CHARACTERISTICS OF INJURY AND FATALITY OF RUN-OFF-ROAD CRASHES ON OHIO ROADWAYS

Thesis

Submitted to

The School of Engineering of the

UNIVERSITY OF DAYTON

In Partial Fulfillment of the Requirements for

The Degree of

Master of Science in Civil Engineering

By

Omar Eid Almutairi

Dayton, Ohio

August, 2013
CHARACTERISTICS OF INJURY AND FATALITY OF RUN-OFF-ROAD CRASHES
ON OHIO ROADWAYS

Name: Almutairi, Omar

APPROVED BY:

Deogratias Eustace, Ph.D., P.E., PTOE
Advisory Committee Chairperson
Associate Professor, Department of
Civil and Environmental Engineering
and Engineering Mechanics

Peter Hovey, Ph.D.
Committee Member
Associate Professor, Department
of Mathematics

Gary Shoup, P.E.
Committee Member, Senior Engineer
Montgomery County Engineering Department

John G. Weber, Ph.D.
Associate Dean
School of Engineering

Tony E. Saliba, Ph.D.
Dean, School of Engineering
& Wilke Distinguished Professor
ABSTRACT

CHARACTERISTICS OF INJURY AND FATALITY OF RUN-OFF-ROAD CRASHES ON OHIO ROADWAYS

Name: Almutairi, Omar
University of Dayton

Advisor: Dr. Deogratias Eustace

A run-off-road (ROR) crash or a roadway departure crash is a non-intersection crash which occurs after a vehicle crosses an edge line or a center line (i.e., leaves its designated traveled way and in the process the vehicle collides with a non-traversable obstacle or another vehicle travelling in the opposite direction or hits a pedestrian, or the vehicle overturns. The main objective of this thesis study was to determine the factors that contribute significantly to the levels of injury severity when ROR crashes occur. This study used a 5-year crash data for years 2008 - 2012 obtained from the Ohio Department of Public Safety. The decision tree model in conjunction with generalized ordered logit model was used to investigate characteristics of injury and fatality of run-off-road crashes in Ohio.

The decision tree modeling was used for exploratory data analysis identified eight factors that explain a large amount of the variation in the response variable, injury
severity. These important predictors for injury severity include road condition, run-off-road (ROR) crash types, posted speed limit, vehicle type, gender, alcohol-related, road contour, and drug-related. Also, complex interactions between parameters were identified. The results from the generalized ordered logit regression show that the following are significant factors in increasing the likelihood of ROR injury severity levels: alcohol and drugs use, curves and grades, female victims, overturn/rollover crashes, ROR crashes on dry roadway surfaces. Additionally, buses, truck, and emergency vehicles, and ROR crashes on roadways with posted speed limits of 40 mph or higher increase the probability of injury severity.
ACKNOWLEDGEMENTS

My first thanks go to the Almighty God, without whose provisions and guidance, my participation in this program of study would have been futile. I would like to express my heartfelt gratitude to my principal advisor, Dr. Deogratias Eustace, who read, criticized and provided necessary support and encouragement to accomplish this research. To Dr. Peter Hovey, I extend special thanks for his immense contribution and support, particularly, in the area of statistics applied in this research and Mr. Gary Shoup for his thorough reading of my thesis manuscript and his technical advice are highly recognized. I count myself blessed to have all of you guiding my thesis. To friends and classmates that contributed in diverse ways to my success in this program, I say a big thank you. Finally, my thanks go to my family, both home and abroad for their continued support, prayers, contributions and bearing with me throughout this program of study.
# TABLE OF CONTENTS

ABSTRACT .................................................................................................................. iii

ACKNOWLEDGEMENTS ............................................................................................. v

TABLE OF CONTENTS .............................................................................................. vi

LIST OF FIGURES ...................................................................................................... x

LIST OF TABLES ......................................................................................................... xii

CHAPTER ONE - INTRODUCTION ............................................................................. 1

1.1 Introduction ........................................................................................................... 1

1.2 Problem Statement .............................................................................................. 2

1.3 Research Objectives ........................................................................................... 3

1.4 Organization of the Thesis .................................................................................. 4

CHAPTER TWO - LITERATURE REVIEW ................................................................. 5

2.1 Introduction .......................................................................................................... 5

2.2 Driver Behavioral Factors .................................................................................. 6

2.2.1 Alcohol Use .................................................................................................... 6

2.2.2 Fatigue and Drowsiness ................................................................................ 7

2.2.3 Speeding ....................................................................................................... 8
3.3.2.1 Introduction................................................................. 27
3.3.2.2 Pruning and Validation ............................................... 29
3.3.2.3 Statistical Analysis...................................................... 30
  3.3.2.3.1 Node Splitting Criteria ........................................... 30
  3.3.2.3.2 Tests for Goodness of Fit ....................................... 32
  3.3.2.3.3 Scatter Plot of Actual by Predicted Values .................... 33
  3.3.2.3.4 Receiver Operating Characteristics (ROC) Curve .......... 34
3.3.3 Generalized Ordered Logit Modeling.................................... 36

CHAPTER FOUR - RESULTS .......................................................... 40
4.1 Introduction ........................................................................ 40
4.2 Descriptive Results of Run-off-Road-Related Traffic Crashes............ 41
4.3 Results of Decision Tree Modeling for Predictor Parameter Screening .... 46
  4.3.1. General ....................................................................... 46
  4.3.2 Interaction of Road Condition and ROR Crash Type .................. 52
  4.3.3 Interaction of Road Condition, ROR Crash Type, Vehicle Type and Alcohol 53
  4.3.4 Interaction of Road Condition, ROR Crash Type and Posted Speed Limit .... 54
  4.3.5 Interaction of Posted Speed Limit, Vehicle Type and Road Contour .......... 54
  4.3.6 Interaction of Posted Speed Limit, Vehicle Type, and Gender ............... 55
  4.3.7 Interaction of Speed Limit, Vehicle Type, Gender, and Drug Related .... 55
  4.3.8 Interaction of Alcohol Use, Vehicle Type, and Speed Limit ............... 55
4.3.9 Interaction of Road Condition, Gender, and ROR Crash Type .................. 56

4.3.10 Interaction of Road Condition, Gender, and Posted Speed Limit ............... 56

4.3.11 Interaction of Road Condition, Gender, and Alcohol Use ...................... 56

4.4 Results of Generalized Ordered Logit Model for Predictor Parameter Estimation 57

CHAPTER FIVE - CONCLUSIONS AND RECOMMENDATIONS ......................... 63

REFERENCES .................................................................................................. 69
LIST OF FIGURES

Figure 2.1: Roadway Departure Crashes Contribution to 2011 Fatal Crashes in The United States (FHWA, 2013) .......................................................................................................................... 14

Figure 2.2: 2011 First Harmful Events in Roadway Departure Fatal Crashes in The United States (FHWA, 2011) .......................................................................................................................... 15

Figure 3.1: The Process of Creating a Joint Crash-Unit-People File ........................................ 21

Figure 3.2: An Example of A Flow-Chart Structure of a Decision Tree .............................. 29

Figure 3.3: The Definition Pertaining to The Parent Node and Its Two Children Nodes 31

Figure 3.4: An Example of a Scatter Plot from Jmp Software Output ................................ 34

Figure 3.5: An Example of a ROC Curve For a Three-Level Response Variable from Jmp Software Output .......................................................................................................................... 35

Figure 4.1: Rates of Fatality and Injury Due to Run-off-Road Crashes by Road Contour Type ........................................................................................................................................ 42

Figure 4.2: Rates of Fatality and Injury Due to Run-off-Road Crashes by Alcohol Use Status ........................................................................................................................................ 43

Figure 4.3: Number of Fatalities Due to Run-off-Road Crashes by Day of The Week ... 44

Figure 4.4: Number of Total Injuries Due to Run-off-Road Crashes by Day of The Week ........................................................................................................................................ 44

Figure 4.5: Number of Vehicles Involved In Run-off-Road Crashes by Day of The Week ........................................................................................................................................ 45
Figure 4.6: Rates of Fatality and Injury Due to Run-off-Road Crashes by Road Condition Type........................................................................................................................................... 46
Figure 4.7: Split History for Training and Validation Datasets.............................................. 48
Figure 4.8: The Column Contributions Report for Selecting Significant Predictor Variables......................................................................................................................................................... 49
Figure 4.9: ROC Curve for Training Sample Dataset........................................................... 50
Figure 4.10: ROC Curve for Validation Sample Dataset ...................................................... 50
Figure 4.11: Final Decision Tree Diagram after 20 Splits.................................................... 51
LIST OF TABLES

Table 3.1 Description of Crash and Vehicle Characteristics ........................................ 25
Table 3.2 Description of Human and Driver Characteristics ...................................... 25
Table 3.3 Description of Roadway and Environmental Characteristics .................... 26
Table 4.1 Parameter Estimate Results from STATA Output ....................................... 58
CHAPTER ONE

INTRODUCTION

1.1 Introduction

The Federal Highway Administration (FHWA) defines a run-off-road (ROR) crash, also known as a roadway departure crash as a non-intersection crash which occurs after a vehicle crosses an edge line or a center line, or otherwise leaves the traveled way (FHWA, 2013). In such a crash, the vehicle may collide with a non-traversable obstacle or another vehicle travelling in the opposite direction or hit a pedestrian. An ROR crash may also end up with the vehicle overturning. A significant portion of ROR crashes are fatal. The National Highway Traffic Safety Administration (NHTSA, 2013a) reports that in 2011, 32,367 people were killed and 2,217,000 were injured in 5,338,000 police-reported motor vehicle traffic crashes in the United States. Furthermore, the Federal Highway Administration (FHWA)’s Roadway Departure Safety Program reports that 51% of all fatal crashes that occurred in the United States in 2011 involved the run-off-road crash types (FHWA, 2013). This is clear evidence that the likelihood of an ROR crash becoming fatal is very high, which makes these types of motor vehicle crashes be one of the major public health problems. Several factors contribute to whether an ROR traffic crash becomes fatal. These include roadside features to be hit by an errant vehicle, vertical and horizontal curving of the road, high speed of an errant vehicle, width of the
shoulder, and the behavior of the driver. Based on the above mentioned factors, the severity of an ROR traffic crash can range from one of the following: fatal injury crash, injury crash, and property damage only (PDO) crash.

1.2 Problem Statement

Many factors may be responsible for causing a vehicle to leave its designated travel lane. These factors include driver behavior, vehicular defects, and the condition of the roadway. After a vehicle leaves the roadway travel lane, factors such as roadside features play a major role in the severity of the traffic crash. As noted above, an ROR crash occurs either when an errant vehicle leaves its designated travel lane to the right and encroaches the roadside or when it leaves the travel lane to the left and encroaches the travel lane(s) designated for the opposite traffic. The roadside encroachment can lead an errant vehicle to collide with non-traversable obstacles (e.g., trees, mail boxes, utility poles, etc.), steep roadside slopes, or fall into inadequate configuration of guardrails or deep ditches. The presence of these various roadside features usually increases the level of injury or cause deaths. The opposite lane(s) encroachment (crossing the centerline or median) can lead an errant vehicle to collide with other vehicles traveling on the opposite direction with a possibility of causing head-on, side-swipe opposite direction, and angle type of collisions.

The Ohio Department of Transportation (ODOT) has been vigorously campaigning for the reduction of roadway departure crashes in the state of Ohio. ODOT (2013) reports that in the period of 2007-2011, in the state of Ohio there were a total of 292,446 ROR crashes that resulted into 2,920 fatalities and 124,491 injuries. In the state
of Ohio for the same time period, about 65% of all motor vehicle fatal crashes involved vehicles that left the roadway and hit fixed objects such as trees and utility poles (ODOT, 2013) Consequently, there is a need to identify important factors that contribute highly or increase the likelihood of the occurrences of ROR crashes in the state of Ohio, as they threaten the lives of drivers and passengers, and the economy of the state. This research was aimed at identifying factors and conditions that influence ROR crashes on Ohio roadways, and to determine which factors pose major influences in increasing the occurrences of fatalities and serious injuries.

1.3 Research Objectives

The main objective of this thesis study was to determine the factors contributing to fatalities and serious injuries when ROR crashes occur in the state of Ohio. Based on this objective, a 5-year crash data for years 2008-2012 was obtained from the Ohio Department of Public Safety (ODPS). This time period was chosen because during the time of this study it was the period that encompassed the most recently available 5-year traffic crash dataset. This research was set to develop a statistical model that quantifies the level of contribution by environmental, traffic, geometric, and driver behavior related factors that can be easily extracted from available crash data records. Traffic-related injuries are usually categorized into fatal injury, incapacitating injury, non-incapacitating injury, possible/invisible injury, and no injury. In addition, a traffic crash severity is normally categorized based on the injury severity sustained by the most injured person in that crash, which may be named as a fatal crash, an injury crash, or property damage only (PDO) crash. A fatal crash is defined as a crash in which at least one death occurs, an
injury crash is defined as a crash in which at least one injury occurs, and a property
damage crash is defined as a crash in which no injury occurs but only involving damage
to a vehicle. This study was designed to highlight factors that have an important effect in
causing or contributing to ROR crashes likely to cause fatalities and injuries in the state
of Ohio.

1.4 Organization of the Thesis

The rest of the thesis report is organized into four chapters. Chapter two presents a
literature review that provides a summary of related research efforts performed by
previous researchers. Chapter three outlines the methodology and data collection. It
describes the data and provides an explanation of the variables. Additionally, this chapter
discusses the methodology used in this study. Chapter four contains the results and
discussion of the results. Finally, Chapter five presents conclusions and
recommendations.
CHAPTER TWO

LITERATURE REVIEW

2.1 Introduction

Traffic crashes are usually caused by a number of factors. Crash factors are conditions or actions that accompany a crash, whether or not they are determined to have contributed to the occurrence of the crash. Factors that are determined to have actually contributed to the occurrence of the crash are considered causes. A crash may be attributed to more than one cause. Crash factors are usually broadly grouped into the following categories:

- Driver behavior factors: driving errors made by the driver, speed choices, driver age, gender, other driver’s states such as intoxication, drowsiness, etc.
- Geometric factors: width of lanes, width of shoulders, road curvature, number of lanes, other features and obstacles, etc.
- Environmental factors: weather condition, road condition, light condition, time of day, day of the week, etc.
- Traffic factors: collision type, speed related, traffic volume, speed limit, etc.

Researchers usually use various statistical models to examine which of the above factors play significant role in causing motor vehicle traffic crashes and injuries. The literature
review indications that various models and variables have been used in analyzing causes and severity of crashes and injuries.

2.2 Driver Behavioral Factors

The behavior of driver is the foremost factor in motor vehicle traffic crash problems. The discussion below just includes some of the examples of behavioral factors that contribute to crashes.

2.2.1 Alcohol Use

When performing a literature review on driving under the influence (DUI), Pisani (1985) found that approximately one-third of motor vehicle related injuries and approximately 50% of motor vehicle related deaths were related to alcohol use. In 2007, roadside surveys conducted in 48 states found out that 4.8% of the vehicles on the highways during the late nighttime were driven by drivers with illegal blood alcohol concentrations (BACs) (Compton and Berning, 2009). Out of 1030 motor vehicle traffic crash-related deaths that occurred in the state of Ohio for the year 2010, 341 (33%) were caused by drivers who were determined to be alcohol impaired (NHTSA, 2012). In comparison to the entire nation, in 2011 there were about 10,000 deaths caused by traffic crashes involving drivers with blood alcohol concentrations (BAC) equal to 0.08 grams per deciliter (g/dL) or higher, which was equivalent to 31% of all motor vehicle traffic deaths for that year (NHTSA, 2013a).

Available statistics support the notion that alcohol impairment is one of the major contributors in the occurrences of motor vehicle fatalities and major injuries. For
example, in 2011 alcohol-impaired drivers contributed for 31% of the total motor vehicle traffic fatalities in the United States (NHTSA, 2012). At the state level, in 2011, there were 1016 fatalities for which 316 (31%) were accounted for alcohol-impaired drivers (NHTSA, 2012). Alcohol use has been surfacing at the top of the list of significant factors affecting the occurrences of fatal and severe injuries in motor vehicle traffic crashes in a number of previous studies (e.g., Wang et al., 2009; Roy and Dissanayake, 2011; Mergia et al., 2013)

2.2.2 Fatigue and Drowsiness

A study by Pack et al. (1995) examined approximately 4300 crash reports from 1990 to 1992 in which the driver was determined to have fallen asleep. In the Pack et al. (1995) study, approximately 78.5% of these crashes were ROR crashes. It was found that age of the driver was a significant factor in the occurrence of such types of traffic crashes and they reached peak likelihood when the driver age was 20 years. Additionally, the most common times of day when such types of crashes occurred were between midnight and 7:00 am and during the middle of the afternoon. Sagberg (1998) surveyed 9200 drivers in Norway where he asked them about their last accident reported to their insurance companies. Sagberg’s study found that 3.9% of all accidents were caused by sleep or fatigue. A number of studies have studied and found that fatigue and drowsiness are among of significant factors that contribute to the occurrences of traffic crashes especially ROR crashes (e.g., Roy and Dissanayake, 2011; Peng and Boyle, 2012)
2.2.3 Speeding

NHTSA defines a motor vehicle traffic crash as a speeding-related crash if the driver was charged with a speeding-related offense or if an officer indicated that racing, driving too fast for conditions, or exceeding the posted speed limit was a contributing factor in the occurrence of the crash (NHTSA, 2013a). A study by Aljanahi et al. (1999) report that the likelihood of a traffic crash is positively correlated with the mean speed and the variability of the speeds in the traffic stream. Hauer’s (2009) study establishes that there is no clear relationship between the probability of accident involvement and speed. However, there is strong evidence that the damage will be more as the pre-crash speed increases (Hauer, 2009).

Speeding has been one of the most frequently cited factors contributing to motor vehicle traffic crashes and fatalities (e.g., Roy and Dissanayake, 2011; Peng and Boyle, 2012; Mergia et al., 2013). Statistics show that in 2011, speeding was a contributing factor in 30% of all fatal crashes in the United States where 9,944 (31%) fatalities resulted due to speeding-related motor vehicle traffic crashes (NHTSA, 2013a). At the state level, in 2011, there were 299 speeding-related fatalities, which is equivalent to 29% of all fatalities in the state (NHTSA, 2013b).

2.2.4 Age of the Driver

The age of a driver has also been considered an important factor with regard with the occurrence of traffic crashes and especially young drivers are highly associated with a higher rate of causing fatal crashes and fatalities. A study conducted in Hawaii by Richardson et al. (1996) found that young drivers were more likely to experience rollover
crashes and older drivers were more likely to experience side swipe and rear-end collisions. By using the 1993-95 road traffic crash data in Indonesia, Achwan and Rudjito (1999) found that drivers aged 21 to 30 years were the most likely age group to be injured or killed in motor vehicle traffic crashes where this age group accounted for 44% of all traffic casualties, 20% of all traffic deaths, and 24% of all major and minor injuries. In a review of motor vehicle traffic crash reports for years 1993 through 1995 conducted by Stamatiadis et al. (1999) found that drivers younger than 25 years old and those older than 65 years were more likely to crash on low traffic rural roads in Kentucky and North Carolina when compared with middle aged drivers. They also note that older people were also more likely to suffer injuries than younger people in similar traffic crash situations.

Statistics show that in 2011, young drivers aged 15-20 years old drivers were overrepresented in fatal traffic crashes where they constituted 10% of all the drivers involved in fatal crashes while they just accounted for only 6% of all registered drivers in the United States (NHTSA, 2013a). Besides less driving experience, young drivers are also most likely to be over speeding, using alcohol, and not using restraint devices (NHTSA, 2013a). For the state of Ohio in 2011, young drivers under the age of 21 accounted for about 16% of all fatal traffic crashes (NHTSA, 2013b).

### 2.2.5 Gender of Person

The consequences of gender of a person in traffic crashes have been reported depending on the objectives of the respective studies. For example, gender can be investigated as a driver and its probability in causing or being involved in traffic crashes in general or it
can be reported based on its propensity to sustain injuries in traffic crashes regardless whether the person is a driver or just a vehicle occupant. A study conducted by Massie and Campbell (1993) mention that men experience more fatal crashes than women. In a study by Yan and Radwan (2006) report that men are more likely to get involved in rear-end crashes than women based on 2001 Florida crash database. Vehicles driven by male drivers have a higher probability of being involved in run-off-road crashes than vehicles driven by female drivers (Liu and Subramanian, 2009). A study conducted by Mergia et al. (2013) using the 2006-2009 Ohio database found that females as vehicle occupants (drivers or passengers) have elevated chances of sustaining severe injuries than males when involved in traffic crashes at freeways’ merging junctions.

2.3 Highway Geometric Factors

2.3.1 Width of Lanes and Shoulders

A study conducted by Zegeer et al. (1988) that looked at nearly 5000 miles of two-lane highways found that ROR crashes, head-on collisions, and side-swap collisions were related to the road width parameters. They also report that greater width of lanes, greater width of shoulders, less hilly terrain, lower vehicle density, and a smoother roadside were correlated with fewer traffic crashes. Zegeer et al. (1994) compiled traffic crash results of two-lane roads from seven U.S. states. That study determined that greater width of roads, better road surface finish, fewer hills, and fewer side roads were correlated with a lower traffic crash rates. A study was conducted to identify whether it is safer to widen lane or shoulder for a fixed total width, using geometric, traffic, and crash data for almost 52,000
miles of roadways from Pennsylvania and Washington State and it was found that there is a slight benefit to increase lane width for a fixed total width (Gross et al., 2009). However, benefits were noticed if total paved width, lane width, and shoulder width were increased individually. In a number of studies where shoulder data is available for use in the crash model, it was found to be a significant factor affecting traffic injury severities (e.g., Wang et al. 2009; Wang et al., 2011; Yang et al., 2011).

2.3.2 Vertical Alignment (Grades)

A study conducted by Shankar et al. (1995) found that rear-end crashes were more likely when the road grade was greater than 2%. This study examined 2,225 crashes between 1988 and 1993 on a 61 km long section of I-90 near Seattle, Washington. They had the information on the road grade for this road section, and were able to analyze the frequency of various types of crashes in light of the road geometry. A study by Ogden (1996) also found that steep roadways contribute to a greater likelihood of traffic crashes. In addition, Ogden points out that vertical curves in the form of sags were not a major problem because they did not obscure or limit driver’s vision, but hill crests, on the other hand, limited the sight distance. Crashes were also associated with downhill sections of roads (Ogden 1996). A number of previous research efforts have identified roadway grade as a significant factor in the determination of injury/crash severity (e.g., Wang et al., 2009; Zhu et al., 2010; Wang et al., 2011).
2.3.3 Horizontal Alignment (Curves)

Highway curves have been identified as one of the most significant geometric factors that affect fatal and injuries crashes on highways (Wang et al., 2009; Zhu et al., 2010; Eustace et al., 2011; Roy and Dissanayake, 2011). For example, a study by Lamm and Choueiri (1987) found that a significant majority (more than 70%) of traffic crashes that occurred on curved sections of roads were either fatal or caused injuries. In addition, they note that for crashes that occurred on curved sections, approximately half of them were associated with ice or wet road surfaces. This is despite the fact that far fewer vehicle miles are logged in such conditions.

2.3.4 Number of Lanes

In a study conducted by Kononov et al. (2008) it was found that increasing the number of lanes causes an increase in number of vehicles changing lane as traffic jam (number of vehicles per mile) increases, which results in increased traffic crashes. The results from a study by Mergia et al. (2013) show that increased number of lanes on freeway mainlines and ramps increase the chance of severe injuries at freeways’ merging locations. On other hand, a study by Liu and Subramanian (2009) reports that roadways with fewer lanes (one or two lanes) tend to have a relatively higher likelihood of being locations of run-off-road crashes when compared with roadways with more than two lanes.
2.4 Environment Factors

Environmental factors that may influence the occurrence and severity of traffic crashes such as light condition, weather condition, day of the week, time of day, and roadway condition have been extensively included in various previous studies (e.g., Shankar et al., 1995; Abdel-Aty and Keller, 2005; Wang et al., 2009; Eustace et al., 2011; Roy and Dissanayake, 2011; Mergia et al., 2013). A study by Parsaud and Mucsi (1995) found that single vehicle crashes were more likely to occur during the night, and multiple vehicle crashes were more likely to occur during the daytime. A study by Livine et al. (1995) evaluated the frequency of crashes during various times and dates in 1990 in Honolulu. This study determined that more crashes occurred on Fridays and Saturdays. According to a study conducted in Indonesia by Achwan and Rudjito (1999), 83% of traffic collisions occurred on good and dry road surfaces, while 6% occurred on wet surfaces. Additionally, approximately 65% of collisions happened during daytime hours and 19% happened during nighttime.

2.5 Traffic Factors

2.5.1 Traffic Volume

A number of studies have found a correlation between traffic volume, especially annual average daily traffic (AADT), with increased traffic crashes. A study by Hadi et al. (1995) found a positive correlation between likelihood of collision and a greater traffic volume on all types of highways. A study by Kononov et al. (2008) found that high AADT on a multilane highway is correlated with high crash frequency.
2.6 Run-off-Road Crashes

Run-off-road (ROR) crashes also known as roadway departure crashes normally make up the majority of motor vehicle traffic fatalities because they mainly result into severe crashes (FHWA, 2013). For example, in 2011, nationwide, there were 15,307 ROR fatal motor vehicle crashes, which accounted for 51% of all fatal motor vehicle crashes in the United States of America and resulting into 16,948 fatalities (FHWA, 2013). Figure 2.1 shows the distribution of 2011 fatal motor vehicles crashes by major types of ROR crashes (FHWA, 2013). It is reported that in the state of Ohio for the period 2007-2011 there was a total of 2,920 motor vehicle fatalities related to ROR crashes (ODOT, 2013) while the total number of motor vehicle traffic fatalities in the same time period in Ohio was 5,564 (NHTSA, 2013b), which means that ROR crashes accounted for 52.5% of all traffic-related fatalities in the state. In the state of Ohio most of ROR crashes occur on two-lane undivided highways (Alhomidan, 2006).

Figure 2.1: Roadway Departure Crashes Contribution to 2011 Fatal Crashes in the United States (FHWA, 2013)
When a vehicle strays from the roadway, that vehicle may collide with another moving vehicle, a tree on the side of the roadway, a post pole, a barrier, an embankment, other fixed objects, or other than fixed objects such as pedestrians and animals, or just overturn. Figure 2.2 shows the distribution of these various types of ROR fatal crashes for the year 2011 in the United States.

![Figure 2.2: 2011 First Harmful Events in Roadway Departure Fatal Crashes in the United States (FHWA, 2011)](chart)

A number of researchers have studied factors contributing to the severity and occurrences of ROR crashes in recent years. A study by Hall and Zador (1985) examined the locations of more than 360 fatal rollover crashes in New Mexico and Georgia. In this study they found that sites associated with fatal crashes had sharper curves and steeper
downhill sections. Zegeer et al (1988) recommend that paying better attention during design by increasing the width of lanes, having less steep terrain, and reducing the curviness of two-lane undivided highways resulted in fewer ROR crashes.

Viner (1995) used the 1991 FARS and General Estimating System (GES) data to study the effects rollovers on sideslopes and ditches. He augmented the FARS and GES data with data from Maryland for 4,213 rollover crashes that occurred between August 1987 and December 1988. Viner (1995) found that rollover crashes on sideslopes and ditches were the principal cause of ROR driver fatalities as they contributed about 25% of all ROR driver fatalities and he notes that these types of ROR crashes most often occur on the outsides of horizontal curves. Newman et al. (2003) attribute the occurrences of ROR crashes to factors that include (i) avoiding a vehicle, object or animal, (ii) inattentiveness due to distractions, drowsiness, or drugs, (iii) slippery pavement due to climactic conditions, (iv) too high a speed for the road contour (curve or hill) and various roadway features that can exacerbate driver inattentiveness and thereby contribute to the likelihood of a ROR crash occurring include narrow lanes, excessively curved road, and narrow or steep shoulders.

The above factors can be mitigated. This is an essential endeavor because of the high number of lives impacted and millions of taxpayers’ money involved/lost. Although these type of motor vehicle crashes occur on various types of roads, but for two-lane undivided highways, half of fatal ROR crashes occur on curved sections (Newman, 2003). Based on available data, Newman et al. (2003) recommend the following cost-effective strategies for mitigating the impact of ROR crashes:

- Add rumble strips at the sides and the centers of roads
- Remove potential hazards on roadsides
- Improve skid resistance by improving road surfaces
- Improve painted marks on roads
- Improve the surfaces of shoulders and reduce their drop-offs

Spainhour and Mirsha (2008) used traffic crash data from the state of Florida (for the year 2000) to investigate the effects of driver, roadway, vehicle, and environmental factors on fatal ROR crashes involving overcorrection. Their study tested 23 explanatory variables and they conducted a backward stepwise regression model to identify variables with significant predictive power. They found that among the contributory factors, alcohol use, speeding, inattention, and sleeping or fatigue were the principal factors.

Liu and Subramanian (2009) used FARS crash data for the period 1991-2007 to analyze factors related to fatal single-vehicle ROR crashes. They used descriptive (univariate) analysis with chi-square tests and logistic regression modeling in assessing the impacts of various factors related to driver, roadway, weather, and traffic characteristics. Their study found that alcohol use, speeding, curved road segments, rural roads, higher posted speed limits, fewer number (one or two) of lanes, adverse weather condition, nighttime driving, two or more occupants in a vehicle, vehicle driven by a young driver (15-24 years old) and a male driver were significant factors that predicted higher probabilities of a vehicle being involved in ROR crashes.

Zhu et al. (2010) used data from Fatality Analysis Reporting System database (FARS) for the year 1997 (for states of Alabama, and Georgia) and for 1998 (for state of South Carolina) to study single-vehicle fatal crashes that occurred on two-lane rural highways from those four selected states. They used logit models for analyzing factors
affecting fatal single-vehicle ROR crashes on rural two-lane highways. Their study conclude that, lane width, shoulder width and type, horizontal curves and crest vertical curve and their interactions, roadside hazard rating, traffic volume, driveway type, lighting conditions, and crash time are significant factors for fatal single-vehicle ROR crashes on rural two-lane highways.

Roy and Dissanayake (2011) used traffic crash data from the state of Kansas for the period of 1999-2008 to study and compare factors associated with ROR and non-ROR crashes in Kansas. These researchers used the Bayesian statistical approach in analyzing roadway, driver, crash, and environmental factors. In that study, they found that nighttime, weekends, adverse weather, rural area, gravel and curved roads, higher speed limits, wet and icy road surfaces, utility vehicles were common characteristics associated with ROR crashes in Kansas. In addition, falling asleep, medical condition, alcohol use, driving too fast for condition, strong winds, freezing rain, shoulders, ruts, holes, and bumps were found to be significant factors having a greater role in contributing to ROR crashes than Non-ROR crashes.

Peng and Boyle (2012) investigated factors affecting commercial drivers in single-vehicle ROR crashes. They developed a logistic model of large truck crash data from the state of Washington for years 2006-2009 in analyzing crash-related factors including driver, vehicle, roadway, and environmental factors. The analysis of the data reveal that the effect of truck driver distraction, inattention, speeding, seat belt non-usages, drowsiness and fatigue increase the likelihood of an ROR resulting into injury or fatality.
CHAPTER THREE

METHODOLOGY AND DATA COLLECTION

3.1 Data Collection

This chapter discusses data collection including procedures used in merging datasets and sorting out the subset of ROR crashes only from the main dataset. In addition, in this chapter the statistical procedures and models developed in analyzing the data are described and discussed. In this thesis, motor vehicle crash data were obtained from the Ohio Department of Public Safety (ODPS) Microsoft Access ODPS crash databases consist all police-reported crashes that occurred on Ohio’s public roadways and streets.

3.1.1 Crash Data

In this study, traffic crash data for five years, from 2008 through 2012, were downloaded from the ODPS website. Crash data in ODPS website are organized in relational format into four related files with crash records compiled together by calendar year. These four files include crash records, unit records, people records, and ODOT records, respectively. Each of these files contains a variable known as DOCNO, which is a variable that relates records in all four files to their respective crash incidents. The unit records file and people records file contain an additional variable called UNITNO, which together with the DOCNO they relate all people involved in traffic crashes to the correct vehicles they
were traveling in and their specific crash incidents they were involved in. Therefore, these are two very important variables that are used to combine the four files together into one file by properly linking all related records together.

3.1.2 Merging Files

The four related files mentioned above in ODPS traffic crash database are can be briefly explained as follows:

a. The “crash records” file contains information specific to each crash that occurred such as crash severity, vehicle in error, date of crash, time of crash, name of city, village or township where the crash occurred, FIPS place code, crash location, type of road, if alcohol or drug was involved, if speeding was involved, etc.

b. The “unit records” file contains information on each unit/vehicle that was involved in a particular crash incident. Information recorded includes unit type (e.g., motor vehicle, motorcycle, bicycle, pedestrian, etc.), point of impact, number of occupants in the unit, etc.

c. The “people records” file contains information on each person involved in each crash with the exception of hit-and-run cases where information is always not available. Information recorded in this file includes person type (e.g., driver, occupant, or pedestrian), age, gender, severity of injury sustained by an individual, safety equipment used, etc.

d. The “ODOT records” file contains information such as county code, route type, latitude, longitude, crash type, etc.
For the scope of the current study, only records from three files were needed, that is, the crash records, unit records, and the people records files. A single-to-many merging technique in SPSS (Version 19.0) software was used to merge the three files mentioned above. As explained earlier in Section 3.1.1, corresponding records in the “crash” and “unit” files were joined by using a common variable, DOCNO. The records in the joint “crash-unit” file were then joined together with their corresponding records in the “people” file by using the two common variables DOCNO and UNITNO to obtain a joint “crash-unit-people” file. The five created joint “crash-unit-people” files for each calendar year were again joined together to obtain a file that contained a five-years traffic crash data for years 2008 through 2012. Records in the final file were systematically checked for consistency and to make sure that all records were correctly joined. The files merging process is schematically illustrated in Figure 3.1.

![Figure 3.1: The Process of Creating a Joint Crash-Unit-People File](image)

### 3.1.3 Creating of Run-off-Road-Related Traffic Crash Database

By using the joined file, a new file that contains ROR crash records only was created by querying and sorting records using a variable known as SEQUENCEEVENT1. This
variable contains information on the events in sequence for the vehicle as it met its first harmful event. A particular crash event was classified as an ROR crash if in the SEQUENCEEVENT1 variable it was recorded as overturn/rollover, run-off-road right, run-off-road left, cross median/centerline, or crash with a fixed object. If a record did not contain either one of the five mentioned events, then it was categorized as a non-ROR crash. Thus, the file was split into two files, i.e., ROR crashes only file and non-ROR crashes only file. Some records with either missing variables or recorded as unknown were deleted from the ROR crashes only file to create a final file of ROR-related traffic crashes with their associated information. Thus a file containing a total of 384,505 records of run-off-road traffic crashes with complete crash-related information from the database of traffic crashes that occurred on Ohio’s public roads and highways between 2008 and 2012 was created.

### 3.2 Description of Selected Variables for Study

The characteristics of run-off-road-related traffic crashes that occurred on Ohio’s public roads and highways between 2008 and 2012 after deleting records with missing, incomplete or unknown values are summarized in Tables 3.1 through 3.3 for crash/traffic characteristics, human/driver characteristics, and roadway/environmental characteristics, respectively. Table 3.1 shows that in terms of crash severity, fatal crashes accounted for 1.1% of all 390,625 ROR crashes analyzed in the current study, injury crashes accounted for 36.7% of all ROR crashes and 62.2% constituted of property damage only (PDO). When looking at the injury severity of the people that were involved in these ROR crashes, we see that 5.3% of them sustained either fatal or incapacitating injuries, 26.2%
of them sustained possible or non-incapacitating injuries, and 68.5% of them were not injured.

ROR crashes were grouped into four major categories and their frequencies of occurrences were analyzed and the data reveal that 1.4% of them involved overturning or rollover crashes, 48.6% constituted crashes where vehicles did run-off-the-road to the right, 37.5% did run-off-the-road to left, crossed median or centerline, and for those that crashed into fixed objects accounted for 12.5% of all ROR crashes. About 91.9% of ROR crashes involved passenger vehicles (cars, SUVs, minivans, and pickup trucks), 8.0% involved trucks and buses, and about 0.2% constituted motorcycles/motorized bicycles and emergency vehicles (police cars, ambulances, and fire trucks). In terms of posted speed limits on roadways where ROR crashes occurred, data shows that crashes that occurred on roadways with speed limits lower than 40 mi/h accounted for 27.6% of all the ROR crashes, 15.1% occurred on posted speed limits of 40-50 mi/h and 57.2% took place on roads with speed limits of 55 mi/h or higher.

The data in Table 3.2 show that while male were involved in 60.3% of the ROR-related traffic crashes, only 39.7% involved females. Drivers made up 75.7% while passengers accounted for 24.3%, and almost negligible pedestrian involvement in ROR-crashes. In terms of age involvement in the ROR-related crashes, 26.5% were people younger than 20 years old, 19.8% involved people in the 20-25 years old range, the 25-64 years old made up 48.6%, and seniors (≥ 65 years old) made up only 5.1% of the total people involved. Alcohol use was estimated to be involved in 88.1% of the ROR-related traffic crashes. Drug involvement was reported in only 2.8% of the ROR-related traffic crashes.
In terms of roadway contour, Table 3.3 shows that 51.8% of the total ROR-related crashes occurred on straight level segments, 18.5% of the total ROR-related crashes occurred on straight graded segments, 13.5% occurred on curved level segments, and curved graded segment accounted for 16.2% of all ROR-related crashes. The data also reveal that 58.7% of the total ROR-related crashes occurred during daylight, down or dusk times, 14.5% occurred on lighted roadways during dark times and 26.7% occurred on unlighted roadways when it was dark. When assessing weather condition, it is found that 41.3% of ROR crashes occurred when the weather was clear, 24.2% during cloudy weather, and 34.5% of the crashes occurred when there was rain, fog, sleet, snow, or strong winds.

The road conditions when ROR crashes occurred show that 51.3% crashed when the roadway surfaces were dry, 24.0% when the surfaces were either wet or with running water, and 24.7% of them occurred when the surfaces were covered with snow, ice, mud, gravel, or slush. The time of the day when the ROR crash occurred was also analyzed and data show that 64.0% occurred during early morning/daylight time (05:00-18:59), 17.1% occurred during early nights (19:00-22:59) and 18.9% occurred during late nights (23:00-04:59). The days when the crashes occurred where grouped into weekdays (Monday through Friday) and weekends (Saturday and Sunday) and data show that although weekends make up only 28.6% of the days of the week, but ROR-crashes that occurred during weekends accounted for 31.0% of all the ROR-crashes analyzed and 69.0% occurred during weekdays.
Table 3.1 Description of Crash and Vehicle Characteristics

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description/Code</th>
<th>Frequency</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Injury severity</td>
<td>No injury = 1</td>
<td>263,433</td>
<td>68.5</td>
</tr>
<tr>
<td></td>
<td>Possible/non-incapacitating</td>
<td>100,782</td>
<td>26.2</td>
</tr>
<tr>
<td></td>
<td>injury = 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Incapacitating/fatal injury = 3</td>
<td>20,290</td>
<td>5.3</td>
</tr>
<tr>
<td>ROR crash type</td>
<td>Overturn/rollover = 1</td>
<td>5,449</td>
<td>1.4</td>
</tr>
<tr>
<td></td>
<td>Run-off-road right = 2</td>
<td>186,798</td>
<td>48.6</td>
</tr>
<tr>
<td></td>
<td>Run-off-road left/cross</td>
<td>144,058</td>
<td>37.5</td>
</tr>
<tr>
<td></td>
<td>median/centerline = 3</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Crash with fixed object = 4</td>
<td>48,200</td>
<td>12.5</td>
</tr>
<tr>
<td>Vehicle type</td>
<td>Passenger vehicles = 1</td>
<td>353,187</td>
<td>91.9</td>
</tr>
<tr>
<td></td>
<td>Trucks/Buses = 2</td>
<td>30,640</td>
<td>8.0</td>
</tr>
<tr>
<td></td>
<td>Motorcycles/motorized</td>
<td>283</td>
<td>0.1</td>
</tr>
<tr>
<td></td>
<td>bicycles = 3</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Emergency vehicles = 4</td>
<td>395</td>
<td>0.1</td>
</tr>
<tr>
<td>Posted speed limit</td>
<td>&lt; 40 mph = 1</td>
<td>106,286</td>
<td>27.6</td>
</tr>
<tr>
<td></td>
<td>40-50 mph = 2</td>
<td>58,110</td>
<td>15.1</td>
</tr>
<tr>
<td></td>
<td>55-70 mph = 3</td>
<td>220,109</td>
<td>57.2</td>
</tr>
</tbody>
</table>

Table 3.2 Description of Human and Driver Characteristics

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description/Code</th>
<th>Frequency</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alcohol-related</td>
<td>No = 0</td>
<td>338,778</td>
<td>88.1</td>
</tr>
<tr>
<td></td>
<td>Yes = 1</td>
<td>45,727</td>
<td>11.9</td>
</tr>
<tr>
<td>Drug-related</td>
<td>No = 0</td>
<td>373,816</td>
<td>97.2</td>
</tr>
<tr>
<td></td>
<td>Yes = 1</td>
<td>10,689</td>
<td>2.8</td>
</tr>
<tr>
<td>Person type</td>
<td>Driver = 1</td>
<td>290,965</td>
<td>75.7</td>
</tr>
<tr>
<td></td>
<td>Occupant/Pedestrian = 2</td>
<td>93,540</td>
<td>24.3</td>
</tr>
<tr>
<td></td>
<td>Male = 0</td>
<td>231,725</td>
<td>60.3</td>
</tr>
<tr>
<td></td>
<td>Female = 1</td>
<td>152,780</td>
<td>39.7</td>
</tr>
<tr>
<td>Age of person</td>
<td>&lt;20 = 1</td>
<td>101,782</td>
<td>26.5</td>
</tr>
<tr>
<td></td>
<td>20-25 = 2</td>
<td>76,280</td>
<td>19.8</td>
</tr>
<tr>
<td></td>
<td>26-64 = 3</td>
<td>186,779</td>
<td>48.6</td>
</tr>
<tr>
<td></td>
<td>65+ = 4</td>
<td>19,664</td>
<td>5.1</td>
</tr>
<tr>
<td>Youth-related (16-25 years old)</td>
<td>No = 0</td>
<td>226,784</td>
<td>59.0</td>
</tr>
<tr>
<td></td>
<td>Yes = 1</td>
<td>157,721</td>
<td>41.0</td>
</tr>
<tr>
<td>Senior-related (65+ years old)</td>
<td>No = 0</td>
<td>359,732</td>
<td>93.6</td>
</tr>
<tr>
<td></td>
<td>Yes = 24</td>
<td>24,773</td>
<td>6.4</td>
</tr>
</tbody>
</table>
Table 3.3 Description of Roadway and Environmental Characteristics

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description/Code</th>
<th>Frequency</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Road contour</td>
<td>Straight Level = 1</td>
<td>199,058</td>
<td>51.8</td>
</tr>
<tr>
<td></td>
<td>Straight Grade = 2</td>
<td>71,242</td>
<td>18.5</td>
</tr>
<tr>
<td></td>
<td>Curve Level = 3</td>
<td>51,966</td>
<td>13.5</td>
</tr>
<tr>
<td></td>
<td>Curve Grade = 4</td>
<td>62,239</td>
<td>16.2</td>
</tr>
<tr>
<td>Light condition</td>
<td>Daylight/Dawn/dusk = 1</td>
<td>225,863</td>
<td>58.7</td>
</tr>
<tr>
<td></td>
<td>Dark - lighted roadway = 2</td>
<td>55,925</td>
<td>14.5</td>
</tr>
<tr>
<td></td>
<td>Dark - Unlighted roadway/unknown = 3</td>
<td>104,134</td>
<td>26.7</td>
</tr>
<tr>
<td>Weather condition</td>
<td>Clear = 1</td>
<td>158,979</td>
<td>41.3</td>
</tr>
<tr>
<td></td>
<td>Cloudy = 2</td>
<td>93,004</td>
<td>24.2</td>
</tr>
<tr>
<td></td>
<td>Rain/fog/sleet/snow/wind/other = 3</td>
<td>132,522</td>
<td>34.5</td>
</tr>
<tr>
<td>Road condition</td>
<td>Dry = 1</td>
<td>197,100</td>
<td>51.3</td>
</tr>
<tr>
<td></td>
<td>Wet/water = 2</td>
<td>92,377</td>
<td>24.0</td>
</tr>
<tr>
<td></td>
<td>Snow/ice/mud/oil/slush/gravel/other = 3</td>
<td>95,028</td>
<td>24.7</td>
</tr>
<tr>
<td>Time of crash</td>
<td>Early morning/daytime = 1</td>
<td>245,939</td>
<td>64.0</td>
</tr>
<tr>
<td></td>
<td>Early night = 2</td>
<td>65,857</td>
<td>17.1</td>
</tr>
<tr>
<td></td>
<td>Late night = 3</td>
<td>72,709</td>
<td>18.9</td>
</tr>
<tr>
<td>Day of week</td>
<td>Weekends = 1</td>
<td>119,352</td>
<td>31.0</td>
</tr>
<tr>
<td></td>
<td>Weekdays = 2</td>
<td>265,153</td>
<td>69.0</td>
</tr>
</tbody>
</table>

3.3 Methodology

3.3.1 Introduction

This section discusses the research methodology implemented to achieve the objectives of the current study. Classification tree (also known as decision tree) modeling was used in conjunction with generalized ordered logit (Gologit) model and the basis of the selected methods and their suitability to injury severity data are discussed. Classification tree model was used for exploratory data analysis (selection of significant independent variables) and generalized ordered logit model was used for model prediction (using only the variables identified by the decision tree procedure). In this study, both classification tree modeling and generalized ordered logit modeling were used to investigate
characteristics of injury and fatality of run-off-road crashes. A decision (classification) tree is a multivariate technique that traditionally has been used for data exploration (mining) and prediction and often used in business to model customer behaviors and in medicine to test best diagnosis of a disease (Lavery, 2012).

According to Lavery (2012): “Tree algorithms, or simply trees, split a dataset (assign observations in the data set to groups) hierarchically (groups are then divided into subgroups) based on the ability of the X variables, associated with the observations, to predict the Y variable. Tree analysis can be used in conjunction with, or as a replacement for, logistic or multiple regression, correspondence analysis, ANCOVA and neural nets.”

According to Myles et al. (2004) decision tree modeling has been widely used for both exploratory data analysis and predictive modeling applications because it is suitable for use in determining features and extracting patterns in large databases, which are essential in predictive and discrimination modeling.

3.3.2 Decision Tree Modeling

3.3.2.1 Introduction

Classification tree modeling works by dividing the dataset into small and more homogeneous subgroups by a set of “if-statements.” A decision tree is a hierarchical model composed of discriminant functions, or decision rules, that are applied recursively to partition the entire dataset into pure, single class subsets. It divides the dataset based on the most predictive independent variable for the response variable. Trees use some statistical measurements in order to split the dataset into small and more homogeneous
subgroups. The classification tree method selects an appropriate measurement based on the type of the response variable.

Both response and predictor variables can be either continuous or categorical. If a predictor variable is continuous, it is split into two partitions according to cutting value of the predictor variable. If the predictor variable is categorical (nominal or ordinal), the model splits the predictor variable categories (levels) into two groups of levels. If the response variable is continuous, the measure of the difference in the two groups is computed as the sum of squares of the differences between means. Both the variable to be split at a given level and the cutting value for the split are determined by maximizing LogWorth, a quantity related to the p-value associated with the sum of squares based on the difference in means. For continuous response variable the fitted values are the means within the two groups but for the case of a categorical response the splits are computed by maximizing a LogWorth statistic, which is related to the likelihood ratio chi-square statistic, known as $G^2$, also called log-likelihood ratio or Deviance (D). For the categorical response the fitted values are the estimated proportions within the groups.

An example of a decision (classification) tree is depicted in Figure 3.2. A decision tree consists of two types of nodes, branch nodes (including the root node) and leaf (terminal) nodes. Node 1 (the top most node) in the decision tree also known as root node contains the entire sample dataset. Each of the remaining nodes (referring to Figure 3.2, nodes 2 through 5) contains a subset of the entire dataset. Each branch node is a “parent” to two “children” nodes. For example, node 1 is split to produce nodes 2 and 3 and thus node 1 becomes their parent and then node 2 is the parent to nodes 4 and 5.
Decision pathways are represented by lines linking a parent node to its children nodes. In essence, the decision tree starts by splitting the original dataset (at root node) into two subsets based on a particular attribute value test. This process is repeated on each derived subset (branch node) in a recursive manner known as recursive partitioning. This repetitive procedure stops when the subset at a node has all the same value of the response variable, or when splitting no longer improves the predictions and this node becomes a leaf or terminal node (e.g., nodes 3, 4, and 5 in Figure 3.2).

### 3.3.2.2 Pruning and Validation

Pruning is a technique used in decision tree procedures to reduce the size of a decision tree or eliminate the less important splits that provide little power in classifying the occurrences or outcomes. Usually, this step is performed after the initial tree is built.
Consequently, pruning reduces the complexity of the final model and improves its predictive accuracy by decreasing the issue of overfitting that is, minimizing the chances of the model describing the random errors (noises) instead of the true underlying relationships. Pruning helps in detecting and removing splits of a classification tree, which may be based on noisy or erroneous data. Usually, overfitting is a result of a model being overly complex, including having too many predictor variables. An overfitted model is generally expected to have a poor predictive power because it will likely overstress small fluctuations in the data as if they were major relationships.

Validation one of the common techniques used to examine how strong the independent (predictor) variables predict the response variable. Under this procedure, the entire dataset is divided into two sets. The first set (called training dataset) is used to build the tree model and the second set (called validation dataset) is used to evaluate the performance of the model built by using the first set of data. That is, testing the capability of the model by assessing its performance on a set of data not used for training (or building it).

### 3.3.2.3 Statistical Analysis

#### 3.3.2.3.1 Node Splitting Criteria

The LogWorth statistic is a parameter usually used to grow as well as prune the decision tree model. It is used to indicate whether a particular predictor variable is significant or not. The larger the LogWorth value, the more significant the predictor variable is. The model generally splits the node based on the larger LogWorth statistic and is computed as shown in Equation 3.1:
\[ \text{LogWorth} = -\log_{10}(p\text{-value}) \] 
………………………………………………. (3.1)

Where the adjusted \( p \)-value is calculated by taking into account the number of different ways splits can be made.

For a continuous response variable, another parameter used to split the nodes is the sum of squares (SS), which is computed for each node. Essentially, this is the change in the error sum-of-squares due to the split. A candidate SS that has been chosen for splitting is computed as shown in Equation 3.2:

\[ S_{\text{test}} = S_{\text{parent}} - (S_{\text{right}} + S_{\text{left}}) \] 
……………………………………………… (3.2)

Where:

\( S_{\text{parent}} \) = sum of squares error in parent node

\( S_{\text{right}} \) = sum of squares error in the right-hand side child node (refer to Figure 3.3)

\( S_{\text{left}} \) = sum of squares error in the left-hand side child node

Figure 3.3: The Definition Pertaining to the Parent Node and its Two Children Nodes
It is noteworthy to mention that the value of the sum of squares (SS) in a node is computed as \( s^2(n - 1) \) where \( s^2 \) is the sample variance for the observations in the node and \( n \) is the number of observations in the node. Additionally, for continuous response variables the difference statistic is also computed, which is the difference between the predicted values for the two child nodes of the same parent node.

For the case of a categorical response variable, the log-likelihood-ratio chi-square, \( G^2 \), is computed instead. This is essentially twice the change in the entropy. This entropy is computed as shown in Equation 3.3:

\[
G^2 = 2 \sum \left[ f_o \log \left( \frac{f_o}{f_e} \right) \right] \quad \text{............................................... (3.3)}
\]

Where

\( f_o = \) observed frequency in a node

\( f_e = \) expected frequency in a node

A candidate \( G^2 \) that has been chosen for splitting is computed as shown by Equation 3.4:

\[
G_{test}^2 = G_{parent}^2 - \left( G_{right}^2 + G_{left}^2 \right) \quad \text{............................................... (3.4)}
\]

**3.3.2.3.2 Tests for Goodness of Fit**

The method used for assessing the decision tree model goodness of fit depends on the type of response variable in the dataset. A scatter plot of actual by predicted values is used for a continuous response variable and the Receiver Operating Characteristics
(ROC) curve is used if the response variable is categorical. Each of these are discussed below

3.3.2.3.3 Scatter Plot of Actual by Predicted Values

For a continuous response variable, the scatter plot shows how well the model fits the data. Figure 3.4 shows an example of an actual by predicted plot from JMP (version 10) software output. Two plots are developed, one for the training set and the other for the validation set. In the plots, the X-axes are the predicted means and the Y-axes consist of actual means for each leaf node. The actual values form a scatter of points around each leaf mean. A diagonal line represents the locus of where predicted and actual values are the same. The distance of a point from this diagonal line indicates how well or how poorly the prediction performed. For a perfect fit, all the points would be on this diagonal.
3.3.2.3.4 Receiver Operating Characteristics (ROC) Curve

The receiver operating characteristics (ROC) curve is used for assessing categorical responses. An ROC curve is a statistical tool that provides a complete and visually attractive way to assess the accuracy (power) of the accuracy of predictions (Agresti, 2007).

According to Agresti (2007), the area under the ROC curve is used to summarize the accuracy (predictive performance) of the analysis data. In essence, the area under the ROC curve estimates the probability that the predictions and the results are in agreement. It takes values from 0 to 1, where a value of 0 indicates a perfectly inaccurate test and a value of 1 reflects a perfectly accurate test (SAS, 2003) and the ROC curve usually takes
a concave curve connecting the points (0,0) on the left lowest corner and (1,1) at the right top most as shown in Figure 3.5. A value of 0.5 specifies that the predictions are essentially based on a random guess, that is, a model includes an intercept term only (Agresti, 2007). The area inside the ROC curve the curve measures discrimination, that is, for case of the current study, the ability of the classification tree analysis to correctly identify the risk factors of ROR crashes.

For a polytomous response variable (i.e., a variable with more than two response levels), such as the injury severity in this study, which has three levels of injury severity, the software packages such as JMP create an ROC curve for each response level. Therefore, the model performance for each response level is assessed separately. Two ROC curves are produced, one for the training (model building) dataset and the other for the validation (testing) dataset.

![ROC Curve](image)

Figure 3.5: An Example of a ROC Curve for a Three-level Response Variable from JMP Software Output
3.3.3 Generalized Ordered Logit Modeling

Crash injury severity represents a typical ordinal categorical data. The response variable, crash injury severity, as used in this thesis study, it consists three levels after combining some of them together as shown in Table 3.1. The three levels in increasing severity were coded as 1 = no injury, 2 = possible injury and non-incapacitating injury, 3 = incapacitating injury and fatal injury. Therefore, let’s define \( k = 1 \) as the lowest level value of injury severity variable, i.e., no injury.

Based on the existing literature, this thesis attempts to model injury severity of ROR crashes using the generalized ordinal logit model, a procedure that has been recognized to be a more flexible modeling approach (Williams, 2006; Wang and Abdel-Aty, 2008; Wang et al., 2009; Mergia et al., 2013). This approach is capable to go overcoming the limitations with both the conventional ordered logit/probit and the unordered methods (Savolainen et al., 2011). The suitability of generalized ordered logit model in modeling injury severity data is due to its flexibility in its procedure as it is capable of relaxing the parallel-line assumption in the ordered logit model by allowing the variability of the regression parameter \( \beta \) across outcome levels, while maintaining the ordinal nature of the response variable (injury severity). This method is also known as the partial proportional odds model. The generalized ordered logit model can be expressed as shown in Equation 3.5:

\[
P(Y \leq y_j | x) = \frac{\text{Exp}(\eta_j - x^T \beta_j)}{1 + \text{Exp}(\eta_j - x^T \beta_j)} \tag{3.5}
\]

Where:
\( \beta_j \) = a vector of unknown regression coefficients

\( x \) = a vector of observed explanatory variables

\( x^T \) = a transpose vector of observed explanatory variables

\( \eta_j \) = unknown threshold or intercept parameters, satisfying the condition \( \eta_1 \leq \eta_2 \)

\( \ldots \leq \eta_k \)

\( j = \) comparison groups; \( j = 1, 2, \ldots, k-1 \) (in this study \( j = 1, 2 \))

\( y_j \) = outcome in comparison group \( j \)

\( Y \) = ordinal response variable with \( k \) outcomes

\( k \) = the number of outcome levels (categories) of injury severity (in this study \( k = 1, 2, 3 \))

\[ P (Y \leq y_j) = \text{cumulative probability of the event (} Y \leq y_j\text{)} \]

Some of the commonly known models are in fact special cases of the generalized ordered logit model presented by Equation 3.5. When \( k = 2 \), the model becomes the logistic regression model; when \( k > 2 \), the model becomes equivalent to a series of binary logistics regressions where categories of the dependent variable are combined; and if the \( \beta' \)'s are the same for all values if \( j \), it becomes the traditional ordered logit model. From Equation 3.5, the probabilities that \( Y \) (injury severity) will take on each of the values 1, 2, or 3 (i.e., the individual outcome groups) can be written using cumulative probability distribution as depicted in Equations 3.6 through 3.8:

\[
P(Y = y_1 | x) = F(\eta_1 - x^T \beta_j) \]  
………………………………………………(3.6)

\[
P(Y = y_2 | x) = F(\eta_2 - x^T \beta_j) - F(\eta_1 - x^T \beta_j) \]  
…………………………………….(3.7)
\[ P(Y = y_3 \mid x) = 1 - F(\eta_2 - x^T \beta_2) \] .................................................. (3.8)

The weakness of the ordered logit model is that the regression coefficient, \( \beta \), does not depend on the outcome comparison group, \( j \), because the model restricts it to be the same value for all outcome levels (which means all lines have the same slope, i.e., parallel lines) regardless of the possibility of variations (Williams, 2006). For the generalized ordered logit model, some of the regression coefficients, \( \beta_j \), may be the same for some outcome levels but it allows others to vary if they violate the parallel line assumption. According to Williams (2006) the parallel-lines assumption is more often violated as it does not hold up most of the time.

The results from this procedure are interpreted as explained here. For a variable of three outcome levels (as is in this thesis study), the three outcome levels are grouped into two comparison groups. As a result, two sets of outcome groups for each model are developed. Since \( k = 3 \), then for \( j = 1 \), outcome level 1 is compared with outcome levels 2 and 3; and for \( j = 2 \), outcome levels 1 and 2 are compared with outcome level 3. A positive regression coefficient estimate indicates that higher values on the predictor variable increase the likelihood of an injury being in the higher injury severity levels than the current one. Similarly, a negative coefficient estimate indicates that higher values on the predictor variable increase the likelihood of an injury being in the current or lower injury severity levels, meaning that it reduces the likelihood of being in higher injury severity levels. Alternatively, a negative coefficient can simply be interpreted as increasing the likelihood of being in the current or lower injury groups. The results from the generalized ordered logit model can be interpreted as explained in the following
example. For instance, let’s assume that a predictor variable “alcohol-related” is coded with two outcomes (coded as 0 for No and 1 for Yes). If the results show that this predictor variable “alcohol-related” has a positive regression coefficient in the outcome group $j = 2$ (comparing no-injury & possible/non-incapacitating injuries against incapacitating/fatal injuries), it means that a person injured in a traffic crash where alcohol use is involved is likely to sustain higher injury levels (incapacitating and fatal injuries) than just no-injury or possible/invisible injuries.
CHAPTER FOUR

RESULTS

4.1 Introduction

Several geometric, environmental, traffic, and behavioral factors, which are considered to have an effect on the occurrence of run-off-road-related traffic injuries were analyzed by using some powerful statistical modeling techniques in order to determine the most significant ones. A total of fourteen variables were selected for exploratory analysis to investigate characteristics of predictor variables of ROR traffic injuries and screen out the most promising ones for use in the next step. A decision tree procedure in JMP software version 10 was used for developing the classification tree modeling. The generalized ordinal logit regression model which uses maximum likelihood estimation (MLE) method was applied to estimate statistically the effects of these variables in contributing to the occurrence of run-of-road traffic-related injury severity levels. The gologit2 procedure in the STATA software release 12 was used in this estimate. Predictor variables were tested at a 95% significance level. The first part of this chapter presents the descriptive results based on the data analyzed in this study; the second part discusses the results of the decision tree modeling; and lastly, in the third part, the results from the generalized ordered logit modeling are discussed.
4.2 Descriptive Results of Run-off-Road-Related Traffic Crashes

Motor vehicle traffic crashes are usually complicated events normally contributed by a number of various types of interactions. Examining the characteristics of crash data provides a fairly decent inference at some underlying aspects of the motor vehicle traffic crashes and thus can assist the analysts in devising possible safety countermeasures. Data used in the current study reveal that run-off-road traffic crashes accounted for 21% of all police-reported traffic crashes that occurred on public highways and streets in the state of Ohio for the period of 2008-2012. For the objectives of this study, a total of 384,505 observations of complete data points were used whereby, 20,290 observations involved incapacitating and fatal injuries, 100,782 were possible and non-incapacitating injuries, and 263,433 included individuals who were not injured when they were involved in the ROR crashes. The main objective of the descriptive results is to provide a better view of characteristics of run-off-road crashes. In particular, this section pays more attention to traffic crashes that resulted into injuries, that is, possible, non-incapacitating, incapacitating and fatal injuries.

Figure 4.1 shows a relationship between the likelihood of road users sustaining injuries (this includes minor, major, and fatal injuries) and the road contour (alignment) where ROR crashes occurred. The results show that about 40.5% of the people who were involved in ROR crashes at curved level road segments in Ohio sustained injuries. Therefore, relatively, the higher injury rates are associated with curved road segments with the curved grade roads causing the second highest rate of 39.0%. Therefore, curved roads were more hazardous than straight ones, overall, in terms of the rates of injury severity sustained.
The results of ROR-related injury rates when alcohol use was involved as a contributing factor in the occurrence of crashes are depicted in Figure 4.2. These results show that while only 34.6% of those involved in ROR crashes for which alcohol was not involved, 49.5% sustained injuries in crashes when alcohol use was involved. More specifically, the fatality rates between crashes when alcohol was involved and when was not involved, was at a ratio of about 5.3 to 1, indicating that run-off-road traffic crashes involving drunk drivers tend to be more dangerous.
Figure 4.2: Rates of Fatality and Injury Due to Run-off-road Crashes by Alcohol Use Status

The distribution of fatalities due to run-off-road motor vehicle traffic crashes is displayed in Figure 4.3. The results show that ROR crashes occurring on weekends sustained much more fatalities than those occurring on weekdays. In the five years of study, 2008-2012, Saturdays produced a tally of 551 ROR-related fatalities, followed closely by Sundays with 545 fatalities. Tuesdays and Wednesdays were the relatively safer days in terms of sustaining fatalities with a total of 303 and 313 fatalities in five years, respectively. However, an interesting image appears when inspecting Figure 4.4, which depicts the total number of injuries by day of the week. In terms of total injuries sustained from run-off-road crashes, Tuesdays are not viewed as the safety days of the week, as they sustain the fourth highest number of injuries just behind Saturdays, Sundays, and Fridays. When looking in terms of number of vehicles involved in run-off-road crashes by day of the week in Figure 4.5, a similar trend shown in Figure 4.4 is
shown again with Tuesdays being the fourth again behind Saturdays, Sundays, and Fridays.

Figure 4.3: Number of Fatalities Due to Run-off-Road Crashes by Day of the Week

Figure 4.4: Number of Total Injuries Due to Run-off-Road Crashes by Day of the Week

44
The results of injury rates of ROR-related traffic crashes by roadway conditions where the crashes occurred are depicted in Figure 4.6. It is clearly indicated from Figure 4.6 that ROR crashes on dry roadways have the highest rates of sustaining injuries. This may be due to the fact that most people drive when the weather is good and most likely the majority of them drive fast due to conducive conditions and this may lead to some of the vehicle getting out of control and run-off-the-road and hence sustain higher injuries due to higher speeds. Though it would be expected that adverse roadway conditions would be more dangerous, but it is most likely that the majority of the drivers slow down when the roads are slippery or covered with snow and ice and they take more caution and as a result they don’t end up sustaining major injuries when they get out of control and run-off-the-road.
4.3 Results of Decision Tree Modeling for Predictor Parameter Screening

4.3.1. General

The response variable in the dataset was the injury severity, which consisted of three levels of injury and it was modeled as an ordinal variable. Sixteen predictor variables (refer to Table 3.1) were retained in the run-off-road traffic crash dataset. Seven of these were coded as binary variables and other nine were coded as multilevel nominal (polytomous) variables. The main objective under this task was to investigate the complex relationships between the injury severity and the sixteen selected predictor variables by utilizing the decision tree’s capability of identifying such relationships regardless of the form, i.e., either linear or nonlinear, including multilevel interactions. The final product from this decision tree analysis was to identify the most significant...
predictor variables, which should be used in the regression modeling of predicting the injury severity of run-off-road crashes.

The dataset inputted in the JMP program consisted of 384,505 observations of run-off-road crashes, out of which 345,795 (90%) were randomly selected (automatically by the program) to form a training sample set and 38,710 observations (10%) were set aside as the validation sample set. The validation sample set is actually used to test the validity of the model developed using the training sample set. In the training sample, there were 236,822 no injury cases, 90,697 possible and non-incapacitating injury cases, and 18,276 incapacitating and fatal injury cases. The JMP program was first allowed to run a full tree splitting where a total of 116 were observed. By assessing the $R^2$ plot depicted in Figure 4.7, it can be observed be that the value of $R^2$ remains about the same after 20 splits for both training and validation datasets. Therefore, the best tree size for found to be attained after about 20 splits and node splitting has to be stopped here. By stopping the node splitting correctly, is one of the ways that help in protecting the model against both underfitting and overfitting. These 20 splits for the model developed and tested using the ROR crashes data can detect the most important interactions between independent variables and determine the most powerful predictors of levels of injury severity.
The column contributions report as shown in Figure 4.8 suggests that there are eight independent variables which explain a large amount of the variation in the response variable in terms of the $G^2$ statistic values. These are the predictor variables that contributed in fitting the model. The eight predictor variables identified under this analysis include road condition, alcohol-related, ROR crash type, vehicle type, gender, posted speed, road contour, and drug-related. Based on this analysis, these are the predictor variables that were passed to the generalized ordered logit model to determine their effects on levels of injury severity of ROR crashes.
The training and evaluation models were evaluated by using the ROC curves as depicted in Figures 4.19 and 4.10, respectively. Both curves show that the model can predict well the levels of injury severity since all the curves of the levels of injury in both plots are above a 45-degree line (i.e., each of them has a rate larger than 0.5, which is a point that specifies predictions taken as based on a random guess). As expected, both training and validation models predict better the injury group level 3 (i.e., incapacitating and fatal injuries), explaining the variations in the response variable by almost 71%.
Figure 4.9: ROC Curve for Training Sample Dataset

Figure 4.10: ROC Curve for Validation Sample Dataset
Figure 4.11: Final Decision Tree Diagram After 20 Splits
As pointed out earlier, a decision tree model has the ability to detect or catch the interactions involving more than two independent variables. Figure 4.11 shows the final decision tree diagram of the run off road crashes, which specified to stop after performing 20 splits as this was assessed to produce the best tree model. Each node box shows the three injury group levels and their probabilities and counts (i.e., number of observations). In addition, each node box also includes the LogWorth value, the value for which the data split is based on.

### 4.3.2 Interaction of Road Condition and ROR Crash Type

The model first splits the data into two groups based on roadway condition (refer to Figure 4.11) as a predictor variable, which shows that this is the most significant factor in predicting injury severity. This is revealed in Figure 4.8 as road condition is the variable with the largest $G^2$ value. The two resulting (child) nodes are grouped as follows: (1) dry roadway condition, and (2) wet condition and snow/ice/sand/dirt conditions. It is clearly shown that the risk of ROR-related injuries of all kinds are overrepresented on dry roadways with a conditional probability of injuries (combining injury groups 2 and 3 together of 0.383 (67,873 observations) in Node 2 as compared with the other roadway conditions combined with a probability of 0.2438 (41,100 observations). The most likely reason is that drivers tend to be more cautious on roads with adverse conditions but over speed or drive carelessly when roads are dry and in good condition elevating chances of sustaining injuries when involved in run-off-road crashes.
4.3.3 Interaction of Road Condition, ROR Crash Type, Vehicle Type and Alcohol

ROR crash types show up in the decision tree diagram in two locations. Under node 2, the data is split into two groups of ROR crash types: overturn ROR crash type and the other types of ROR crash (run right, run left/cross median/centerline, and crash with fixed objects). The vehicle type variable appears under overturn ROR (node 5) but not under other types of ROR crash (node 4). Meanwhile, the alcohol related does not show up under overturn ROR (node 5) but it shows up under other types of ROR (node 4). This is an indication that there is an interaction between overturn ROR type and vehicle type and between other types of ROR crash and alcohol-related crashes. This means that if an emergency vehicle or truck or bus runs off a dry road and overturns, the probability of sustaining incapacitating and fatal injury levels is 0.246 (which is 0.8545 when considering all injuries). On the other hand, if a passenger car or motorcycle/motorized bicycle runs off a dry road and overturns, the probability of sustaining incapacitating and fatal injury levels is only 0.0751 (which is 4.4799 when considering all injury levels). This is an indication that large vehicles are likely to cause more severe injuries in ROR crashes as it can be seen that vehicle type as a variable shows up in a number of nodes (see nodes 11, 16, 27, and 30). Likewise, if a vehicle runs off a dry road to the right or left or crashes with a fixed object and the influence of alcohol is involved, then the probability of sustaining incapacitating and fatal injury levels is higher at 0.1379 than when alcohol is not involved, which is only 0.0599 (please refer to nodes 6 and 7 in Figure 4.11). This means alcohol use mostly increases the severity of injuries in ROR crashes that occur on dry pavement conditions.
4.3.4 Interaction of Road Condition, ROR Crash Type and Posted Speed Limit

The decision tree model shows an interaction between crashes involving trucks, buses and emergency vehicles running off (to the right or left or crashes with a fixed object) and posted speed limits of 40 mph or higher. If a vehicle runs off (to the right or left or crashes with a fixed object) a road with posted speed limit over 40 mph and no influence of alcohol involved, the conditional probability of sustaining injuries is 0.3808 (please see node 12). Meanwhile, if the posted speed limit is below 40 mph, then the conditional probability of sustaining injuries is only 0.2884 (node 13). This is an indication that high posted speed limits interact with other variable in enhancing the occurrences of severity of injuries in ROR crashes. The severity of injuries is strongly correlated with the speed of the vehicle (Miltner and Salwender, 1995).

4.3.5 Interaction of Posted Speed Limit, Vehicle Type and Road Contour

The model shows a strong interaction between posted speed limit, road contour and vehicle type. This interaction appears in two locations in the decision tree (under node 16 and 30). If an emergency vehicle, truck, or bus crashes on a curved (curve level or curve grade alignment) road with a dry surface and with posted speed limits of 40 mph or higher, the probability of fatal and incapacitating injuries is 0.2299 (see node19). By observing child nodes branching from parent nodes 16 and 30, it can be clearly observed that curved road segments (both curve level and curve grade alignments) tend to increase probabilities of severe injuries when compared with straight level and straight grade segments. This means that ROR crashes occurring at curved sections of roads generally cause more severe injuries than those occurring on straight (but either graded or level)
sections of roads. This observation agrees with previous findings where the majority of crashes occurring on curved sections of roads were either fatal or caused injuries (Lamm and Choueiri, 1987).

4.3.6 Interaction of Posted Speed Limit, Vehicle Type, and Gender

There is an interaction between posted speed limits, vehicle type, and gender. Females in ROR crashes involving cars or motorcycles occurring on higher posted speed limit (40 mph or higher) roads have higher chances of sustaining injuries when compared with males in the same situations. When these interactions occur, females have a conditional probability of sustaining injuries of 0.4822 compared with 0.3429 for males (please refer to nodes 22 and 23).

4.3.7 Interaction of Speed Limit, Vehicle Type, Gender, and Drug Related

There is an interaction between posted speed limit, vehicle type, gender, and drug-related variables. A flow chart connecting nodes 12, 17, 22, and 41 together, reveal that if males are involved in run-off-road crashes where drug involvement is suspected and involving cars or motorcycles on roads with posted speed limits of 40 mph or higher tend to sustain severe injuries.

4.3.8 Interaction of Alcohol Use, Vehicle Type, and Speed Limit

There is an interaction between alcohol use, vehicle type and posted speed limit. A flow chart connecting nodes 6, 10, and 29 together indicate that ROR crashes involving
passenger cars, motorcycles, and emergency vehicles and alcohol use on roads with posted speed limits over 40 mph have higher probabilities of causing higher levels of injury when compared with similar situations on roads with posted speed limits lower than 40 mph.

**4.3.9 Interaction of Road Condition, Gender, and ROR Crash Type**

There is an interaction between road condition, gender, and ROR crash type. A flow chart connecting a series of nodes 8, 14, 24, and 36 together indicate that overturning/rolling over type of ROR crashes involving males occurring on wet road condition tend to lead into higher chances of sustaining injuries.

**4.3.10 Interaction of Road Condition, Gender, and Posted Speed Limit**

There is an interaction between road condition, gender, and posted speed limit. A flow chart connecting a series of nodes 9, 20, 34, and 39 together indicate that crashes occurring when the road surface is in snowy condition, involving males on roads with posted speed limits of 55 mph or higher have higher chances of sustaining major injuries.

**4.3.11 Interaction of Road Condition, Gender, and Alcohol Use**

There is an interaction between road condition, gender, and alcohol use. A flow chart connecting a series of nodes 9, 20, and 35 together indicate that crashes occurring when
the road surface is in snowy condition, involving males and alcohol use as a factor have higher chances of sustaining injuries.

4.4 Results of Generalized Ordered Logit Model for Predictor Parameter Estimation

The eight predictor factors that were identified by the decision tree procedure that explain most of the variations in the injury severity levels of the ROR crashes were analyzed using the generalized ordinal logit regression, which uses a maximum likelihood estimation (GLE) technique. The gologit2 procedure in the STATA software release 12 was used to produce the predictor parameter estimates. All predictor variables tested in this model were found to be significant in predicting the injury levels. The results from STATA output are presented in Table 4.1. Road contour categories coded 2 through 4 (straight grade, curve level, and curve grade, respectively) appear in both with positive parameter estimates) in Table 4.1 indicating that these geometric features increase the likelihood of injuries of all levels. However, the effect of straight grade segments to ROR injuries is modest increase (due to small coefficient value). The curve level segments are the most hazardous segments, besides having the largest coefficient value, its value increases from panel one \((j = 1)\) to panel \(2\) \((j = 2)\), meaning that these features tend to increase higher levels of injury severity, i.e. incapacitating and fatal injuries. Alcohol-related crashes (coded as 1) have a very strong effect on injury severity of crashes. This parameter has positive coefficients in both estimate panels with the larger coefficient in the second panel indicating that alcohol related crashes increase more the likelihood of sustaining incapacitating and fatal injuries.
Table 4.1 Parameter Estimate Results from STATA Output

| Injury severity | Coeff.  | Std. Err | z     | P>|z| | [95% Conf. Interval] |
|-----------------|---------|----------|-------|------|---------------------|
| J = 1 (no-injury vs. possible, non-incapacitating, incapacitating and fatal injuries) |         |          |       |      |                     |
| Road Contour 2  | 0.073   | 0.010    | 7.49  | 0.000| 0.054 - 0.093       |
| Road Contour 3  | 0.241   | 0.011    | 22.52 | 0.000| 0.220 - 0.262       |
| Road Contour 4  | 0.198   | 0.010    | 19.53 | 0.000| 0.178 - 0.218       |
| Alcohol Related 1 | 0.544 | 0.011    | 49.33 | 0.000| 0.522 - 0.565       |
| Drug Related 1  | 0.573   | 0.021    | 27.37 | 0.000| 0.532 - 0.614       |
| Gender 1        | 0.328   | 0.007    | 44.27 | 0.000| 0.314 - 0.343       |
| ROR Crash Type 2 | -1.702 | 0.033    | -52.27| 0.000| -1.766 - -1.638     |
| ROR Crash Type 3 | -1.676 | 0.033    | -51.18| 0.000| -1.74 - -1.612      |
| ROR Crash Type 4 | -1.948 | 0.034    | -57.68| 0.000| -2.014 - -1.881     |
| Road Condition 2 | -0.401 | 0.009    | -45.14| 0.000| -0.418 - -0.384     |
| Road Condition 3 | -0.844 | 0.010    | -87.16| 0.000| -0.863 - -0.825     |
| Vehicle Type 2  | 0.131   | 0.014    | 9.49  | 0.000| 0.104 - 0.158       |
| Vehicle Type 3  | -0.634  | 0.153    | -4.15 | 0.000| -0.934 - -0.335     |
| Vehicle Type 4  | 1.036   | 0.109    | 9.47  | 0.000| 0.821 - 1.250       |
| Posted Speed 2  | 0.230   | 0.012    | 19.56 | 0.000| 0.207 - 0.253       |
| Posted Speed 3  | 0.370   | 0.009    | 42.03 | 0.000| 0.353 - 0.387       |
| Constant        | 0.633   | 0.033    | 19.06 | 0.000| 0.568 - 0.698       |
| J = 2 (no-injury, possible, non-incapacitating vs. incapacitating and fatal injuries) |         |          |       |      |                     |
| Road Contour 2  | 0.072   | 0.021    | 3.48  | 0.000| 0.031 - 0.113       |
| Road Contour 3  | 0.283   | 0.020    | 13.81 | 0.000| 0.242 - 0.323       |
| Road Contour 4  | 0.190   | 0.020    | 9.41  | 0.000| 0.151 - 0.23        |
| Alcohol Related 1 | 0.966 | 0.018    | 53.52 | 0.000| 0.931 - 1.001       |
| Drug Related 1  | 0.715   | 0.029    | 24.45 | 0.000| 0.658 - 0.773       |
| Gender 1        | 0.145   | 0.016    | 9.23  | 0.000| 0.114 - 0.175       |
| ROR Crash Type 2 | -0.720 | 0.041    | -17.54| 0.000| -0.801 - -0.64      |
| ROR Crash Type 3 | -0.580 | 0.042    | -13.94| 0.000| -0.662 - -0.499     |
| ROR Crash Type 4 | -1.009 | 0.047    | -21.32| 0.000| -1.102 - -0.916     |
| Road Condition 2 | -0.582 | 0.020    | -29.68| 0.000| -0.620 - -0.544     |
| Road Condition 3 | -1.169 | 0.026    | -45.28| 0.000| -1.219 - -1.118     |
| Vehicle Type 2  | 0.863   | 0.022    | 38.70 | 0.000| 0.82 - 0.907        |
| Vehicle Type 3  | -0.153  | 0.310    | -0.49 | 0.622| -0.76 - 0.455       |
| Vehicle Type 4  | 1.048   | 0.141    | 7.45  | 0.000| 0.772 - 1.324       |
| Posted Speed 2  | 0.190   | 0.025    | 7.69  | 0.000| 0.141 - 0.238       |
| Posted Speed 3  | 0.427   | 0.018    | 23.28 | 0.000| 0.391 - 0.463       |
| Constant        | -2.644  | 0.044    | -60.29| 0.000| -2.73 - -2.558      |

<table>
<thead>
<tr>
<th>Model Goodness-of-Fit Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Observations</td>
</tr>
<tr>
<td>Likelihood Ratio (LR)</td>
</tr>
<tr>
<td>Chi-Square</td>
</tr>
<tr>
<td>Prob &gt; Chi-Square</td>
</tr>
<tr>
<td>Log likelihood</td>
</tr>
<tr>
<td>Pseudo R^2</td>
</tr>
</tbody>
</table>
Drug-related crashes (coded as 1) have also positive coefficients in both panels with increasing values indicating that drug use is significant in increasing the likelihood of injuries especially sustaining higher levels of injuries such as incapacitating and fatal injuries. Drug-related predictor parameter shows the same trends as those shown by the alcohol-related parameter to ROR traffic crashes.

Gender is significant with positive parameter estimates in both the first and second panels (but with a higher coefficient in the first panel) indicating that females are more likely to sustain injuries compared to males but they sustain more of possible/non-incapacitating than incapacitating/fatal injuries. Run-off-road crash type was modeled as a four-category class variable. The “overturn/rollover” category (Type 1) was used as the reference level. The results in Table 4.1 show that run-off-road right (Type 2), run-off-road left/cross median/centerline (Type 3), and crash with fixed object (Type 4) all are found to increase the likelihood of no-injury crashes because all of them have negative coefficients in both panels. Likewise, vehicles involved in overturn/rollover crashes are the ones that increase the likelihood of injuries. In terms of road condition, dry roadway condition (code = 1) is the only one among the three condition categories that were considered in this study, which increases the likelihood of injury crashes. For the other two conditions, i.e., wet/standing or running water (code = 2) and snow/ice/sand/mud/etc. (code = 3), they have negative coefficients in both panels indicating that they simply increase no injury crashes. Vehicle type (Type 2) variable is significant in both panels with positive coefficients and the larger value in the second panel. This indicates that vehicle type 2 (buses and trucks) increase the likelihood of sustaining all kinds of injuries but especially incapacitating and fatal injuries. But motorcycles and motorized bicycles
tend to increase the possibility of no injury ROR crashes. However, a very interesting result is for emergency vehicles (Type 4), with large positive coefficients on both panels indicating that ROR crashes involving emergency vehicles increase the likelihood of sustaining all kinds of injuries almost equally, ranging from possible/non-incapacitating injuries to incapacitating/fatal injuries.

The parameter estimate for posted speed limit is found to be significant for posted speed limit of 40 mi/h and higher (code = 2 for 40-50 mi/h and code = 3 for 55+ mi/h) in both panels with positive coefficients (also increasing from code 2 to code 3) indicating that higher posted speed limits increase the likelihood of sustaining injuries related to ROR crashes.

Comparing results from the current study with other previous studies of run-off-road traffic crashes, we observe fairly notable agreements of significant contributing factors and some few differences as well. However, sometimes it is important to note the differences in the objectives of the studies being compared because these also affect the findings and methods used in the analysis. For example, some researchers may be studying significant factors affecting the injury severity of drivers only or all occupants; others may be studying only fatal crashes versus all crashes, while others may be investigating factors affecting the occurrences of crash frequencies. In addition, the statistical experimental study unit may be different as well due to the setup and objective of the study, for example, the study unit can be a vehicle occupant, a driver, a vehicle or the crash incident. Significant factors that increase the likelihood of run-off-road injury severity, which have been identified in the current study and are in agreement with some previous studies include curves (Hall and Zador, 1985; Zeeger et al., 1988; Viner, 1995;
Newman et al., 2003; Liu and Subramanian, 2009; Zhu et al., 2010; Roy and Dissanayake, 2011), alcohol involvement (Spainhour and Mirsha, 2008; Liu and Subramanian, 2009; Roy and Dissanayake, 2011), higher posted speed limits (Liu and Subramanian, 2009; Roy and Dissanayake, 2011). Weather condition was not significant in our study, but it was significant in some previous studies (Liu and Subramanian, 2009; Roy and Dissanayake, 2011), adverse roadway condition such as slippery roads, snow and ice on the roadway surfaces, etc., were significant factors in Liu and Subramanian (2009), Roy and Dissanayake (2011), but they were not significant in the current study, but it is dry roadway condition, which was found to be significant factor in increasing the likelihood of injuries especially severe injuries (incapacitating and fatal injuries). The run-off-road crash type was considered as a variable in the current study and found significant but it was not included in most of the previous studies, which were reviewed. An interesting finding from the current study is that only overturning/rollover as an ROR type of crash is the only one that increases the chances of injuries while all others mostly end up into property damage only crashes (i.e., no injury to people involved in the crashes). Another interesting finding that was not studied previously is the inclusion of emergency vehicles as a separate vehicle type in the analysis. The current study found that when emergency vehicles (i.e., police cruiser cars, fire trucks, and ambulances) get involved into run-of-the-road crashes, the mainly cause higher levels of injury severity. This may make sense due to the high speeds that these kinds of vehicles are always involved in when they are called for emergency incidents.

The time of crash, which did not even pass the decision tree analysis, i.e., and therefore it was not a significant factor in the current study, however it happened to be a
significant factor in some previous studies of ROR crashes (e.g., Zhu et al., 2010; Roy and Dissanayake, 2011). Although weather condition was not significant in the current study, but it may be correlated with roadway condition, which was determined to be significant factor in the current study but the effect of roadway condition in this study was found to be the opposite of what some previous studies reported.
A run-off-road (ROR) crash or a roadway departure crash is a non-intersection crash which occurs after a vehicle crosses an edge line or a center line (i.e., leaves its designated traveled way and in the process the vehicle collides with a non-traversable obstacle or another vehicle travelling in the opposite direction or hits a pedestrian, or the vehicle overturns. The main objective of the current study was to determine the factors that contribute significantly to the levels of injury severity when ROR crashes occur. Based on this objective, a 5-year crash data for years 2008 - 2012 obtained from the Ohio Department of Public Safety was used for this analysis. In this study, decision tree model in conjunction with generalized ordered logit model were used to investigate characteristics of injury and fatality of run-off-road crashes in Ohio.

The decision tree model was used because it has the ability to detect the important predictors for injury severity and easily discover the interactions between two or more predictors. Therefore, it reduces the number of predictor variables to be used in the regression model, which is a significant step in reducing noises and correlation between the regression variables. The generalized ordered logit model was used because it is capable of overcoming the limitations inherent in both the conventional ordered logit/probit and the unordered models in modeling injury severity data is due to its flexibility in its procedure by relaxing the restrictive parallel-line assumption in the
ordered logit models while maintaining the ordinal nature of the response variable (injury severity).

The decision tree procedure identified only eight factors, which are considered to explain a large amount of the variation in the response variable (i.e., injury severity levels) considered in the current study. These important predictors of injury severity include road condition, run-off-road crash type, posted speed limit, vehicle type, gender, alcohol-related, road contour, and drug-related only. The decision tree model determined that the most severe ROR-related injuries occurred on roads when their surfaces were dry. The presumable reason is that drivers tend to over speed or drink and drive when the weather is good and road surfaces are in good/dry condition and if they get involved in run-off-road crashes that increases the chances of causing higher levels of injuries. Likewise, drivers tend to be more cautious when driving on wet or snow/ice covered road surfaces and when they run-off-road, they are already driving at lower speeds and the collision impacts end up causing no-injury or minor injuries.

Specifically, the decision tree modeling analysis, besides identifying eight significant factors of injury severity levels; it also uncovered the following complex interactions, which were mainly confirmed by the parameter estimates in gologit2 model:

1- There is an interaction between the overturned run-off-road crash and vehicle type and between run-off-road crash for a vehicle leaving the road to the right or left or for a vehicle crashing into a fixed object and alcohol use. An ROR crash involving an overturned large vehicle (bus and truck) or emergency vehicle (police car, fire truck and ambulance) lead into higher probabilities of sustaining higher levels of injury severity.
2- There is an interaction between posted speed limit and vehicle type. ROR crashes involving large vehicles on roads with posted speed limits over 40 mph have higher probabilities of causing injuries while passenger cars (cars, pickup trucks, SUVs, and minivans) and motorcycles have elevated likelihood of injuries on roadways with posted speed limits under 40 mph. Generally, higher posted speeds increase injury severity.

3- There is an interaction between posted speed limits, vehicle type and road contour. Specifically, large vehicles and emergency vehicles crashing on curved level and curved graded road segments with high posted speed limits have elevated likelihood of causing injuries.

4- There is an interaction between posted speed limits, vehicle type, and gender. Specifically, females in cars or on motorcycle crashing on higher posted speed limits have higher chances of sustaining injuries when compared with males in the same situations.

5- There is an interaction between posted speed limit, vehicle type, gender, and drug-related variables. Specifically, if males have drug involvement in cars or on motorcycle crashing on higher posted speed limits they tend to have higher chances of sustaining injuries.

6- There is an interaction between alcohol use, vehicle type and posted speed limit. ROR crashes involving passenger cars, motorcycles, and emergency vehicles and alcohol use on roads with posted speed limits over 40 mph have higher probabilities of causing injuries.
7- There is an interaction between road condition, gender, and ROR crash type. Specifically, overturning/rolling over crashes involving males occurring on wet road conditions tend to lead into higher chances of sustaining injuries.

8- There is an interaction between road condition, and alcohol use. Specifically, crashes occurring on wet surface conditions and also involving alcohol have an elevated likelihood of sustaining injuries.

9- There is an interaction between road condition, gender, and posted speed limit. Specifically, crashes occurring when the road surface is in snowy condition, involving males on roads with posted speed limits of 55 mph or higher have higher chances of sustaining injuries.

10- There is an interaction between road condition, gender, and alcohol use. Specifically, crashes occurring when the road surface is in snowy condition, involving males and alcohol use have higher chances of sustaining injuries.

All the predictor variables that the decision tree modeling found significant were also confirmed to be significant by the generalized ordered logit model. The results of the generalized ordered logit show that roadway curves and grades are features that increase the likelihood of injuries of all levels while grades tend to have moderate influence. The horizontal curves on level road segments are the most hazardous locations. Alcohol-and drugs-related crashes have very strong effects on injury severity of crashes as they tend increase more the likelihood of sustaining incapacitating and fatal injuries. Alcohol-related crashes (coded as 1) have positive coefficients in both panels with increasing values indicating that alcohol use is significant in increasing the likelihood of injuries especially sustaining higher levels of injuries such as incapacitating and fatal injuries.
In terms of gender, females were found to be more likely to sustain injuries compared to males but they sustain more of possible/non-incapacitating than incapacitating/fatal injuries. Also, in terms of run-off-road crash types, overturn/rollover crashes were found to increase the chances of injuries while run-off-road left/cross median/centerline and crash with fixed objects all increase the likelihood of no-injury crashes. For roadway condition, dry roadway condition was the only one that increased the likelihood of injury crashes while wet/standing or running water and snow/ice/sand/mud/etc. simply increase no injury crashes. In terms of vehicle type, buses, trucks and emergency vehicles increase the likelihood of sustaining all kinds of injuries especially incapacitating and fatal injuries. For roadways with posted speed limits of 40 mi/h or higher tend to increase the likelihood of sustaining injuries related to ROR crashes.

Significant factors that increase the likelihood of run-off-road injury severity, which have been identified in the current study and are in agreement with some previous studies include curves (e.g., Hall and Zador, 1985; Zeeger et al., 1988; Viner, 1995; Newman et al., 2003; Liu and Subramanian, 2009; Zhu et al., 2010; Roy and Dissanayake, 2011), alcohol involvement (e.g., Spainhour and Mirsha, 2008; Liu and Subramanian., 2009; Roy and Dissanayake, 2011), higher posted speed limits (e.g., Liu and Subramanian, 2009; Roy and Dissanayake, 2011). Although weather condition was not significant in the current study, however, it was significant in some previous studies (e.g., Liu and Subramanian, 2009; Roy and Dissanayake, 2011), adverse roadway conditions such as slippery roads, snow and ice on the roadway surfaces, etc., were
significant factors in Liu and Subramanian (2009), Roy and Dissanayake (2011), but they were not significant in the current study.

The current study analyzed injuries sustained by all victims of run-off-road crashes that occurred in the state of Ohio during the five-year period, from 2008 to 2012. The results from the current study are important in identifying factors that increase the likelihood of people who get involved in ROR crashes sustaining various injury severity levels. It is recommended that characteristics of drivers who get involved, or more correctly, drivers who cause the occurrences of such crashes and other pertinent contributing factors should be studied in order to identify vital countermeasures of reducing the occurrences of ROR crashes on Ohio’s roadways and streets.
REFERENCES


“Characteristics of CrashesAttributed to the Driver Having Fallen Asleep.” In
_Accident Analysis and Prevention_, Vol. 27, 1995, pp. 769-773.

Peng, Y. and Boyle, L.N. “Commercial Driver Factors in Run-off-road Crashes.” In
_Transportation Research Record: Journal of the Transportation Research Board_,

Persaud, B.N. and Mucsi, K. “Microscopic Accident Potential Models for Two-Lane
Rural Roads.” In _Transportation Research Record: Journal of the Transportation
Research Board_, No. 1485, 1995, pp. 134-139

Pisani, V.D. “Multi-dimensional Assessment and Variably Intense Interventions: A
Systems Approach to DUI.” In _Transportation Research Record: Journal of the
Transportation Research Board_, No. 1047, 1985, pp. 93-96.

Richardson, J., Kim, K., Li, L., and Nitz, L. “Patterns of Motor Vehicle Crash
Involvement by Driver Age and Sex in Hawaii.” In _Journal of Safety Research_,

Roy, U. and Dissanayake, S. “Comparison of Factors Associated with Run-off-road and
Non-run-off Road Crashes in Kansas.” In _Journal of the Transportation Research


