 38% respectively. This finding is consistent with the finding of other researchers (Chang and Mannering 1999; Wang et al. 2009). This finding is quite understandable because the higher the number of occupants the higher the probability of at least one occupant sustaining injury.

2. Crash Environment at crash location

AADT/lane

Crashes that occur at roadways with higher AADT/lane had decreasing impacts on injury severity (OR: .97 to .93). The predicted probabilities of level 3 and level 5 are impacted negatively by (-0.0056, -0.0013) respectively. This finding is consistent with the finding of other researches (Abdel –Aty and Abdelwahab 2002; Donnell and Mason 2004). Perhaps as AADT/lane increases, travel speed decreases and that could have a reduction on the intensity of the crash and consequently reduction in the level of injury.

Speed limit

This finding is consistent with the finding of other studies (Duncan et al. 1998; Renski et al. 1999; Chang and Mannering 1999; Khattak et al. 2002; Abdel –Aty and Abdelwahab 2002). From the data, increase in posted speed limit is more likely to result an increase in the odds of a fatal/incapacitating injury by 6% (OR: 106) and an increase on the predicted probability by (+ 0.0012), while it does not have any statistically significant impact on the evident injury.

Dark not lighted

In dark roads with no lighting present, the odds of level 3 and level 5 are more likely to increase by 65% and 84% with increase on the predicted probabilities by (+0.1050 and +0.0112) respectively. Other words, dark roads with no lights have an increasing impact on the injury severity sustained by PVs occupants. This finding is consistent with the finding of other researches (Duncan et al. 1998; Khattak et al. 2002; Abdel –Aty and Abdelwahab 2002; Angel and Hickman 2009).

2-lane road

Referring to table 5-4, crashes on 2-lane roadways are more severe, with 97%-782% greater odds of no-injury crashes (OR: 1.97-8.82). From the data, 2-lane facilities are more likely to have an increasing impact on injury severity and the largest OR and probability values (0.1182 and 0.0678) in crash environment at crash location category. This finding is consistent with the finding of other researches (Khattak et al. 2002; Khorashadi et al. 2005).

Multilane divided

Crashes on multilane-divided roadways are more likely to result in level 5 injuries (OR: 3.89), that is a significant increase in the odds (almost 300%). Driving on these type of facilities increases the predicted probability of level 5 by (+.0422), but do not have any statistically significant impact on the probability of evident injury. This finding is consistent with the finding of other research (Khattak et al. 2002).

3. Crash Characteristics

Travel speed

This finding is consistent with the finding of other researches (O'Donnell and Connor 1996; Abdel –Aty and Abdelwahab 2002). The model shows that, higher PV-travel speed has higher impact on the injury severity. Increasing the PV-travel speed by one mile/hour increase the odds of suffering an evident and fatal injury by 1% to 4% and the predicted probabilities of these levels by (+0.0028 and +0.0007) respectively. Furthermore, if travel speed is increased by 30 miles/ hour the odds of suffering an evident and fatal injury increases by 34% to 99%.

Head on

In the category of crash type, the head on has the highest values of OR for injury severity. This finding is consistent with the finding of other studies (O'Donnell and Connor 1996; Kockelman and Kweon 2002). Head on crashes increase the odds of evident and fatal/incapacitating injury, 5 and 23 times, respectively (OR: 4.58-23.25). Head on crashes increase the predicted probability of level 3 and 5 by (+.2378 and +.1778) correspondingly. Due to the crash type (head on), the large magnitude of these numbers are expected.

Rear end

The model indicated that, the rear-end crashes increase the odds of suffering evident and fatal/incapacitating injury by 150% and 100%. The predicted probability of these levels increase by (+0.1922 and +0.0085) respectively. This finding is consistent with the finding of other studies (O'Donnell and Connor 1996; Kockelman and Kweon 2002; Donnell and Mason 2004).

Angle

The results indicated that, the angle collisions are the second most severe crashes in the category of the crash type. Angle crashes increase the odds of evident and fatal/incapacitating injury, 4 and 8 times, respectively (OR: 3.57-8.40). The variable has a rising impact on the predicted probability of level 3 and 5 (+0.2553 and +0.0652). This finding is consistent with the finding of other studies (O'Donnell and Connor 1996; Wang and Abdel-Aty (2008).

Turn

Turn crashes are more likely to result in evident injury, but do not have any statistically significant effect on the probability of level 5. These type of crashes increase the odds and predicted probability of the evident injury by (91% and +0.14) correspondingly. This finding is consistent with the finding of other studies (O'Donnell and Connor 1996; Abdel-Aty and Keller 2005; Wang and Abdel-Aty 2008; Xie et al. 2009).

4. Vehicle characteristics

Weight

In the category of vehicle characteristics, PV-curb weight (weight) has the highest reduction in the odds of level 3. Other words, by increasing the weight, odds of evident injury reduce by 43% (OR: 0.57), but increasing the weight does not have statistically significant impact on the probability of sustaining a level 5 injury. By increasing the weight of PVs by one unit (1000 lbs) will reduce the predicted probability of evident injury by (-0.1167). This finding is not consistent with the finding of other studies (Farmer et al. 1997; Angel and Hickman 2009). In their studies, an increase in weight would decrease the injuries.

Front overhang

By increasing the front overhang of PVs by one unit (one foot) will reduce the predicted probability of evident injury by (-0.0704) and the odds of sustaining a level 3 injury by 29%. Since this independent predictor was not included in the reviewed literature, thus it was not possible to compare the finding of this study with the others.

Rear overhang

From the data, rear-overhang has opposing impacts on the odds of suffering level 3 and level 5. Increasing the rear overhang increases the odds of level 3 by 55% while has decreasing impact on sustaining a level 5 injury (61% reduction). The marginal effect of increasing the rear overhang by one foot has opposite effect on the probability of level 3 and level 5 (-0.0704 and +0.0014). Since this independent predictor was not included in the reviewed literature, thus it was not possible to compare the finding of this study with the others.

Width

PV-overall width (width) has no statistical significance on the probability of sustaining a level 5 injury, but it does have 100% increase in the odds of suffering a level 3 injury. By increasing the width by one foot the predicted probability of level 3 increases by (+0.1447).

Figure 5-2 provides a summary of the DOR for all the seventeen statistically significant variables. As shown in the figure, the odds of sustaining a level 5 decrease with increase on the size of the rear overhang. Moreover, the head-on crashes (compared to sideswipe) have the highest DOR of level 5 injury.

Figure 5-3 represents the marginal impacts of different significant variables on the predicted probability of injury severity levels. As illustrated in the figure, compared to other crash types, turn has the smallest impact on the predicted probability of injury severity levels.

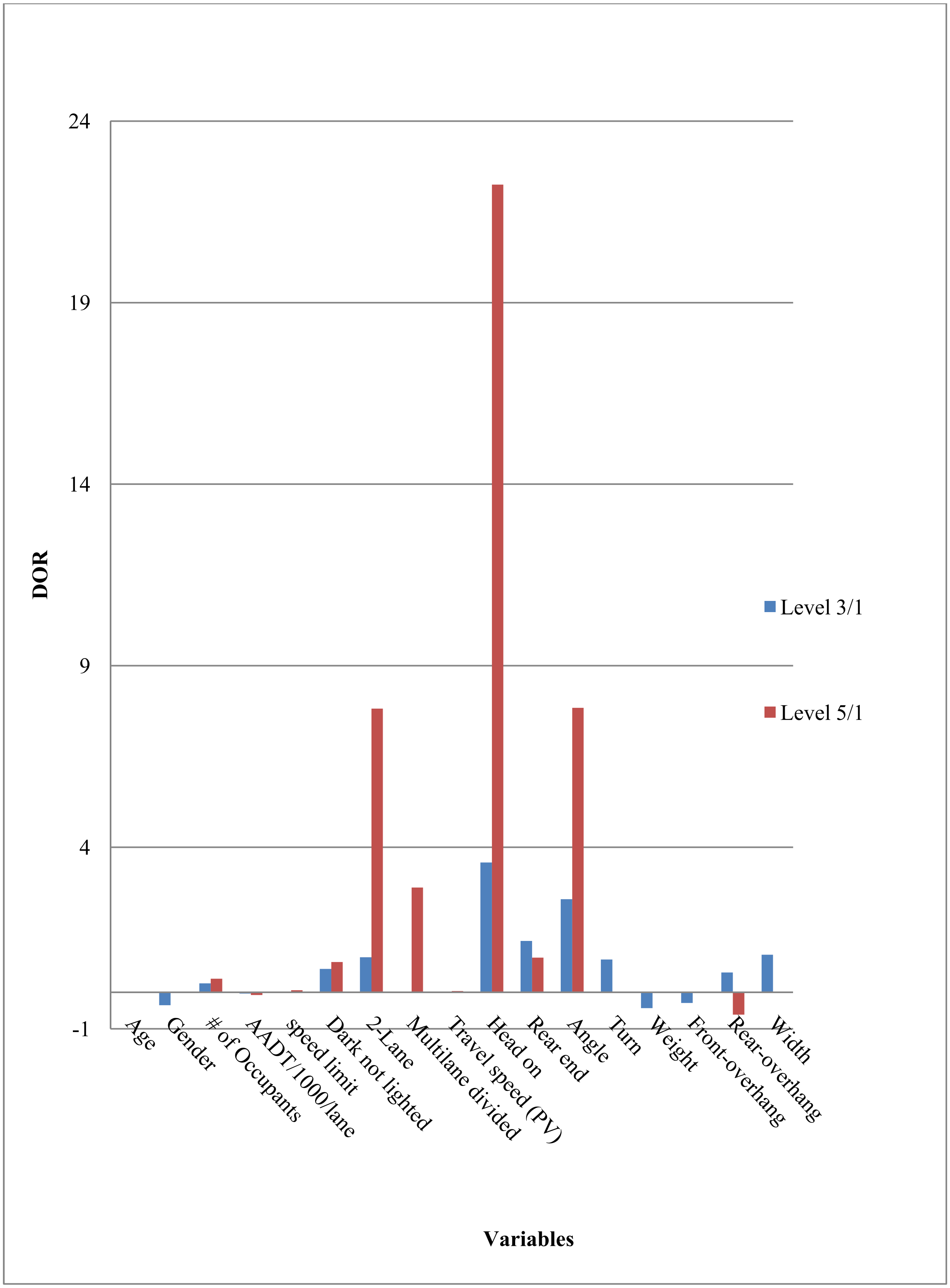

Figure 5-2 DOR for the general model

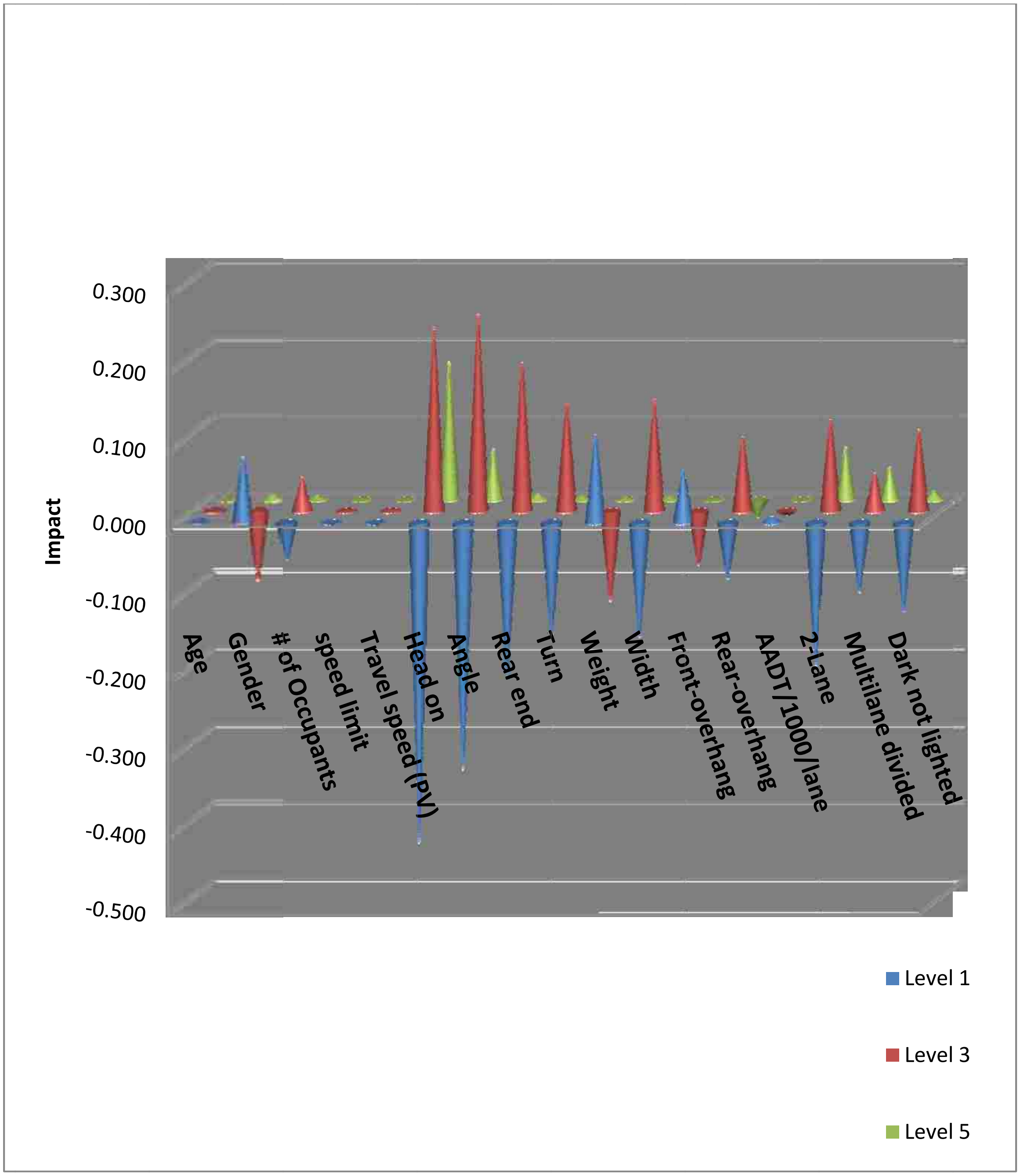

Figure 5-3 Marginal impact on the predicted probability for the general model

5.3.2 Gender Models

The data was divided into two datasets by the gender of the driver of the PVs. Two separate models were calibrated using the same approach as the general model.

1. Final models

Table 5-10 presents the final models conditioned on PV-drivers gender. Out of forty one variables, fifteen were statistically significant in male drivers model and twelve in female drivers model.

2. Scalar measures of model fit

Table 5-10 and 5-11 present the scalar measures of fit for both models. The same reasoning explained in chapter three and applied in achieving the final general model in the previous section were applied in evaluating and accepting these models.

3. Model assumption test

The results of SH tests for both models are presented in Table 5-12 and 5-13. The null hypotheses were accepted for both models. Other words, the ORs for Level 3/ Level 1 do not depend on Level 5/ Level 1.

Analysis of results

1. Predicted probability for each injury level

The results of using equation 3-31 and calculating the Pr(y=m|X) are tabulated in Table 5-15. As shown in the table, the predicted probability for level 1 has the highest and level 5 has the lowest value (as expected). Furthermore, having a male driver reduces the probability of sustaining a level 3 injury. Moreover, having a female driver reduces the probability of sustaining a level 5 injury. In the following sections, the impact of significant variables on the predicted probability for each outcome is analyzed.

Table 5-10 Final PV-male and female drivers models

VARIABLES	Male Drivers model				Female Drivers model			
	Level 3/Level 1		Level 5/ Level 1		Level 3/Level 1		Level 5/ Level 1	
	Odds ratio	P value	Odds ratio	P value	Odds ratio	P value	Odds ratio	P value
Drivers age	*NA*	*NA*	1.02	0.042				
# Of occupants	1.18	0.044	1.34	0.072	1.33	0.002	*NA*	*NA*
Dark-not lighted	1.64	0.018	2.30	0.027	1.56	0.088	*NA*	*NA*
2-lane					1.54	0.039	4.39	0.019
Multilane-divided	1.82	0.018	9.96	0.000				
Median-divided	0.32	0.000	0.06	0.000				
AADT/1000/lane					0.97	0.018	0.80	0.003
Speed limit	*NA*	*NA*	1.07	0.009	*NA*	*NA*	1.09	0.015
Head on	3.35	0.05	17.20	0.000	8.45	0.011	43.27	0.000
Rear-end	2.02	0.000	2.51	0.032	3.29	0.000	*NA*	*NA*
Angle	3.22	0.000	11.47	0.000	4.35	0.000	4.05	0.021
Turn	2.10	0.002	*NA*	*NA*	1.99	0.007	*NA*	*NA*
Travel speed	1.02	0.000	1.04	0.000	1.01	0.07	*NA*	*NA*
Weight	0.58	0.003	*NA*	*NA*	0.77	0.056	2.14	0.01
Front-overhang	0.67	0.091	*NA*	*NA*				
Rear-overhang	1.92	0.002	*NA*	*NA*	*NA*	*NA*	0.16	0.002
Width	1.91	0.087	*NA*	*NA*				
Constant	-5.681076		-7.292185		-2.334371		-4.922855	

Table 5-11 Scalar measures of fit for PV-male drivers model

Scalars	Initial Model	Final model
Likelihood Ratio (Degrees of Freedom, P=0.000)	260.42 (82)	194.70 (30)
Log Likelihood (Intercept Only)	-798.389	-798.389
Log Likelihood (Full Model)	-668.17826	-701.039
Akaike Information Criteria(AIC)	1.462	1.425
Pseudo ρ^2	0.163	0.122
Count R^2 (% of Correctly Predicted)	0.699	0.693
Number of Observations	1029	1029

Table 5-12 Scalar measures of fit for PV-female drivers model

Scalars	Initial Model	Final Model
Likelihood Ratio (Degrees of Freedom, P=0.000)	238.90(82)	173.240(24)
Log Likelihood (Intercept Only)	-609.097	-609.097
Log Likelihood (Full Model)	-489.647	-522.477
Akaike Information Criteria(AIC)	1.480	1.415
Pseudo ρ^2	.196	.142
Count R^2 (% of Correctly Predicted)	.68	.67
Number of Observations	775	775

Table 5-13 Model assumption test for PV-male drivers model

Injury level	lnL(full)	lnL(omit)	df	evidence
Evident	-77.767	-71.551	16	for H_o
Fatal/Incapacitating	-289.210	-277.268	16	for H_o

Table 5-14 Model assumption test for PV-female drivers model

Injury level	lnL(full)	lnL(omit)	df	evidence
Evident	-34.479	-31.575	13	for H_o
Fatal/Incapacitating	-227.174	-221.341	13	for H_o

Table 5-15 Predicted probability for the male and female drivers models

Pr (y=m|X)			
Gender models	Pr(y=1|X)	Pr(y=3|X)	Pr(y=5|X)
Male drivers	0.7139	0.2592	0.0269
Female drivers	0.6521	0.3397	0.0082

2. Odds Ratio and marginal impact

The OR for all the significant variables for male and female drivers models are listed in Table 5-10. These results are compared across the two gender models.

Table 5-16 presents marginal effects (elasticity) of the significant variables on the predicted probability of each injury severity level. In the following sections the OR and marginal effects of all the significant variables are analyzed.

Table 5-16 Marginal effects for the male and female drivers models

Variables	Discrete effect	Male drivers model			Female drivers model		
		Level 1	Level 3	Level 5	Level 1	Level 3	Level 5
Drivers age		0.0008	-0.0013	0.0005			
# of occupants		-0.0362	0.0297	0.0065	-0.0640	0.0625	0.0015
Dark-not lighted	0->1	-0.1158	0.0930	0.0227	-0.1025	0.1051	-0.0026
Speed limit		-0.0042	0.0025	0.0017	-0.0041	0.0034	0.0007
2-Lane	0->1				-0.1075	0.0930	0.0145
Multilane-divided	0->1	-0.1970	0.0804	0.1166			
Median divided	0->1	0.3029	-0.1874	-0.1155			
AADT/1000/lane					0.0089	-0.0072	-0.0018
Head on crash	0->1	-0.3626	0.1686	0.1940	-0.4866	0.4051	0.0815
Angle crash	0->1	-0.3163	0.2002	0.1161	-0.3507	0.3433	0.0073
Rear-end crash	0->1	-0.1599	0.1374	0.0226	-0.2796	0.2855	-0.0059
Turn	0->1	-0.1582	0.1545	0.0037	-0.1631	0.1617	0.0014
Travel speed		-0.0044	0.0035	0.0009	-0.0023	0.0022	0.0001
Weight		0.1062	-0.1041	-0.0020	0.0551	-0.0621	0.0070
Front-overhang		0.0772	-0.0746	-0.0026			
Rear-overhang		-0.1111	0.1289	-0.0178	-0.0155	0.0308	-0.0154
Width		-0.1242	0.1222	0.0020			

1. Driver/ Occupant characteristics

Drivers age

As age increases by one year the odds of level 5 increase by 2% for male drivers while for female drivers this variable has no statistical significance in either of injury severities. It is possible females are more cautious as they get older.

Number of occupants

As number of occupants increases the odds of level 3 increase by 33% for female drivers, whereas for male drivers this variable had increasing impact on the odds of injury severity (18% and 34%) level 3 and level 5 correspondingly. This may be an indication that female drivers are more inclined toward interacting with others and being focused on the driving task than male drivers.

2. Crash Environment at crash location

AADT/lane

The model indicated that, AADT/lane has decreasing impact on the odds of level 3 and level 5 by (3% and 20%) for female drivers. This variable did not have any statistically significant impact on odds of either of injury levels for male drivers.

Speed limit

From the data, increase in posted speed limit is more likely to result increase in the odds of a fatal/incapacitating injury by 7% and 9% (OR: 1.07 and 1.09) for male and female drivers, while it does not have any statistically significant impact on the evident injury.

Dark not lighted

According to the data, dark roads with no lighting are more likely to increase the OR of level 3 by 56% and a positive impact on the predicted probability by (+.105) for the female, whereas this significant variable has increasing impact on the odds of both levels for male. Perhaps, female driver are more aware of their surroundings than male drivers when driving on dark roads.

2-lane road

According to the results, crashes on 2-lane facilities with female drivers are more severe with 55%-350% greater odds of level 3 and 5 respectively. Whereas, with male drivers, crashes on these facilities were not statistically significant. Perhaps, male drivers are more inclined toward navigating vehicles in narrower roads than female drivers.

Multilane divided

Crashes on multilane-divided roadways had no statistical significance on either level for female drivers; however, this variable was statistically significant for both levels for male drivers. There may be behavioral differences between male and female drivers when driving on these types of facilities.

Median divided

The results of the male drivers model indicated that, roadways with median dividers reduce the OR of level 3 and level 5 by 68% and 94% correspondingly. However, this variable has no statistical significance in the female drivers model. There may be different perceptions of danger due to presence or not presence of median dividers between male and female drivers. The impact of the median on the injury severity based on the gender is a different finding from that of Abdel-Aty and Keller (2005) and Khorashadi et al. (2005).

3. Crash Characteristics

Travel speed

The data indicated that, increasing the PV-travel speed by one mile/hour increase the odds of suffering an evident injury by 1% for female drivers, however, the data for male drivers model showed that, the odds were increasing across the injury severity levels. Perhaps, female drivers are not inclined toward speeding as much as male drivers are.

Head on

Head-on crashes have the highest values of OR for injury severity regardless of the gender. It is worth mentioning that the magnitude of the OR values are much higher in both levels for female versus male drivers. As an example, having a female driver increases the OR of level 5 injury by 4200% versus a male driver with 1600%. Perhaps, male-drivers navigational skills are different from female drivers.

Rear end

The results of the models indicated that, the rear-end crashes increase the odds of suffering level 3 and level 5 by 100% and 150% for male drivers, whereas, the odds of sustaining level 3 increase by 230% for female drivers.

Angle

Angle crashes increase the odds of evident injury by 335% and level 5 by 300% for female drivers. On the other hand, the data for male drivers model indicated that, angle crashes increase the odds of level 5 by 1050% and level 3 by 220% respectively. Perhaps, female drivers are more skilled in navigating their vehicles better than their counterparts.

Turn

Turn crashes are more likely to result in evident injury but do not have any statistically significant effect on the probability of level 5 (both models). The magnitude of the ORs in both models conveys that being involved in a turn crash and having a male or female driver does not impact the OR of sustaining a level 3 injury.

4. Vehicle characteristics

Weight

PV-curb weight had opposing impacts on the OR of injury severity for the different genders. By increasing the weight the OR of evident injury reduce by 23% and the OR of level 5 increase by 115% for female drivers. Whereas, for male drivers it reduces the OR of evident injury by 42% and it is not statistically significant for level 5. Perhaps, as the vehicles get heavier controlling it will be different for female than male drivers.

Front-overhang

From the presented data, PVs front-overhang had no statistical significance on either of injury severity levels for female drivers; however this explanatory variable had statistical significance on the OR of level 3 for male drivers. Other words, increasing the front overhang by one foot reduces the likelihood of level 3 for male drivers about 33%.

Rear-overhang

From the data, an increase in rear-overhang by one foot reduces the odds of sustaining a level 5 by 84% for female drivers. Whereas, for male drivers it has no statistical significance at level 5. On the other hand, increasing the rear overhang increases the OR of level 3 by 92% for male drivers.

Width

PV-overall width (width) had no statistical significance on either of injury severity levels in female drivers model. Whereas, it had an increasing effect on the OR of level 3 (91%) in the male drivers model.

Figure 5-4 and Figure 5-5 provide a summary of the DOR for all the fifteen and twelve statistically significant variables for male and female drivers models respectively. As shown in the figures, driving at roadways with no lights pose higher risk of sustaining a level 5 for male drivers and not for female drivers.

Figure 5-6 and Figure 5-7 represent the marginal impacts of different significant variables on the predicted probability of injury severity levels. As illustrated in these figures, median reduces the predicted probability of level 5 by almost 11% in male drivers model, whereas, it does not exist in the female-drivers model.

In summary, the preceding sections illustrated the major differences between male and female drivers, given a crash occurred.

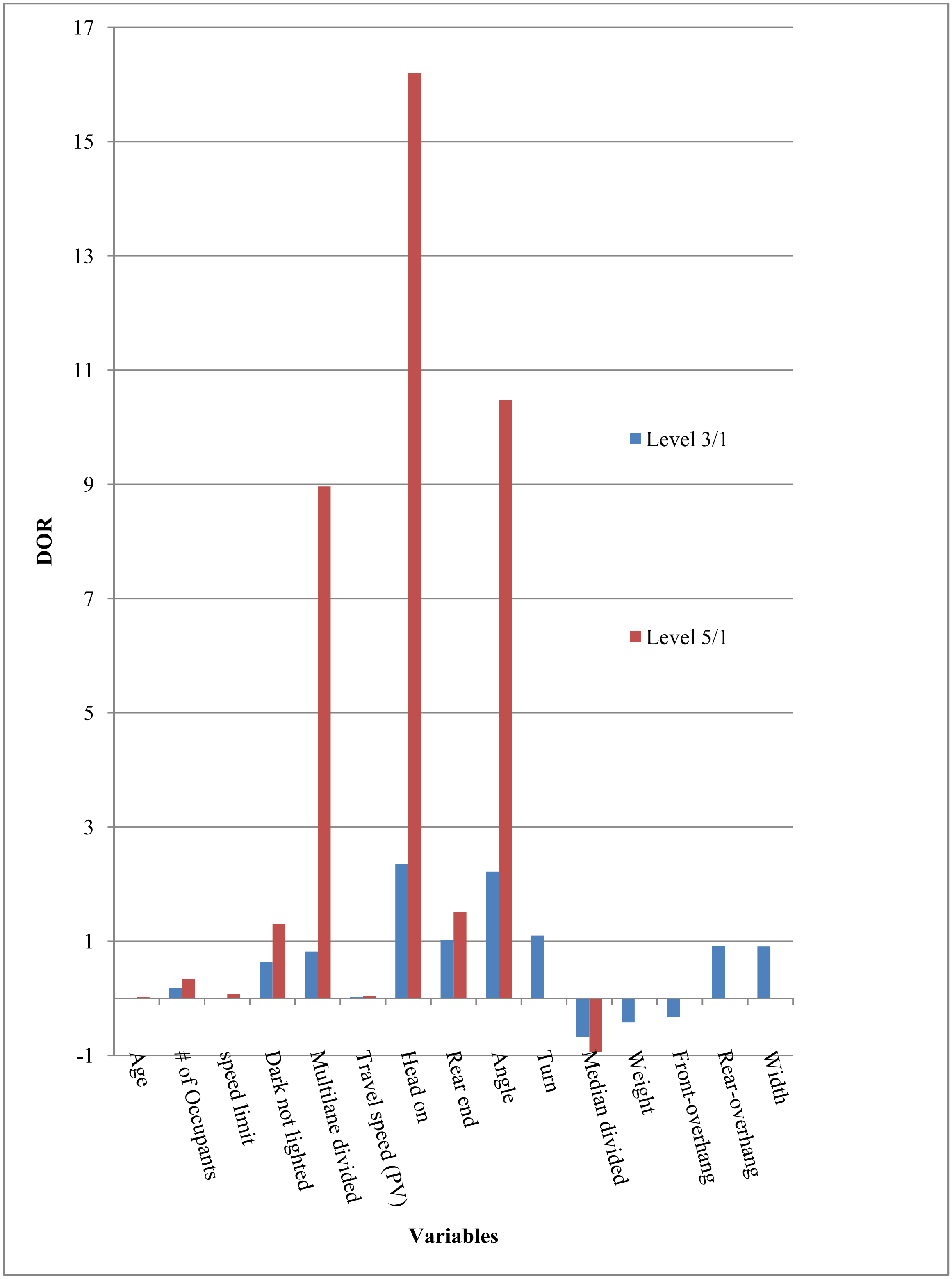

Figure 5-4 DOR for male drivers model

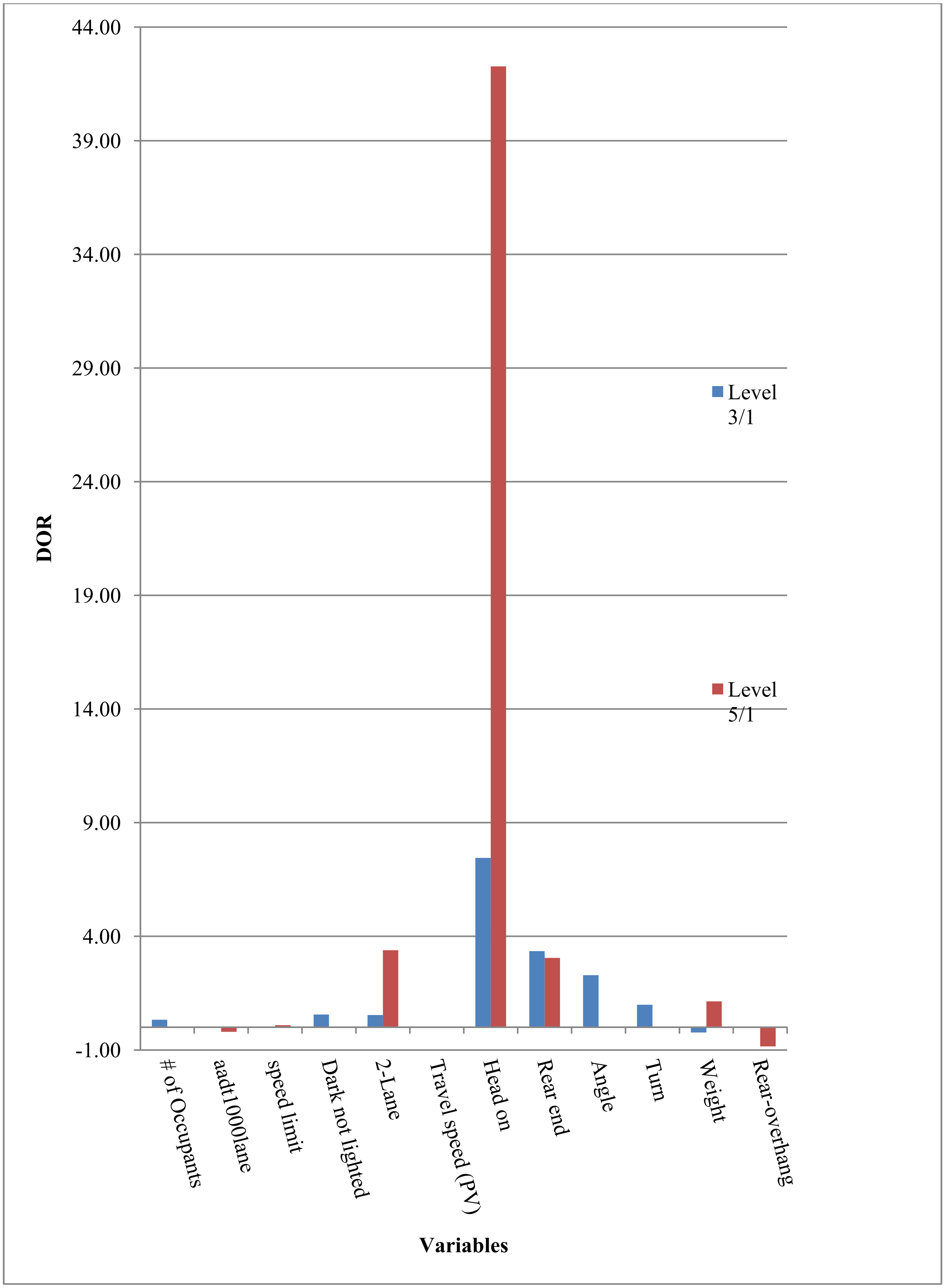

Figure 5-5 DOR for female drivers model

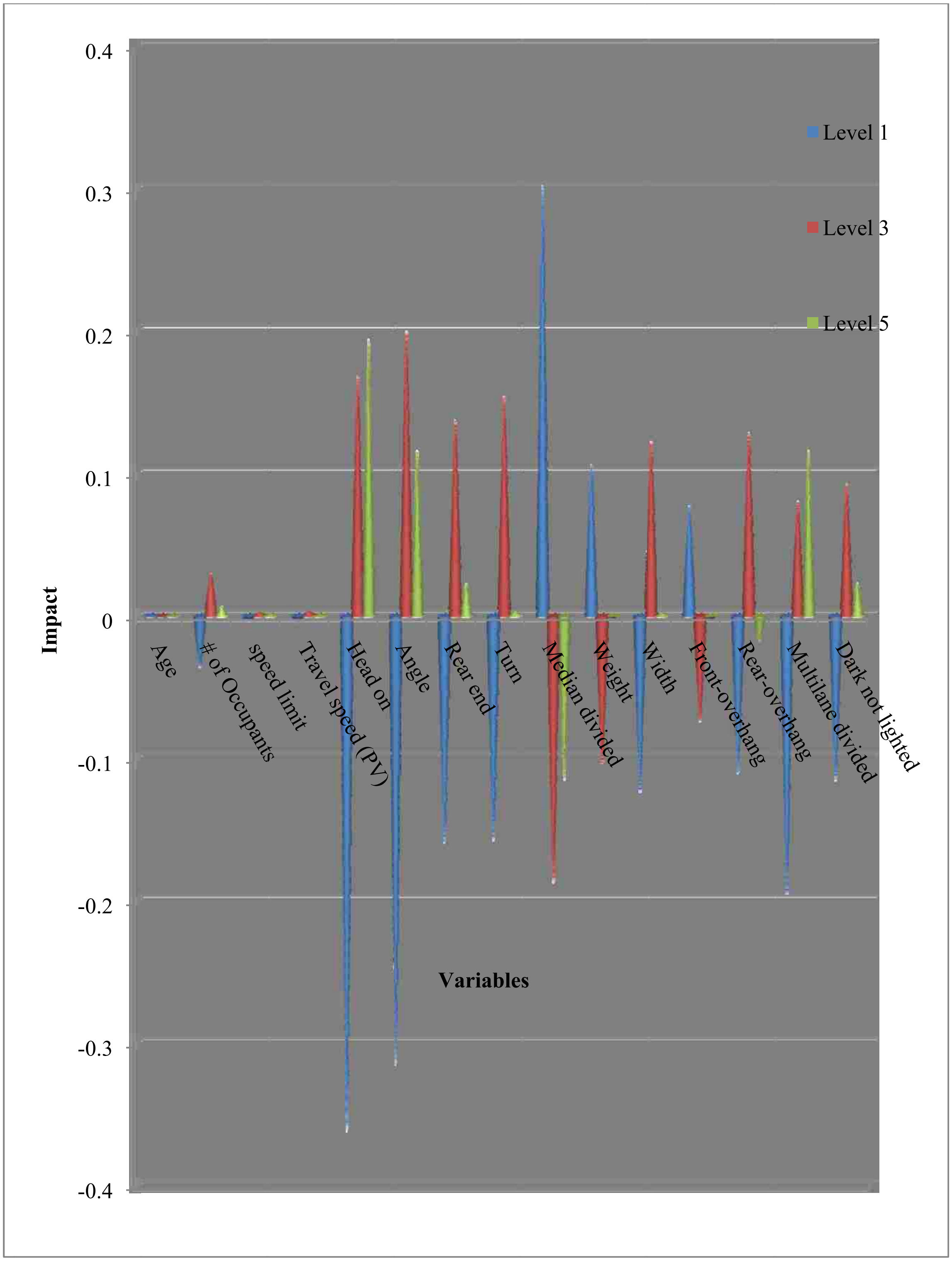

Figure 5-6 Marginal impact on the predicted probability for the male drivers model

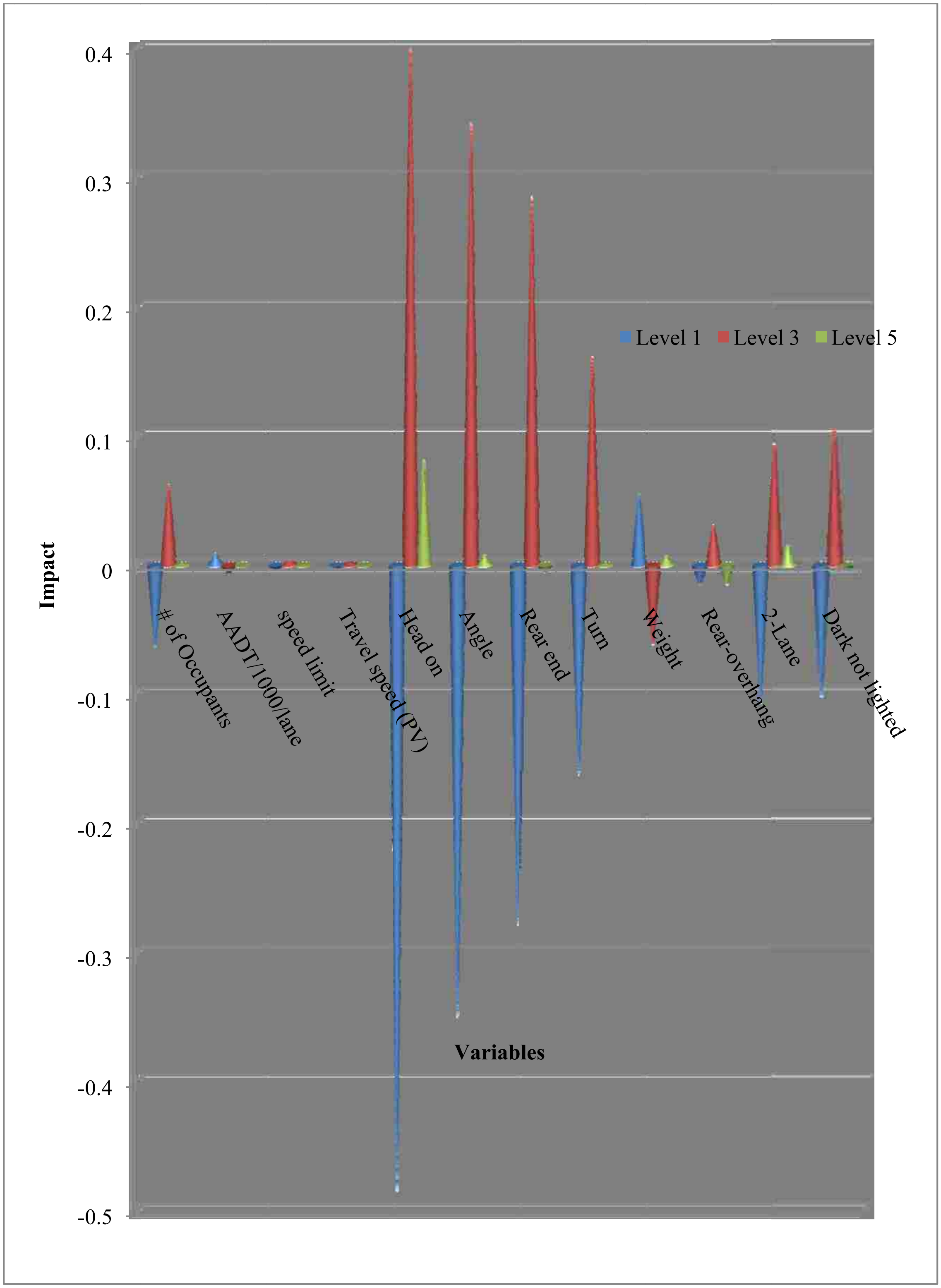

Figure 5-7 Marginal impact on the predicted probability for the female drivers model

5.3.3 Age Models

The data was divided into four groups based on the age of the drivers of the PVs. These four groups are:

- Drivers age <=25
- Drivers age from 26-45
- Drivers age from 46-65
- Drivers age from 66 and older

For each age group a separate multivariate multinomial logit model was calibrated by applying the same procedure mentioned in previous sections.

1. Final models

Table 5-17 presents all the final models conditioned on PV-drivers age. Out of forty one variables, thirteen variables were statistically significant for drivers age <= 25, ten for drivers age group 26-45, twelve for drivers age group 46-45, and seven for drivers age 66-older.

2. Scalar measures of model fit

Table 5-18, 5-19, 5-20, and 5-21 present the scalar measures of fit for the four models.

3. Model assumption test

The results of SH tests for four models are presented in Table 5-22, 5-23, 5-24, and 5-25.

Table 5-17 Final models for all four drivers age-groups

VARIABLES	Drivers' age <=25		Drivers' age 26-45		Drivers' age 46-65		Drivers' age 66 & older	
	Level 3/1	Level 5/1	Level 3/1	Level 5/1	Level 3/1	Level 5/1	Level 3/1	Level 5/1
	OR	OR	OR	OR	OR	OR	OR	OR
Driver's gender			0.66	NA	0.45	NA		
# Of occupants	1.21	1.79	1.26	1.48	1.33	NA		
Dark-not lighted	NA	7.46			2.22	NA	4.06	NA
2-lane			1.74	3.24	4.6	25.47	1.92	NA
Multilane-divided	2.34	5.26			2.41	12.06		
Multilane-undivided					3.4	NA		
Median-divided	0.39	0.1						
AADT/1000/lane			0.97	0.76				
Speed limit					1.05	1.09		
Head on	NA	722.61	3.69	20.48	9.45	36.52		
Rear-end	NA	7.62	2.75	NA	3.04	3.51		
Angle	3.57	44.32	3.41	3.91	4.86	13.28		
Turn	2.18	NA			2.07	NA	2.04	NA
Travel speed	1.02	NA	1.01	1.04	1.02	1.06		
Overall height	NA	0.85						
Weight	NA	6.03	0.57	NA			0.27	0.05
Front-overhang	NA	0.01					0.26	NA
Rear-overhang	NA	0.16	2.47	NA			NA	0.04
Width							25.71	606.71

Note: Complete table with the Logit coefficients and P values are in Appendix C.

Table 5-18 Scalar measures of fit for initial & final models for drivers age<=25

Scalars	Initial model	Final model
Likelihood Ratio (Degrees of Freedom, P=0.000)	170.856(82)	100.370 (26)
Log Likelihood (Intercept Only)	-357.352	-357.352
Log Likelihood (Full Model)	-271.924	-307.167
Akaike Information Criteria (AIC)	1.554	1.464
Pseudo ρ^2	0.239	0.140
Count R^2 (% of Correctly Predicted)	0.707	0.653
Number of Observations	458	458

Table 5-19 Scalar measures of fit for initial & final models for drivers age 26-45

Scalars	Initial model	Final model
Likelihood Ratio (Degrees of Freedom, P=0.000)	183.013(38)	163.789(20)
Log Likelihood (Intercept Only)	-523.103	-532.103
Log Likelihood (Full Model)	-440.597	-450.209
Akaike Information Criteria (AIC)	1.379	1.355
Pseudo ρ^2	.172	.152
Count R^2 (% of Correctly Predicted)	.70	.69
Number of Observations	697	697

Table 5-20 Scalar measures of fit for initial & final models for drivers age 46-65

Scalars	Initial model	Final model
Likelihood Ratio (Degrees of Freedom, P=0.000)	181.774 (82)	121.725 (24)
Log Likelihood (Intercept Only)	-372.455	-372.455
Log Likelihood (Full Model)	-281.569	-311.593
Akaike Information Criteria (AIC)	1.593	1.471
Pseudo ρ^2	0.244	0.163
Count R^2 (% of Correctly Predicted)	0.749	0.693
Number of Observations	459	459

Table 5-21 Scalar measures of fit for initial & final model for drivers age 66 and older

Scalars	Initial model	Final model
Likelihood Ratio (Degrees of Freedom, P=0.000)	84.488 (46)	53.075 (14)
Log Likelihood (Intercept Only)	-147.385	-147.385
Log Likelihood (Full Model)	-105.141	-120.847
Akaike Information Criteria (AIC)	1.612	1.440
Pseudo ρ^2	0.287	.180
Count R^2 (% of Correctly Predicted)	0.758	.70
Number of Observations	190	190

The null hypotheses were accepted for all the four models. Other words, the ORs for level 3/ level 1 do not depend on level 5/ level 1.

Table 5-22 Model assumption test for drivers age<=25

Injury level	lnL(full)	lnL(omit)	df	evidence
Evident	24.836	-13.978	14	for H_o
Fatal/Incapacitating	-122.223	-115.221	14	for H_o

Table 5-23 Model assumption test for drivers age 26-45

Injury level	lnL(full)	lnL(omit)	df	evidence
Evident	-39.430	-32.206	11	for Ho
Fatal/Incapacitating	-207.063	-203.776	11	for H_o

Table 5-24 Model assumption test for drivers age 46-65

Injury level	lnL(full)	lnL(omit)	df	evidence
Evident	-25.307	-17.767	13	for H_o
Fatal/Incapacitating	-129.787	-120.855	13	for H_o

Table 5-25 Model assumption test for drivers age 66 and older

Injury level	lnL(full)	lnL(omit)	df	evidence
Evident	-3.612	-0.000	8	for H_o
Fatal/Incapacitating	-43.351	-40.969	8	for H_o

Analysis of results

1. Predicted probability for each injury level

The results of calculating the Pr(y=m|X) are tabulated in Table 5-26. As depicted in Figure 5-8, drivers age 45-65 has the highest predicted probability of level 1. Moreover, the smallest predicted probability of level 5 belongs to drivers age group 26-45. Moreover, drivers age<=25 and the largest predicted probability of evident injury and the second highest probability of level 5 injury. In the following sections, the impacts of significant variables on the predicated probability for each injury level are analyzed.

Table 5-26 Predicted probabilities for all drivers age-groups

Pr (y=m|X)			
Drivers age	Pr(y=1|X)	Pr(y=3|X)	Pr(y=5|X)
Age <=25	0.657	0.337	0.006
Age 26-45	0.709	0.291	0.000
Age 45-65	0.726	0.269	0.005
Age 66-older	0.699	0.291	0.011

2. Odds Ratio and marginal impact

The OR for all the significant variables for four drivers age-groups are listed in Table 5-17. These results were compared to each other across the four models. Table 5-27 presents marginal effects of the significant variables on the predicted probability of each injury severity level for all the drivers age-groups. In the following sections the OR and marginal effects of all the significant variables are analyzed.

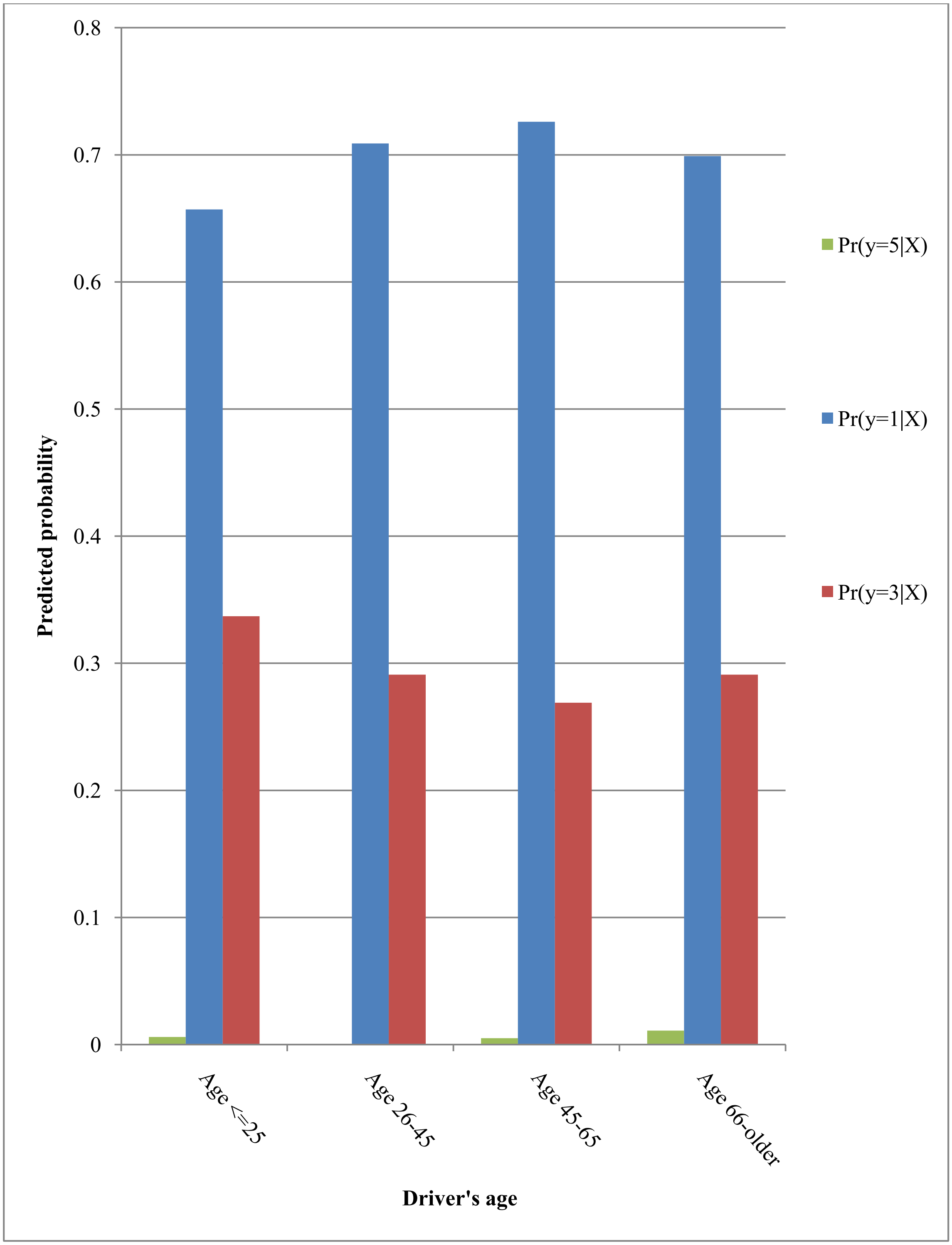

Figure 5-8 Predicted probabilities for drivers age-groups

Table 5-27 Marginal effects for the drivers age-groups

Variables	Discrete effect	Drivers age<=25			Drivers age 26-45			Drivers age 46-65			Drivers age 66 and older		
		Level 1	Level 3	Level 5	Level 1	Level 3	Level 5	Level 1	Level 3	Level 5	Level 1	Level 3	Level 5
Drivers gender	0->1				0.086	-0.086	0.000	0.158	-0.159	0.001			
# of occupants		0.041	0.003	-0.044	-0.048	0.048	0.000	-0.056	0.055	0.002			
Dark-not lighted	0->1	0.070	0.081	0.024				-0.177	0.174	0.003	-0.324	0.329	-0.005
2-Lane	0->1				-0.120	0.120	0.000	-0.341	0.313	0.028	-0.145	0.136	0.009
Multilane-divided	0->1	0.138	0.194	0.012				-0.206	0.181	0.026			
Multilane Undivided	0->1							-0.261	0.291	-0.030			
Median divided	0->1	-0.209	-0.017	0.226									
AADT/1000/lane					0.007	-0.007	0.000						
Speed limit								-0.009	0.009	0.000			
Head on	0->1	-0.142	0.689	-0.547	-0.310	0.310	0.000	-0.516	0.470	0.047			
Rear-end	0->1	0.093	0.022	-0.115	-0.221	0.232	-0.010	-0.248	0.242	0.006			
Angle	0->1	0.255	0.080	-0.335	-0.285	0.285	0.000	-0.370	0.348	0.022			
Turn	0->1	0.177	0.013	-0.190				-0.157	0.157	0.000	-0.152	0.160	-0.007
Travel speed		0.005	0.000	-0.006	-0.003	0.003	0.000	-0.003	0.003	0.000			
Overall height		-0.007	-0.001	0.008									
Weight		0.048	0.010	-0.058	0.117	-0.117	0.000				0.287	-0.257	-0.030
Front-overhang		-0.142	-0.024	0.165							0.293	-0.275	-0.018
Rear-overhang		0.000	-0.011	0.011	-0.187	0.187	0.000				0.072	-0.037	-0.035
Width											-0.709	0.648	0.061

1. Driver/ Occupant characteristics

Drivers gender

The results of drivers age 26-45 and 46-65 models indicated that, having a male driver reduces the odds of sustaining a level 3 injury by 34% and 55% respectively. However, drivers gender does not affect the likelihood of any injury severity for the younger and older drivers populations. Perhaps, younger drivers lack of driving skills makes the gender difference unimportant. Moreover, in the older population physiological degradation due to aging process makes the gender an irrelevant issue.

Number of occupants

As the number of occupants increases, there is a positive impact on the injury severity for all the age groups except the older population and female drivers. Among the influenced groups, the younger drivers had the largest OR of sustaining a level 5 injury severity. The older driver population seemed to be not influenced by this significant variable. Perhaps, this group is the most experienced and less influenced by the surroundings. This finding was different from that of Chang and Mannering (1999), and Farmer et al. (1997).

2. Crash Environment at crash location.

AADT/lane

The data showed that, AADT/lane has decreasing impact on the odds of level 3 and level 5 for drivers age 26-45. This variable did not have any statistically significant impact on odds of either of injury levels of other age groups.

Speed limit

From the data, increase in posted speed limit is more likely to result an increase in the odds of evident and fatal/incapacitating injury by 5% and 9% for 46-65 age group. As shown in the general model, increase in posted speed limit was more likely to result an increase in the odds of level 5 by 6% (OR: 106) and it did not have any statistically significant impact on the evident injury.

Dark not lighted

Dark roads with no lighting are more likely to increase the OR of level 5 by 650% for the young-age group and OR of level 3 by 300% for the older group. Perhaps, lack of driving skills combined with youth behavior is the detriments of such high magnitude. This variable poses extra risk of sustaining a fatal injury for drivers age <=25 and the lowest risk of injuries for the female drivers. These finding were different from that of Dunken et al. (1998), Khattak et al. (2002), and Abdel-Aty (2003).

2-lane road

From the data, 2-lane facilities are more likely to have an increasing impact on injury severity and the largest OR of level 5 (25.47) for the 46-65 age group. Perhaps, 2-lane roadways (comparing to freeways) are mostly located in rural area with no median-dividers and less police controls.

Multilane undivided

Crashes on multilane-undivided roadways are more likely to result in level 3 injury for 46-65 age group.

Multilane divided

Crashes on multilane-divided roadways are more likely to increase the OR values of both injury levels for drivers age<=25 and 46-65 groups.

Median divided

According to the data, this variable reduces the odds of level 3 by 61% and level 5 by 90% for the young-driver population and it does not have any statistical significance for other age groups. Perhaps, roads equipped with medians can compensate for lack of driving experience for this age group. This finding is consistent with that of Abdel-Aty and Keller (2005) and Khorashadi et al. (2005). However, it is different by exclusion of other age groups.

3. Crash Characteristics

Travel speed

Travel speed was not statistically significant for drivers age 66 and older. Perhaps, this age group drives conservatively.

Head on

Drivers age <=25 had the highest OR (722.61) among other groups (excluding 66 and older). This variable was not statistically significant for drivers age 66 and older. The magnitude of the OR for the young population is an alarming value. Head on crashes impose higher risk of a fatal injury for drivers age<=25 and female drivers. Perhaps, young drivers due to their lack of driving experience cannot maneuver out of a potential fatal crash. Although, these findings were consistent with the findings of O'Donnell and Connor (1996), Kockelman and Kweon (2002), Abdel-Aty and Keller (2005), Hill and

Boyle (2006), and Xie et al. (2009) but there were differences in the magnitude and the gender of the drivers.

Rear end

Drivers age group <=25 had the highest OR (7.62) among other groups (excluding 66 and older). This variable was not statistically significant for drivers age 66 and older.

Angle

The results indicated that, angle crashes increase the odds of evident and fatal/incapacitating injury for all the age groups except the older population. The magnitude of the OR for level 5 was the highest for the younger population (OR: 44.32). Other words, being in this age group and involved in an angle collision increases the OR of sustaining a level 5 injury by 4300%.

Turn

Turn crashes are more likely to increase the OR of level 3 for all the drivers-age groups except drivers age 26-45. Turn crashes are not statistically significant at level 5 injury.

4. Vehicle characteristics

Overall height

From the data, PV-overall height was statistically significant for the younger population. It had decreasing impact on the odds of suffering a level 5 injury by 15%. Perhaps, the higher the height of a vehicle would create a better sight distance for this age group.

Weight

Increasing the weight of vehicle by one unit (1000 lbs) decrease the OR for 26-45 and 66-older driver-age groups. However, increasing the weight would increase the OR by 500% for the younger drivers. Perhaps, navigating heavier cars are more cumbersome for the younger age group. These results were different from that of Wang and Kockelman (2005) and Angel and Hickman (2009).

Front overhang

From the presented data, PVs front-overhang had the most decreasing impact on the odds of suffering a level 5 injury (99%) for the younger driver group. Moreover, increasing the front overhang, reduce the OR of sustaining level 3 injury for older drivers group. Since vehicles with longer front overhangs have farther distance between point of impact and the occupant compartment which may contribute to the reduction of the OR.

Rear overhang

From the data, rear-overhang decreasing impact on the odds of suffering a level 5 injury (84%), but increasing the rear overhang does not have statistically significant impact on the probability of sustaining a level 3 injury. Perhaps, vehicles with longer rear overhangs create longer distance between point of impact at time of crash and the occupant compartment.

Width

PV-overall width (width) was statistically significant at both injury levels for the older driver group (OR: 25.71-606.76). This variable had the largest OR values compared to others for this age group. Perhaps, wider cars are heavier and larger in size and that would make the navigation of these types of vehicle harder.

In summary, Figure 5-9, 5-10, 5-11, and 5-12 provide a summary of the DOR for each of the four drivers-age groups. The bar representing the DOR value corresponding to head on crashes in Figure 5-9 (drivers age<=25) was truncated due to the magnitude of the DOR (722). Moreover, the bar representing the DOR value corresponding to width crashes in Figure 5-12 (drivers age 66 and older) was truncated due to the magnitude of the DOR (606).

Figure 5-13, 5-14, 5-15, and 5-16 represent the marginal impacts of different significant variables on the predicted probability of injury severity levels. As illustrated in these figures, increasing front overhang reduces the OR values of sustaining a level 3 injury for age groups <=25 and 66 and older.

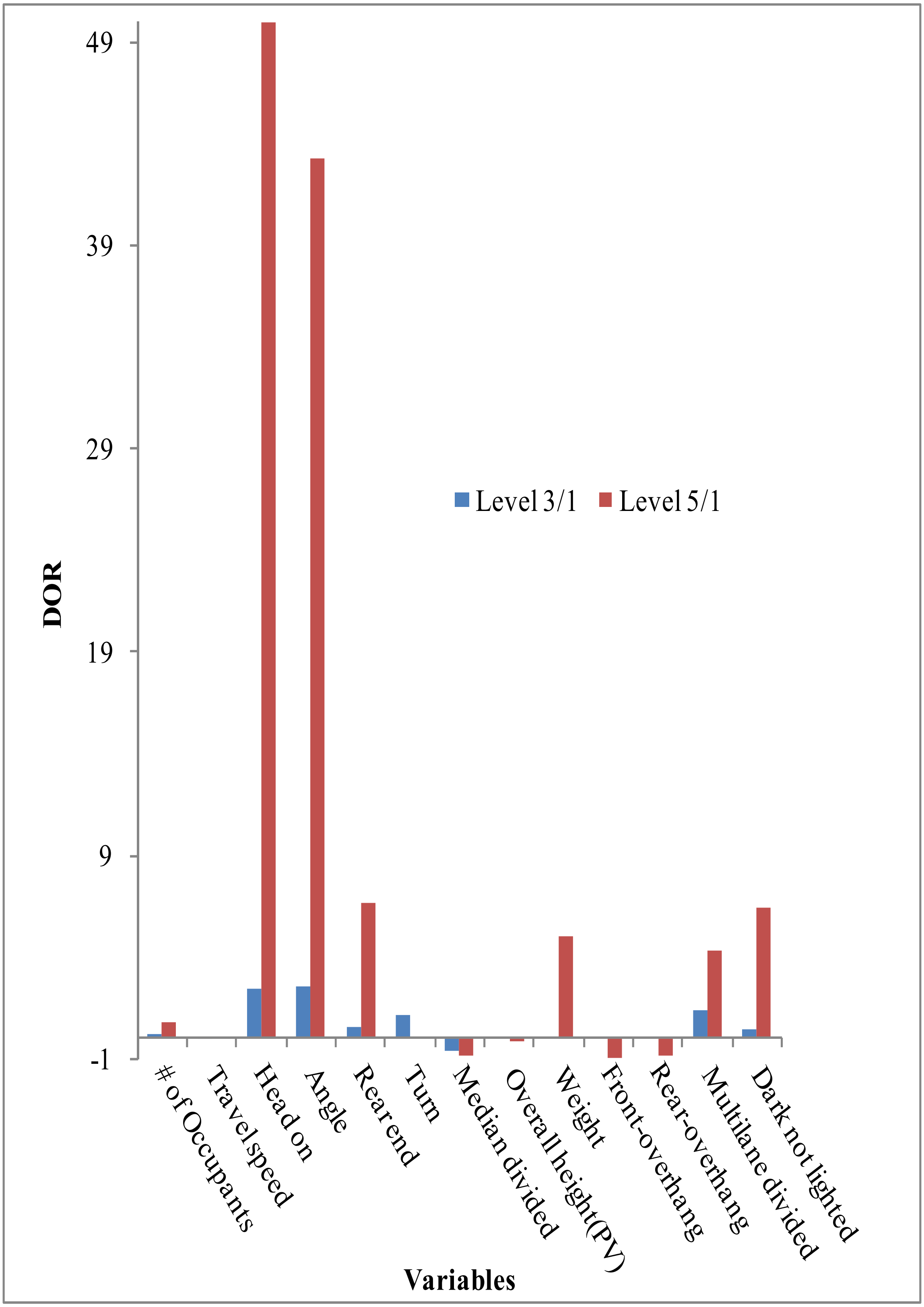

Figure 5-9 DOR for drivers age<=25 model

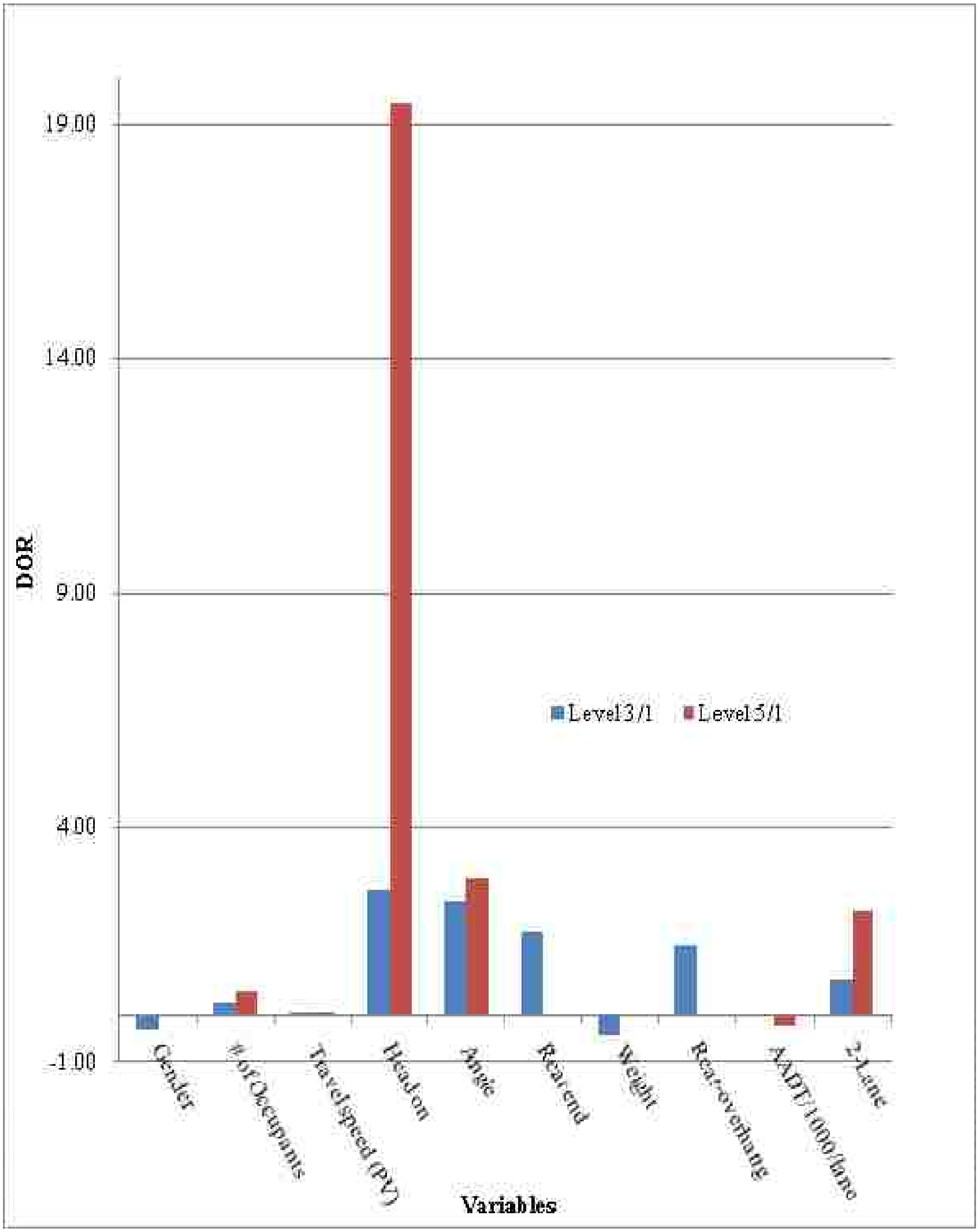

Figure 5-10 DOR for drivers age 26-45 model

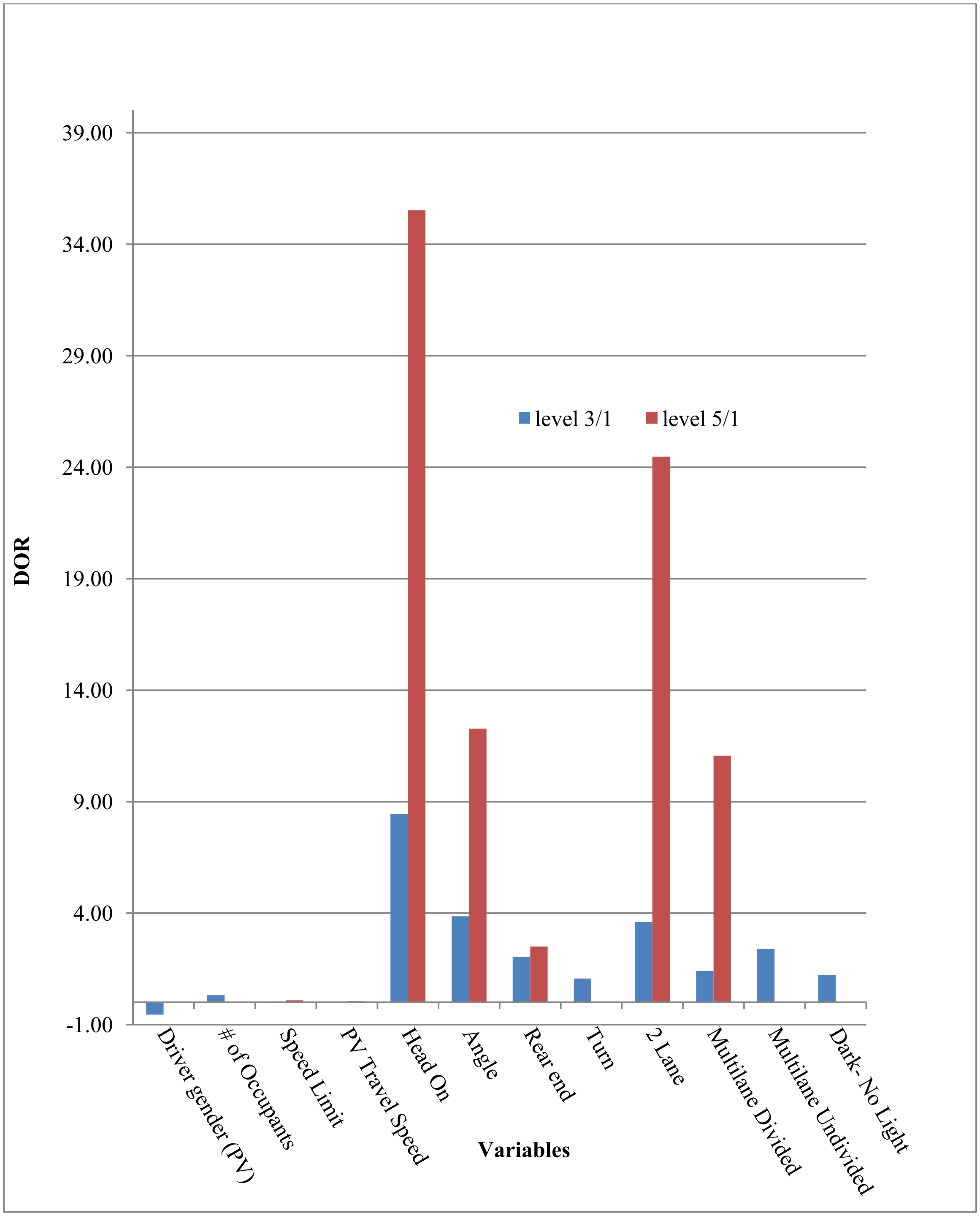

Figure 5-11 DOR for the drivers age 46-65 model

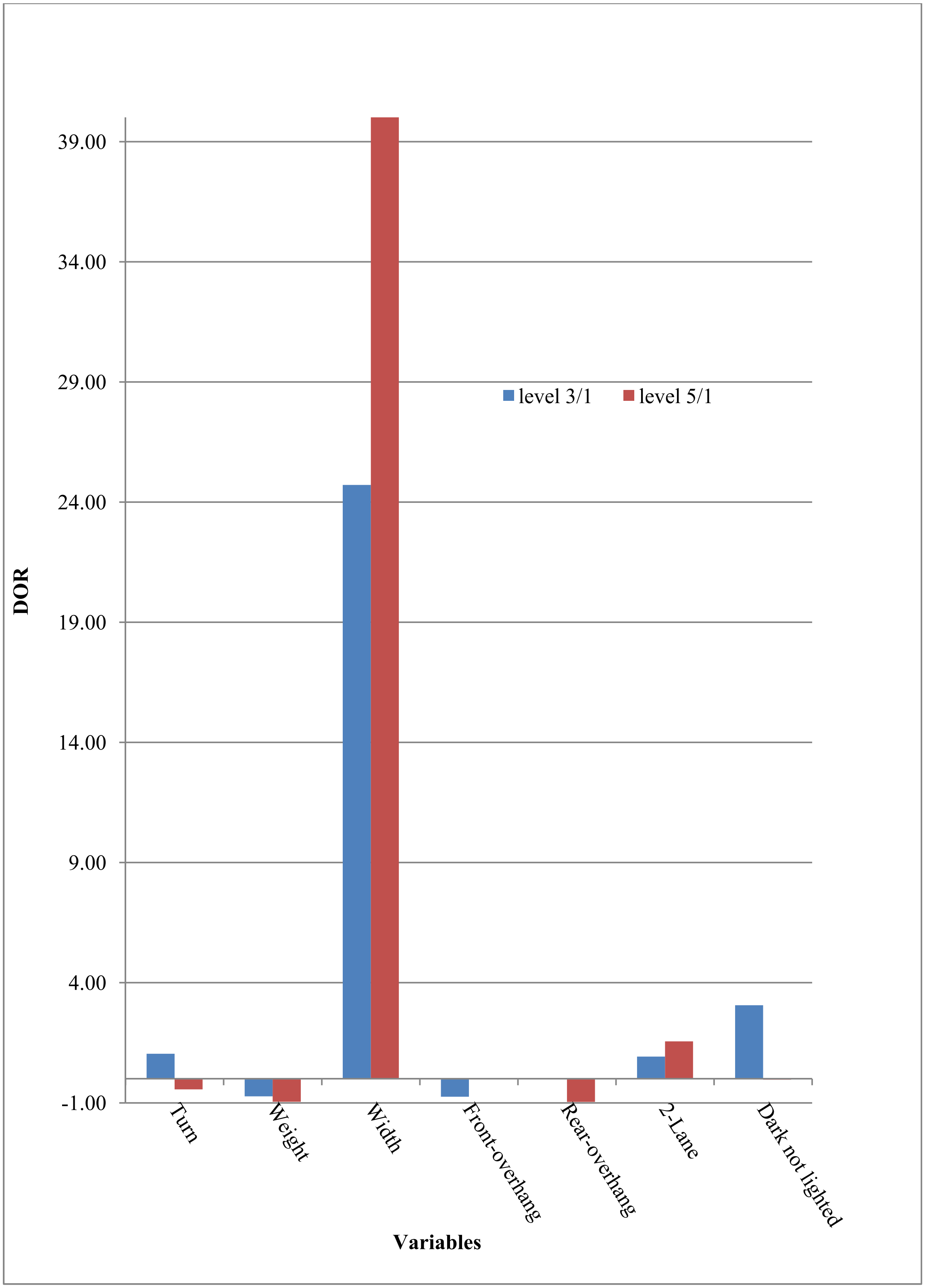

Figure 5-12 DOR for drivers age 66 and older model

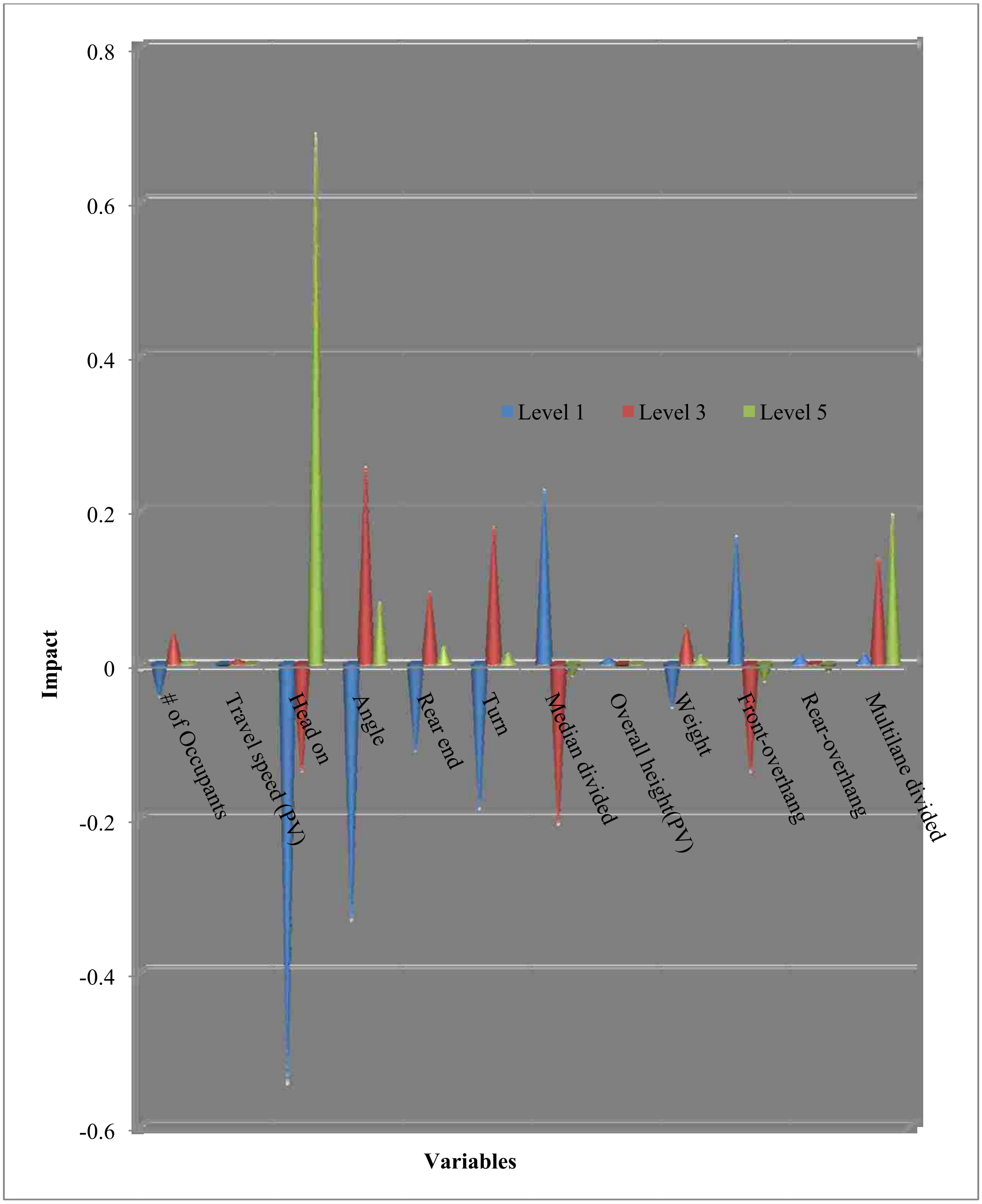

Figure 5-13 Marginal impact on the predicted probability for drivers age<=25

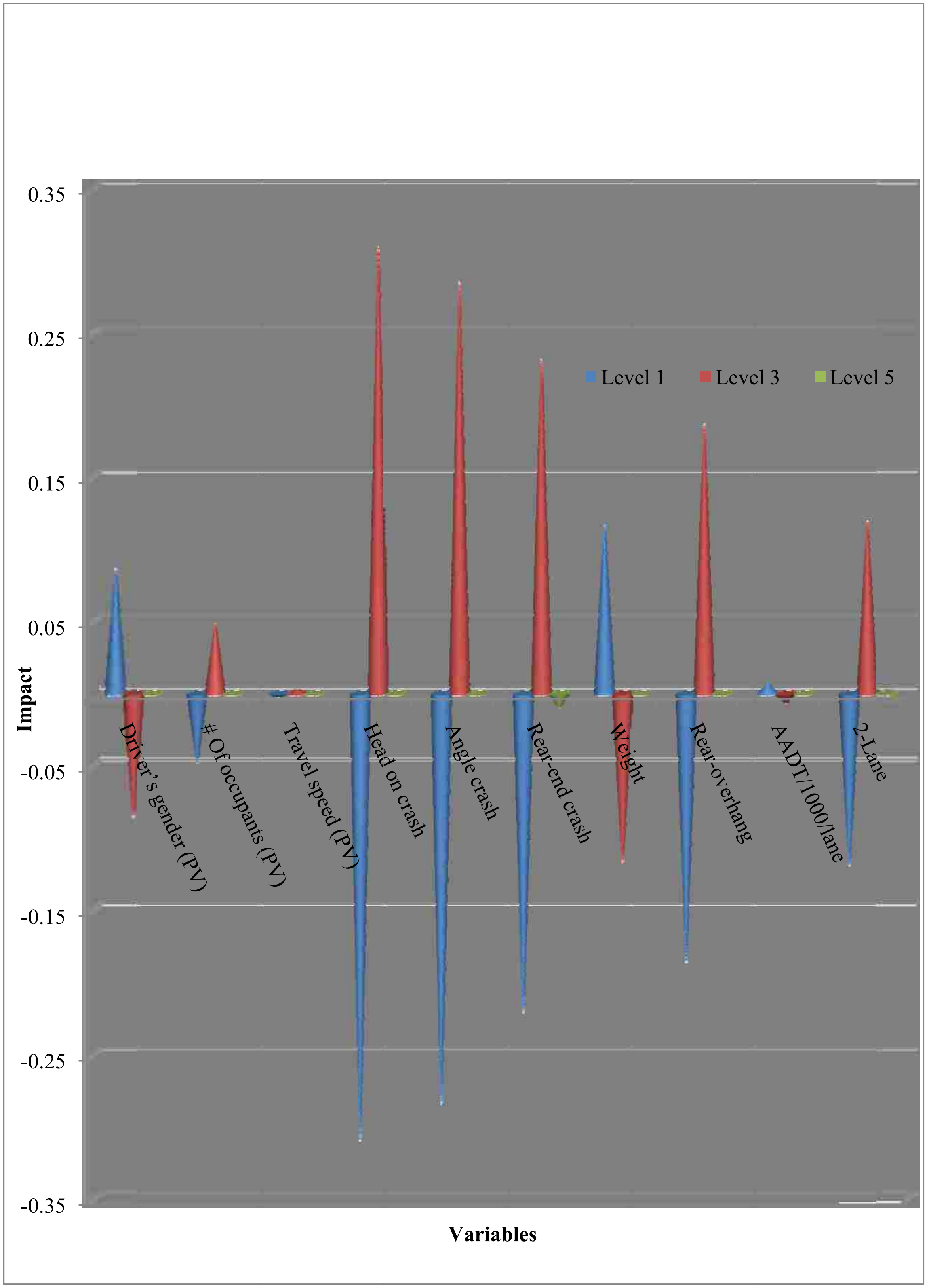

Figure 5-14 Marginal impact on the predicted probability for drivers age 26-45 model

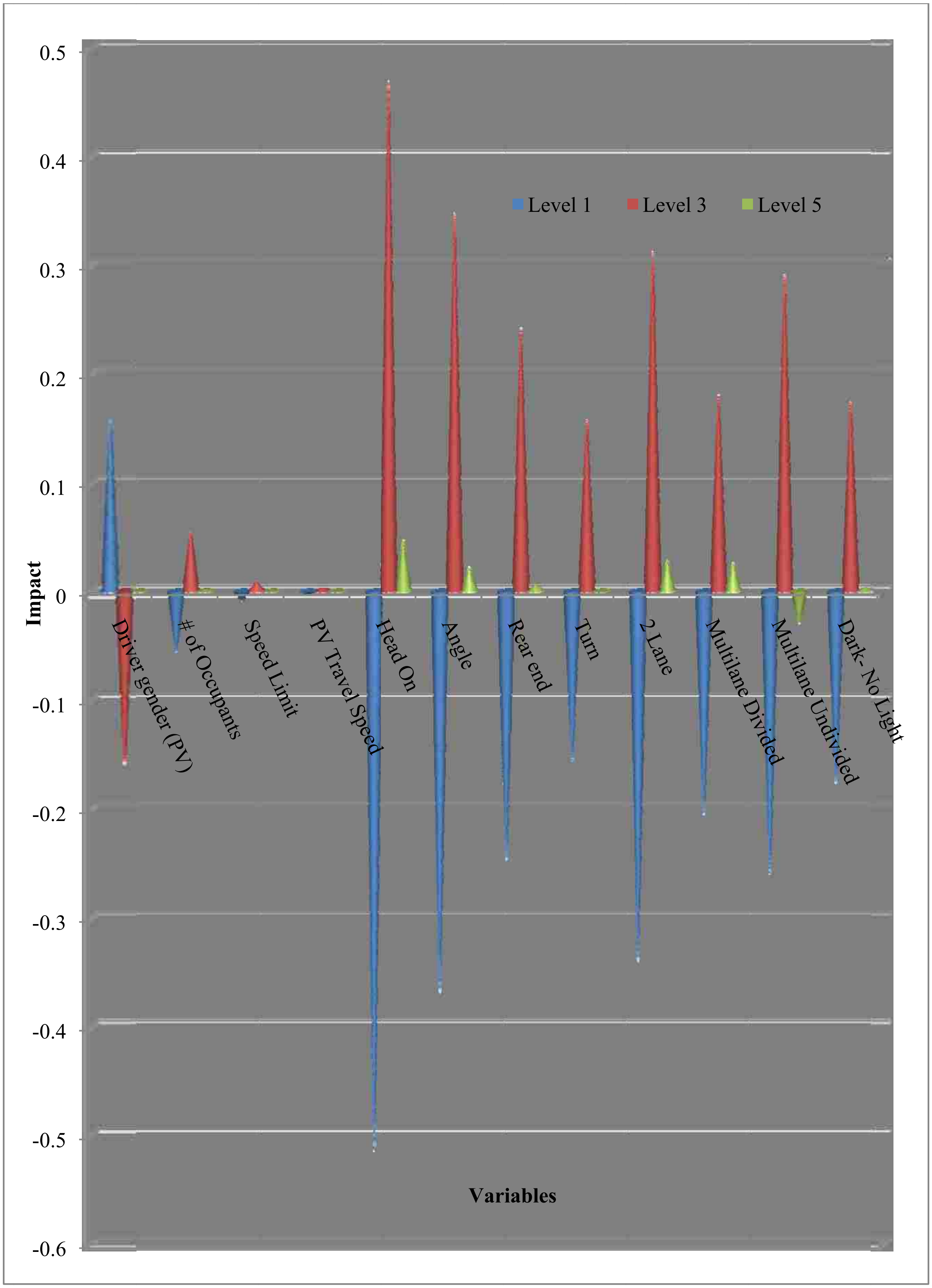

Figure 5-15 Marginal impact on the predicted probability for drivers age 46-65 model

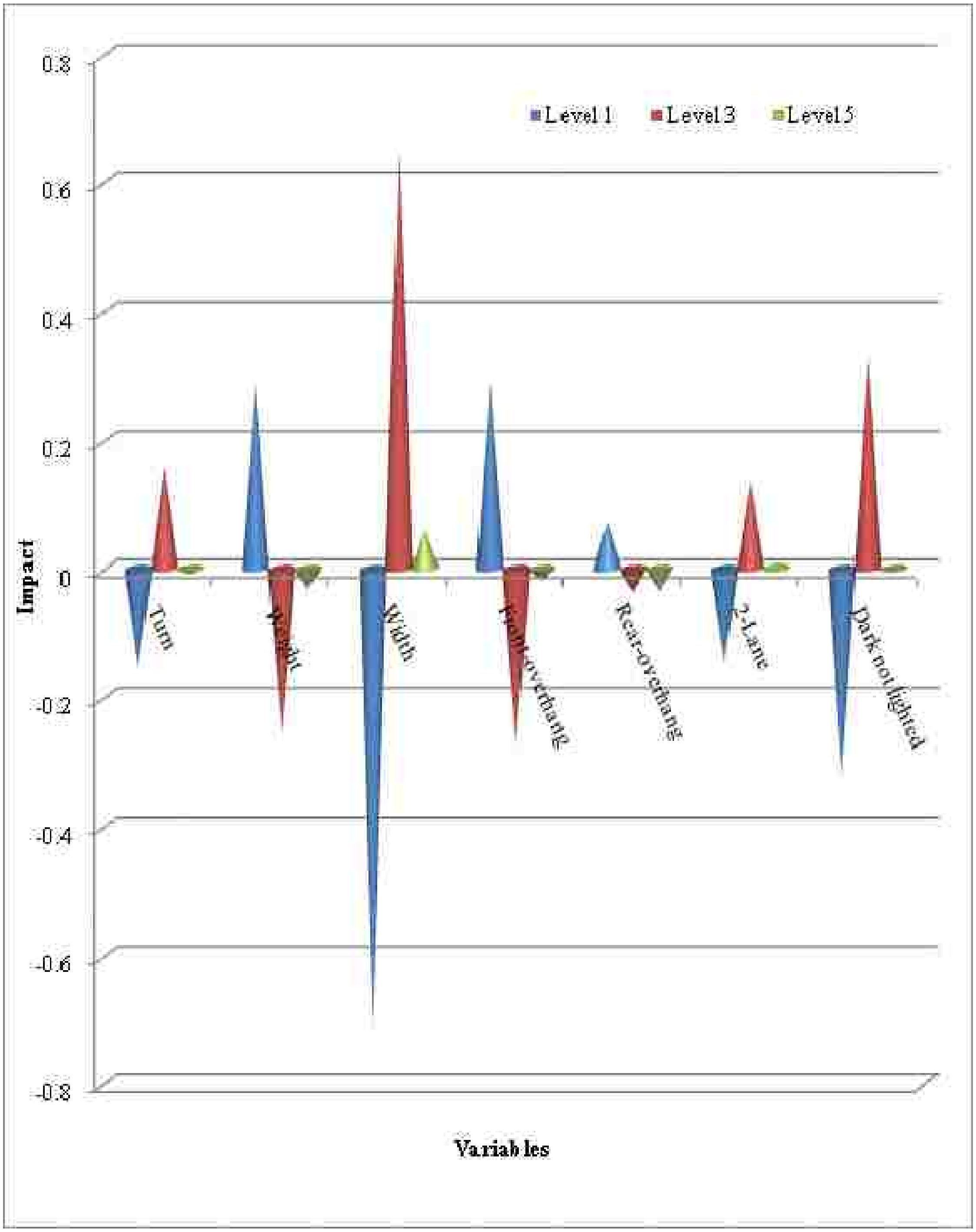

Figure 5-16 Marginal impact for drivers age 66 and older

5.4 Discussion

In total, twenty statistically significant explanatory variables were identified. Among the seven calibrated models, general model had the highest number of the independent predictors (17) and drivers age 66 and older had the lowest number of explanatory variables (7). In this section, variables of interest are discussed

1. Driver/ Occupant characteristics

Number of occupants

Figure 5-16 illustrates the impact of number of occupants across seven models. As shown in the figure, there are three groups with distinct differences. Drivers age<=25 had the highest OR of all groups. Perhaps, due to lack of driving experience this is the highest risk group. Excluding the older-drivers population, increasing the number of occupants had an increasing impact on the OR of both injury levels except for female drivers and drivers age 46-65. For these two groups as the number of occupants increases, the OR of level 5 was not likely to be affected. Perhaps, female drivers by nature are more focused in coping with higher number of occupants than the male drivers. The latter group (46-65 drivers age) seemed to be the most experienced and less physiologically challenged as the older group (66 and older), therefore, due to their experience (almost 39 years of average experience) they are not likely to be affected by this significant variable in sustaining a level 5 injury.

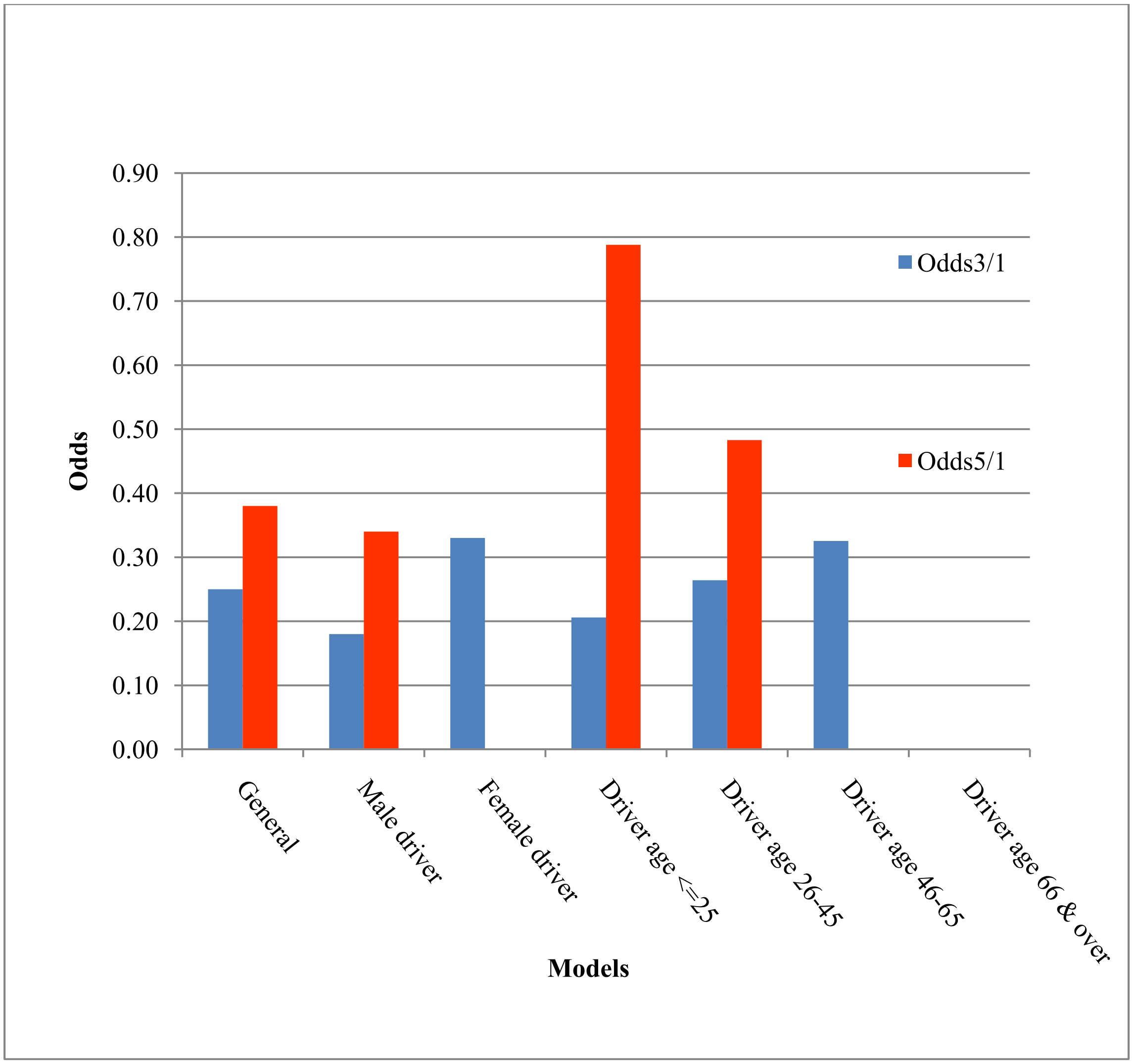

Figure 5-17 Impact of number of occupants on injury severity levels

2. Crash Environment at crash location

Dark not lighted

By referring to Figure 5-17, it is shown that, roadways with no lights (during dark hours) increase the odds of sustaining a level 3 injury across seven models. There are two groups with major differences. Drivers age<=25 had the highest OR of level 5 and lowest in level 3. Other words, having a driver from this age group increase the odds of a fatal injury while it reduces the odds of a level 3 injury. On the other hand, having a female driver had the lowest overall odds of injuries. Perhaps, female drivers are more cautious of their surrounding when driving during the night time than the male drivers.

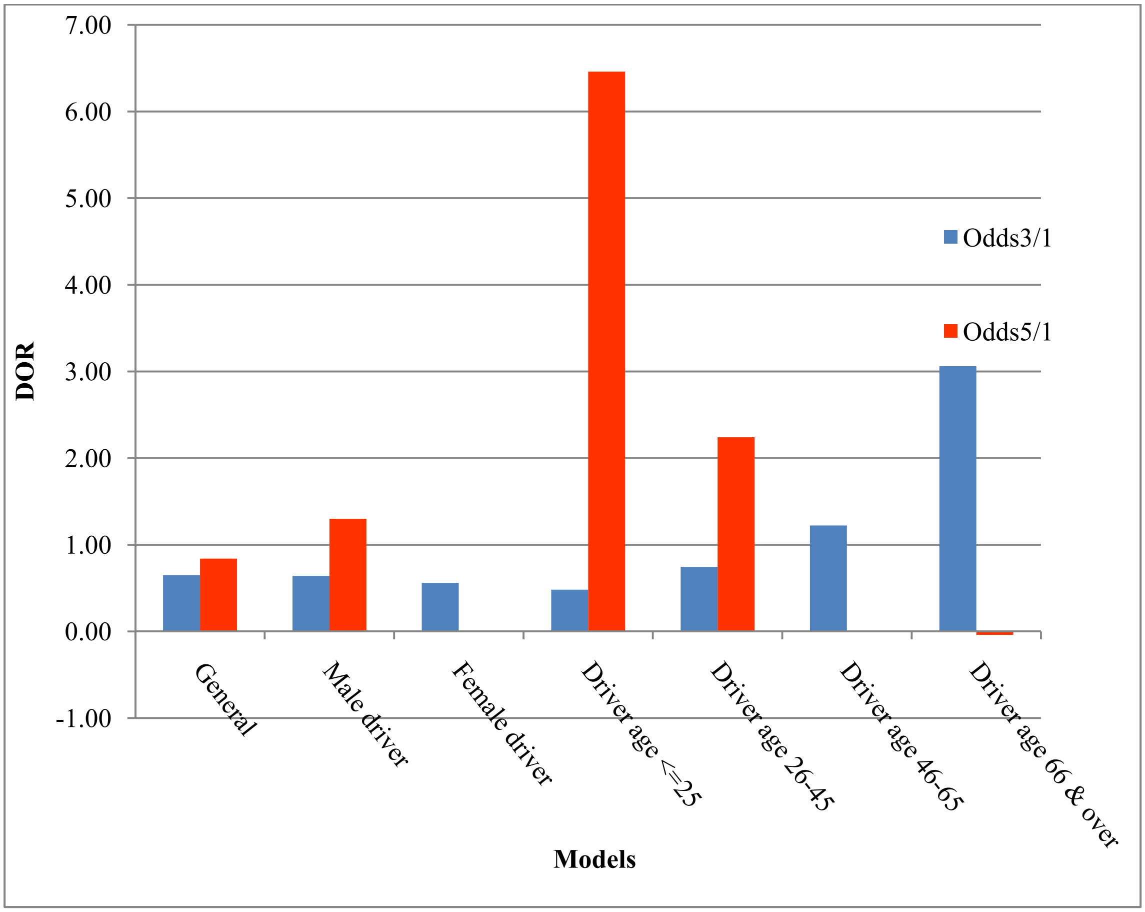

Figure 5-18 Impact of dark not lighted on injury severity levels

Head on

Excluding the older-drivers population, head-on crashes had an increasing impact on the OR of both injury levels across the six models. There are two groups with the highest odds of sustaining level 5 injuries; namely, drivers age<=25 (OR: 722) and female drivers (OR: 43). Figure 5-18 represents the OR values across the models. The bar representing the OR of level 5 for drivers age <=25 was truncated due to the magnitude (OR: 722). As shown before, female drivers were more cautious at dark roads with no lights and more concentrated when there are more occupants present , however, the odds of sustaining a level 5 injury was much higher with female drivers than the male drivers given a head on crash occurred. The odds of suffering a level 5 injury would be astronomical when a drivers age<=25 would be in control of a vehicle given a head on crash has occurred.

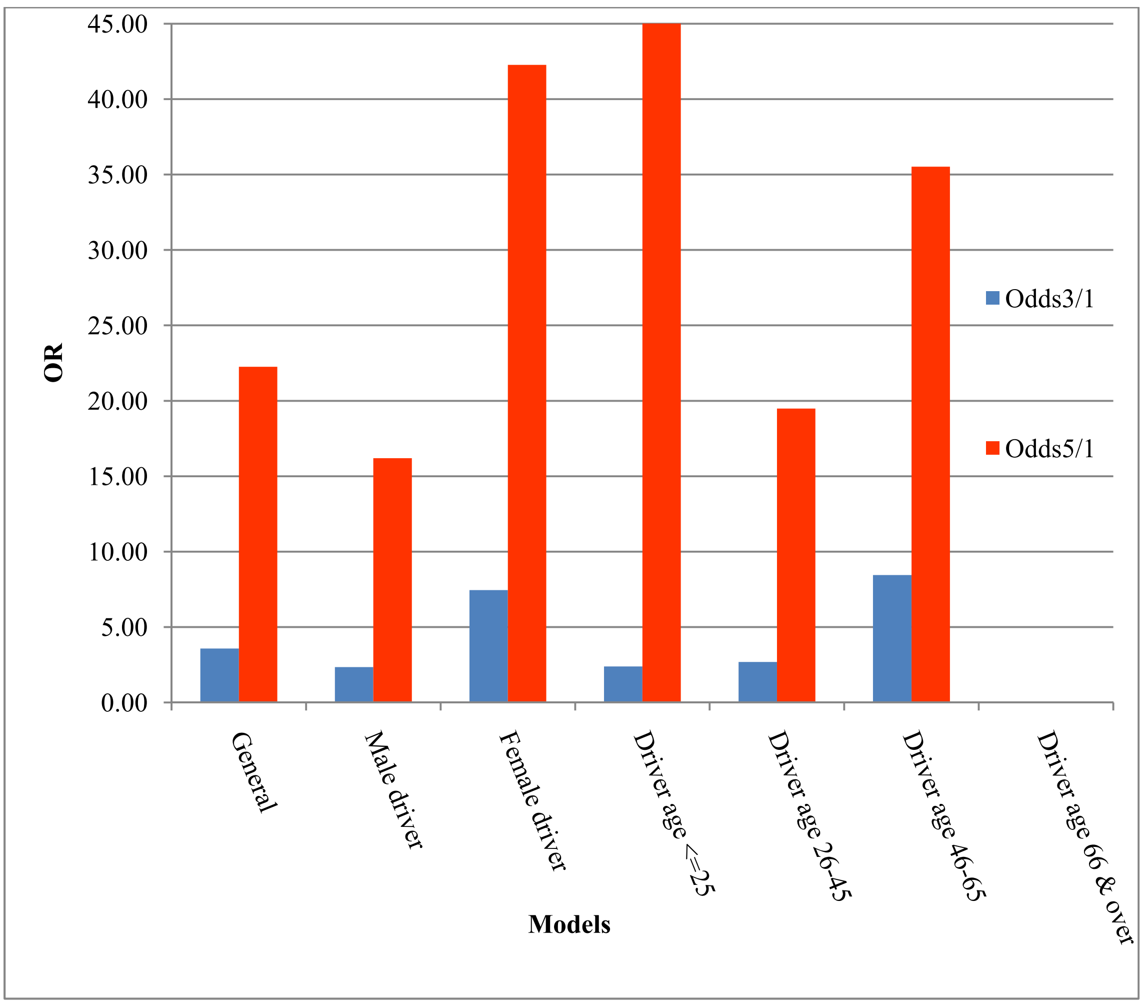

Figure 5-19 Impact of head on crashes on injury severity levels

Median divided

Roads with medians have reducing impacts on the ORs of both injuries in drivers age<=25 and male drivers models. Perhaps, the male drivers and drivers age<=25 would require medians more than any other groups to compensate for their lack of driving experience and or behavioral patterns.

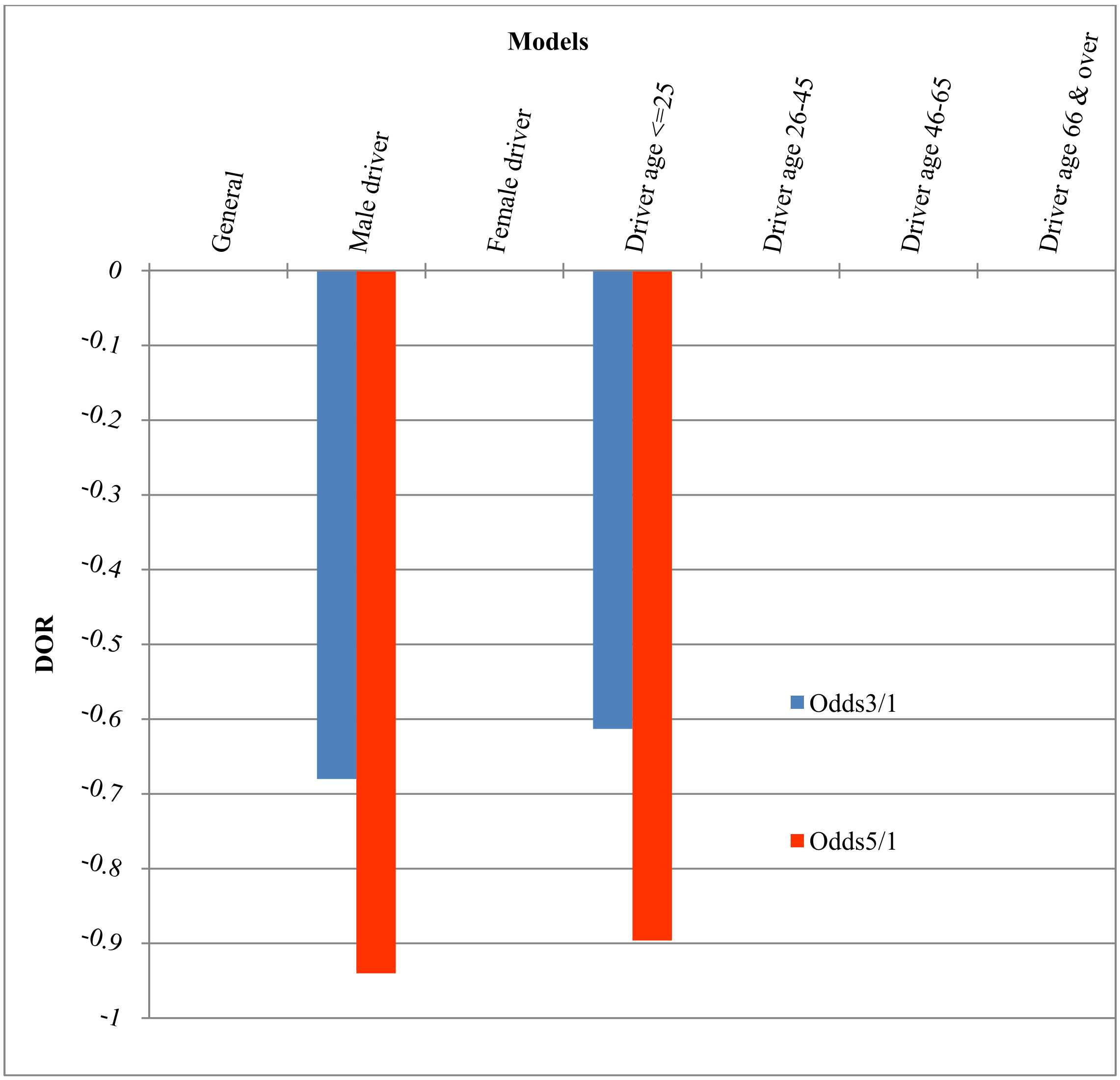

Figure 5-20 Impact of median divided on injury severity levels

3. Vehicle characteristics

Weight

Vehicle weight had opposing impact on the injury severities. As shown in Figure 5-20, there are three groups with distinct differences. Excluding drivers age 46-65 group, increasing the vehicle weight would decrease the OR of level 3 for all the groups except drivers age <=25. Heavier vehicles would increase the odds of fatal injury if driven by drivers age <=25. For female drivers the heavier cars had opposing impacts, it may reduce the OR of sustaining a level 3 injury severity but it increase the OR of sustaining a level 5 injury. Perhaps, female drivers are not as skilled as the male drivers in navigating the heavier cars. The third group was the older driver population, for this group, increasing the vehicle weight would decrease the OR for both injury severities.

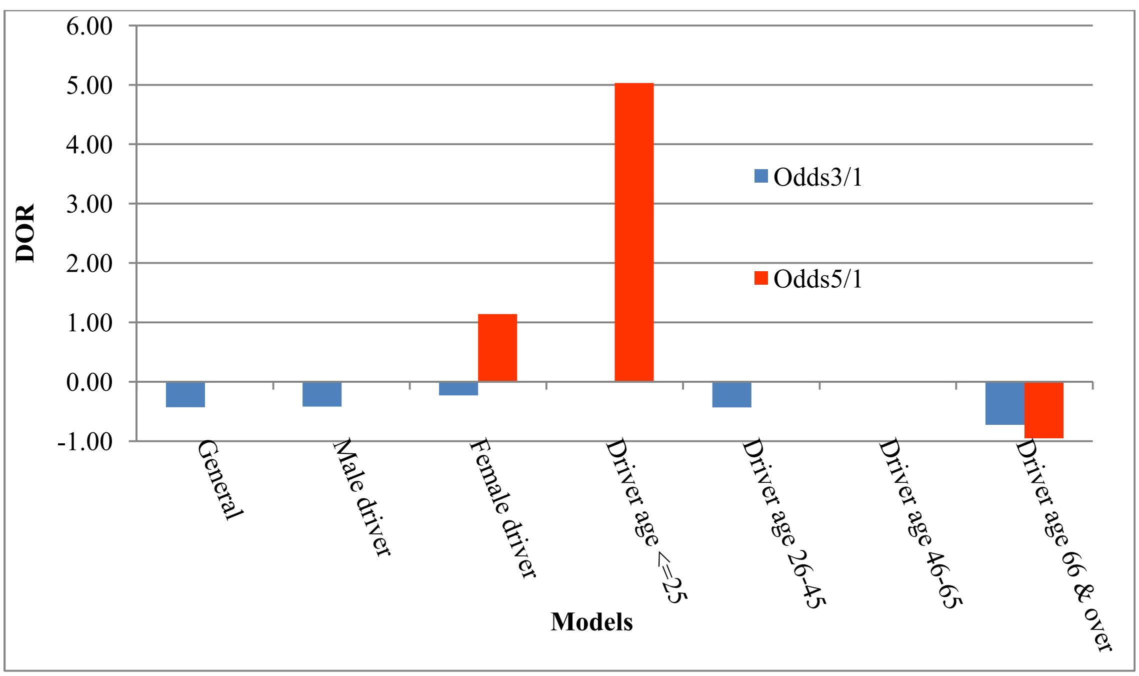

Figure 5-21 Impact of weight on injury severity levels

Front overhang

Figure 5-21 illustrates that increasing the front overhang reduces the OR of level 3 in general model. There are three groups with specific differences. The increase in front overhang would decrease the OR of suffering a level 3 for the male drivers and drivers

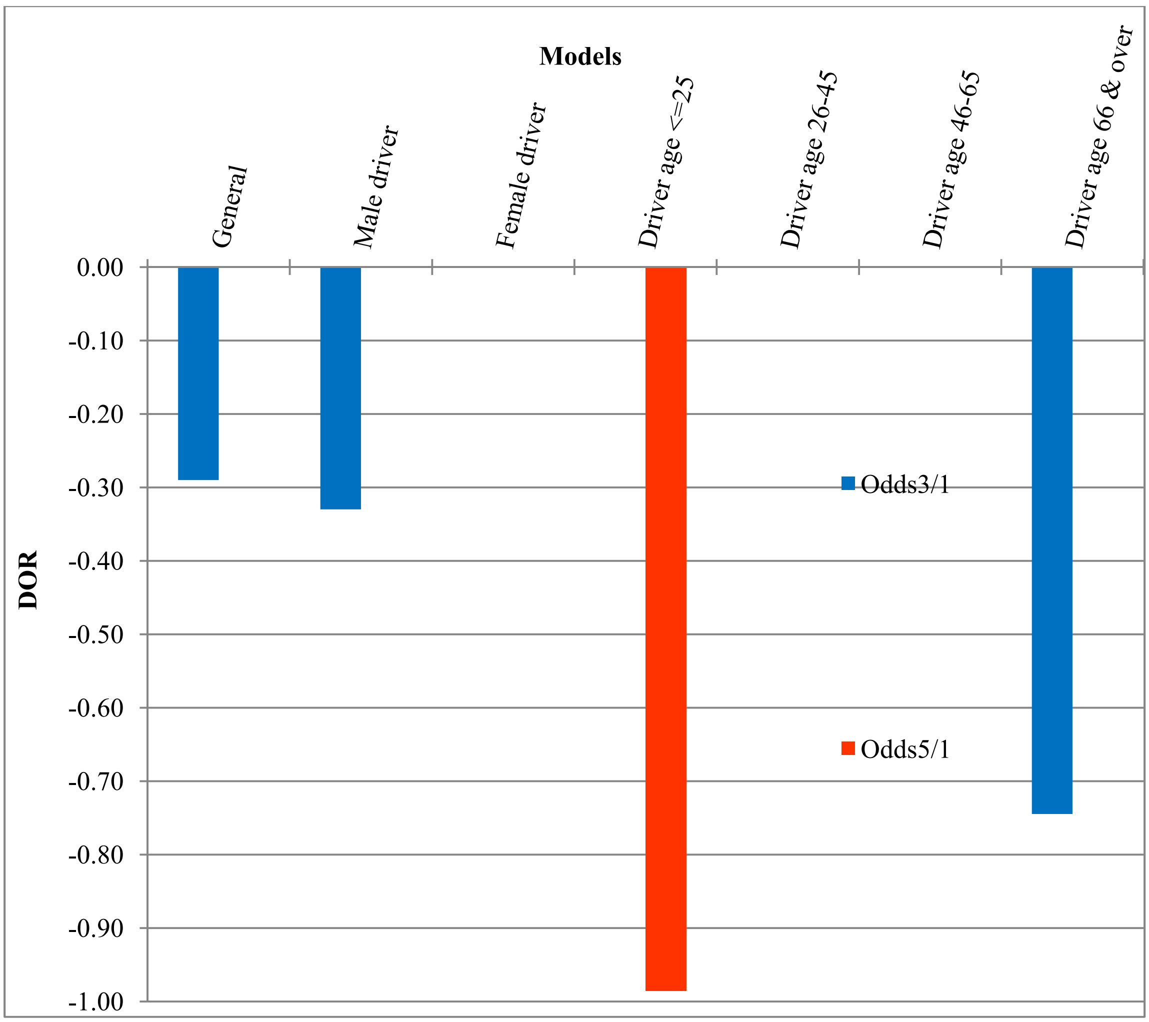

Figure 5-22 Impact of front overhang on injury severity levels

age 66 and older (consistent with the general model) models. However, the same increase will not affect the OR of level 3 for drivers age <= 25. This significant variable reduces the OR of level 5 by almost 100% for the latter age group. Other words, increase in front overhang has reducing impacts of injury severity levels more on one group compared to the other groups.

Rear overhang

Figure 5-22 illustrates three groups with reducing impacts of rear overhang on the ORs of fatal injury levels. Increasing the rear overhang by one unit reduces the ORs of level 5 for female drivers, drivers age <= 25, and drivers age 66 and older. Perhaps, these three groups are more involved in rear end-crashes (struck) than other groups.

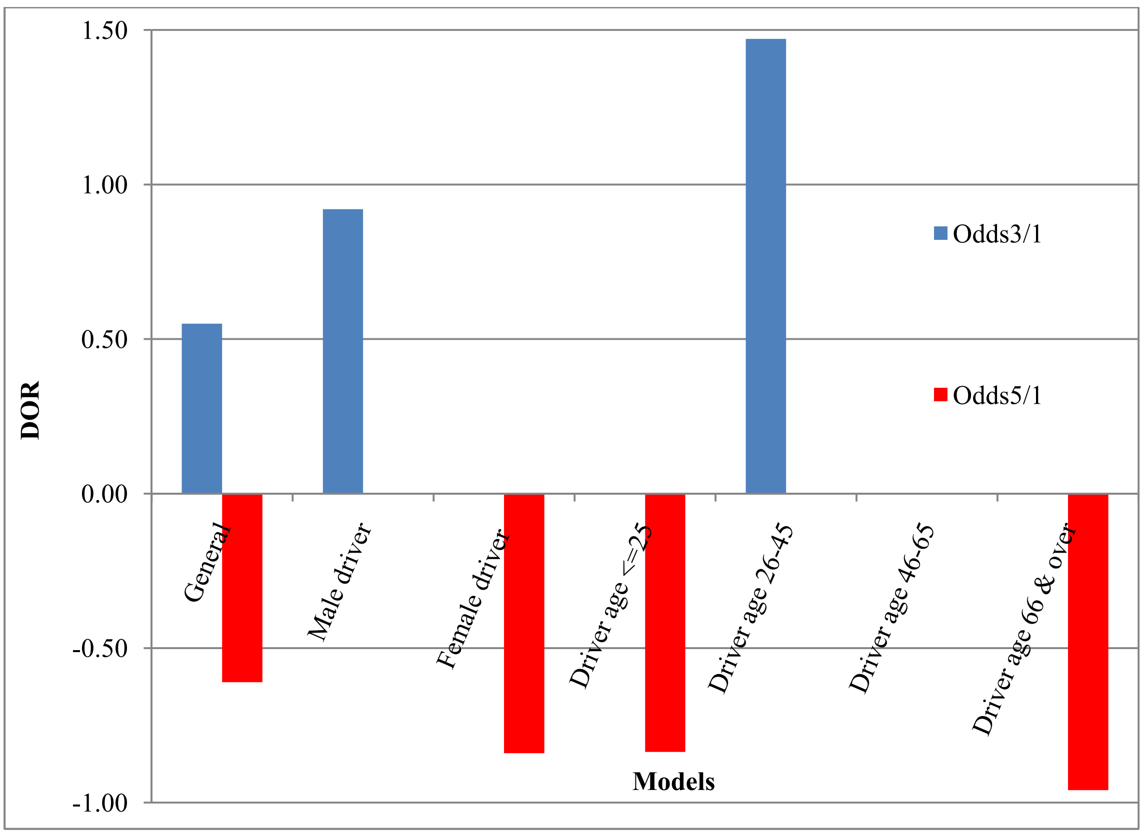

Figure 5-23 Impact of rear overhang on injury severity levels

Width

As shown in Figure 5-23, vehicle width has the highest impact on the ORs of level 3 and level 5 for drivers age-group 66 and older. As previously mentioned, the bar corresponding to OR of level 5 (OR: 607) was truncated at 30 on the y axis. Perhaps, this age group is not physiologically capable of handling wider cars. Although heavier cars were reducing the ORs of level 3 and level 5 for this age group, however, as the vehicles get wider they get heavier and this phenomenon may increase the ORs of suffering injury level 3 and level 5. Based on the data, it appears the wider and heavier vehicles are not the best combination of the vehicle characteristics for this age group.

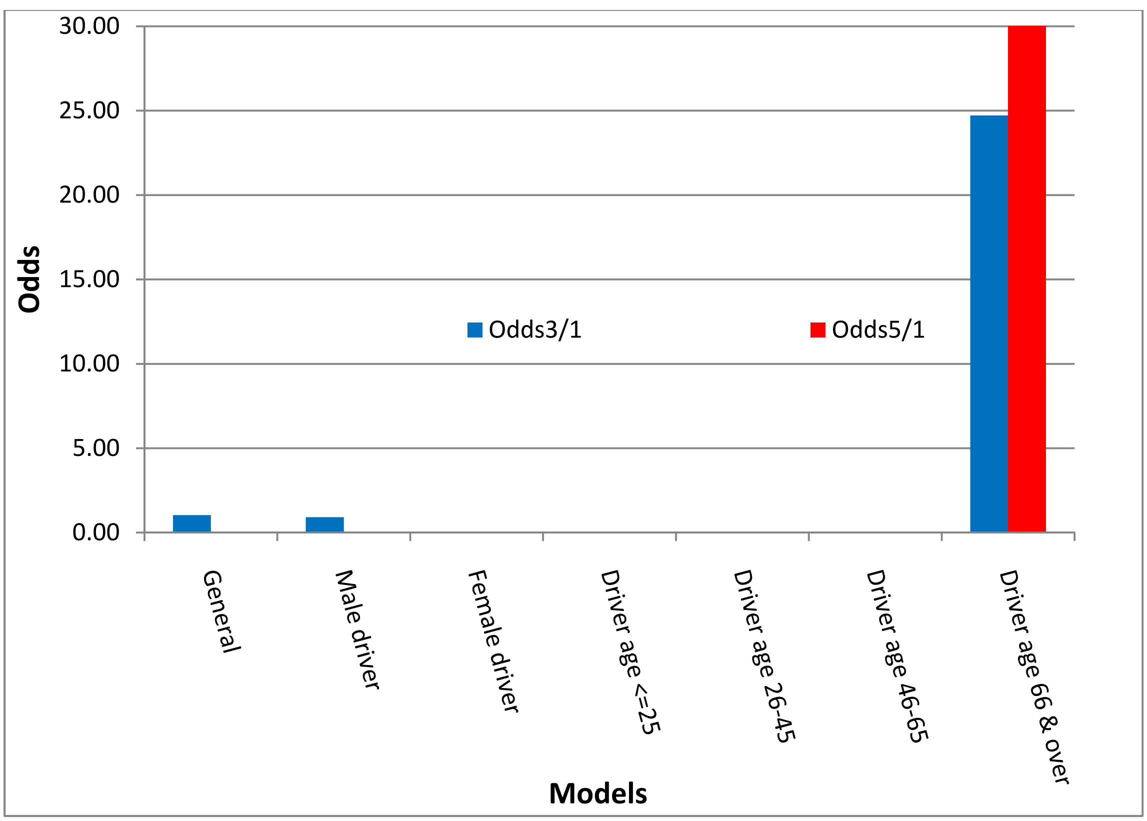

Figure 5-24 Impact of width on injury severity levels

Height

The vehicle height was statistically significant for drivers age <=25. By increasing vehicle height the OR of sustaining a level 5 injury reduce by 15%. The data suggest that this age group would benefit from an increase in a combined set of vehicles characteristics such as: front overhang, rear overhang, and higher heights.

CHAPTER 6

CONCLUSIONS AND RECOMMENDATIONS

6.1 Conclusions

This research calibrated three sets of models using multinomial logit and one model using ordered probit. To calibrate these models, a systematic approach was employed. The ordered probit model proved to be inappropriate methodology due to the failure of the model's assumption.

The calibrated models included variables representing driver/occupants characteristics, crash environment at the crash location, crash characteristics, and vehicle characteristics. In total, twenty variables were found to be statistically significant. The general model had the largest number of statistically significant variables (17) and the drivers age 66 and older model had the smallest number of statistically significant predictors (7).

The data was from the state of North Carolina and obtained from the Department Of Transportation Highway Information Safety System and supplemented with data from other sources, as described in chapter 4. The data included all the crash records for the year 2003.

The results of this study conveyed very interesting findings and shed new light on this area of research. The contributions of this study are twofold:

The first contribution of this study was to identify that the vehicle dimensions were important contributors to injury severity levels. It is worth mentioning that since PV-dimensions were significant in the crashes with LT, they may remain significant in PV-PV crashes.

The second contribution was that, although the other findings of this research were consistent with findings from other researchers, some differences were identified and are summarized below:

It is a well established fact that as the number of occupants increases the probability of injury increases (Chang and Mannering 1999; Farmer et al. 1997). However, the results of this study indicated that this variable does not pose an extra risk of fatal injury for two groups: drivers age 66 and older, and female drivers. Perhaps, female drivers are more focused in the presence of higher number of occupants than the male drivers. Moreover, in the case of driver age group 66 and older, perhaps, this group is the most experienced compared to other groups.

Dark roads with no lighting were identified (Dunken et al. 1998; Khattak et al. 2002; Abdel-Aty 2003) as significant contributors to injury severity. In this study it was shown that the risk associated with this variable was different, given a crash occurred. This variable poses (OR: 7.46, level 5/level 1) an extra risk of sustaining a fatal injury for drivers age <=25 and the lowest risk of injuries for the female drivers (OR: 1.56, level 3/ level 1). These differences may point out a fundamental behavioral pattern between these two groups.

This study and other studies (O'Donnell and Connor 1996; Kockelman and Kweon 2002; Abdel-Aty and Keller 2005; Hill and Boyle 2006; Xie et al. 2009) have emphasized on the degree of the importance of crash type, specifically head on crashes. Head on crashes impose a higher risk of a fatal injury severity for drivers age<=25 and female drivers. Perhaps, young drivers due to their lack of driving experience cannot maneuver out of a potential fatal crash.

In this study and others (Abdel-Aty and Keller 2005; Khorashadi et al. 2005) roadways with medians were important predictors of injury severity. However, this significant predictor posed a lower risk of injury for male drivers (OR: 0.32, level 3/level 1, OR: 0.06 level 5/level 1) and not for female drivers. Moreover, roadways with medians had a reducing impact on the OR for drivers age <= 25 (OR: 0.39, level 3/level1, OR: 0.10 level 5/level1).

Increasing vehicle weight (Wang and Kockelman 2005; Angel and Hickman 2009) had a reducing impact on the severity of injury. However, the results of this study indicated that for each additional 1000 lbs of the vehicle weight the risk of sustaining a fatal injury would increase for two groups: female drivers (OR:2.14, level 5/level 1) and drivers age<= 25 (OR: 6.03, level 5/level 1). Perhaps, since the bigger cars are heavier female drivers and young drivers are not as skilled as other groups in navigating the heavier cars.

Additional findings of this study: increasing the front overhang by one foot reduces the risk of sustaining a fatal injury by almost 100% for drivers age<=25 (OR: 0.01, level 5/level 1). Moreover, increasing the front overhang decreases the OR of suffering a level 3 injury for male drivers (OR: 0.67, level 3/level 1) and drivers age 66 and older (OR: 0.27, level3/ level1). This finding may be due to the fact that vehicles with longer front overhangs have a longer distance between the point of impact and the occupant compartment. In most crash types this longer distance may reduce the risk of injury.

Rear overhang was significant across six of the calibrated models. An increase in the length of this variable by one foot would more likely reduce the risk associated to fatal injury for female drivers (OR: 0.16, level 5/level 1), drivers age <= 25 (OR: 0.16, level 5/level 1), and drivers age 66 and older (OR: 0.04, level 5/level 1). Perhaps, due to the longer distance between the point of impact and the occupant compartment may cause a negative impact on the risk of sustaining injury.

Width had the highest impact on the ORs of sustaining level 3 and level 5 injuries for drivers age 66 and older (OR: 25.71 level 3/level 1, OR: 606.76 level 5/level 1). Perhaps, wider cars are heavier and larger in size and that would make the navigation of these types of vehicle harder for this age group. As an example, vehicles with larger width are harder to navigate between the two lanes of a roadway than the ones with smaller width. Moreover, as drivers get older, the perception reaction time may increase. Adding these two phenomena together may explain why wider vehicles may pose a higher risk for this age group.

Height was statistically significant for drivers age <=25 (OR: 0.85, level 5/level 1). Increasing the vehicle height had a reducing impact on the risk of sustaining a level 5 injury by 15%. Perhaps, the higher the height of a vehicle would create a better sight distance for this age group.

In summary, this study's contributions are: quantification of the impact of vehicle dimensions on the predicted probability of injury severity and expansion of the existing knowledge about the significant factors contributing to injury severity.

6.2 Recommendations

This section is a summary of recommendations based on the findings and shortcomings of this study.

1. Policy implications:

Elected officials can have additional information for evaluating and ranking project priorities based on the safety implications. Similarly, policy makers may use the findings of this study to develop age and or gender- specific educational programs. In addition, DMV examination boards may incorporate the findings of this study into driver licensing questions.

2. Traffic Engineers:

To reduce the risk associated with injury severity, counter measures such as installation of medians on multilane highways, installation of lightings on roadways, and whenever possible discourage large trucks on 2-lane highways.

3. Vehicle designers and vehicle buyers:

As shown in this study, vehicles dimensions play a significant role in reducing the risk associated with injury severity. Vehicles with longer front overhang, longer rear overhang, and higher height were likely to reduce the risk of sustaining injury for younger drivers. Therefore, this age group may consider purchasing vehicles that satisfy these conditions.

4. Insurance companies:

Insurance companies may use the results of this study to examine their calculated risk for different gender and age groups. The majority of variables did not increase the risk of sustaining injury for drivers age 46-65. On the other hand, for drivers age <= 25 various variables were likely to increase the risk of sustaining injury. Moreover, the results of the impact of vehicle dimensions may be used for adjusting premiums for different age and gender groups. As an example, if the drivers age<=25 drive vehicles with dimensions that reduce the impact on the injury risk, they could get a better insurance rate.

Study limitations and recommendations for future research:

One of the limitations of this study was the use of one year crash data from the state of North Carolina. However, the crash data from the state of North Carolina may not represent the entire nation's crash characteristics. Therefore, it is recommended that in the future studies, several years of crash data from different locations be used.

This study may have opened a new angle in injury severity research by including vehicle dimensions as explanatory variables. The importance of vehicle dimensions should be further studied to include different crash scenarios such as: PV-PV, multi-vehicle, PV-stationary objects, different crash types, and different crash types at different roadways. Moreover, interactions among various significant variables such as, gender versus age are highly recommended.

APPENDIX A CRASH STATISTICS

Table A- 1 All Motor Vehicle Injury Crash Statistics

Year	Injury Crashes	Vehicles Involved	Persons Injured	Million Vehicle Miles Traveled
1990	2,122,000	3,775,000	3,231,000	2,144,362
1991	2,008,000	3,581,000	3,097,000	2,172,050
1992	1,991,000	3,587,000	3,070,000	2,247,151
1993	2,022,000	3,647,000	3,149,000	2,296,378
1994	2,123,000	3,865,000	3,266,000	2,357,588
1995	2,217,000	4,094,000	3,465,000	2,422,696
1996	2,238,000	4,120,000	3,468,000	2,485,848
1997	2,149,000	3,966,000	3,348,000	2,561,695
1998	2,029,000	3,757,000	3,192,000	2,631,522
1999	2,054,000	3,773,000	3,236,000	2,691,056
2000	2,070,000	3,783,000	3,189,000	2,746,925
2001	2,003,000	3,663,000	3,033,000	2,797,287
2002	1,929,000	3,520,000	2,926,000	2,855,756
2003	1,925,000	3,536,000	2,889,000	2,890,450
2004	1,862,000	3,415,000	2,788,000	2,964,788
2005	1,816,000	3,287,000	2,699,000	2,989,430
2006	1,746,000	3,181,000	2,575,000	3,014,116
Average number of people injured from 1990-2006=*3,095,353*				

Source BTS (2008)

http://www.bts.gov/publications/national transportation statistics/html/table_02_19.html

ACCESSED 12-16-2008

Table A- 2 Large Truck Injury Crash Statistics

Year	Injury Crashes	Vehicles Involved	Persons Injured	Million Vehicle Miles Traveled
1990	102,000	107,000	150,000	146,242
1991	75,000	78,000	110,000	149,543
1992	91,000	95,000	139,000	153,384
1993	93,000	97,000	133,000	159,888
1994	91,000	96,000	133,000	170,216
1995	80,000	84,000	117,000	178,156
1996	89,000	94,000	129,000	182,971
1997	92,000	96,000	131,000	191,477
1998	85,000	89,000	127,000	196,380
1999	95,000	101,000	142,000	202,688
2000	96,000	101,000	140,000	205,520
2001	86,000	90,000	131,000	209,032
2002	90,000	94,000	130,000	214,603
2003	85,000	89,000	122,000	217,917
2004	83,000	87,000	116,000	220,811
2005	78,000	82,000	114,000	222,523
2006	77,000	80,000	106,000	223,037
Average number of people injured in large truck crashes from 1990-2006=*127,647*				

http://ai.fmcsa.dot.gov/CarrierResearchResults/HTML/2006Crashfacts/tbl1.htm

ACCESSED 12-16-2008

Table A- 3 Passenger vehicle occupants killed in large truck crashes

YEAR	passenger cars	Light truck	TOTAL PVs' Occupants
1990	2876.00	987.00	3,863
1991	2535.00	986.00	3,521
1992	2419.00	916.00	3,335
1993	2615.00	1077.00	3,692
1994	2639.00	1197.00	3,836
1995	2546.00	1153.00	3,699
1996	2683.00	1270.00	3,953
1997	2674.00	1426.00	4,100
1998	2556.00	1510.00	4,066
1999	2524.00	1493.00	4,017
2000	2475.00	1487.00	3,962
2001	2269.00	1539.00	3,808
2002	2206.00	1505.00	3,711
2003	2206.00	1515.00	3,721
2004	2240.00	1577.00	3,817
2005	2070.00	1646.00	3,716
2006	2015.00	1527.00	3,542
Average number of PVs' occupants died in large truck crashes from 1990-2006=3786			

http://ai.fmcsa.dot.gov/CarrierResearchResults/HTML/2006Crashfacts/tbl1.htm

ACCESSED 12-16-2008

Due to the scope of the study the above table excludes the occupants of large truck, Motor cycles, buses, and other unknowns.

Table A- 4 Motor Vehicle Fatalities from 1990 to 2006

Year	Fatal Crashes	Vehicles Involved	Occupant Fatality	Total Fatality
1990	39,836	59,292	37,134	44,599
1991	36,937	54,765	34,740	41,508
1992	34,942	52,227	32,880	39,250
1993	35,780	53,777	33,574	40,150
1994	36,254	54,911	34,318	40,716
1995	37,241	56,524	35,291	41,817
1996	37,494	57,347	35,696	42,065
1997	37,324	57,060	35,725	42,013
1998	37,107	56,922	35,382	41,501
1999	37,140	56,820	35,875	41,717
2000	37,526	57,594	36,348	41,945
2001	37,862	57,918	36,440	42,196
2002	38,491	58,426	37,375	43,005
2003	38,477	58,877	37,341	42,884
2004	38,444	58,729	37,304	42,836
2005	39,252	59,751	37,727	43,510
2006	38,588	58,152	36,947	42,642
Average	37,570	57,005	35,888	42,021

Average number of people died due to vehicle crashes from 1990-2006=42,021

ACCESSED 12-16-2008

http://ai.fmcsa.dot.gov/CarrierResearchResults/PDFs/LargeTruckCrashFacts2006.pdf

APPENDIX B CRASH DATA

Table B- 1 Accident file

cnty_rte	milepost	begmp	endmp	aadt	access	inv_cntl	med_type	no_lanes
0010000040	0	0	0.93	83000	3	1	6	8
0010000040	0	0	0.93	83000	3	1	6	8
0010000040	0	0	0.93	83000	3	1	6	8
0010000040	0	0	0.93	83000	3	1	6	8
0010000040	0.019	0	0.93	83000	3	1	6	8
0010000040	0.02	0	0.93	83000	3	1	6	8
0010000040	0.057	0	0.93	83000	3	1	6	8
0010000040	0.1	0	0.93	83000	3	1	6	8
0010000040	0.1	0	0.93	83000	3	1	6	8
0010000040	0.1	0	0.93	83000	3	1	6	8
0010000040	0.1	0	0.93	83000	3	1	6	8
0010000040	0.11	0	0.93	83000	3	1	6	8
0010000040	0.12	0	0.93	83000	3	1	6	8
0010000040	0.15	0	0.93	83000	3	1	6	8
0010000040	0.22	0	0.93	83000	3	1	6	8
0010000040	0.22	0	0.93	83000	3	1	6	8
0010000040	0.32	0	0.93	83000	3	1	6	8
0010000040	0.39	0	0.93	83000	3	1	6	8
0010000040	0.42	0	0.93	83000	3	1	6	8
0010000040	0.42	0	0.93	83000	3	1	6	8
0010000040	0.5	0	0.93	83000	3	1	6	8
0010000040	0.52	0	0.93	83000	3	1	6	8
0010000040	0.52	0	0.93	83000	3	1	6	8
0010000040	0.6	0	0.93	83000	3	1	6	8
0010000040	0.72	0	0.93	83000	3	1	6	8
0010000040	0.72	0	0.93	83000	3	1	6	8
0010000040	0.75	0	0.93	83000	3	1	6	8
0010000040	0.8	0	0.93	83000	3	1	6	8
0010000040	0.82	0	0.93	83000	3	1	6	8

Original data from HSIS in Excel (2003) format.

Table B- 2 Vehicle File

caseno	Vehno	make--me	amtdamg	spdlim	impactsp	vin_id
100781689	2	FORD	5000	45	45	1FMZU62X2YZB60828
100781689	1	CHEV	5000	45	15	1GNDU23E51D225685
100781689	3	VOLKS	2000	45	0	3VWSC29M7XM038721
100782957	2	FORD	1500	45	0	1FTJW35F0SEA80066
100782957	1	DODG	1500	45	30	1B7HW14T6DS401638
100783951	1	TOYO	10000	45	35	4TARNB1A4RZ197241
100784933	2	CHEV	1000	55	55	1G1AW69K2BD432166
100784933	1	BUIC	3000	55	40	1G4HP52L0RH409404
100784933	3	FORD	500		0	1FALP6537TK234973
100784936	1	CHEV	3000	35	30	1GNEV18K7KF162900
100784936	2	PONT	4000	35	30	1G2NE52T8VM561868
100784936	3	FORD	200	35	35	1FMCA11UXRZA19029
100784937	1	OLDS	1000	20	10	1GHDT13W6Y2220736
100784937	2	FORD	500	20	10	1FTEF14YX--22464
100790589	1	NISSAN	2500	55	50	1N4BU31D1RC244659
100790590	2	GMC	2000	35	40	2GTEC19K051518396
100790590	1	DODGE	1200	35	35	1B4GK44R4MX555696
100790591	1	DODGE	3000	70	70	1B4HS28Z84F144736
100790591	2	FORD	2500	70	70	1FTSW31FX2ED54609
100790593	1	CHEV	0	45	1	1GBM6P1F6LV104306
100790593	2	PLYMOUTH	200	45	0	1P4GP44G3WB754849
100790594	1	FORD	1200	55	35	1FABP52U6JG145317
100790594	2	FORD	1000	55	10	1FDHF25H7VEA60245
100790595	1	PLYMOUTH	1000	55	30	1P3EJ46C3VN686220
100790595	2	VOLKS	1800	45	5	3VWFA81H4VM018254
100790597	2	FORD	400	45	5	1FTDF15Y7RNB57495
100790597	1	HONDA	1700	45	20	1HGCB767X--141231
100790598	1	ACURA	10000	45	20	JH4DB7556WS001383
100790599	1	TOYOTA	10000	45	55	JT3VN39WXP0103621

Original data from HSIS in Excel (2003) format.

Table B- 3 Occupant file

caseno	vehno	prsn_nbr	prsn_cty	prsn_zip	prsn_dob	age
100781689	2	1	MOORESVILLE	281159739	9/12/1954	48
100781689	1	2	CORNELIUS	27612	2/19/1954	48
100781689	3	2	TAMPA	33647	2/22/1981	21
100781689	1	1	WINSTON SALEM	271036256	3/19/1969	33
100781689	3	1	BERNYN	19312	11/8/1982	20
100782957	2	1	CHARLOTTE	282141455	8/6/1941	61
100782957	1	1	CHARLOTTE	282162769	4/30/1962	40
100783951	1	1	GARNER	275298308	3/13/1970	32
100784933	2	4	MORGANTON	28655	11/1/1976	26
100784933	2	2	MORGANTON	28655	4/20/1983	19
100784933	3	1				
100784933	1	1	MORGANTON	286559330	2/8/1980	22
100784933	2	1	MORGANTON	286559077	2/16/1980	22
100784933	2	3	MORGANTON	28655		19
100784936	3	1	MEBANE	273020000	9/29/1972	30
100784936	2	2	MEBANE	273029631	10/4/2001	1
100784936	1	1	BURLINGTON	272171715	2/3/1944	58
100784936	2	3	SWEEPSONVILLE		12/7/1984	18
100784936	2	1	MEBANE	273029631	3/23/1984	18
100784937	1	2	VALDESE	286902652	6/22/1972	30
100784937	1	3	VALDESE	286902652	4/11/2001	1
100784937	2	1	HICKORY	286011034	9/22/1972	30
100784937	1	1	VALDESE	286902652	9/13/1967	35
100790589	1	1	WENDELL	27591	########	19
100790590	2	1	KNIGHTDALE	27545	4/1/1946	57
100790590	1	1	RALEIGH	27604	11/4/1943	59
100790590	2	2	KNIGHTDALE	27545	5/25/1944	58
100790591	1	1	RALEIGH	276037908	2/18/1954	49
100790591	2	1	BENSON	275046893	9/11/1969	33

Original data from HSIS in Excel (2003) format.

APPENDIX C MODELS, TESTS, AND PROBABILITIES

Table C- 1 Full and Final models using Ordered Probit Model

Variables	FULL MODEL		FINAL MODEL	
	Coefficient	P>z	Coefficient	P>z
Alcohol Flagged	-0.368331	0.634	--	--
PV Driver Age	0.0010745	0.556	--	--
PV Driver Gender	-0.1399988	0.029	-0.1238591	0.047
PV Total # of Occupants	0.126454	0.000	0.1225737	0.000
Posted Speed Limit	0.0126228	0.019	0.0172529	0.000
PV Impact Speed	0.0090524	0.008	--	--
PV Travel Speed	0.002932	0.378	0.0102864	0.000
LT Impact Speed	-0.00265	0.488	--	--
LT Travel Speed	0.0066274	0.063	--	--
Crash Type Other	0.1236757	0.245	--	--
Crash Type Head On	1.324162	0.000	1.284812	0.000
Crash Type Angle	0.8949504	0.000	0.8236835	0.000
Crash Type Rear	0.5852907	0.000	0.4590484	0.000
Crash Type Turn	0.4228096	0.000	0.3165104	0.001
Geometric Straight	-0.0118663	0.893	--	--
Median Divided	-0.0072098	0.972	--	--
PV Overall Height	-0.0022673	0.794	--	--
PV Curb weight	-0.0001695	0.079	-0.0001832	0.007
PV overall Width	0.0365665	0.01	0.0205192	0.086
PV Wheel Base	-0.0053071	0.15	--	--
PV Front Overhang	-0.0179853	0.093	--	--
PV Rear Overhang	0.0020022	0.792	--	--
LT Weight/1000	0.0005961	0.356	--	--
LT Driver Age	0.0010648	0.682	--	--
LT Driver Gender	-0.1158713	0.465	--	--
LT Percentage	0.0094339	0.219	--	--
AADT/1000	0.0083528	0.049	--	--
No of Lanes	-0.136477	0.045	--	--
AADT/1000/Lane	-0.0428174	0.019	-0.0123404	0.058
Million Miles Travelled	-0.0007431	0.635	--	--
Road Class 2 Lane	0.6823454	0.002	0.7448693	0.000
Road Class Multilane Divided	0.6061717	0.000	0.4334241	0.001
Road Class Multilane Undivided	0.5931116	0.031	0.4505366	0.022
Surface Condition Others	-0.093096	0.708	--	--
Surface Condition Wet	0.0071166	0.95	--	--
Weather Condition Others	-0.0360128	0.671	--	--
Weather Condition Rain	-0.0731549	0.631	--	--
Pavement Type Others	0.102763	0.769	--	--
Pavement Type Concrete	0.1154474	0.427	--	--

Table C- 2 Full and Final models using Ordered Probit Model (continued)

Lighting Condition Others		0.2049584	0.192	--	--
Lighting Condition Dark-Lighted		0.1807468	0.209	--	--
Lighting Condition Dark- No Light		0.3088761	0.001	0.2822316	0.001
Thresholds	τ_1	2.870979		3.2955712	
	τ_2	4.350846		4.751982	

Table C- 3 Parallel regression test for OP

Variables		**Coefficient**	**P>z**
PV Driver Gender		-0.1233	0.034
PV Total # of Occupants		0.1227	0
Posted Speed Limit		0.0173	0
PV Travel Speed		0.0103	0
Crash Type Head On		1.2837	0
Crash Type Angle		0.8231	0
Crash Type Rear		0.4586	0
Crash Type Turn		0.3156	0.001
PV Curb weight		-0.0002	0.003
PV overall Width		0.0211	0.2
AADT/1000/Lane		-0.0123	0.037
Road Class 2 Lane		0.7460	0.056
Road Class Multilane Divided		0.4339	0
Road Class Multilane Undivided		0.4510	0.001
Lighting Condition Dark- No Light		0.2832	0.021
Thresholds	τ_1	2.870979	
	τ_2	4.350846	

Table C- 4 Scalar measure of approximate likelihood ratio test of equality of coefficients

Likelihood Ratio (Degrees of Freedom) , P=0.000	293.92(15)
Log Likelihood (Intercept Only)	-1414.3461
Log Likelihood (Full Model)	-1267.3878
Chi ^{2} (Degrees of freedom), P=.0140	29.46 (15)
Number of Observations	1804

Table C- 5 Full general model using multinomial logit methodology

Variables	Logit coefficient (3/1)	P-value	Logit coefficient (5/1)	P-value
Alcohol Flagged	-0.3293757	0.800	-28.48193	1.000
PV Driver Age	-0.0032684	0.331	0.0188359	0.013
PV Driver Gender	-0.4372826	0.000	0.1532121	0.577
PV Total # of Occupants	0.2162706	0.000	0.377437	0.003
Posted Speed Limit	0.0054873	0.572	0.0781606	0.003
PV Impact Speed	0.0091021	0.148	0.0364197	0.012
PV Travel Speed	0.0067727	0.264	0.0064468	0.65
LT Impact Speed	-0.0003266	0.964	-0.0140693	0.302
LT Travel Speed	0.0092395	0.164	0.0212187	0.103
Crash Type Other	0.3014613	0.121	-0.1113734	0.828
Crash Type Head On	1.6090900	0.001	3.123441	0.000
Crash Type Angle	1.3676550	0.000	2.333894	0.000
Crash Type Rear	1.1219660	0.000	0.9028756	0.035
Crash Type Turn	0.8144854	0.000	0.7215255	0.124
Geometric Straight	-0.0450347	0.780	-0.1487397	0.683
Median Divided	0.2569901	0.485	-1.02585	0.297
PV Overall Height	0.0013732	0.932	-0.0244205	0.49
PV Curb weight	-0.5211272	0.005	0.1896499	0.635
PV overall Width	0.8649364	0.007	0.6096986	0.359
PV Wheel Base	-0.1059386	0.196	-0.0511338	0.768
PV Front Overhang	-0.4369514	0.069	-0.2941459	0.543
PV Rear Overhang	0.4535531	0.007	-0.9911375	0.008
LT Weight/1000	0.0001364	0.910	0.0017127	0.498
LT Driver Age	0.0054665	0.253	-0.0083772	0.457
LT Driver Gender	-0.1796169	0.538	-0.2331526	0.726
LT Percentage	0.0203112	0.147	0.0011597	0.971
AADT/1000	0.0161429	0.041	-0.0081599	0.757
No of Lanes	-0.2379393	0.056	-0.0933959	0.808
AADT/1000/Lane	-0.0868792	0.012	0.0047277	0.961
Million Miles Travelled	0.0002927	0.917	-0.004903	0.564
Road Class 2 Lane	1.0039930	0.012	1.856372	0.069
Road Class Multilane Divided	0.7594278	0.004	2.203567	0.001
Road Class Multilane Undivided	0.9295970	0.061	1.470744	0.247
Surface Condition Others	0.0363136	0.937	-1.178805	0.326
Surface Condition Wet	0.0398653	0.845	-0.0013131	0.998
Weather Condition Others	-0.0218116	0.888	-0.1829203	0.622
Weather Condition Rain	-0.1256227	0.652	-0.1795032	0.786
Pavement Type Others	0.5694652	0.351	-28.80123	1.000
Pavement Type Concrete	-0.0330899	0.906	0.658804	0.177
Lighting Condition Others	0.0476273	0.876	1.129052	0.03
Lighting Condition Dark-Lighted	0.2278592	0.387	0.8994219	0.151
Lighting Condition Dark- No Light	0.5358209	0.001	0.7526805	0.031
Constant	-5.4376870	0.002	-8.913533	0.027

Table C- 6 Sample of predicted versus observed injury severity for MNLM

Injury level	Predicted Probability				Injury level
Observed	Level 1	Level 3	Level 5	Max	Predicted
1	0.8324749	0.1654637	0.0020614	0.8324749	1
1	0.4627782	0.4665621	0.0706597	0.4665621	3
1	0.7110415	0.2854185	0.00354	0.7110415	1
3	0.5401489	0.4226272	0.0372239	0.5401489	1
1	0.5024267	0.4287897	0.0687836	0.5024267	1
3	0.4692748	0.471471	0.0592542	0.471471	3
1	0.6680188	0.25491	0.0770712	0.6680188	1
1	0.5711082	0.4104347	0.0184571	0.5711082	1
1	0.5097119	0.3642362	0.1260519	0.5097119	1
1	0.6463367	0.195571	0.1580923	0.6463367	1
1	0.5895805	0.3902364	0.020183	0.5895805	1
1	0.840607	0.1560483	0.0033447	0.840607	1
1	0.7618032	0.2360915	0.0021053	0.7618032	1
3	0.5675779	0.4257964	0.0066258	0.5675779	1
1	0.8047844	0.1836443	0.0115712	0.8047844	1
1	0.686662	0.2704226	0.0429154	0.686662	1
1	0.8016344	0.1219499	0.0764157	0.8016344	1
3	0.7466714	0.2463552	0.0069734	0.7466714	1
1	0.7197043	0.2520952	0.0282005	0.7197043	1
1	0.5984651	0.3713178	0.0302171	0.5984651	1
3	0.6414947	0.3387057	0.0197995	0.6414947	1
1	0.7072168	0.2805545	0.0122287	0.7072168	1
3	0.4667831	0.4869345	0.0462824	0.4869345	3
3	0.4789609	0.4268584	0.0941807	0.4789609	1
1	0.8228148	0.1616709	0.0155143	0.8228148	1
5	0.1138061	0.1741511	0.7120428	0.7120428	5
1	0.409453	0.2381766	0.3523705	0.409453	1
1	0.7119882	0.2504127	0.0375991	0.7119882	1
3	0.7494504	0.2387829	0.0117667	0.7494504	1
1	0.5191849	0.4560249	0.0247902	0.5191849	1
3	0.4791559	0.400769	0.1200751	0.4791559	1
1	0.8949062	0.1043858	0.000708	0.8949062	1
3	0.531553	0.4502917	0.0181552	0.531553	1
1	0.6529781	0.3325621	0.0144598	0.6529781	1
1	0.6503984	0.3349447	0.0146569	0.6503984	1
1	0.6361299	0.3439301	0.0199401	0.6361299	1
1	0.8031163	0.1893432	0.0075405	0.8031163	1
1	0.6632334	0.3046604	0.0321062	0.6632334	1
1	0.7531087	0.2304387	0.0164525	0.7531087	1

Table C- 7 Sample of predicted versus observed injury severity for MNLM (continued)

1	0.9331845	0.0584522	0.0083633	0.9331845	1
1	0.6821609	0.2664687	0.0513704	0.6821609	1
1	0.8180977	0.1694868	0.0124156	0.8180977	1
1	0.585867	0.4008683	0.0132647	0.585867	1
3	0.5827805	0.404707	0.0125125	0.5827805	1
3	0.558329	0.4004678	0.0412031	0.558329	1
3	0.7067471	0.2680451	0.0252078	0.7067471	1
1	0.5653011	0.4206292	0.0140697	0.5653011	1
1	0.7775582	0.1742869	0.0481549	0.7775582	1
1	0.6673421	0.2825731	0.0500847	0.6673421	1
1	0.3674412	0.5569942	0.0755646	0.5569942	3
3	0.715295	0.2776845	0.0070205	0.715295	1
1	0.7610555	0.1954143	0.0435301	0.7610555	1
3	0.511736	0.4659209	0.0223431	0.511736	1
1	0.7872372	0.2046286	0.0081342	0.7872372	1
1	0.9179547	0.0799992	0.002046	0.9179547	1
1	0.890018	0.1094398	0.0005422	0.890018	1
1	0.9190902	0.0800399	0.00087	0.9190902	1
3	0.7184513	0.2475845	0.0339642	0.7184513	1
1	0.7340094	0.2131996	0.0527911	0.7340094	1
1	0.7255765	0.2707355	0.003688	0.7255765	1
1	0.276406	0.5314456	0.1921484	0.5314456	3
1	0.7522317	0.2088497	0.0389186	0.7522317	1
1	0.3662742	0.5071237	0.1266021	0.5071237	3
3	0.7739742	0.2071069	0.0189189	0.7739742	1
1	0.9188687	0.0792115	0.0019198	0.9188687	1
3	0.230022	0.4546601	0.3153179	0.4546601	3
3	0.4883639	0.3016221	0.2100141	0.4883639	1
1	0.6072166	0.3747194	0.018064	0.6072166	1
5	0.4763167	0.3552339	0.1684494	0.4763167	1
3	0.7198005	0.2416149	0.0385846	0.7198005	1
3	0.836467	0.1462041	0.0173289	0.836467	1
3	0.3210977	0.6040947	0.0748076	0.6040947	3
1	0.8367934	0.1594528	0.0037538	0.8367934	1
1	0.7422175	0.2254588	0.0323237	0.7422175	1
1	0.8467517	0.1402842	0.0129641	0.8467517	1
1	0.9357014	0.0601759	0.0041227	0.9357014	1
1	0.6015401	0.3416662	0.0567937	0.6015401	1
1	0.8628063	0.134947	0.0022467	0.8628063	1
3	0.6681332	0.2887967	0.0430701	0.6681332	1
3	0.5474182	0.3723566	0.0802253	0.5474182	1

Table C- 8 Full model for PV-male drivers using MNL

Variables	Logit coefficient (3/1)	P value	Logit coefficient (5/1)	P value
Alcohol Flagged	-34.26772	1.000	-32.99568	1.000
PV Driver Age	-0.0063631	0.166	0.0222746	0.023
PV Total # of Occupants	0.1528205	0.076	0.4699401	0.008
Posted Speed Limit	0.0057663	0.669	0.0800023	0.015
PV Impact Speed	0.004478	0.612	0.0571067	0.005
PV Travel Speed	0.015701	0.072	-0.0068113	0.73
LT Impact Speed	0.0131733	0.197	-0.0419198	0.015
LT Travel Speed	0.0005215	0.957	0.0383051	0.02
Crash Type Other	0.2975024	0.27	-0.642769	0.437
Crash Type Head On	1.333856	0.037	2.853234	0.000
Crash Type Angle	1.288365	0.000	2.622042	0.000
Crash Type Rear	1.014726	0.000	1.002394	0.055
Crash Type Turn	0.9491963	0.001	0.3090011	0.625
Geometric Straight	0.0112779	0.961	-0.67999	0.136
Median Divided	-0.7537534	0.135	-1.250194	0.311
PV Overall Height	-0.0086968	0.671	-0.0450822	0.336
PV Curb weight	-0.405395	0.096	0.12645	0.816
PV overall Width	0.8617524	0.035	0.3136864	0.721
PV Wheel Base	-0.1351061	0.17	0.0437366	0.84
PV Front Overhang	-0.5943048	0.064	-0.5566058	0.387
PV Rear Overhang	0.6884351	0.002	-0.6544168	0.158
LT Weight/1000	-0.0003654	0.848	-0.0002312	0.942
LT Driver Age	0.0091713	0.171	-0.0003401	0.981
LT Driver Gender	-0.3947168	0.285	-0.4819498	0.57
LT Percentage	-0.0207886	0.276	-0.0304972	0.442
AADT/1000	-0.0015888	0.889	0.0102697	0.737
No of Lanes	-0.0544217	0.747	-0.3851957	0.432
AADT/1000/Lane	-0.0124827	0.788	0.001137	0.992
Million Miles Travelled	-0.0006469	0.871	-0.0120761	0.286
Road Class 2 Lane	0.1512176	0.777	1.160132	0.367
Road Class Multilane Divided	0.4206692	0.247	2.474334	0.002
Road Class Multilane Undivided	-0.2168685	0.748	1.192006	0.452
Surface Condition Others	-0.5097599	0.413	-1.366809	0.283
Surface Condition Wet	0.4612697	0.092	-0.1845804	0.785
Weather Condition Others	-0.0217389	0.921	-0.1576037	0.759
Weather Condition Rain	-0.5294272	0.158	-0.010938	0.99
Pavement Type Others	1.064609	0.164	-33.46594	1
Pavement Type Concrete	-0.1021669	0.787	1.093182	0.063
Lighting Condition Others	0.0553301	0.89	1.566751	0.009
Lighting Condition Dark-Lighted	0.5051166	0.163	0.9423339	0.222
Lighting Condition Dark- No Light	0.4568539	0.036	1.014337	0.017
Constant	-4.958241	0.029	-5.397709	0.314

Table C- 9 Full PV-female drivers model

Variables	Logit coefficient (3/1)	P-value	Logit coefficient (5/1)	P-value
Alcohol Flagged	3.46E+01	1.00E+00	4.36E-01	1.00E+00
PV Driver Age	0.0011958	0.822	-0.0039184	0.788
PV Total # of Occupants	0.2563029	0.007	0.2993492	0.224
Posted Speed Limit	0.008431	0.571	0.0850808	0.135
PV Impact Speed	0.0189648	0.052	0.0035309	0.886
PV Travel Speed	-0.0056827	0.532	0.020929	0.373
LT Impact Speed	-0.017278	0.121	0.0505542	0.121
LT Travel Speed	0.021075	0.033	-0.0099163	0.756
Crash Type Other	0.2723333	0.352	0.8700014	0.273
Crash Type Head On	2.18129	0.014	4.229676	0.001
Crash Type Angle	1.642246	0.000	1.646992	0.033
Crash Type Rear	1.396765	0.000	0.3711833	0.727
Crash Type Turn	0.7849281	0.008	1.059792	0.213
Geometric Straight	-0.0119866	0.96	0.6837496	0.363
Median Divided	1.509898	0.015	-3.024903	0.321
PV Overall Height	-0.0124861	0.664	0.0228886	0.76
PV Curb weight	-0.4107977	0.164	0.0093685	0.989
PV overall Width	0.4542681	0.427	1.812707	0.237
PV Wheel Base	0.0532587	0.747	-0.1868943	0.658
PV Front Overhang	-0.4056556	0.301	0.513575	0.579
PV Rear Overhang	0.0523002	0.859	-2.278167	0.003
LT Weight/1000	0.0009611	0.546	0.0016669	0.786
LT Driver Age	0.0033221	0.649	-0.018021	0.381
LT Driver Gender	0.0342974	0.947	-0.3522327	0.797
LT Percentage	0.0755663	0.001	0.0919811	0.283
AADT/1000	0.0313105	0.01	-0.2068902	0.051
No of Lanes	-0.4319816	0.031	1.78454	0.061
AADT/1000/Lane	-0.1615658	0.004	0.3284324	0.253
Million Miles Travelled	0.0020723	0.634	0.0109365	0.515
Road Class 2 Lane	2.140127	0.002	2.304231	0.442
Road Class Multilane Divided	1.216605	0.004	1.021152	0.489
Road Class Multilane Undivided	2.399396	0.003	-0.0838634	0.98
Surface Condition Others	1.154022	0.173	-33.08205	1.000
Surface Condition Wet	-0.4773879	0.143	0.4862581	0.605
Weather Condition Others	-0.0541622	0.814	-0.8371776	0.254
Weather Condition Rain	0.3684003	0.402	-1.15819	0.396
Pavement Type Others	-0.4243413	0.746	-32.28773	1.000
Pavement Type Concrete	0.0100271	0.983	-0.8650937	0.549
Lighting Condition Others	0.0525505	0.918	0.1134081	0.938
Lighting Condition Dark-Lighted	0.040702	0.92	0.2953091	0.854
Lighting Condition Dark- No Light	0.6019484	0.032	0.1624852	0.854
Constant	-4.296171	0.131	-19.30053	0.017

Table C- 10 Final PV-male drivers model

VARIABLES	LEVEL 3/1			LEVEL 5/1		
	Logit Coeffcient	Odds ratio	P-value	Logit Coeffcient	Odds ratio	P-value
Drivers age	*NA*	*NA*	*NA*	0.0178419	1.02	0.042
# Of occupants	0.1653736	1.18	0.044	0.2934816	1.34	0.072
Dark-not lighted	0.4933184	1.64	0.018	0.8327113	2.30	0.027
Multilane-divided	0.5969163	1.82	0.018	2.298165	9.96	0.000
Median-divided	-1.151085	0.32	0.000	-2.87634	0.06	0.000
Speed limit	*NA*	*NA*	*NA*	0.06719	1.07	0.009
Head on	1.208807	3.35	0.05	2.84471	17.20	0.000
Rear-end	0.7045802	2.02	0.000	0.9211934	2.51	0.032
Angle	1.170791	3.22	0.000	2.440095	11.47	0.000
Turn	0.7432704	2.10	0.002	*NA*	*NA*	*NA*
Travel speed	0.0198896	1.02	0.000	0.0397694	1.04	0.000
Weight(PV)	-0.5504859	0.58	0.003	*NA*	*NA*	*NA*
Front-overhang	-0.3958779	0.67	0.091	*NA*	*NA*	*NA*
Rear-overhang	0.6530257	1.92	0.002	*NA*	*NA*	*NA*
Width (PV)	0.6454384	1.91	0.087	*NA*	*NA*	*NA*
Constant	-5.681076	*NA*	0.001	-7.292185	*NA*	0.045

Table C- 11 Final PV-female drivers model

VARIABLES	LEVEL 3/1			LEVEL 5/1		
	Logit Coeffcient	Odds ratio	P-value	Logit Coeffcient	Odds ratio	P-value
# Of occupants	0.2821724	1.33	0.002	*NA*	*NA*	*NA*
Dark-not lighted	0.4453593	1.56	0.088	*NA*	*NA*	*NA*
2-lane	0.4329767	1.54	0.039	1.48037	4.39	0.019
AADT/1000/lane	-0.0348384	0.97	0.018	-0.2272573	0.80	0.003
Speed limit	*NA*	*NA*	*NA*	0.0886579	1.09	0.015
Head on	2.134519	8.45	0.011	3.767453	43.27	0
Rear-end	1.191257	3.29	0	*NA*	*NA*	*NA*
Angle	1.470744	4.35	0	1.398832	4.05	0.021
Turn	0.6891962	1.99	0.007	*NA*	*NA*	*NA*
Travel speed	0.0100547	1.01	0.07	*NA*	*NA*	*NA*
Weight(PV)	-0.2672348	0.77	0.056	0.7586793	2.14	0.01
Rear-overhang	*NA*	*NA*	*NA*	-1.838094	0.16	0.002
Constant	-2.334371	NA	0.013	-4.922855	NA	0.062

Table C- 12 Full model for drivers age<=25

Variables	Logit coefficient (3/1)	P value	Logit coefficient (5/1)	P value
Alcohol Flagged	0.1781755	0.918	-29.80394	1.000
PV Driver gender	-0.345634	0.153	1.102676	0.31
PV Total # of Occupants	0.1624928	0.172	1.137011	0.033
Posted Speed Limit	0.0016615	0.937	0.2578517	0.03
PV Impact Speed	0.0163306	0.162	-0.0098206	0.826
PV Travel Speed	0.0101421	0.385	0.0104842	0.791
LT Impact Speed	0.0127305	0.357	0.0461039	0.229
LT Travel Speed	-0.0034179	0.793	-0.0855425	0.079
Crash Type Other	0.6755281	0.093	-2.393912	0.578
Crash Type Head On	2.467092	0.171	15.89906	0.004
Crash Type Angle	1.385972	0.000	6.831883	0.004
Crash Type Rear	0.8355337	0.022	4.210114	0.049
Crash Type Turn	0.9132714	0.017	2.643298	0.231
Geometric Straight	0.0788373	0.822	-1.913433	0.17
Median Divided	0.5845854	0.424	-6.295015	0.626
PV Overall Height	-0.0499432	0.165	-0.2124941	0.114
PV Curb weight	0.1207628	0.739	3.329461	0.005
PV overall Width	0.5914976	0.405	1.234868	0.638
PV Wheel Base	0.1542773	0.394	-1.004004	0.233
PV Front Overhang	-1.152459	0.033	-6.670527	0.016
PV Rear Overhang	-0.3818004	0.331	-5.19835	0.016
LT Weight/1000	0.0055491	0.16	0.020308	0.295
LT Driver Age	-0.0035904	0.727	-0.0564369	0.236
LT Driver Gender	-0.2819014	0.702	-7.469486	0.021
LT Percentage	0.0803504	0.02	0.2890793	0.054
AADT/1000	0.0242134	0.152	-0.0090335	0.889
No of Lanes	-0.5942556	0.023	0.7096006	0.496
AADT/1000/Lane	-0.125581	0.077	-0.1898791	0.483
Million Miles Travelled	0.0026071	0.648	-0.0411817	0.311
Road Class 2 Lane	1.418732	0.048	-0.2557945	0.985
Road Class Multilane Divided	1.841621	0.001	3.773412	0.156
Road Class Multilane Undivided	3.07506	0.002	3.030706	0.82
Surface Condition Others	-0.3144576	0.785	-29.96497	1.000
Surface Condition Wet	-0.0976663	0.818	-0.925007	0.727
Weather Condition Others	-0.2824338	0.401	-2.890759	0.146
Weather Condition Rain	-0.0287208	0.96	1.227943	0.643
Pavement Type Others	0.4667912	0.644	-33.01049	1.000
Pavement Type Concrete	-0.4844591	0.406	-32.61286	1.000
Lighting Condition Others	-0.3209898	0.63	5.553426	0.007
Lighting Condition Dark-Lighted	0.9335771	0.056	-4.483601	0.166
Lighting Condition Dark- No Light	0.6085485	0.073	2.602589	0.092
Constant	-0.3450567	0.92	26.52861	0.158

Table C- 13 Final model for drivers age<=25

VARIABLES	LEVEL 3/1			LEVEL 5/1		
	Logit Coeffcient	Odds ratio	P value	Logit Coeffcient	Odds ratio	P value
# Of occupants	0.1873222	1.21	0.08	0.580958	1.79	0.046
Dark-not lighted	NA	NA	NA	2.00961	7.46	0.004
Multilane-divided	0.8503311	2.34	0.006	1.660502	5.26	0.089
Median divided	-0.949285	0.39	0.001	-2.264502	0.10	0.011
Head on	NA	NA	NA	6.582869	722.61	0.001
Rear-end	NA	NA	NA	2.031285	7.62	0.061
Angle	1.273257	3.57	0	3.791476	44.32	0.001
Turn	0.778097	2.18	0.017	NA	NA	NA
Travel speed	0.024528	1.02	0	NA	NA	NA
Overall height	NA	NA	NA	-0.1625394	0.85	0.022
Weight	NA	NA	NA	1.796975	6.03	0.006
Front-overhang	NA	NA	NA	-4.251625	0.01	0.002
Rear-overhang	NA	NA	NA	-1.805135	0.16	0.045
Constant	1.110464	NA	0.614	15.61172	NA	0.002

Table C- 14 Full model for drivers age 26-45

Variables	Logit coefficient (3/1)	P value	Logit coefficient (5/1)	P value
PV Driver gender	-0.4026654	0.034	0.2805917	0.555
PV Total # of Occupants	0.2145641	0.022	0.2539228	0.22
Posted Speed Limit	0.007839	0.556	0.146125	0.001
Crash Type Head On	1.376726	0.038	2.86473	0.001
Crash Type Angle	1.245971	0.000	1.143123	0.06
Crash Type Rear	1.009311	0.000	-35.12494	1.000
Crash Type Turn	0.1601381	0.602	-0.4554057	0.463
Median Divided	-0.0726796	0.904	-3.163536	0.405
PV Overall Height	0.0041783	0.873	0.0063947	0.918
PV Curb weight	-0.734339	0.009	-0.1308968	0.843
PV overall Width	0.8509326	0.078	0.2708846	0.823
PV Wheel Base	-0.149958	0.241	-0.2553332	0.403
PV Front Overhang	-0.416288	0.290	0.9588903	0.284
PV Rear Overhang	0.9150249	0.001	0.0401105	0.955
AADT/1000/Lane	-0.0336992	0.087	-0.1801326	0.102
Road Class 2 Lane	0.4035084	0.536	0.8441545	0.83
Road Class Multilane Divided	0.0669012	0.858	2.875576	0.048
Road Class Multilane Undivided	-0.9510303	0.273	1.117792	0.781
Lighting Condition Dark- No Light	0.4334331	0.096	-0.5986963	0.460
Constant	-4.972742	0.050	-12.26884	0.089

Table C- 15 Final model for drivers age 26-45

VARIABLES	LEVEL 3/1			LEVEL 5/1		
	Logit Coeffcient	Odds ratio	P-value	Logit Coeffcient	Odds ratio	P-value
Driver's gender	-0.412978	.66	0.027	NA	NA	NA
# Of occupants	0.2344274	1.26	0.01	0.393976	1.48	0.047
2-Lane	0.5564109	1.74	0.019	1.175705	3.24	0.053
AADT/1000/lane	-0.0340718	0.97	0.032	-0.2726905	0.76	0.002
Head on	1.304673	3.69	0.045	3.019504	20.48	0.000
Rear-end	1.01045	2.75	0.000	NA	NA	NA
Angle	1.226827	3.41	0.000	1.364702	3.91	0.013
Travel speed	0.0128913	1.01	0.014	0.0419749	1.04	0.007
Weight	-0.5683939	0.57	0.000	NA	NA	NA
Rear-overhang	0.9047949	2.47	0.000	NA	NA	NA
Constant	-3.053947	NA	0.000	-5.619828	NA	.013

Table C- 16 Full model for drivers age 46-65

Variables	Logit coefficient (3/1)	P value	Logit coefficient (5/1)	P value
Alcohol Flagged	-40.99781	1.000	-42.22683	1.000
PV Driver gender	-0.8895521	0.001	0.0842107	0.880
PV Total # of Occupants	0.4176204	0.009	0.4882972	0.103
Posted Speed Limit	0.037895	0.076	0.0759207	0.157
PV Impact Speed	0.0183597	0.231	-0.0015478	0.949
PV Travel Speed	0.0008408	0.953	0.062631	0.028
LT Impact Speed	0.0010547	0.946	-0.016712	0.445
LT Travel Speed	0.016525	0.236	0.0394034	0.083
Crash Type Other	1.044428	0.015	1.292037	0.132
Crash Type Head On	3.029829	0.035	6.183501	0.009
Crash Type Angle	1.954044	0.000	3.581661	0.000
Crash Type Rear end	1.638363	0.000	1.603025	0.039
Crash Type Turn	1.213029	0.008	0.5404036	0.610
Geometric Straight	-0.1866634	0.611	0.4267272	0.556
Median Divided	1.290243	0.109	0.7620618	0.643
PV Overall Height	0.0112515	0.712	0.0379536	0.566
PV Curb weight	-0.5500413	0.192	0.7368434	0.389
PV overall Width	0.5702743	0.439	-1.146401	0.397
PV Wheel Base	-0.0354223	0.837	-0.1583433	0.643
PV Front Overhang	-0.1324603	0.786	0.8554944	0.322
PV Rear Overhang	0.5580547	0.116	-0.0000382	1.000
LT Weight/1000	0.0013612	0.664	0.0220262	0.006
LT Driver Age	0.0103556	0.312	-0.0120526	0.582
LT Driver Gender	-0.8098434	0.228	-1.988065	0.105
LT Percentage	0.0092781	0.759	-0.0472751	0.407
AADT/1000	0.0504219	0.007	0.0155742	0.732
No of Lanes	-0.6561353	0.034	-0.5143456	0.431
AADT/1000/Lane	-0.1932397	0.019	-0.0660482	0.709
Million Miles Travelled	-0.0035152	0.615	-0.0017816	0.895
Road Class 2 Lane	2.295983	0.017	2.734104	0.145
Road Class Multilane Divided	1.459367	0.015	2.557141	0.021
Road Class Multilane Undivided	2.950056	0.008	-36.34738	1.000
Surface Condition Others	-1.920749	0.197	-42.36039	1.000
Surface Condition Wet	0.105985	0.810	-1.58594	0.263
Weather Condition Others	0.3452078	0.316	-0.2419304	0.765
Weather Condition Rain	-0.1428195	0.821	0.3357305	0.847
Pavement Type Others	1.497525	0.276	-36.46921	1.000
Pavement Type Concrete	0.2095356	0.742	0.0140129	0.990
Lighting Condition Others	-0.3782841	0.625	-51.50089	1.000
Lighting Condition Dark-Lighted	0.8653099	0.125	1.552196	0.239
Lighting Condition Dark- No Light	1.086946	0.006	1.027284	0.131
Constant	-8.63438	0.035	-11.03281	0.175

Table C- 17 Final model for drivers age 46-65

VARIABLES	LEVEL 3/1			LEVEL 5/1		
	Logit Coeffcient	Odds ratio	P-value	Logit Coeffcient	Odds ratio	P-value
Driver's gender(PV)	-0.7935969	0.45	0.001	NA	NA	NA
# Of occupants	0.2816833	1.33	0.041	NA	NA	NA
2-Lane	1.526675	4.6	0.000	3.23752	25.47	0.000
Multilane-divided	0.8807712	2.41	0.030	2.490181	12.06	0.001
Multilane Undivided	1.222431	3.40	0.054	NA	NA	NA
Dark-not lighted	0.7985827	2.22	0.023	NA	NA	NA
Speed limit (Posted)	0.0457442	1.05	0.009	0.0881298	1.09	0.027
Head on	2.246277	9.45	0.090	3.597833	36.52	0.015
Rear-end	1.112766	3.04	0.000	1.25508	3.51	0.028
Angle	1.581823	4.86	0.000	2.585901	13.28	0.000
Turn	0.727368	2.07	0.051	NA	NA	NA
Travel speed (PV)	0.0150069	1.02	0.058	0.0551173	1.06	0.003
Constant	-5.345591	NA	0.000	-12.78215	NA	0.000

Table C- 18 Full model for drivers age 66 & older

Variables	Logit coefficient (3/1)	P value	Logit coefficient (5/1)	P Value
PV Driver gender	-0.2363448	0.572	1.358409	0.403
PV Total # of Occupants	0.2798159	0.278	0.9363147	0.394
Posted Speed Limit	0.0280297	0.378	0.1008511	0.341
PV Impact Speed	-0.003586	0.844	0.1886528	0.288
PV Travel Speed	0.0140542	0.424	-0.1583299	0.319
Crash Type Head On	2.784628	0.055	-27.73501	1.000
Crash Type Angle	0.7547692	0.288	6.738365	0.097
Crash Type Rear end	0.638542	0.271	1.792311	0.403
Crash Type Turn	1.369281	0.027	5.073535	0.172
Median divided	-0.373376	0.782	2.308427	0.568
PV Overall Height	0.0544045	0.467	-0.6539702	0.142
PV Curb weight	-1.501932	0.092	-0.7341008	0.823
PV overall Width	3.228802	0.022	8.045037	0.054
PV Wheel Base	-0.3185652	0.425	2.968626	0.090
PV Front Overhang	-1.051364	0.291	-8.467318	0.098
PV Rear Overhang	0.2296607	0.719	-6.056789	0.012
AADT/1000/Lane	-0.0294631	0.572	-0.1107404	0.569
Million Miles Travelled	-0.0298083	0.034	-0.0156096	0.652
Road Class 2 Lane	-0.3482433	0.792	-1.422949	0.715
Road Class Multilane Divided	-0.9090404	0.284	-1.839204	0.521
Road Class Multilane Undivided	-0.6710272	0.663	-35.46266	1.000
Lighting Condition Dark-Lighted	0.5184445	0.683	-31.30279	1.000
Lighting Condition Dark- No Light	1.255212	0.047	-1.48667	0.502
Constant	-14.57056	0.037	-0.07185	0.998

Table C- 19 Final model for drivers age 66 & older

VARIABLES	LEVEL 3/1			LEVEL 5/1		
	Logit Coeffcient	Odds ratio	P value	Logit Coeffcient	Odds ratio	P Value
2-Lane	0.654453	1.92	0.083	NA	NA	NA
Dark-not lighted	1.401242	4.06	0.012	NA	NA	NA
Turn	0.713874	2.04	0.097	NA	NA	NA
Weight (PV)	-1.296227	0.27	0.035	-3.07955	0.05	0.067
Width (PV)	3.246993	25.71	0.005	6.408133	606.76	0.014
Front-overhang	-1.365714	0.26	0.008	NA	NA	NA
Rear-overhang	NA	NA	NA	-3.204596	0.04	0.008
Constant	-11.16149	NA	0.018	-13.81054	NA	0.212

Table C- 20 Significant variables across models

Categories	Variables	MODELS						
		General	Male	Female	<=25	26-45	46-65	>=66
Driver/ Occupants Characteristics	1. Drivers age	X	X	NA	NA	NA	NA	NA
	2. Drivers gender	X	NA	NA	NA	X	X	NA
	3. # Of occupants	X	X	X	X	X	X	NA
Crash Environment at the Crash Location	4. AADT/lane	X	NA	X	NA	X	NA	NA
	5. Speed limit	X	X	X	NA	NA	X	NA
	6. Dark not lighted	X	X	X	X	NA	X	X
	7. 2-Lane	X	NA	X	NA	NA	X	X
	8. Multilane undivided	NA	NA	NA	NA	NA	X	NA
	9. Multilane divided	X	X	NA	X	X	X	NA
	10. Median divided	NA	X	NA	X	NA	NA	NA
Crash Characteristics	11. Travel speed	X	X	X	X	X	X	NA
	12. Head on	X	X	X	X	X	X	NA
	13. Rear end	X	X	X	X	X	X	NA
	14. Angle	X	X	X	X	X	X	NA
	15. Turn	X	X	X	X	NA	X	X
Vehicle Characteristics	16. Weight	X	X	X	X	X	NA	X
	17. Front-overhang	X	X	NA	X	NA	NA	X
	18. Rear-overhang	X	X	X	X	X	NA	X
	19. Height	NA	NA	NA	X	NA	NA	NA
	20. Width	X	X	NA	NA	NA	NA	X
Total number of significant variables	20	17	15	12	13	10	12	7

REFERENCES

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

Abdel-Aty, M. & Keller, J. (2005). Exploring the overall and specific crash severity levels at signalized intersections. *Accident Analysis and Prevention, 37, (pp 417-425).*

Abdelwahab, H.T. & Abdel-Aty, M. (2002). Investigating Driver Injury Severity in Traffic Accidents Using Fuzzy ARTMAP, *Computer-Aided Civil and Infrastructure Engineering, 17, (pp 396-408).*

Abdel-Aty, M. & Abdelwahab, H.T. (2002). Configuration Analysis of Two-Vehicle Rear-end Crashes, *Transportation Research Board annual meeting (CD-ROM, TRB 2003).*

Aitchison, J. & Silvey, S.D. (1957). The generalization of probit analysis to the case of multiple responses. *Biometrika, 44, (pp 131-140).*

Akaike, H. (1973). Information theory and an extension of the maximum likelihood Principle. *Second International Symposium on Information Theory. Budapest: Akademiai Kiado.*

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

Angel, A. & Hickman, M. (2009). Estimating Occupant Injury Severity in Two-Vehicle Crashes, *Transportation Research Board annual meeting (CD-ROM, TRB 2009).*

Auto Check. http://www.autocheck.com .*(First accessed on 02/09/2009).*

Ben-Akiva, M., & Lerman, S.R. (1985). Discrete Choice Analysis (pp 20-167). *The MIT Press, Cambridge, Massachusetts.*

Blincoe, L., Seay, A., Zaloshnja, E., Miller, T., Romano, E., Luchter, S., Spicer, R. (2002). The Economic Impact of Motor Vehicle Crashes, 2000. http://www-nrd.nhtsa.dot.gov/Pubs/809446.PDF (accessed on 01/22/2009).

Brant, R. (1990). Assessing proportionality in the proportional odds model for ordinal logistic regression. *Biometrics*,46, 1171-1178.

Cars Direct. http://www.carsdirect.com/research/model-year *(First accessed on 3/17/2009).*

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

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

Delen D., R. Sharda, and M. Bessonov (2006). Identifying significant predictors of injury severity in traffic accidents using a series of artificial neural networks, *Accident Analysis and Prevention, 38, (pp 434-444).*

Department of Transportation, Research and Innovative Technology Administration (RITA), Bureau of Transportation Statistics. http://www.bts.gov/publications/national_transportation_statistics/html/table_02_19.html*(Accessed on 12/16/2008).*

Department of Transportation, Highway Traffic Safety Administration, Automotive Sampling System, General Estimates System (GES), *Analytical User's Manual 1988-2007.*

Donnell, E.T. & Mason, J.M. (2004). Predicting the Severity of Median-Related Crashes in Pennsylvania using Logistic Regression, *Transportation Research Board annual meeting (CD-ROM, TRB 2004).*

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

Farmer, C.M., Braver, E.R., & Mitter, E.T. (1997). Two-vehicle side impact crashes: The relationship of vehicle and crash characteristics to injury severity, *Accident Analysis and Prevention, volume 29, No. 3, (pp 399-406).*

Hartog, J., Rider, G., & Visser, M. (1994). Allocation of individuals to job levels under rationing. *Journal of applied econometrics, 9, (pp 437-451).*

Hedström, P. (1994). Local employment contexts and job attainment in Swedish manufacturing industry. *Work and Occupations, 21, (pp 355-368).*

Highway Safety Information System. http://www.hsisinfo.org. (*First accessed on 11/18/2008).*

Highway Safety Information System. http://www.hsisinfo.org/pdf/nc_vol1_06.pdf (*First accessed on 01/15/2009).*

Hill, J.D., & Boyle, L.N. (2006). Assessing the relative risk of severe injury in automotive crashes for older female occupants. *Accident Analysis and Prevention, volume 38, (pp 148-154).*

Holdridge, J.M., Shankar, V. N., & Ulfarsson, G.F. (2005). The crash severity impacts of fixed roadside objects, *Journal of Safety Research volume 36, (pp 139-147).*

Kim, K., L. Nitz, J. Richardson, and L. Li (1995). Personal and behavioral predictors of automobile crash and injury severity, *Accident Analysis and Prevention, volume 27, n 4, (pp 469-481).*

Khattak, A.J., Pawlovich, M.D., Souleyrette, R.R., & Hallmark, S.L. (2000). Factors related to more severe older driver traffic crash injuries, *Journal of Transportation Engineering ,volume 128, no. 3, (pp 243-249).*

Khattak, A. (2002). Injury severity in multi-vehicle rear-end crashes. *Transportation Research Board annual meeting (CD-ROM, TRB 2000).*

Khorashadi, A., Niemeier, D., Shankar, V. & Mannering, F. (2005). Differences in rural and urban driver-injury severities in accidents involving large-trucks: An explanatory analysis, *Accident Analysis and Prevention, 37, (pp 910– 921).*

Kockelman, K.M., Y.J. Kweon (2002). Driver injury severity: an application of ordered probit models, *Accident Analysis and Prevention, 34, (pp 313– 321).*

Kweon, Y.J., Kockelman, K.M., (2003). Overall injury risk to different drivers: combining exposure, frequency, and severity models, *Accident Analysis and Prevention, 35, (pp 441– 450).*

Long, J. S. (1997). Regression Models for Categorical and Limited dependent Variables. *Thousand oaks, Ca: Sage.*

Long, J. S., & Freese, J. (2006). Regression Models for Categorical dependent Variables using STATA). *College Station, Texas.*

Ma, J. & Kockelman, K.M., (2004). Anticipating injury and death: Controlling for new variables on southern California highways. *Transportation Research Board annual presentation(2004).*

Malyshkina, N., & Mannering, F. (2008). Analysis of the effect of speed limit increases on accident injury severities. *Transportation Research Board annual meeting (CD-ROM, TRB 2008).*

Marcus, A., & Greene, W.H. (1985). The determinants of rating assignments and performance. *Working paper, No CRC528. Alexandria, VA, Center for naval analysis.*

McKelvey, R.D., and Zavonia, W. (1975). A statistical model for the analysis of ordinal level dependent variables. *Journal of mathematical sociology, 4, 103-120.*

Meng, X., & Miller, P. (1995). Occupational segregation and its impact on gender wage discrimination in China's rural industrial sector. *Oxford Economic Papers, 47, (pp 136-155).*

Milton, C.J., Shankar, V.N. & Mannering, F.L. (2007). Highway accident severities and the mixed logit model: An explanatory empirical analysis. *Accident Analysis and Prevention, 40(2008), (pp 260– 266).*

National Safety Council, 1990. Manual on Classification of Motor Vehicle Traffic Accidents, fifth ed. *(ANSI D-16.1-1989). Itasca, IL.*

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

O'day J. Accident data quality. National Cooperative Highway Research Program: Synthesis of Highway Practice 192. *Transportation Research Board. Washington DC: national Academy Press; 1993.*

Powers, A.D., & Xie, Y. (2000). Statistical Methods for Categorical Data Analysis. *San Diego, Ca: Academic Press.*

Rifaat, S.M. & Chor, C.H. (2005). Analysis of severity of single-vehicle crashes in Singapore. *Transportation Research Board annual meeting.*

Renski, H., Khattak, A.J. & Council, M.F. (1999). Impact of speed limit increases on crash injury severity: Analysis of single-vehicle crashes on North Carolina interstate highways. *Transportation Research Board 78th annual meeting.*

Rumar, K., L. Evans, and R.C. Schwing (1985). The Role of Perceptual and Cognitive Filters in Observed Behavior, *Human behavior and traffic safety, Plenum press, New York, N.Y. (pp 151-165).*

Schneider IV, W.H., Savolainen, P.T., Zimmerman, K. (2009). An examination of driver injury severity resulting from single-vehicle crashes along horizontal curves on rural two-lane highways. *Transportation Research Board annual meeting.*

Ulfarsson, G.F., & Mannering, F. (2004). Differences in male and female injury severities in sport-utility vehicle, minivan, pickup and passenger cars accidents. *Accident Analysis and Prevention, Vol. 36, (pp 135-147).*

Visual Statement inc. http://www.visualstatement.com/Products/components/vehicle-spec-db.aspx. *(First accessed on 02/09/2009).*

Wang, Z., Chen, H, & Lu, J.J. (2009). Exploring Impacts of Factors Contributing to Injury Severity at Freeway Diverge Areas, *Transportation Research Board annual meeting (CD-ROM, TRB 2009).*

Wang, X., & Kockelman, K.M. (2005). Occupant Injury Severity using a Heteroscedastic Ordered Logit Model: Distinguishing the Effects of Vehicle Weight and Type, *Transportation Research Board annual meeting (CD-ROM, TRB 2005).*

Wang, X., & Abdel_Aty, M. (2008). Left-turn crash injury severity analysis using partial proportional odds models: significant factors and varying effects, *Transportation Research Board annual meeting (CD-ROM, TRB 2008).*

Washington, S., & Karalaftis, M., & Mannering, F. (2003). Statistical AND Econometric Methods FOR Transportation Data Analysis. CHAPMAN & HALL/CRC.

Winship, C., & Mare, R. D. (1984). Regression models with ordinal variables. *American Sociological Review, 49, (pp 512-525).*

William L. Carlson (1979). Crash Injury Prediction Model. Accident Analysis and Prevention, Vol. 11, (pp 137-153).

Xie, Y., Y. Zhang, and F. Liang (2009). Crash Injury Severity Analysis Using Bayesian Ordered Probit Models, *Journal of Transportation Engineering, Vol. 135, No 1, (pp 18-25).*

Zaloshnja, E., & Miller T. (2004). Costs of large truck-involved crashes in the United States. *Accident Analysis and Prevention 36, (pp 801-808).*

VITA

Graduate College
University of Nevada, Las Vegas

Ahmad Kouhpaenejad

Degrees:
Bachelor of Science in Economics, 1977
College of Economics and Banking

Bachelor of Science in Industrial Technology (Mfg), 1981
California State University

Master of Science in Industrial and Systems Engineering (OR), 1982
University of Southern California

Dissertation Title: Statistical Modeling and Analysis of Injury Severity Sustained by Occupants of Passenger Vehicles Involved in Crashes with Large Trucks

Dissertation Examination Committee:
Chair, Dr. Mohamed S. Kaseko, Ph. D.
Committee Member, Dr. Moses Karakouzian, Ph. D.
Committee Member, Dr. Nader Ghafoori, Ph. D.
Committee Member, Dr. Alexander Paz, Ph.D.
Committee Member, Graduate Faculty Representative, Dr. Ashok Singh, Ph. D.