CHARACTERISTICS OF DRIVERS WHO CAUSE 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
Abdullah Faleh Alruwaished
Dayton, Ohio
August, 2014
CHARACTERISTICS OF DRIVERS WHO CAUSE RUN-OFF-ROAD-CRASHES ON OHIO ROADWAYS

Name: Alruwaished, Abdullah

APPROVED BY:

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

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

_________________________________
Paul Goodhue, P.E., PTOE
Committee Member,
Senior Traffic Engineer & Project Manager,
IBI Group

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

_________________________________
Eddy M. Rojas, Ph.D., M.A., P.E.
Dean, School of Engineering
ABSTRACT

CHARACTERISTICS OF DRIVERS WHO CAUSE RUN-OFF-ROAD-CRASHES ON OHIO ROADWAYS

Name: Alruwaished, Abdullah
University of Dayton

Advisor: Dr. Deogratias Eustace

A vehicle that leaves its travel lane at a non-intersection location and collides with another vehicle or with a fixed object or overturns is considered to be involved in a run-off-road (ROR) crash. ROR crashes also known as roadway departure crashes, and these include head-on crashes, crashes that occur due to lane shifts, and crashes where the vehicle leaves its designated travel lane. The main objective of this study was to identify the significant factors that lead to these types of crashes. Crash data used in this study were obtained from the Ohio Department of Public Safety for a five-year period from 2008 to 2012.

The classification tree modeling was used in this study to investigate the significant predictor variables of crash severity of ROR crashes. In addition, this thesis study developed two models, the ROR crashes model and the non-run-off-road (NROR) crashes model. The NROR crashes model used crash data for drivers who were at fault when their crash incidents occurred and for ROR crashes; it was assumed
that all drivers in this category were at fault of causing the crashes. The ROR model identified nine variables, which include road condition, collision type, alcohol related, posted speed limit, speed related, crash type, vehicle type, gender, and age. The NROR crashes model has six significant predictor variables including collision type, posted speed limit, speed related, road condition, alcohol related, and vehicle type.
Dedicated to my parents, wife, and daughter.
ACKNOWLEDGEMENTS

First of all, without the power that our God gave me at the time I was working on this thesis, I couldn’t have successfully completed it; so I would like to thank the Almighty God for the power he gave me as well as the guidance, and support I got from him. I would like to extend my warmest appreciation to my advisor, Dr. Deogratias Eustace, for all his assistance and support. I considered it an honor to work with Dr. Peter Hovey, who helped me create the model for this study and supported me with statistics applications pertaining to my research. I am indebted to my many colleagues who supported me and motivated me to complete my study on time. In addition, I would like to thank Jubail University College and Aramco chair at Dammam University for their support. Finally, I would like to thank my family and particularly my parents who prayed for me to achieve my goals.
TABLE OF CONTENTS

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

DEDICATION ...................................................................................................................... v

ACKNOWLEDGEMENTS ...................................................................................................... vi

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

LIST OF TABLES ................................................................................................................ xiii

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 Characteristics ................................................................................................. 5

2.2.1 Alcohol Crashes Related ......................................................................................... 6

2.2.2. Fatigue and Drowsiness ......................................................................................... 8

2.2.3. Speeding ................................................................................................................ 9
2.2.4. Age of the Driver ......................................................... 10
2.2.5. Gender of the Driver .................................................. 11
2.2.6. Highway Geometric Factors ........................................ 11
  2.2.6.1. Horizontal Alignment Curves .................................. 11
  2.2.6.2. Vertical Alignment Curves .................................... 12

CHAPTER THREE - METHODOLOGY AND DATA COLLECTION .......... 13
3.1 Data Collection .............................................................. 13
  3.1.1 Crash Data ............................................................. 13
3.1.2. Merging Files .......................................................... 14
    3.1.2.1 Crash Records .................................................... 14
    3.1.2.2 Unit Records ..................................................... 14
    3.1.2.3 People Records .................................................. 14
    3.1.2.4 Process of Merging the Crash Files ......................... 15
3.1.3 Creating Driver Database ............................................. 15
3.1.4 Creating ROR and NROR Crash Database ......................... 16
3.2. Description of Selected Study Variables ................................ 16
3.3. Methodology ............................................................... 11
  3.3.1. Introduction ......................................................... 11
  3.3.2. Classification Tree Modeling .................................... 20
    3.3.2.1. Introduction ................................................... 20
    3.3.2.2 Tree Pruning ..................................................... 23
    3.3.2.3 Validation ....................................................... 24
    3.3.2.4. Statistical Analysis ......................................... 24
3.3.2.4.1 Node Splitting Criteria .............................................. 24

CHAPTER FOUR – RESULTS .......................................................... 26

4.1. Introduction ................................................................................. 26
4.2 Descriptive Results of ROR and NROR Related Crashes............... 62
4.3 Results of Classification Tree Modeling for Predictor Parameter Screening
......................................................................................................................... 43
4.3.1 General .................................................................................. 43
4.3.2. Classification Tree for ROR Crash Model ......................... 43
4.3.3. The Classification Tree for NROR Crash Model ................. 48
4.4. Discussion.................................................................................. 52

CHAPTER FIVE - CONCLUSIONS AND RECOMMENDATIONS .............. 59

5.1 Conclusions ............................................................................... 59
5.2 Recommendations ....................................................................... 60

REFERENCES ................................................................................... 62
LIST OF FIGURES

Figure 2.1. Alcohol-Impaired Fatality Rates for Montana, 2001-2007 .................. 7

Figure 2.2. Percentage of Fatality and Injury Due to Run-Off-Road Crashes by Road Contour .......................................................... 12

Figure 3.1. The Process of Creating a Joint File ............................................ 15

Figure 3.2. An Example of a Classification Tree Flowchart .............................. 22

Figure 4.1. Crash Severity Percentage by Crash Type .................................... 27

Figure 4.2. Relative Occurrences of Different Types of ROR Crashes .................... 28

Figure 4.3. Relative Crash Severity for Different ROR Crash Type .................... 28

Figure 4.4. Crash Types for Drivers in Error by Posted Speed Limit ................... 29

Figure 4.5. ROR Crash Types by Speed Limit .............................................. 30

Figure 4.6. Percent of Crash Types by Vehicle Type ..................................... 30

Figure 4.7. Percent of ROR Crashes by Vehicle Type ...................................... 31

Figure 4.8. Percent of Crash Types by Weather Condition ............................. 32

Figure 4.9. Percent of Crash Types by Gender of Driver Who Caused the Crash ... 32

Figure 4.10. Percent of Crash Types by Age of the Driver Who Caused the Crash .. 33
Figure 4.28. Splitting of Collision Type in the ROR Model Tree …………………. 54

Figure 4.29. Splitting of Alcohol Use in the ROR Model Tree …………………. 55

Figure 4.30. The First Split of NROR Crashes from the Root Node ……………… 56

Figure 4.31. Splitting of Collision Type in the NROR Model Tree ………………… 57

Figure 4.32. Splitting of Speed Limit in the NROR Model Tree ………………….. 58
# LIST OF TABLES

3.1. Description of crash and vehicle characteristics .............................................. 17

3.2. Description of human and driver characteristics .............................................. 18

3.3. Description of roadway and environmental characteristics ............................. 19

4.1. Percent of crash distribution based on some variables ..................................... 42

4.2. Column contributions results for ROR crashes model ..................................... 46

4.3. Column contributions results for NROR crashes model .................................... 50
CHAPTER ONE

INTRODUCTION

1.1 Introduction

Due to the ever-increasing population, there are more motor vehicles on the road than ever before and this trend is expected to continue to grow. Unfortunately, as the number of motor vehicles traveling on our roadways increases, so too will the number of traffic crashes unless ways can be found to prevent them. Run-off-road (ROR) crashes have become a major concern in the state of Ohio as they continue to cause fatalities and major injuries to motorists. Run-off-road crashes are defined as types of crashes where a vehicle leaves the lane of the road or the transport way it is traveling on and hits one or more objects or overturns. ROR crashes are also known as roadway departure crashes. An ROR crash may involve a single vehicle or it may involve more than one vehicle such as a head on collision. Hitting another vehicle traveling in the opposite direction, hitting trees and other roadside objects, hitting a pedestrian or overturning are examples of ROR crashes. ROR crashes are non-intersection crashes. NHTSA reports that single-vehicle crashes cause 60 percent of all fatal crashes and run-off-road crashes contribute about 71 percent of these fatal single-vehicle crashes (LeRoy et al., 2008).

In 2008, American Association of State Highway and Transportation Officials (AASHTO) reported that out of all crashes in the state of Idaho, 24 percent were ROR crashes, and 73 percent of the ROR crashes happened on rural roadways. In addition, in
2008, about 39 percent of vehicle fatalities were due to ROR crashes (AASHTO, 2008). Several reasons have been identified as the contributing factors for the occurrences of fatal single-vehicle ROR crashes and these may include alcohol use, weather condition, rural roads, traveling at high speed, design of the roadway, and curved road segments (Liu et al., 2009).

1.2 Problem Statement

There are many reasons that may cause a vehicle to leave its designated lane these reasons can be due to driver characteristics, roadway characteristics, and vehicles characteristics. One of the most important factors, however, is the roadside features because when the vehicle leaves its lane these factors play a major role in the severity of the traffic crash.

There are two types of ROR crashes, which can be defined as follows:

1- Run-off-road crashes to the right occur when a vehicle leaves its lane to the right and departs the roadway and hits objects on the roadside or simply overturns.

2- Run-off-road crashes when a vehicle leaves its lane to the left and gets through the opposite traffic, a run-off-road crash to the left occurs.

Run-off-road crashes have become such a major concern for the state of Ohio that the Ohio Department of Transportation (ODOT) is attempting to reduce the number of roadway departure crashes. ODOT (2013) reports between 2006 and 2010, there were 293,085 ROR crashes that occurred throughout the state and most of those crashes took
place in rural areas, and that most of the vehicles crashed into a ditch or a fixed object. In addition, these crashes resulted into 3,023 fatalities, and 126,204 injuries. Between 2007 and 2011, according to ODOT (2013), 65 percent of ROR fatal crashes had run to the left side of the road and hit a static object. ROR crashes to the left are less common, but still they are very deadly. In the state of Ohio, ODOT considers the trees and utility poles as static objects (ODOT, 2013). The foregoing evidence shows that ROR crashes are among the major public health problems. It is important to classify factors that contribute to the occurrence of ROR crashes in order to discover which of these are the most prevalent in causing fatalities so that targeted programs can be designed to eliminate or minimize them. The main purpose of this research was to identify the most common factors that likely contribute to the occurrence of ROR crashes versus those that contribute to the occurrence of NROR crashes.

1.3 Research Objectives

The main purpose of this thesis study was to use the information gathered from state of Ohio crash database to study characteristics of drivers who cause the occurrences of ROR crashes in order to determine the primary factors that increase the probability of ROR crashes. This study compared three types of crashes by creating a model using the characteristic of the drivers. Developing a statistical model that estimates the level of impact by traffic, environment, geometric, and driver characteristics usually recorded in crash databases, was the aim of this research. The driver’s age, driver errors, and gender of driver are examples of driver characteristics that can be analyzed from such crash databases.
Fatal injury, incapacitating injury, non-incapacitating injury, possible/invisible injury, and no injury are the five injury levels used to classify injuries individuals sustain in traffic crashes. Likewise, fatal crashes, injury crashes and property damage only (PDO) crashes are used in classifying traffic crash severity. Crash severity is classified based upon the condition of the most severely injured individual involved in the crash. ODOT classifies a fatal crash as a crash in which at least one person is killed; an injury crash as a crash in which at least one injury occurs and a property damage crash as a vehicle damage only crash in which no injury occurs (ODOT, 2008).

1.4 Organization of the Thesis

This thesis is organized into four chapters. The second chapter presents the literature review and summarizes some of the topics that relate to this study. The third chapter puts forth the methodology used in this thesis, describing the data and providing an explanation of each variable studied. Chapter four presents results and discussion of results. Chapter five includes conclusions and recommendations.
CHAPTER TWO
LITERATURE REVIEW

2.1 Introduction

Many factors contribute to the occurrence of traffic crashes. We cannot say that only one factor leads to a traffic crash because multiple factors may be involved. Traffic crash factors can be grouped into four categories:

- **Driver characteristics:** Factors related to the driver, such as driver errors, speed choice, age of driver, driver gender, status of the driver’s condition such as intoxication, drowsiness, etc.

- **Geometric factors:** Includes the road design such as the lane width, the width of shoulders, the curvature of the road, how many lanes in each direction, and other features and obstacles.

- **Environmental factors:** These are factors, which are very difficult to control; for example, the condition of the weather, the condition of the road, the light condition, and the time of the day and the day of the week.

- **Traffic factors:** These factors include collision type, speed related, traffic volume, speed limit.

2.2 Driver Characteristics

Driver characteristics are considered as one of the major factors contributing to motor
vehicle crashes. The following sections discuss some of the findings from previous research efforts where driver characteristics are significant contributors to traffic crashes.

### 2.2.1 Alcohol Crashes Related

In North Carolina, researchers studied the percentage of crashes caused by the drunk drivers (Brewer et al., 1994). They used a dataset covering ten years between 1980 and 1989 and concluded that drunk drivers caused 83 percent of all fatal crashes. Of the drivers who died in these crashes, 7,499 had blood alcohol concentrations above the legal limit. They used unconditional logistic-regression analysis to calculate the crude odds ratios and they reported 95 percent confidence intervals (Brewer et al. 1994).

Liu and Subramanian found that the possibility of being involved in ROR crashes for the driver with alcohol use is higher than the sober drivers (Liu et al. 2009). Montana Department of Transportation (MDOF) reports that there were 38.3 percent alcohol related fatalities in 2007 caused by ROR crashes. Figure 2.1 shows alcohol-impaired fatality rates for the state of Montana between 2001 and 2007 (NHTSA, 2010).
NHTSA reports that 17,602 people died in vehicle accidents around the U.S because they were under the influence of alcohol (NHTSA, 2006). In the same year, they reported that 41 percent of the fatal crashes around the world were related to alcohol use (LeRoy et al., 2008). In the state of Ohio, out of all traffic crashes that resulted in deaths, one out of three involved alcohol use (NHTSA, 2012). In 2011, there were 1,016 fatal crashes in Ohio, out of these crashes, 31 percent involved drivers who were driving under
the influence of alcohol (NHTSA, 2012). A study by Massie and Campbell (1993) found that 2.0 percent of the daytime traffic crashes involved drivers who were driving under the influence of alcohol, but nearly 22.8 percent of the traffic crashes that occurred at night involved the use of alcohol. Of all fatal nighttime single-vehicle crashes where the driver was considered to be under the influence of alcohol, 37.5 percent were females, 49.4 percent were males, and only 13.1 percent were non-alcohol related crashes (Massie and Campbell, 1993).

### 2.2.2 Fatigue and Drowsiness

A study by Knipling and Wang (1994) used 280,000 crashes that were caused by drowsiness and fatigue (DAF) drivers between 1989 and 1993. The number of DAF-related crashes represented 1 percent of the total number of crashes during that time. They also found that 55 percent of crashes caused by drowsiness occurred between the hours of 12:00 AM and 7:59 AM while 18 percent of these crashes occurred between the hours of 1:00 PM and 4:59 PM. Most of these crashes occurred in non-urban areas that have a 55-65 mph posted speed limits (Knipling and Wang, 1994).

Around the United States, the National Sleep Foundation (NSF) reports that in the fall of 2004, 1,456 adult drivers fell asleep while they were driving (NSF, 2005). The NSF further reports that 60 percent of those drivers said that they drove their vehicles at least once when they were feeling drowsy and 37 percent said that they drove one time while they were drowsy (NSF, 2005). In Canada, Burns et al. studied 1,209 crashes where they found that 57 percent admitted to driving while they were tired, and 20 percent had
slept on the road (Burns et al. 2005). NHTSA reports that 11 percent of ROR crashes had fallen asleep in the past year while they were driving (Royal, 2003).

Masa et al. (2000) studied automobile crashes caused by sleeping drivers in the western Spanish city of Cáceres. They found that 31 percent of the sleepy drivers were females and male drivers made up 69 percent. Out of all sleepy female drivers, almost 52 percent of them were between the ages of 18 and 34 and 4.3 percent were between the ages of 55 and 84, and almost 44 percent were between the ages of 35 and 54. Out of all sleepy male drivers, the ages between 35 and 54 had the largest percent of crashes at almost 43 percent. Sleepy males between 18 and 34 made up the second largest group of male driver’s crashes at almost 34 percent. Sleepy male drivers between the ages of 55 and 84 had the lowest crash rate, which was almost 24 percent (Masa et al., 2000).

2.2.3 Speeding

A study by Knipling and Wang (1994) found that rural areas having speed limits of 55-65 mph were the areas that most of the speeding-related crashes occurred. A study by Liu et al. (2005), which analyzed speeding-related motor vehicle traffic crashes using NHTSA’s Fatality Analysis Reporting System (FARS) reports that almost one-third of traffic-related fatalities occur to vehicles crashed deemed that speeding was involved in their respective traffic crashes. Roy and Dissanayake (2012) using data from Kansas found that driving too fast for the conditions (speeding) was one of the major factors contributing to ROR crashes (Roy and Dissanayake, 2012).
Liu and Subramanian report that 90 percent of the speeding vehicle crashes involved ROR crashes while only 59.5 percent of non-speeding vehicle crashes involved ROR crashes (Liu et al., 2009). In a study done in Adelaide, Australia reports that in 1992 there were 151 fatal vehicle crashes on roads with speed limits of 60 Km/h (Kloeden et al., 1997). Out of the 151 crashes, there were 14 single vehicle ROR crashes, while 8 out of the 14 were involved in collisions with objects (Kloeden et al., 1997). One of the most cited factors in motor vehicle crashes and fatalities is speeding (Roy et al. 2012).

2.2.4 Age of the Driver

Chen et al. (2000) conducted a nationwide study that used crash data collected between 1992 and 1997. From this study, they found that adult drivers had significantly lower risks for crashes than 16 and 17 years old drivers because the younger drivers speed more often and they are less experienced. They also found that between midnight and 5:59 AM drivers between 16 and 17 years old were responsible for the highest driver crashes. Data of crashes for older drivers (age older than 64 years) was reported in the state of Iowa between 1990 and 1999. Over this ten-year period, they found that 17,045 accidents involved drivers over the age of 65 and out of this a total 9,107 (53.4 percent) involved male drivers. Drivers between the ages of 76 and 85 accounted for 62.1 percent of the total number while those drivers 85 and older caused only 5.6 percent of the accidents. In addition, 87.8 percent of these accidents happened during daylight hours. Ordered probit and ordered logit models were used in this study because the dataset involved ordinal data (Khattak et al. 2002). Liu et al. (2009) found that young drivers 15 to 24 have a
higher likelihood of getting involved in fatal ROR crashes and drivers over 65 years of age come on the second place. The possibility that young drivers got involved in fatal crashes was 75.2 percent, and the possibility that drivers between age 25 to 44 to get involved in fatal crashes was 68.6 percent. Drivers between the ages of 45 and 64 came out last at 66.4 percent (Liu et al. 2009).

2.2.5 Gender of the Driver

Based on the 2001 Florida crash database, Yan and Radwan (2006) found that male drivers have a greater probability of being involved in run off road crashes and also they have a higher likelihood of being involved in rear end crashes (Liu and Subramanian, 2009). McEvoy et al. (2005) studied the effect of cell phones on road safety in Perth, Western Australia. The data for this study was taken between April 2002 and July 2004, and involved drivers aged 17 years and older. They found that 456 drivers involved in crashes were using cell phones (McEvoy et al. 2005). They also found that 423 (93 percent) of the drivers had at least one injury, 201 (44percent) had more than one injury, and just 33 drivers had no injury. Furthermore, their research indicates that female drivers had more crashes than males where 264 crashes involved females, and 192 crashes involved males (McEvoy et al. 2005).

2.2.6 Highway Geometric Factors

2.2.6.1 Horizontal Alignment Curves

A study by Eustace et al. (2013) investigated characteristics of ROR injury severity by using Ohio crash data. In that study, they found that 21 percent of the total traffic injuries
that occurred on Ohio’s public roads and streets between 2008-2010 involved run-off-
road crashes. In terms of injuries, curved level roads and curved graded roads were more
dangerous than straight roads as depicted in Figure 2.2 (Eustace et al. 2013).

Figure 2.2. Percentage of Fatality and Injury Due to Run-Off-Road Crashes by Road

Contour

Between 2003 and 2008, Bauer et al. report states that 43.5 percent of the crashes
that occurred on horizontal curves were fatal and injury crashes and 56.5 percent of the
crashes that occurred on the horizontal curves were property damage only (Bauer et al.,
2013).

2.2.6.2 Vertical Alignment Curves

Bauer et al. summary report further notes that between 2003 and 2008, about 41.9 percent
of the crashes that occurred on vertical alignment curves were fatal and injury crashes
while 58.1 percent of the crashes involved property damage only (Bauer et al., 2013).
CHAPTER THREE
METHODOLOGY AND DATA COLLECTION

3.1 Data Collection
The Ohio Department of Public Safety (ODPS) provided five years of motor vehicle crash data for 2008 to 2012. These are statewide police reported crash data for traffic crashes that occurred on Ohio’s public roads and streets.

3.1.1 Crash Data
Five years of traffic crash data were downloaded from the Ohio Department of Public Safety (ODPS) website, and the latest complete datasets were for 2008 through 2012. ODPS website has a specific way of organizing their data because they store them into four different crash files and they use the calendar year to relate them together. Crash records, unit records, people records, and ODOT records are the four names of the files that ODPS use in storing their data. DOCNO (document number) is a variable, which identifies specific crash events, and it is available in each file. UNITNO (unit number) is available in both unit record file and people file, and it represents each vehicle and it is used to relate each person to his/her vehicle. By using UNITNO and DOCNO variables all the files can be properly merged together. UNITNO and DOCNO are very important variables because without them we cannot have all the crash information properly related together in one complete file. For the purpose of this study, only three files, the crash
records file, unit records file, and people records file were used since all the required variables are stored in these three files.

3.1.2 Merging Files

Each of the files used in this study, that is, crash records, unit records, and people records, are described below.

3.1.2.1 Crash Records

This file covers all the information about the crash, and reports all the details needed for the crash. Crash specific information recorded in this file include crash record number, crash severity, vehicle in error, date of crash, time of crash, crash location, type of road, if alcohol or drugs were involved, if speeding was involved, etc.

3.1.2.2 Unit Records

This file holds all the information that is related to the unit or to a specific vehicle, which was involved in the crash. Units information recorded in this file include crash record number, unit number, unit type (e.g., motor vehicle, motorcycle, bicycle, pedestrian, etc.), point of impact, number of occupants in the unit, etc.

3.1.2.3 People Records

All the information about each person involved in the crash is recorded in the people file. Information recorded in people record file includes crash record number, unit number, person type (e.g., driver, occupant, or pedestrian), age, gender, severity of injury sustained by an individual, safety equipment used, etc.
3.1.2.4 Process of Merging the Crash Files

SPSS software (version 20.0) was used to merge the three files together into one file. As described in Section 3.1.1 DOCNO was used to merge the units file with the crash file because DOCNO is the only unique variable in both files that connects all entries, which are related to each particular crash event. Then the DOCNO and UNITNO were used to join the new file (crash + unit) to people file by connecting people to their correct vehicles and crash events; then a single file (crash + unit + people) with complete crash information for each year was created. The final step was to join all the five years data into one file. At the end, a final check was needed to be sure that all the files were joined properly. Figure 3.1 illustrates the file merging process.

3.1.3 Creating Driver Database

The joined file has a variable known as PERSONTYPE; this variable provides information occupant person involved in a traffic crash. This variable has three classes indicating different ways a person can be recorded, namely the driver (D), vehicle passenger (occupant) (O), and the pedestrian (P). From the joined file, querying and sorting records using the PERSONTYPE variable equal to D the driver) created a new file. Therefore, a new file, which contains drivers-related database only, was created.
3.1.4 Creating ROR and NROR Crash Database

The crash data in the driver database file was divided into two crash types. The first type was defined as run-off-road (ROR) crashes and the other type was defined as non-run-off-road (NROR) crashes. A variable known as SEQUENCEEVENT1 was used in defining which crash cases are ROR and NROR. If the vehicle in the SEQUENCEEVENT1 variable was described as overturned/rollover, run-off-road right, run-off-road left, cross median/centerline, or crash with a fixed object, then it was categorized as an ROR crash, otherwise it was termed as an NROR crash. For ROR crashes, it was assumed that all drivers were in error when the crash occurred. The aim here was to create a crash data file containing only drivers, who were in error, i.e., drivers who were judged responsible in causing the crash events. For NROR crash cases, another variable known as VEHICLEINERROR, a variable recorded with a “yes” if the driver was cited for being responsible and with a “no” if the driver was not cited was used in determining the vehicle in error. At the end, a final file was created that contained ROR-related and NROR-related crashes in which drivers were in error. As a result, the final file contained a total number of 1,652,375 records of crashes in which drivers were in error that occurred between 2008 and 2012 on Ohio’s public roads and streets. The final records indicate that 24.5 percent (308,232) of the total crash records involved ROR crashes, and 75.5 percent (948,143) were NROR crashes whose drivers were in error.

3.2 Description of Selected Study Variables

Tables 3.1 through 3.3 summarize data characteristics for traffic crashes that occurred on Ohio’s public roads and streets and these data are only for drivers who were deemed
responsible for the crash occurrences. Table 3.1 summarizes the crash and vehicle characteristics. Variables in this category include crash severity, which was treated as a dependent variable and independent variables are vehicle type and posted speed limit.

Table 3.1 Description of crash and vehicle characteristics

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
<th>Frequency (Percent)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Crash severity</td>
<td>Fatal</td>
<td>4,459 (0.35 percent)</td>
</tr>
<tr>
<td></td>
<td>Injury</td>
<td>350,412 (27.9 percent)</td>
</tr>
<tr>
<td></td>
<td>PDO</td>
<td>901,504 (71.75 percent)</td>
</tr>
<tr>
<td>Vehicle type</td>
<td>Passenger car</td>
<td>1,187,681 (95.5 percent)</td>
</tr>
<tr>
<td></td>
<td>Med/heavy trucks</td>
<td>57,797 (4.6 percent)</td>
</tr>
<tr>
<td></td>
<td>Bus/Van/Limo</td>
<td>10,897 (0.9 percent)</td>
</tr>
<tr>
<td>Posted speed limit</td>
<td>&lt; 40 mph</td>
<td>699,202 (55.7 percent)</td>
</tr>
<tr>
<td></td>
<td>40-50 mph</td>
<td>213,765 (17 percent)</td>
</tr>
<tr>
<td></td>
<td>55-70 mph</td>
<td>343,408 (27.3 percent)</td>
</tr>
</tbody>
</table>

Table 3.2 summarizes human and driver characteristics. Variables included in this category are gender, age, alcohol related use, drug use, and speeding.
Table 3.2 Description of human and driver characteristics

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
<th>Frequency (Percent)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alcohol-related</td>
<td>Yes</td>
<td>65,872 (5.2 percent)</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>1,199,503 (94.8 percent)</td>
</tr>
<tr>
<td>Drug related</td>
<td>Yes</td>
<td>15,614 (1.2 percent)</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>1,240,131 (98.8 percent)</td>
</tr>
<tr>
<td>Age of driver</td>
<td>&lt;20</td>
<td>202,000 (16.1 percent)</td>
</tr>
<tr>
<td></td>
<td>20-25</td>
<td>239,249 (19 percent)</td>
</tr>
<tr>
<td></td>
<td>26-64</td>
<td>695,580 (55.4 percent)</td>
</tr>
<tr>
<td></td>
<td>65+</td>
<td>119,546 (9.5 percent)</td>
</tr>
<tr>
<td>Gender of driver</td>
<td>Female</td>
<td>540,774 (43 percent)</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>715,599 (57 percent)</td>
</tr>
<tr>
<td>Speeding related</td>
<td>No</td>
<td>1,102,991 (87.8 percent)</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>153,384 (12.2 percent)</td>
</tr>
</tbody>
</table>

Table 3.3 summarizes roadway and environmental characteristics. Variables in this category include light condition, weather condition, roadway condition, time of day, day of week, and road contour.
Table 3.3 Description of roadway and environmental characteristics

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
<th>Frequency ( Percent )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Light condition</td>
<td>Daylight/dawn/dusk</td>
<td>949,811 (75.6 percent)</td>
</tr>
<tr>
<td></td>
<td>Dark - lighted roadway</td>
<td>166,595 (13.3 percent)</td>
</tr>
<tr>
<td></td>
<td>Dark - unlighted roadway</td>
<td>139,969 (11.1 percent)</td>
</tr>
<tr>
<td>Weather condition</td>
<td>Clear</td>
<td>651,638 (51.9 percent)</td>
</tr>
<tr>
<td></td>
<td>Cloudy</td>
<td>318,061 (25.3 percent)</td>
</tr>
<tr>
<td></td>
<td>Rain/fog/sleet/snow/wind</td>
<td>286,676 (22.8 percent)</td>
</tr>
<tr>
<td>Roadway contour</td>
<td>Straight level</td>
<td>894,934 (71.2 percent)</td>
</tr>
<tr>
<td></td>
<td>Straight grade</td>
<td>217,317 (17.3 percent)</td>
</tr>
<tr>
<td></td>
<td>Curve level</td>
<td>66,924 (5.3 percent)</td>
</tr>
<tr>
<td></td>
<td>Curve grade</td>
<td>77,200 ( 6.1 percent)</td>
</tr>
<tr>
<td>Road condition</td>
<td>Dry</td>
<td>838,980 (66.7 percent)</td>
</tr>
<tr>
<td></td>
<td>Wet/water</td>
<td>277,200 (22.1 percent)</td>
</tr>
<tr>
<td></td>
<td>Snow/ice/mud/oil/slush/gravel/other</td>
<td>140,223 11.2 percent)</td>
</tr>
<tr>
<td>Time of crash</td>
<td>Morning</td>
<td>222,943 (17.9 percent)</td>
</tr>
<tr>
<td></td>
<td>Midday</td>
<td>467,284 (37.5 percent)</td>
</tr>
<tr>
<td></td>
<td>Evening</td>
<td>299,556 (24 percent)</td>
</tr>
<tr>
<td></td>
<td>Early night</td>
<td>126,379 (10.1 percent)</td>
</tr>
<tr>
<td></td>
<td>Late night</td>
<td>131,213 (10.5 percent)</td>
</tr>
<tr>
<td>Day of week</td>
<td>Weekends</td>
<td>279,805 (22.3 percent)</td>
</tr>
<tr>
<td></td>
<td>Weekdays</td>
<td>976,570 (77.7 percent)</td>
</tr>
</tbody>
</table>

3.3 Methodology

3.3.1 Introduction

This section discusses the methodology used in order to achieve the objectives of this study. The classification tree model, which is also known as classification tree model was used in analyzing the data in order to determine variables, which have significant effect
on the severity of ROR crashes. This method is frequently used in medicine and business fields to examine customer behaviors and to test the diagnosis of a disease (Lavery et al., 2012). In recent years, classification tree methods have been successfully applied in the field of traffic safety to analyze traffic injury/crash severity (e.g., Yan and Radwan, 2006; Montella et al. 2011, 2012; Eustace et al., 2012; Griselda, et al., 2012)

According to Lavery et al. (2012): “Tree algorithms, or simply trees, split a dataset (assign observations in the data set of 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.” Because it is suitable for use in determining features and extracting patterns in large databases, the classification tree modeling has been widely used to explore data analysis and to predict the modeling applications (Myles et al.2004).

3.3.2 Classification Tree Modeling

3.3.2.1 Introduction

Based on the context of decision-making, a classification tree is a hierarchical model, which assigns a probability to each of the possible choices. The classification tree model works by setting up IF-statements; the aim of these IF-statements is to divide the dataset into smaller homogeneous subgroups. Based on the type of the response variable, the classification tree model selects an appropriate measurement. The response and the predictor variables can be either continuous or categorical.
• For a continuous predictor variable, the model splits the predictor variable into two partitions according to a cutting value of the predictor variable.

• For a categorical predictor variable, the model splits the predictor variable categories (levels) into two groups of levels.

• For a continuous response variable, the model splits the data at a known level and the cutting value made by the model is determined by maximizing the log Worth value. In addition, the quantity related to the p-value based on the difference in means is used in conjunction with the Log Worth.

• For a categorical response variable, the model splits the data at a known level and the cutting value made by the model is determined by maximizing the Log Worth. In addition, the quantity related to the likelihood ratio chi-square statistic is used in conjunction with the Log Worth value.

It is important note that the fitted values of the continuous response variable are the means within the two groups, and the fitted values of the categorical response are the estimated proportions within the groups. There several advantages of using the classification tree model, some of these advantages include:

• It is quicker to make predictions.

• Understanding the important variable in making the decision is very easy.

• Making the prediction by averaging all the leaves in the sub-tree if some of the data is missing.

• There are fast and reliable algorithms.

• A jagged response will be given at the end of the model, so it can work when the true regression surface is not smooth.
Figure 3.2 shows an example of a classification tree model. Each classification tree model contains nodes, those nodes can either be branch nodes, internal nodes (including the root node), leaf nodes, or terminal nodes. In the classification tree literature, there are many names used for the highest node, either they name it as node A, node 1, or root node, and the highest node contains the whole dataset. The different nodes are defined as follows:

- A root node is the node that does not have any incoming edge and one or more outgoing edges (does not have a parent).
- An internal node is a node that has only one incoming edge and two or more outgoing edges (it’s not leaf node).
- A leaf or terminal node is a node that has only one incoming edge and zero outgoing edge (node without children).

The remains node contains a subset of the entire dataset. Every branch node is considered a parent node, and each parent node has a node known as a child node. For an example illustrated in Figure 3.2, node A is divided into two nodes, node B and node C, so node A is a parent of nodes B and C and node C is a parent of nodes and E.
The pathways of a classification tree model are symbolized by using lines. These lines link the parent nodes to the child nodes. In essence, the first step in the classification tree model is splitting the root node or the original dataset into two subgroups based on the attribute’s test value. This process is repeated on each derived subgroup until the last result or the last decision is reached. This process is also known as recursive partitioning. The process will not stop until the subgroups at that node have the same value of the response variable in all subgroups; or when the model reaches a point that will not have any improvement in the prediction. At this point, the node becomes a leaf or terminal node (e.g., nodes B, D, and E in Figure 3.2).

### 3.3.2.2 Tree Pruning

By merging disjuncts (Individual components) that are adjacent in instance space, the pruning will simplify the classifier. One of the advantages of pruning is to improve the classifier performance by eliminating the error-prone components, and to reduce the size of the classification tree model.

The important thing is that pruning must not eliminate the predictive parts of the classifier. For that reason, the way of deciding that the set of disjuncts is productive or not, or whether the disjuncts should be merged into a single, larger disjunct need the pruning procedure to have a mechanism. Commonly, the step of reducing the size of the model is done after the original tree is built. Another advantage of the pruning is to make the final model easy, to improve the accuracy of the predictions by decreasing the problem of over fitting the model, and to minimize the chances of random errors. In addition, based on the noise or erroneous data, pruning helps in detecting and removing
splits of the classification tree. The result of the model being overly complex or having too many predictive variables is over fitting. Over fitting model will result in to a model with a poor predicting power because, there will be a small variation in the data.

3.3.2.3 Validation

There are many techniques used to examine the model if it is strong or not. Validation is one of the commonly used techniques to check the predictor variables in predicting the response variable. Under this procedure, the dataset is split into two subsets of data. The first subset, known as training dataset, is used to build the tree model. The second subset, known as validation dataset, is used to evaluate the performance of the model built the training dataset.

3.3.2.4 Statistical Analysis

3.3.2.4.1 Node Splitting Criteria

To grow and prune the classification tree model, the Log Worth statistic is used. This statistic is used to indicate whether the predictor variable is significant or not. When the Log Worth value becomes larger, that means the predictor variable is more significant. It is important thing to note that the model splits the node based on the larger value of the Log Worth statistic and it is calculated as shown in Equation 3.1:

\[
\text{LogWorth} = -\log_{10}(p\text{-value})
\]

Where the adjusted \( p \)-value is calculated by taking into account the number of different ways splits can be made.
For a categorical response variable, the log-likelihood-ratio, $G^2$ is also used in splitting the notes. The log-likelihood-ratio is equal to the entropy times two. Equation 3.2 shows how to compute this statistic:

$$G^2 = 2 \sum [f_o \log \left( \frac{f_o}{f_e} \right)] \tag{3.2}$$

Where:

$f_o$ = the observed frequency in a node.

$f_e$ = the expected frequency in a node.

A $G^2$ value for a candidate noted that is needed for splitting is computed as shown in Equation 3.3:

$$G_{test}^2 = G_{parent}^2 - (G_{right}^2 + G_{left}^2) \tag{3.3}$$
CHAPTER FOUR
RESULTS

4.1 Introduction

Several variables considered to have effect on traffic crash severity were analyzed in this study. This study investigated seventeen variables as possible predictors of ROR and NROR crash severity models, with the aim of capturing the most significant ones. JMP software version 11 was used in developing decision (classification) tree models. Control of significant values for splitting nodes and merging categories were tested at the default significance level of 0.05. The next sections discuss descriptive statistics and classification tree modeling results.

4.2. Descriptive Results of ROR and NROR Related Crashes

Between 2008 and 2012, 1,256,375 drivers were cited for causing traffic crashes. The data shows that 24.5 percent were involved in ROR crashes and 75.5 percent caused NROR crashes. For ROR crashes, there were 2,696 fatality crashes, 103,281 injury crashes and 202,285 PDO crashes. On the other hand, the NROR crashes resulted into 1,763 fatal crashes, 247,161 injury crashes and 699,217 PDO crashes. The crash characteristics that occurred between 2008 and 2012 on Ohio’s public roads and streets for drivers who were cited due to their errors being responsible in causing the crashes are
summarized in Figures 4.1 through 4.19. Figure 4.1 summarizes the ROR and NROR crashes by their crash severity. Figure 4.1 illustrates the seriousness of severity of ROR crashes in the Ohio crash data where ROR crashes constituted only 24.5 percent of all the crashes but they accounted for 60.5 percent of fatal crashes.

Figure 4.1. Crash Severity Percentage by Crash Type

Figure 4.2 provides an insight of different types of ROR crashes in the database. These crashes include overturn/rollover, run-off-road right, run-off-road left/crossing median/centerline, and crash with fixed objects. Figure 4.2 shows that the most prevalent ROR crash type was run-off-road to the right, which made up 49.1 percent of all ROR crashes and run-off-road to the left/cross median/centerline came up second at 37.3 percent.
ROR crashes to the left/crossing median/centerline tend to be more dangerous where 1.2 percent of them were fatal and 36.3 percent resulted into injuries as shown in Figure 4.3. Although overturning and rollover type of ROR crashes seem to be rare, constituting of only 0.6 percent of all ROR crashes, however, 1.0 percent of them were fatal and a whopping 55.6 percent of them resulted into injuries.

Figure 4.2. Relative Occurrences of Different Types of ROR Crashes

Figure 4.3. Relative Crash Severity for Different ROR Crash Type
Figure 4.4 shows the distribution of all crashes in which drivers were in error (i.e., they were judged to be responsible for the crashes they were involved in) by posted speed limits on the highway where the crashes occurred.

![Figure 4.4. Crash Types for Drivers in Error by Posted Speed Limit](image)

For ROR crashes, slightly more than half (50.6 percent) of them occurred on high speed limit roadways, i.e., with 55-70 mi/h posted speed limits whereas 21.9 percent and only 12.6 percent occurred on 40-50 mi/h and lower than 40 mi/h speed limit roads, respectively. However, according to Figure 4.5, on lower speed limit roads (less than 40 mi/h posted speed limits), drivers are more prone to hitting fixed objects than on major roads. Specifically, Figure 4.5 shows that 25.6 percent of ROR crashes that occurred on roadways with speed limits lower than 40 mi/h did hit fixed objects while only 10.2 percent and 7.4 percent of ROR crashes on roads with speed limits of 40-50 mi/h and 55-70 mi/h did hit fixed objects, respectively.
The crash data analyzed in this study show that passenger cars made up 94.9 percent of all crashes, medium and heavy trucks made up 4.6 percent and buses, limos and vans accounted for just 0.5 percent of the crashes. Figure 4.6 shows the percent distribution of crash types by vehicle types, which indicates that buses, limos, and vans are less likely to be involved in ROR crashes when compared with the other two types of vehicles.
Although buses, limos and vans have the lowest percentage of ROR crashes compared with passenger cars and medium/large trucks, but according to Figure 4.7 they tend to hit fixed objects more often (36.5 percent) once they run-off-the-road when compared with other vehicle types (12.5 percent and 24.3 percent for trucks and passenger cars, respectively). An interesting observation is that passenger cars are more likely be involved in run-off-road left cross median/centerline type of ROR crashes than the other vehicle types and trucks are more likely to overturn and rollover than the two type of vehicles.

![Figure 4.7. Percent of ROR Crashes by Vehicle Type](image)

Figure 4.7. Percent of ROR Crashes by Vehicle Type

Figure 4.8 illustrates the impact of weather condition on the type of crashes that drivers in error cause. As expected, ROR crashes made the highest proportion (38.0 percent) of the crashes that occurred when the weather was bad such as in
rainy/foggy/snowy or windy conditions while ROR crashes made only 23.2 percent and 19.2 percent of those occurring when the weather was cloudy and clear, respectively.

Figure 4.8. Percent of Crash Types by Weather Condition

Figure 4.9 shows the percent of crash types by gender of the driver who was in error when the crash occurred. For the crashes they caused, male drivers tended to have higher percent of ROR crashes (26.7 percent) when compared with female drivers whose ROR crashes accounted for 21.6 percent of the crashes they were responsible.

Figure 4.9. Percent of Crash Types by Gender of Driver Who Caused the Crash
Figure 4.10 shows the percent of crash types by age of the driver who was responsible for the crash. The data show older drivers (aged 65 and above) were less likely to be involved in ROR crashes for the crashes they were cited as liable drivers.

Figure 4.10. Percent of Crash Types by Age of the Driver Who Caused the Crash

Figure 4.11 shows crash types by light condition. It is not surprising that ROR crashes tend to be more frequent when it is dark and especially on unlighted roadways as Figure 4.11 illustrates. ROR crashes made up the highest proportion (60.4 percent) for crashes that occurred on dark, unlighted roadways, followed by dark but lighted roadways they made up 28.2 percent of crashes that occurred during that light condition. However, ROR crashes just made up 18.6 percent of all crashes that occurred during daylight, dawn or dusk.
Figure 4.11. Percent of Crash Types by Light Condition

Figure 4.12 shows a breakdown of ROR crash types by light condition. For either lighting condition, the most frequent ROR crash type involves run-off-road to the right. It is interesting to note that crashing with fixed objects was more frequent during darkness on lighted roadways when compared with the two light conditions considered.

Figure 4.12. Percent of ROR Crash Types by Light Condition
Figure 4.13 shows the percent distribution of crash types by type of day of the weekday, which are grouped into weekdays and weekends. ROR crashes were more frequent (33.1 percent) for crashes that occurred during weekends when compared with those that occurred during weekdays for which ROR accounted for only 22.1 percent. This trend may be expected due to weekend reckless driving and generally lighter traffic during weekends compared to busy commuting weekday travels.

![Figure 4.13. Percent of Crash Types by Day of the Week](image)

Figure 4.14 shows the percent distribution of crash type by time of the day. Time of the day was divided into five time groups namely morning, midday, evening, early night, and late night. It is surprising that late nighttime has the largest percent of ROR crashes accounting for 56.9 percent of all crashes that occurred during that time group. Meanwhile evening and midday times had the least percent of ROR crashes (16.1 percent and 17.1 percent, respectively).
Figure 4.14. Percent of Crash Types by Time of the Day

Figure 4.15 illustrates the percent distribution of crash types by road condition. The road condition was divided into four condition groups. These groups include dry, wet/water, snow/ice/oil/gravel, and unknown conditions. Figure 4.15 shows that ROR crashes are by far most likely to occur during adverse road condition. Specifically, about 43.7 percent of crashes that occurred when the road surface was covered with snow, ice, mud, oil, slush or gravel involved ROR crashes, which is much higher percent when compared with the other roadway conditions considered in this study.
Figure 4.15 shows percent of crash types in relation to speeding. Police officers who attended the crash scenes to determine whether speed was a contributing factor judged speeding related crashes. Figure 4.16 reveals that for crashes where speeding was an issue, ROR crashes were more frequent making up of 66.4 percent of such crashes. On the other hand, ROR crashes just made up of only 18.7 percent of all non-speed related crashes. This result also makes sense due to known relationship between vehicle control problems and speeding.
Figure 4.16. Effect of Speeding on Crash Types

Figure 4.17 illustrates the effect of seat belt use behaviors on the crash severity of ROR crashes. As expected, most of the drivers who caused fatal ROR crashes (62.3 percent) did not wear seat belts. On the other extreme end, for drivers who caused the PDO crashes, just a mere of them, 3.7 percent, did not use seat belts. This shows that drivers who cause major crashes are likely not to be belted, which is a dangerous combination.

Figure 4.17. Effect of Seat Belt Use on Crash Severity
Figure 4.18 shows the percent of crash types by roadway contour. Figure 4.18 illustrates how horizontal curves highly influence the occurrence of ROR crashes. ROR crashes composed of about 71 percent of all crashes that occurred on curved (both level and graded) segments.

![Figure 4.18. Percent of Crash Types by Road Contour](image)

Figure 4.18 shows the effect of drivers’ alcohol use to the type of crashes they are likely to cause. Majority of the drivers who were determined to have used alcohol when they got involved in the crashes tended to cause ROR crashes. Specifically, 61.1 percent of the drivers who had used alcohol caused ROR crashes as opposed to only 24.1 percent of non-drunk drivers who caused ROR crashes.
Figure 4.19. Effect of Alcohol Use to Crash Severity

Table 4.1 shows that ROR crashes are more dangerous than NROR crashes because fatal and injury crashes are 34.4 percent of ROR crashes, but for NROR crashes, fatal and injury crashes make up only 26.3 percent. ROR crashes may be considered more dangerous because when a vehicle leaves its designated lane to the left or to the right, the possibility of colliding with more objects or overturning is higher than NROR crashes. For this reason, colliding with another vehicle or roadside fixed objects increases the possibility of resulting into fatal or injury crashes. Only 65.6 percent of all ROR crashes were property damage only, but the percentage of property damage for NROR crashes consisted of 73.7 percent.

In terms of weather condition, 40.8 percent of ROR crashes occurred during dry weather condition, whereas 55.5 percent of NROR crashes occurred during dry weather condition. Most of the crashes that occur during dry weather conditions happen because drivers feel more comfortable to drive and speed. Furthermore, when the weather is cloudy both ROR crashes and NROR crashes have approximately the same chances of happening. On the other hand, when it is raining or snowing, ROR vehicles have a higher
possibility of being involved in crashes than NROR vehicles because the possibility of losing the control of the vehicle is greater.

More than 50 percent of all ROR and NROR crashes involve drivers between 26 and 64 years old because drivers between those ages drive more often and most generation than the other age groups. Therefore, they have a higher probability of being involved in crashes. The age variable shows a significant difference between the two types of crashes when the drivers’ age is more than 64 years old. Table 4.1 also shows that 5.8 percent of ROR crashes involved drivers more than 65 years old, but 10.7 percent of NROR crashes involved drivers more than 64 years old.

In terms of the day of the week for which the crash occurred, there is a significant difference between ROR and NROR crashes. About 30.1 percent of ROR crashes occurred on weekends, while only 19.7 percent of NROR crashes occurred on weekends. On weekends, drivers drink alcohol more often, so the possibility of losing control of their vehicles and running off the road becomes elevated. About 69.9 percent of ROR crashes occurred on weekdays, but 80.3 percent of NROR crashes occurred on weekdays. During weekdays, the percent of ROR crashes becomes less than the percent of NROR crashes because most drivers are most likely to be sober when making commuter trips and high commuter traffic volumes may generally drive down average travel speeds and hence lowering chances of running-off-the road.

Most of ROR and NROR crashes involve passenger cars because those types of vehicles are the most common types of vehicles on the roads. About 62.0 percent of ROR crashes involved male drivers while 55.3 percent of NROR crashes involved male drivers.
About 38.0 percent of ROR crashes and 44.7 percent of NROR crashes involved female drivers. Generally, this means that female drivers have a lower probability of causing run off road crashes than male drivers.

Table 4.1. Percent of crash distribution based on some variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>ROR crashes</th>
<th>NROR crashes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Crash severity</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fatal</td>
<td>0.9 percent</td>
<td>0.2 percent</td>
</tr>
<tr>
<td>Injury</td>
<td>33.5 percent</td>
<td>26.1 percent</td>
</tr>
<tr>
<td>PDO</td>
<td>65.6 percent</td>
<td>73.7 percent</td>
</tr>
<tr>
<td><strong>Weather condition</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dry</td>
<td>40.8 percent</td>
<td>55.5 percent</td>
</tr>
<tr>
<td>Cloudy</td>
<td>23.9 percent</td>
<td>25.8 percent</td>
</tr>
<tr>
<td>Other</td>
<td>35.5 percent</td>
<td>18.8 percent</td>
</tr>
<tr>
<td><strong>Age</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;20</td>
<td>18.7 percent</td>
<td>15.2 percent</td>
</tr>
<tr>
<td>20 – 25</td>
<td>21.3 percent</td>
<td>18.3 percent</td>
</tr>
<tr>
<td>26 – 64</td>
<td>54.2 percent</td>
<td>55.7 percent</td>
</tr>
<tr>
<td>65+</td>
<td>5.8 percent</td>
<td>10.7 percent</td>
</tr>
<tr>
<td><strong>Day of week</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weekend</td>
<td>30.1 percent</td>
<td>19.7 percent</td>
</tr>
<tr>
<td>Weekday</td>
<td>69.9 percent</td>
<td>80.3 percent</td>
</tr>
<tr>
<td><strong>Road condition</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dry</td>
<td>50.2 percent</td>
<td>72.2 percent</td>
</tr>
<tr>
<td>Wet/water</td>
<td>24.7 percent</td>
<td>21.2 percent</td>
</tr>
<tr>
<td>Other</td>
<td>25.2 percent</td>
<td>6.6 percent</td>
</tr>
<tr>
<td><strong>Vehicle type</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Passenger cars</td>
<td>94.6 percent</td>
<td>94.5 percent</td>
</tr>
<tr>
<td>Medium/heavy trucks</td>
<td>4.8 percent</td>
<td>4.5 percent</td>
</tr>
<tr>
<td>Bus/van/limo/other</td>
<td>0.6 percent</td>
<td>0.1 percent</td>
</tr>
<tr>
<td><strong>Gender</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>62.0 percent</td>
<td>55.3 percent</td>
</tr>
<tr>
<td>Female</td>
<td>38.0 percent</td>
<td>44.7 percent</td>
</tr>
</tbody>
</table>
4.3 Results of Classification Tree Modeling for Predictor Parameter Screening

4.3.1. General

In this study, crash severity was a dependent variable. The dependent variable, crash severity, was grouped into three levels namely, fatal crash, injury crash, and property damage only (PDO). Seventeen predictor variables were used to build the models. The aim of this study was to determine significant factors that significantly affect crash severity of ROR and NROR crashes. Classification tree modeling was used in identifying the most significant predictor variables of crash severity. A crash dataset consisting of 1,256,375 observations of ROR and NROR crashes was used. JMP software version 11 was used to run the classification tree models. JMP was set to randomly select 80 percent of the observations to be used as a training sample and 20 percent of the observations to be used as a validation sample.

4.3.2. Classification Tree for ROR Crash Model

The model results contained split history, receiver-operating characteristic (ROC) for validation and training samples, and column contribution results. The ROR crashes model ran a complete tree and the model stopped after 104 splits. Figure 4.20 shows that the model remains the same after 30 splits, and then the model was pruned to 30 splits.
Figure 4.20. ROR Crash Models Split History

Figure 4.21 shows the receiver-operating characteristic (ROC) curves for validation and training samples. The diagonal line cuts the graph into two portions. The diagonal line confirms if the model is good or not. For that reason, if all the curves are above the line, then the model is good, but if all the curves are below the diagonal line then the model is not good. Figure 4.21 shows that both the training and validation graphs are above the diagonal line, which means that the models good and we can trust the results or the outcomes of the model.
Figure 4.21. Training and Validation ROC Graphs for ROR Crash Model

The last output of the model is the column of contributions, which gives us significant and non-significant predictor variables based on the classification tree model selection criteria discussed earlier. The column distributions table provides information concerning each variable, and how many times the variable was split on the tree. For each variable value of $G^2$ is included. Table 4.2 shows the column contribution values for all predictor variables for ROR model. The most significant variables are road condition,
collision type, alcohol related, posted speed limit, speed related, type of crashes, vehicle

type, gender, and the age of the driver.

Table 4.2. Column contributions results for ROR crashes model

<table>
<thead>
<tr>
<th>Term</th>
<th>Number of Splits</th>
<th>G^2</th>
<th>Portion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Road Condition</td>
<td>2</td>
<td>6679.21272</td>
<td>0.3535</td>
</tr>
<tr>
<td>Collision Type</td>
<td>3</td>
<td>2501.56284</td>
<td>0.1324</td>
</tr>
<tr>
<td>Alcohol Related</td>
<td>3</td>
<td>2058.16463</td>
<td>0.1089</td>
</tr>
<tr>
<td>Posted Speed Limit</td>
<td>5</td>
<td>2029.67933</td>
<td>0.1074</td>
</tr>
<tr>
<td>Speed Related</td>
<td>6</td>
<td>1661.82115</td>
<td>0.0879</td>
</tr>
<tr>
<td>Type of Crashes</td>
<td>3</td>
<td>1515.06559</td>
<td>0.0802</td>
</tr>
<tr>
<td>Vehicle Type</td>
<td>2</td>
<td>1226.97381</td>
<td>0.0649</td>
</tr>
<tr>
<td>Gender</td>
<td>5</td>
<td>1132.79355</td>
<td>0.0599</td>
</tr>
<tr>
<td>Age</td>
<td>1</td>
<td>91.6122622</td>
<td>0.0048</td>
</tr>
<tr>
<td>Work Zone Related</td>
<td>0</td>
<td>0</td>
<td>0.0000</td>
</tr>
<tr>
<td>Drug Related</td>
<td>0</td>
<td>0</td>
<td>0.0000</td>
</tr>
<tr>
<td>Weather Condition</td>
<td>0</td>
<td>0</td>
<td>0.0000</td>
</tr>
<tr>
<td>Light Condition</td>
<td>0</td>
<td>0</td>
<td>0.0000</td>
</tr>
<tr>
<td>Day of Week</td>
<td>0</td>
<td>0</td>
<td>0.0000</td>
</tr>
<tr>
<td>Time of Crash</td>
<td>0</td>
<td>0</td>
<td>0.0000</td>
</tr>
<tr>
<td>Road Contour</td>
<td>0</td>
<td>0</td>
<td>0.0000</td>
</tr>
</tbody>
</table>
Figure 4.22. The ROR Crashes Classification Tree Model.
4.3.3. The Classification Tree for NROR Crash Model

A classification tree model was also developed for the NROR crashes for data that included drivers who were at fault only. Similarly, the model contains information on split history, receiver-operating characteristic (ROC) for validation and training samples, and column contributions results. The NROR crashes model ran a complete tree and the model stopped after 90 splits. The split history graph depicted in Figure 4.23 shows that the model remains the same after 30 splits, so the model was stopped after 30 splits for NROR crashes model.

![Split History](image)

**Figure 4.23. NROR Crashes Model Split History**

Figure 4.24 shows the receiver-operating characteristic (ROC) for the validation and training samples. The diagonal line cuts the ROC graph into two portions. If all the curves are above the diagonal line, then the model is good, but if the model curves are below the diagonal line then the model is not good. The graphs depicted in Figure 4.24 show that both training and validation graphs are good and we can trust the result or the outcomes of this model.
Table 4.3 shows the column contributions results NROR crashes model. Predictor variables, which were found significant, include collision type, posted speed limit, speed related, road condition, alcohol related, vehicle type, and drug related.
Table 4.3. Column contributions results for NROR crashes model

<table>
<thead>
<tr>
<th>Term</th>
<th>Number of Splits</th>
<th>$G^2$</th>
<th>Portion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Collision Type</td>
<td>8</td>
<td>20383.4276</td>
<td>0.5191</td>
</tr>
<tr>
<td>Posted Speed Limit</td>
<td>5</td>
<td>9698.10918</td>
<td>0.2470</td>
</tr>
<tr>
<td>Speed Related</td>
<td>4</td>
<td>4380.1342</td>
<td>0.1116</td>
</tr>
<tr>
<td>Road Condition</td>
<td>4</td>
<td>1899.86406</td>
<td>0.0484</td>
</tr>
<tr>
<td>Alcohol Related</td>
<td>5</td>
<td>1450.31058</td>
<td>0.0369</td>
</tr>
<tr>
<td>Vehicle Type</td>
<td>3</td>
<td>1309.66127</td>
<td>0.0334</td>
</tr>
<tr>
<td>Drug Related</td>
<td>1</td>
<td>142.229206</td>
<td>0.0036</td>
</tr>
<tr>
<td>Work Zone Related</td>
<td>0</td>
<td>0</td>
<td>0.0000</td>
</tr>
<tr>
<td>Weather Condition</td>
<td>0</td>
<td>0</td>
<td>0.0000</td>
</tr>
<tr>
<td>Age</td>
<td>0</td>
<td>0</td>
<td>0.0000</td>
</tr>
<tr>
<td>Light Condition</td>
<td>0</td>
<td>0</td>
<td>0.0000</td>
</tr>
<tr>
<td>Day of Week</td>
<td>0</td>
<td>0</td>
<td>0.0000</td>
</tr>
<tr>
<td>Time of Crash</td>
<td>0</td>
<td>0</td>
<td>0.0000</td>
</tr>
<tr>
<td>Road Contour</td>
<td>0</td>
<td>0</td>
<td>0.0000</td>
</tr>
<tr>
<td>Gender</td>
<td>0</td>
<td>0</td>
<td>0.0000</td>
</tr>
</tbody>
</table>
Figure 4.25. The NROR Crashes Classification Tree Model.
4.4 Discussion

The classification tree results show that ROR crashes model consists of nine significant predictor variables, but NROR crashes model consists of only six significant predictor variables. In addition, these results show that road condition was the most significant variable in ROR crashes model and also significant in NROR crashes model, this agrees with some previous studies’ findings (e.g., Liu and Subramanian 2009; Roy and Dissanayake 2011; Eustace et al. 2013). Collision type was the most significant variable for NROR crashes model and second most significant variable for ROR crashes model. In addition, the results show that age of driver at fault was a relative weakly significant variable for ROR crashes model, but it was not significant for NROR crashes model. Furthermore, the results show that gender of the driver at fault was a significant variable for ROR crashes model and this finding agrees with findings from other previous studies (e.g., Liu and Subramanian 2009; Eustace et al. 2013), but it was not for NROR crashes model. Speed related variable was significant for both ROR crashes and NROR crashes models, which is a variable that is always highly significant in many similar studies (e.g., Liu and Subramanian 2009; Eustace et al. 2013). Crash type was only significant for ROR crashes model. Alcohol related as usual, was a significant variable for both models as is always the case in many related studies (e.g., Eustace et al. 2013) and vehicle type was significant in both models, which also agrees some previous studies (e.g., Eustace et al. 2013).

Figure 4.26 shows the relationship between crash severity and road condition for the ROR crashes model. Figure 4.26 depicts the first split on the tree that divides the ROR crashes into two nodes based on road condition as a predictor variable. In addition,
Figure 4.26 shows that 38.2 percent of the crashes that occurred on dry, wet/water conditions involved fatal and injury crashes. On the other hand, only 22.67 percent of the crashes that occurred on snow/ice/mud/oil/slush/gravel road conditions were fatal and injury crashes. Therefore, crashes that occurred on dry/wet/water conditions were relatively more dangerous than those that occurred on snow/ice/mud/oil/slush/gravel road conditions.

Figure 4.26. The First Split of ROR Crashes from the Root Node

Figure 4.27 shows that the dry, wet/water road condition node splits by collision type into two nodes. Those child nodes are (sideswipe, same direction, opposite direction, angle, rear-end, backing, rear-to-rear, not collision) collision type and head-on collision type. Figure 4.27 shows that 43.5 percent of the head-on crashes that occurred on dry, wet/water road conditions were fatal and injury crashes. On the other hand, 37.5 percent of sideswipe, same direction, opposite direction, angle, rear-end, backing, rear-to-rear, not collision crashes that occurred on dry, wet/water road conditions resulted into fatal and injury crashes.
Figure 4.27. Splitting of Road Condition in the ROR Model Tree

Figure 4.28 shows that collision type (sideswipe, same direction, opposite direction, angle, rear-end, backing, rear-to-rear, not collision) node was split by alcohol related variable. The node splitting results of Figure 4.28 show that 46.5 percent of crashes, which occurred on dry/wet/water conditions for which alcohol was involved were fatal and injury crashes, while for the same types of collisions for which alcohol was not involved, only 35.7 percent of the crashes were fatal and injury crashes.

Figure 4.28. Splitting of Collision Type in the ROR Model Tree
Figure 4.29 shows that alcohol related variable was split by posted speed limit variable. The results illustrate that 38.4 percent of crashes that occurred on dry/wet/water conditions and for which alcohol use was involved that occurred on roadways with posted speed limits less than 40 mi/h were fatal and injury crashes. Whereas, 51.1 percent of crashes involving alcohol use on posted speed limits of 40 mi/h or higher were fatal and injury crashes.

![Figure 4.29. Splitting of Alcohol Use in the ROR Model Tree](image)

Figure 4.30 shows the relationship between crash severity and collision type for the NROR crashes model. Figure 4.30 depicts the first split on the tree that divides the NROR crashes into two nodes based on collision type as a predictor variable. In addition, Figure 4.30 shows that 11.8 percent of sideswipe, same or opposite direction involved fatal and injury crashes. On the other hand, 48.4 percent of angle, rear-to-end, backing, rear-to-rear, head-on, and not collision were fatal and injury crashes. Therefore, crashes that involved angle, rear-to-end, backing, rear-to-rear, head-on, and not collision were relatively more dangerous than those that involved sideswipe, same or opposite direction collisions.
Figure 4.30. First Split of NROR Crashes from the Root Node

Figure 4.31 shows that the angle, rear-to-end, backing, rear-to-rear, head-on, and not collision node splits by collision type into two nodes. The resulting child nodes are (angle, head-on) collision types and (rear-to-end, backing, rear-to-rear, and not collision) collision types. Figure 4.31 shows that 32.9 percent of angle and head-on crashes resulted into fatalities and injuries. Meanwhile, 37.5 percent of the rear-to-end, backing, rear-to-rear, and not collision crashes resulted into fatalities and injuries.
Figure 4.32 shows that angle and head-on collision type variable was split by posted speed limit variable. The results illustrate that 43.1 percent of NROR crashes that involved angle and head-on collisions that occurred on roadways with posted speed limits of 40 mi/h or higher were fatal and injury crashes. Whereas, 28.7 percent of NROR crashes involving angle and head-on collisions that occurred on roadways with posted speed limits lower than 40 mi/h fatal and injury crashes.
Figure 4.32. Splitting of Speed Limit in the NROR Model Tree
CHAPTER FIVE
CONCLUSIONS AND RECOMMENDATIONS

5.1 Conclusions
By using the classification tree model, we were able to identify some of the most significant variables that lead to one of the crash severity levels (i.e., fatal crash, injury crash, and property damage only). The classification tree model identified nine significant predictor variables of crash severity for ROR crashes. These variables listed in order starting with the most significant predictor variable include road condition, collision type, alcohol related, posted speed limit, speed related, crash type, vehicle type, gender, and age. Likewise, the classification identified six significant predictor variables of crash severity for NROR crashes. These variables listed in order starting with the most significant predictor variable include collision type, posted speed limit, speed related, road condition, alcohol related, and vehicle type.

The results from the current study show that ROR crashes were more dangerous than NROR crashes because in most cases the percent of crashes that results into fatal and injury crashes were higher for ROR crashes than NROR crashes. ROR crashes are more likely to result in fatalities and injuries when they occur on roadways with higher posted speed limits. Most ROR crashes happen when the weather is clear, so we can speculate that more people are traveling when the weather is clear. The classification tree identified that the most significant variables for both ROR and NROR crashes
models are road condition, collision type, alcohol related, posted speed limit and speed related. Most significant variables that increase the probability of crash severity of run-off-road crashes identified in the current study agree with findings from previous studies. Variables such as alcohol related (e.g., Liu and Subramanian, 2009; Roy and Dissanayake, 2012; Eustace et al. 2013), posted speed limit (e.g., Liu and Subramanian; Roy and Dissanayake, 2012; Eustace et al. 2013). Others include road condition (e.g., Liu and Subramanian 2009; Roy and Dissanayake 2012; Eustace et al. 2013), gender (e.g., Liu and Subramanian 2009), and speed related (e.g., Liu and Subramanian 2009; Eustace et al. 2013). Run off road crash incidents increase when drivers are under the influence of alcohol and likely to drive at a higher speed during clear condition where overconfidence and reckless driving may also play part.

5.2 Recommendations

Regular campaigns against drunk driving and increased enforcement on major roadways may help in reducing the injury and fatal ROR crashes. It is recommended to improve roadway drainage and use of pavement materials that dry quicker.

For collision type, a further investigation on the interaction between head-on and angle collision types and the snowy and icy road conditions is recommended to confirm their combined influence on the crash severity of ROR crashes. The current study recommends a further investigation on the interaction of high-posted speed limits (40 mi/h and higher) and dry road condition on the crash severity of ROR crashes is suggested.
It is recommended to investigate the interaction between higher posted speed limit roadways (speed more than 40 mph) and angle/head-on collision types to evaluate their influence on the crash severity of NROR crashes.
REFERENCES


Carnegie Mellon University. “Classification and Regression Trees: Data Mining.”
Accessed on January 15, 2014


Eustace, D., Almutairi, O., Hovey, P., and Shoup, G. “Using Decision Tree Modeling to Analyze Factors Contributing to Injury and Fatality of Run-Off-Road Crashes in Ohio.” In *Proceedings of the Transportation Research Board 93rd Annual Meeting 2013*. 63


Ohio Department of Transportation (ODOT). “Roadway Departure Crashes: Stay Alive, Stay on the Road.” Available at:

Rokach, L. and Maimon, O. “Decision Trees.” Department of Industrial Engineering, Tel-Aviv University. Available at:


