AN ABSTRACT OF THE DISSERTATION OF

Hamsa Abbas Zubaidi for the degree of Doctor of Philosophy in Civil Engineering presented on April 09, 2021.

Title: Roundabout Safety: Econometric and Machine Learning Models and Applications.

Abstract approved:

____

Salvador Hernandez

The rapid growth in the number of roundabouts raises some significant research challenges, especially around safety performance and evaluation in the U.S. Despite the safety advantages that roundabout geometric design brings, crashes still occur. While operationally roundabouts are good at preventing severe crashes, they may lead to more less severe crash outcomes. Much attention has been given to injury severity at signalized intersections in the past, but little to no work has been performed on roundabout crash severity. One possible reason for this is the lack of data and given that roundabouts are a growing trend among many states for their safety performance, data is still tricking in slowly. The challenge with current data is its lack of detail that would allow for a clearer understanding of how the complex interactions between roundabout crash related factors, crash types, injury severity and roundabout configurations can be captured with current econometric methods utilized by transportation safety analysts today.

However, there have been considerable advancements over the last few years especially in econometrics methods that account for unobserved heterogeneity. These advancements have been shown to provide a more reasonable understanding of contributing factors to overall safety. In addition, there has been a greater push to improve predictability and performance of these econometric techniques by utilizing complementary approaches such as machine learning. With this in mind, the goal of

this dissertation is to present an exploratory crash-based approaches that utilize both advanced econometric methods and machine learning techniques to better understand the factors that may influence less severe crashes to those of more severe crashes given various configurations and crash types utilizing Oregon and Washington state crash data at roundabouts.

The first manuscript investigates a crash-based analysis to better understand the factors that may influence less severe crashes to those of more severe crashes given various roundabout configurations and crash types in Oregon. Using Oregon's crash database from 2011 to 2015 in which 1,006 crashes occurred at roundabouts. A series of log likelihood ratio tests were conducted to validate that four separate random parameters binary probit models by configuration type were warranted.

The second manuscript develops a machine learning methodology that evaluates crash injury severity at roundabouts and compares this with traditional econometric techniques. This work estimates a Random Parameter Binary Probit model (RPBP) and compares its predictions with those rendered from machine learning techniques, namely, Support Vector Machine (SVM) with linear, radial, polynomial, and sigmoid kernels. This is accomplished by utilizing Oregon crash data from 2011 to 2015 and focuses on both three- and four-leg roundabouts. Two significant variable sets have been conducted by utilizing random forest and binary model.

Finally, the third manuscript investigates risk factors that significantly contribute to driver injury severity at roundabout crashes while systematically accounting for unobserved heterogeneity and the variance in means of the random parameters within the crash data. In this method data from the Washington State Department of Transportation (WSDOT) over a six-year period (2013 to 2018) in which 8548 crashes occurred at roundabouts is used. A random parameter binary probit model with heterogeneity in random parameter means is estimated to explore the effects of a wide range of variables on driver injury severity outcomes. Although the results of this work are exploratory, they provide evidence that crashes are occurring at roundabouts and several factors lead to crashes that result in an injury. In addition, the modeling approaches in this work offer a methodology that can account for unobservable factors in roundabout crash data. The findings of this research underscore the importance of fully accounting for unobserved heterogeneity by considering possible heterogeneity in the means of parameters. With the growing and importance relating to roundabout safety, this work provides some essential initial findings with Washington data, but also hopefully can provide some guidance for the analysis of other roundabouts-crash databases from other geographic locations and time periods. Several low-cost mitigation measures can reduce the number of crashes at roundabouts. First, improving pavement marking and signage to guide the motorist better and enhance driver expectancy. Furthermore, educating the public, including public-private partnerships between law enforcement agencies, driver's education instructors, transportation engineering groups, and insurance companies.

Roundabout Safety: Econometric and Machine Learning Models and Applications

by
Hamsa Abbas Zubaidi

A DISSERTATION

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Doctor of Philosophy

Presented April 09, 2021
Commencement June 2021

Doctor of Philosophy dissertation of Hamsa Abbas Zubaidi presented on April 09, 2021

APPROVED:

Major Professor, representing Civil Engineering

Head of the School of Civil and Construction Engineering

Dean of the Graduate School

I understand that my dissertation will become part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my dissertation to any reader upon request.

Hamsa Abbas Zubaidi, Author

ACKNOWLEDGEMENTS

I would like first to express my sincere gratitude to my supervisor, *Dr. Salvador Hernandez*, for supporting me during my study journey. His expertise was invaluable in formulating this research, and his insightful feedback pushed me to sharpen my thinking and brought my work to a higher level. His advice on both research as well as on my career has been invaluable and developed my professional personality in the way of making me ready and confident to navigate and explore my next chapter.

I would like to extend my sincere thanks to the members of my Ph.D. committee, *Dr. Chris A. Bell, Dr. David S. Hurwitz, Dr. Katharine H. Zaworski, and Dr.Yue Zhang,* for their support, valuable comments, and suggestions. My committee went above and beyond to help me reach my goal. Thank you. I would particularly like to single out *Dr. Katharine H. Zaworski*. I want to thank you for your indefinite support during my toughest days; I am forever grateful for you.

I am also grateful to the Higher Committee for Education Development (HCED) in Iraq for granting this scholarship to pursue my Ph.D. at Oregon State University. I would like to express my deepest thanks to the Family Resource Center at Oregon State, especially *Kristi King*, *Amy Luhan*, and *Erika Woosley*, for their persistent support to my family and me. It was my honor to serve as FRC advisory board member with them.

I cannot begin to express my deepest thanks to my husband’s wonderful family, my parents in low, brothers, and sister-in-law, who all have been supportive and caring during the past years. There are no words to convey how much I love them. Also, I’m deeply indebted, grateful, and appreciative of my amazing family. My parents, my lovely sisters, and their wonderful husbands for the invaluable assistance that they all provided. This would not have been possible without the daily support and nurturing of all of them. Lucky to have you all in my life.

My heart felt regards goes to my parents, my mom and my dad (may his soul rest in peace), there are no words in the world can express my feeling toward the unconditional love, the unlimited encouragement and the moral support that you have both provided for me during my entire life. Thank you both for giving me the strength to reach for the stars and chase my dreams. Dad, I'm sure you can hear and feel me; I have achieved your dream.

I owe all the beautiful words to my amazing kids, *Lana*, *Abdullah*, *Ali*, and *Mohammed*. They made countless sacrifices greater than their age to help and support me during this hard journey. I couldn't have moved on during my toughest days without their smiles, hugs, and unconditional love; they literally got me through. I'm the luckiest mom ever to have such blessed angels. I love you all dearly to the moon and back.

Finally, and above all, I dedicate this to my beloved husband, *Ihsan Ali Obaid*, a companion of my life; thank you for being there for me at the end of the day. Your love and support are what had gotten me through when I wanted to give up during my toughest days. You were always there during my good and bad times; I'm lucky to have you in my life.

I would like to praise and thank God, the almighty, who has granted countless blessings, knowledge, opportunities, and such a wonderful family for me.

CONTRIBUTION OF AUTHORS

The author expresses sincere appreciation to Dr. Hernandez for his extensive efforts, guide, and assistance on three manuscripts with the data analysis and with the unlimited insightful recommendations that led to this dissertation's achievement. Jason Anderson helped edit the first manuscript for English grammar and structure. Also, he helped with rewrite different parts and provided some R Codes in the second manuscript.

TABLE OF CONTENTS

Page

Chapter 1: Introduction 1

1.1 Motivation 1

1.2 Background 2

1.3 Organization of Dissertation 8

Chapter 2: Understanding Roundabout Safety Through the Application of Advanced Econometric Techniques 14

2.1 Introduction 16

2.2 Background 18

2.3 Empirical Setting 20

2.4 Methodology 24

2.4.1 Unobserved Heterogeneity 25

2.4.2 Log-likelihood Test 26

2.5 Discussion 29

2.5.1 Full Model 29

2.5.2 Three and Four leg Combination Model 32

2.5.3 Four leg Model 33

2.5.4 Three leg Model 35

2.6 Summary and Conclusions 37

Chapter 3: Injury Severity Prediction at Roundabouts and Variables Selection with Econometric and Machine Learning Methods 45

3.1 Introduction 47

3.2 Empirical Setting 49

TABLE OF CONTENTS (Continued)

Page

3.3 Research Methodology 51

3.3.1 Variable Selection 51

3.3.2 Prediction methods 53

3.3.3 Evaluation metrics 55

3.4 Modeling Results and Discussion 56

3.4.1 Selected Variables Using the Binary Probit Model 56

3.4.2 Selected Variables Using Random Forest 60

3.4.3 Comparison of prediction performance 63

3.5 Conclusion 73

Chapter 4: The effect of driver age and gender on driver-injury severities at roundabouts: A random parameters binary probit model with a heterogeneity-in-means approach 79

4.1 Introduction 81

4.2 Methodology 83

4.2.1 The statistical model 83

4.3 Model Estimation Results 90

4.3.1 Likelihood ratio test results 90

4.4 Models Result Discussion 91

4.4.1 Temporal Characteristics 91

4.4.2 Spatial Characteristics 91

4.4.3 Roadway Characteristics 92

4.4.4 Driver Action and Contribution 93

TABLE OF CONTENTS (Continued)

Page

4.4.5 Collision Types 95

4.4.6 Junction Relationship 96

4.4.7 Road User Involvement 97

4.4.8 Restraint System Type 98

4.4.9 Weather Conditions 99

4.4.10 Heterogeneity in the Mean of the Random Parameters 99

4.5 Summary 100

Chapter 5: Summary and Conclusion 115

5.1 Summary 115

5.2 Study Findings 116

5.3 Discussion of Study Findings 119

5.4 Study Application 120

5.5 Limitation and Future work 120

Chapter 6: References 123

LIST OF FIGURES

Figure | Page

Figure 2.1: Normalized vs Non-normalized Data of Roundabout Crashes in 21

Figure 2.2: Number of crashes according to the type of roundabout......................... 21

Figure 3.1: Number of Crashes at Different Roundabout Configurations.................. 50

Figure 3.2: Injury Severity Outcomes at (a) Three-leg Roundabouts and (b) 50

Figure 3.3: Methodological Process of Variable Selection and Model Prediction (numbers in parentheses refer to the split in training and test data, e.g., 80% training 51

Figure 3.4: Random Forest Classifier 52

Figure 3.5: Random Forest Variable Importance for (a) Three-leg Roundabouts and (b) Four-leg Roundabouts 62

Figure 3.6: Visual Comparison of Prediction Accuracies by Model and Variable Selection Method 71

LIST OF TABLES

Table Page

Table 1.1:Summary of research accounting for crash frequency modeling for the roundabouts. 3

Table 1.2: Summary of research accounting for Crash Severity Modeling at the roundabouts. 3

Table 1.3: Summary of methodological approach characteristics 4

Table 2.1: Crash Injury Severity at Roundabouts in Oregon from 2011-2015. 20

Table 2.2: Descriptive Statistics of Key Variables in all Models. 22

Table 2.3: Dependent Variable Frequency and percentage distribution in all the Models. 24

Table 2.4: Chi-Square Statistics and Degrees of Freedom for Crash Severity related to the Roundabout Type Transferability Test. 28

Table 2.5: Fixed and Random Parameter Binary Probit Models of Injury Severity for All Roundabout Type. 29

Table 2.6: Fixed and Random Parameter Binary Probit Models Results for Three and Four leg Roundabout Combination. 32

Table 2.7: Random Parameter Binary Probit Model Results for Four leg Roundabout. 33

Table 2.8: Random Parameter Binary Probit Model Results for Roundabout with Three leg. 35

Table 3.1: SVM Kernel Functions 54

Table 3.2: Random Parameter Binary Probit Model Specifications 58

Table 3.3: Descriptive Statistics of Significant Variables 59

Table 3.4: Variable Descriptions for Random Forest 61

Table 3.5: Prediction models Results for Three-leg Roundabouts Using Selected Variables from a Random Forest. 67

Table 3.6: Prediction Results for Four-leg Roundabouts Using Selected Variables from a Random Forest. 68

LIST OF TABLES (Continued)

Table Page

Table 3.7: Prediction Results for Three-leg Roundabouts Using Selected Variables from Binary Probit Model.. 69

Table 3.8: Prediction Results for Three-leg Roundabouts Using Selected Variables from Binary Probit Model... 70

Table 3.9: Comparison of Prediction Accuracy Across Models and Variable Selection 72

Table 4.1: Injury severity frequency and percentage distribution by different categories 87

Table 4.2: Descriptive statistics of the significant variables in the injury severity models... 88

Table 4.3: Chi-square statistics and degrees of freedom for driver injury severity regarding driver age and gender transferability test. 90

Table 4.4: Estimation results for random parameter binary probit model with/without heterogeneity in the mean. 102

Table 4.5: Averaged marginal effects for the random parameter binary probit model with /without heterogeneity in the mean.. 106

DEDICATION

To those who lost their lives due to car accidents.

Chapter 1: Introduction

1.1 Motivation

A roundabout is a type of circular intersection or junction in which road traffic is permitted to flow in one direction around a central island, and priority is typically given to traffic already in the junction. The major characteristics than distinguished the roundabouts from other circular intersections are there is a yield control on all the entries whereas there is stop sign or no control at the traffic circle. Additionally, No parking is allowed within the circulatory roadway or at the roundabout's entries whereas some traffic circles allow parking within the circulatory roadway. Finally, all vehicles circulate counter-clockwise and pass to the right of the central island and the roundabout, whereas some traffic circles allow left-turning vehicles to pass to the left of the central island.

Roundabouts are increasingly popular in the United States due to the advantages of improving safety and reducing delays at intersections. Although modern roundabouts were first designed in the United Kingdom in the 1960s, their prevalence in the United States did not begin until the 1990s (Qin et al. 2011). Roundabout implementation has been on the rise in the United States, with less than 100 in 1997 to about 1000 in 2007 (Montella 2011). Roundabouts are generally used as a solution in some cases of some intersections with different leg types. At the roundabout, vehicles coming from each road, heading towards the roundabout, move in one direction when entering the roundabout and around a central island in a circle. Traffic in the roundabout is continuous but at a relatively slow speed.

The conversion of intersections into roundabouts produces a substantial 27% growth in the number of injury accidents involving bicyclists on or nearby the roundabouts in Flanders. (Daniels et al. 2008) tried to figure out if roundabouts have an impingement on the safety of different cases of road users to develop adequate decision-making criteria for places when the structure of a carousel is being taken. So, compared to other types of intersections, roundabouts have some intrinsic properties favoring traffic safety: they reduce speeds considerably, and they decrease the number of possible conflict points between road users (Daniels et al. 2011). For instance, many intersections have been converted into roundabouts to enhance the capacity and to reduce the number of more severe crashes; that is, roundabouts have fewer conflict points compared to the more traditional intersection reducing the crash potential (Montella 2011). FHWA formally defines a conflict as an "observable situation in which two or more road users approach each other in time and space to such an extent that there is a risk of collision if their movements remain unchanged." Three types of conflicts are possible: merge, diverge, and crossing. Crossing conflicts are often the most severe in terms of vehicular injuries and fatalities. Recent studies of intersections converted to roundabouts indicate a steady reduction in injury crashes, particularly for crashes with fatal or more severe injuries (Daniels et al. 2010a). Moreover, these studies have reported a larger decrease in the number of severe crashes (fatalities and crashes involving severe injuries) compared to the reduction of the total number of injury-related crashes. However, the effects of property-damage-only crashes are highly uncertain (Daniels et al. 2008).

The construction of roundabouts as alternatives to signalized or STOP-controlled intersections has increased in Oregon due to their safety performance characteristics. The rapid growth in the number of roundabouts in Oregon and the United States raises some significant research challenges, especially in the area of safety performance and evaluation for U.S. specific data under varying conditions. Despite the advantages in the roundabout geometric design, crashes still occur. While roundabouts are great at preventing severe crashes, they may bring on more non-fatal wrecks. Although the increase of using roundabouts as a superior alternative to other intersection types in urban and rural areas, there is still a high need for further research to develop advanced injury severity prediction models that account for unobserved heterogeneity and to get the contributing factors that lead to a specific type of injury.

1.2 Background

Modern roundabouts have become increasingly popular in the United States due to their innovative design, benefits on traffic operations, and increased safety.

Considerable research has been conducted in better understanding design and improved traffic operations of roundabouts (Chen et al. 2013; Coelho et al. 2006; Flannery 2001; Pratelli 2006; Valdez et al. 2011), and from a safety perspective, roundabouts have also been studied quite extensively (De Brabander et al. 2005; Kamla et al. 2016; Montella 2007; Persaud et al. 2001). Yet, literature that attempts to capture the complex interactions of crash factors, injury severity, crash types, and configuration is sparse. There have been several efforts that attempt to capture these complex interactions through advanced econometric techniques (Lord and Mannering 2010, Mannering and Bhat 2014, and Mannering et al. 2016) from the perspective of analyzing signalized and unsignalized intersections, but not much on roundabouts. Some of these methods have been confined to the development of crash rates/frequency models. Table 1.1 summarizes the most commonly used econometric techniques for crash rate/frequency as found in the literature. Still, from an advanced econometric methodology perspective, Table 1.2 illustrates the sparseness of these methods as applied to roundabout injury severity.

Table 1.1:Summary of research accounting for crash frequency modeling for the roundabouts.

Methodological Approach	**Previous Research**
Linear regression	(Taekratok 1998), (Ambros et al. 2016)
Poisson	(Daniels et al. 2010b), (Dixon 2012), (Daniels et al. 2011),
Gamma probability	(Daniels et al. 2010b), (Daniels et al. 2011),
Negative binomial (NB)	(Chen et al. 2013), (Dixon and Zheng 2013), (Chiu 2014), (Dixon 2012), (Qin et al. 2011a), (Kamla et al. 2016)
Zero-Inflated Poisson (ZIP)	(Chen et al. 2013),
Zero-inflated negative binomial (ZINB)	(Anjana and L. R. Anjaneyulu 2015)

Table 1.2: Summary of research accounting for Crash Severity Modeling at the roundabouts.

Methodological Approach	**Previous Research**
Logit	(Daniels et al. 2010a), (Polders et al. 2015)

To deal with the data and methodological issues associated with crash severity and crash frequency data, a variety of methods have been used over a long time. Table 1.3 provides a summary of advantages and disadvantages for some of these methodological approaches used to analyze crash severity and crash frequency at roundabouts.

Table 1.3: Summary of methodological approach characteristics

Methodological Approach	**Advantages**	**Disadvantages**
Linear regression	• when relationships between the independent variables and the dependent variable are almost linear, it shows optimal results	• Linear regression is often inappropriately used to model non-linear relationships. • Linear regression is limited to predicting the numeric output. • The dependent variable must be continuous. • Linear regression only looks at the mean of the dependent variable. • Regression is sensitive to outliers.
Poisson	• Counts of events that occur randomly in each interval of time (or space). • Maintains the constancy of the sums	• Poisson has support only on the positive integers. • It does not account for over dispersion because it assumes that the mean and variance of the errors are equal
Negative Binomial (NB)	• It loosens the highly restrictive assumption that the variance is equal to the mean so it can easily handle the crash data over dispersion	• The negative binomial has support only on the positive integers. • The constancy of sums is not maintained • Cannot handle under-dispersion
Zero-Inflated Poisson (ZIP)	• Use to model count data that has an excess of zero counts	• It does not account for over dispersion problems • Can create theoretical inconsistencies
Zero-inflated negative binomial (ZINB)	• Count data that exhibit over-dispersion and excess zeros	• Can create theoretical inconsistencies • can be adversely influenced by the low sample-mean and small sample size bias
Gamma	• Account for under dispersion	• Dual-state model with one state having a long-term mean equal to zero
Logit	• Use to model dichotomous outcome variables • Allows properties of a linear regression model to be exploited	• Cannot predict continuous outcomes • Data should be independent • High bias

	• The logit itself can take values between - ∞ and + ∞ • The probability remains constrained between 0 and 1	

The current econometric models, as previously shown (see Table 1.4), were good 10 to 15 years ago; however, since that time, there has been considerable advancement in econometrics, especially econometrics methods that account for unobserved heterogeneity. These advancements have been shown to provide a more reasonable understanding of contributing factors to overall safety issues.

Machine learning algorithms for crash severity predictions have been addressed so far at different roadway locations and become popular due to their good predictive performance. Through methods like Nearest Neighbor Classification (NNC), Support Vector Machines (SVM), Random Forests (RF), and deep learning model using a Recurrent Neural Network Objectives, there is high potential to highlight significant factors as they relate to crash injury severities in addition to increased predictive power over more conventional econometric techniques. Many different studies provided invaluable insights for the use of predictive analytics in this domain and exposed the relative importance of crash related risk factors with the changing levels of injury severity in different locations except at the roundabouts (Aghayan et al. 2015; Ahmadi et al. 2020; Delen et al. 2017; Iranitalab and Khattak 2017a; Jeong et al. 2018; Kashani and Mohaymany 2011; Li et al. 2018a; Mafi et al. 2018; Pal et al. 2016; Rezapour et al. 2020; Sameen and Pradhan 2017a; Tang et al. 2019a; Tay 2015a; Toran n.d.; Yu and Abdel-Aty 2014; Zhang et al. 2018a).

In addition, from an injury severity analysis, crash type, and configuration, there is much more that can be done, as seen from the lack of literature that captures the complex interactions of these variables. Considerable research has been conducted in better understanding design and improved traffic operations of roundabouts (Chen et al. 2013; Coelho et al. 2006; Flannery 2001; Pratelli 2006; Valdez et al. 2011), and from a safety perspective, roundabouts have also been studied quite extensively (De Brabander et al. 2005; Kamla et al. 2016; Montella 2007; Persaud et al. 2001). Yet, literature that attempts to capture the complex interactions of crash factors, injury severity, crash types, and configuration is sparse. There have been several efforts that attempt to capture these complex interactions through advanced econometric techniques (Lord and Mannering 2010, Mannering and Bhat 2014, and Mannering et al. 2016) from the perspective of analyzing signalized and unsignalized intersections, but not much on roundabouts.

With this in mind, what is still not clearly understood is the relationship between roundabout crash-related factors, crash types, injury severity, and roundabout configurations. A reason for this may stem from the lack of available detailed crash-related data. Recent studies (Al-Bdairi et al. 2018; Al-Bdairi and Hernandez 2017; Anderson and Hernandez 2017a; Pahukula et al. 2015; Romo et al. 2014) have illustrated the use of limited crash data sources to discover relationships between crash-related factors and injury severities through the use of advanced unobserved heterogeneity based econometric techniques. Hence, the objective of this study is to conduct crash-based analyses to better understand the factors that may influence less severe crashes to those of more severe crashes given various configurations and crash types. In addition, used an artificial intelligent method to predict better injury severity models with different outcome ratio. This will be accomplished by exploring relatively new techniques applied to roundabout crash data to fill the gap in the literature. Both advanced econometric techniques and machine learning methods were applied to predict traffic injury severity models at the roundabouts to accomplish the following aims:

- What are the contributing factors and the unobserved heterogeneity that lead to a specific type of injury severity? Is this related to the configuration of the roundabout?
- Is the machine learning method could perform better than the econometric technique in predicting injury severity outcomes?
- Is there a relationship between specific driver injury severity and the gender and the age of the driver at the roundabout crashes?

To answer the previous questions, three extensive studies have been conducted to present clearer insight about the crash's characteristics at the roundabout and distinctive of applying specific advance statistical methods. This is accomplished through exploring advanced econometric techniques applied to roundabout crash data that account for unobserved heterogeneity (unobservable in the data). These advanced econometric techniques have been shown to provide a more accurate understanding of contributing factors to overall safety issues (Mannering et al. 2016; Mannering and Bhat 2014). Specifically, this work utilizes the random parameters binary probit model. The random parameters binary probit model is used here to gain a better understanding of the complex interactions between factors found to be significant and those unobserved factors that may be influencing estimated outcomes. So, the study conducts crash-based analyses to better understand the factors that may influence less severe crashes to those of more severe crashes given various roundabout configurations. Then developed a machine learning methodology that evaluates crash severity at roundabouts and conducts two datasets as a significant variable by using both traditional method and machine learning techniques and compare these methods with traditional econometric techniques by using the two conducting datasets. Precisely, this work will estimate a random parameter binary probit model (RPBP) and compare its results with those rendered from machine learning techniques, namely, Support Machine Vector (SVM) with linear, nonlinear, polynomial, and sigmoid kernels. The comparison will focus on predictions of the two crash outcomes for three- and four-leg roundabouts.

Finally, to understand the relationship between specific driver injury severity and the gender and the age of the driver at the roundabout crashes. The majority of research focusing and intends to contribute to a better understanding of driver characteristics on specific injury severity at the roundabouts with heterogeneous mean specified as a function of age and gender. This has been done by conducting crash-based analyses by taking into consideration the unobserved heterogeneity, and the variance in random parameter means.

1.3 Organization of Dissertation

The current dissertation is a component of three journal manuscripts that briefly establish a better understanding of the dissertation's scope. Chapter 2 represents the first journal manuscript published in the "International Journal of Transportation Science and Technology." This manuscript investigated the applying of fixed and random parameters binary probit models to model the probability of two possible crash severity outcomes. These outcomes represent the aggregation of injury-type crashes and fatal crashes, and no injury crashes.

Chapter 3 describes the development of a machine learning methodology that evaluated crash injury severity at roundabouts and compared this method with traditional econometric techniques. Precisely, this work estimated a random parameter binary probit model (RPBP) and compared its predictions with those rendered from machine learning techniques, namely, support vector machine (SVM) with linear, radial, polynomial, and sigmoid kernels. To compare the two methods, two variables selection techniques were used. First, variables identified as being significant in the econometric method were used to predict crash severity in both methodologies. Second, variables were identified through a random forest analysis and used to predict crash severity in both methodologies. Regardless of the variable selection technique, results demonstrated that the SVM models outperformed the econometric models in crash severity predictions.

Chapter 4 presents the effect of the driver age and gender on driver injury severity outcome at the roundabouts, a random parameter binary probit model with heterogeneity in means developed. An additional layer of heterogeneity has been added that is associated with the mean of the distribution of the estimated random parameter, in other words allowing the random parameter to vary by the explanatory variables. Chapter 5 includes the conclusion of this work and future work, and finally, Chapter 6 illustrates the references.

Reference

Aghayan, I., Hosseinlou, M.H., Kunt, M.M., 2015. Application of Support Vector Machine for Crash Injury Severity Prediction : A Model Comparison Approach. J. Civ. Eng. Urban. 5, 193–199.

Ahmadi, A., Jahangiri, A., Berardi, V., Machiani, S.G., 2020. Crash severity analysis of rear-end crashes in California using statistical and machine learning classification methods. J. Transp. Saf. Secur. 12, 522–546.

Al-Bdairi, N.S.S., Hernandez, S., 2017. An empirical analysis of run-off-road injury severity crashes involving large trucks. Accid. Anal. Prev. 102, 93–100.

Al-Bdairi, N.S.S., Hernandez, S., Anderson, J., 2018. Contributing Factors to Run-Off-Road Crashes Involving Large Trucks under Lighted and Dark Conditions. J. Transp. Eng. Part A Syst. 144, 1–9.

Ambros, J., Novak, J., Borsos, A., Hoz, E., Kiec, M., MacHcinik, ??tefan, Ondrejka, R., 2016. Central European Comparative Study of Traffic Safety on Roundabouts. Transp. Res. Procedia 14, 4200–4208.

Anderson, J., Hernandez, S., 2017. Roadway classifications and the accident injury severities of heavy-vehicle drivers. Anal. Methods Accid. Res. 15, 17–28.

Anjana, S., L. R. Anjaneyulu, M. V., 2015. Development of Safety Performance Measures for Urban Roundabouts in India. J. Transp. Eng. 141, 04014066.

Chen, Y., Persaud, B., Sacchi, E., Bassani, M., 2013. Investigation of models for relating roundabout safety to predicted speed. Accid. Anal. Prev. 50, 196–203.

Chiu, L., 2014. the Development of Safety Performance.

Coelho, M.C., Farias, T.L., Rouphail, N.M., 2006. Effect of roundabout operations on pollutant emissions. Transp. Res. Part D Transp. Environ. 11, 333–343.

Daniels, S., Brijs, T., Nuyts, E., Wets, G., 2010a. Externality of risk and crash severity at roundabouts. Accid. Anal. Prev. 42, 1966–1973.

Daniels, S., Brijs, T., Nuyts, E., Wets, G., 2010b. Explaining variation in safety performance of roundabouts. Accid. Anal. Prev. 42, 393–402.

Daniels, S., Brijs, T., Nuyts, E., Wets, G., 2011. Extended prediction models for crashes at roundabouts. Saf. Sci. 49, 198–207.

Daniels, S., Nuyts, E., Wets, G., 2008. The effects of roundabouts on traffic safety for bicyclists: An observational study. Accid. Anal. Prev. 40, 518–526.

De Brabander, B., Nuyts, E., Vereeck, L., 2005. Road safety effects of roundabouts in

Flanders. J. Safety Res. 36, 289–296.

Delen, D., Tomak, L., Topuz, K., Eryarsoy, E., 2017. Investigating injury severity risk factors in automobile crashes with predictive analytics and sensitivity analysis methods. J. Transp. Heal. 4, 118–131.

Dixon, K.K., 2012. Developing Safety Performance Functions for 4-leg Single-lane Roundabouts Based on Oregon Data: A Case Study.

Dixon, K.K., Zheng, J., 2013. Developing safety performance measures for roundabout applications in the State of Oregon.

Flannery, A., 2001. Geometric Design and Safety Aspects of Roundabouts. Transp. Res. Rec. J. Transp. Res. Board 1751, 76–81.

Iranitalab, A., Khattak, A., 2017. Comparison of four statistical and machine learning methods for crash severity prediction. Accid. Anal. Prev. 108, 27–36.

Jeong, H., Jang, Y., Bowman, P.J., Masoud, N., 2018. Classification of motor vehicle crash injury severity: A hybrid approach for imbalanced data. Accid. Anal. Prev. 120, 250–261.

Kamla, J., Parry, T., Dawson, A., 2016. Roundabout Accident Prediction Model. Transp. Res. Rec. J. Transp. Res. Board 2585, 11–19.

Kashani, A.T., Mohaymany, A.S., 2011. Analysis of the traffic injury severity on two-lane, two-way rural roads based on classification tree models. Saf. Sci. 49, 1314–1320.

Li, Y., Ma, D., Zhu, M., Zeng, Z., Wang, Y., 2018. Identification of significant factors in fatal-injury highway crashes using genetic algorithm and neural network. Accid. Anal. Prev. 111, 354–363.

Lord, D., Mannering, F., 2010. The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives. Transp. Res. Part A Policy Pract. 44, 291–305.

Mafi, S., AbdelRazig, Y., Doczy, R., 2018. Machine Learning Methods to Analyze Injury Severity of Drivers from Different Age and Gender Groups. Transp. Res. Rec. 2672, 171–183.

Mannering, F.L., Bhat, C.R., 2014. Analytic methods in accident research: Methodological frontier and future directions. Anal. Methods Accid. Res. 1, 1–22.

Mannering, F.L., Shankar, V., Bhat, C.R., 2016. Unobserved heterogeneity and the statistical analysis of highway accident data. Anal. Methods Accid. Res. 11, 1–16.

Montella, A., 2007. Roundabout In-Service Safety Reviews: Safety Assessment Procedure. Transp. Res. Rec. J. Transp. Res. Board 2019, 40–50.

Montella, A., 2011. Identifying crash contributory factors at urban roundabouts and using association rules to explore their relationships to different crash types. Accid. Anal. Prev. 43, 1451–1463.

Pahukula, J., Hernandez, S., Unnikrishnan, A., 2015. A time of day analysis of crashes involving large trucks in urban areas. Accid. Anal. Prev. 75, 155–163.

Pal, C., Okabe, T., Kulothungan, V., Sangolla, N., Manoharan, J., Stewart, W., Combest, J., 2016. Factors influencing specificity and sensitivity of injury severity prediction (ISP) algorithm for AACN. Int. J. Automot. Eng. 7, 15–22.

Persaud, B., Retting, R., Garder, P., Lord, D., 2001. Safety Effect of Roundabout Conversions in the United States: Empirical Bayes Observational Before-After Study. Transp. Res. Rec. J. Transp. Res. Board 1751, 1–8.

Polders, E., Daniels, S., Casters, W., Brijs, T., 2015. Identifying crash patterns on roundabouts. Traffic Inj. Prev. 16, 202–7.

Pratelli, A., 2006. Design of modern roundabouts in urban traffic systems. In: Urban Transport XII: Urban Transport and the Environment in the 21st Century, WIT Transactions on The Built Environment, Vol 89. WIT Press, Southampton, UK, pp. 83–93.

Qin, X., Khan, G., Bill, A., Noyce, D.A., 2011a. Comprehensive Safety Evaluation of Roundabouts in Wisconsin. J. Transp. Saf. Secur. 3, 289–303.

Qin, X., Khan, G., Bill, A., Noyce, D.A., 2011b. Comprehensive Safety Evaluation of Roundabouts in Wisconsin. J. Transp. Saf. Secur. 3, 289–303.

Rezapour, M., Mehrara Molan, A., Ksaibati, K., 2020. Analyzing injury severity of motorcycle at-fault crashes using machine learning techniques, decision tree and logistic regression models. Int. J. Transp. Sci. Technol. 9, 89–99.

Romo, A., Hernandez, S., Cheu, R.L., 2014. Identifying Precrash Factors for Cars and Trucks on Interstate Highways: Mixed Logit Model Approach. J. Transp. Eng. 140, 04013016.

Sameen, M.I., Pradhan, B., 2017. Severity prediction of traffic accidents with recurrent neural networks. Appl. Sci. 7.

Taekratok, T., 1998. Modern roundabouts for Oregon. Oregon Dep. Transp. Res. Unit.

Tang, J., Liang, J., Han, C., Li, Z., Huang, H., 2019. Crash injury severity analysis using a two-layer Stacking framework. Accid. Anal. Prev. 122, 226–238.

Tay, R., 2015. A random parameters probit model of urban and rural intersection crashes. Accid. Anal. Prev. 84, 38–40.

Toran, A., n.d. Analysing the pedestrian crash severity at mid-blocks using classification tree models.

Valdez, M., Cheu, R., Duran, C., 2011. Operations of Modern Roundabout with Unbalanced Approach Volumes. Transp. Res. Rec. J. Transp. Res. Board 2265, 234–243.

Yu, R., Abdel-Aty, M., 2014. Analyzing crash injury severity for a mountainous freeway incorporating real-time traffic and weather data. Saf. Sci. 63, 50–56.

Zhang, J., Li, Z., Pu, Z., Xu, C., 2018. Comparing prediction performance for crash injury severity among various machine learning and statistical methods. IEEE Access 6, 60079–60087.

Understanding Roundabout Safety Through the Application of Advanced Econometric Techniques

By Hamsa Abbas Zubaidi, Jason Anderson, and Salvador Hernandez, Ph.D.

Journal: International Journal of Transportation Science and Technology
Address: DOI: 10.1016/j.ijtst.2020.03.001

Chapter 2: Understanding Roundabout Safety Through the Application of Advanced Econometric Techniques

Hamsa Abbas Zubaidi
Ph.D. Candidate, School of Civil and Construction Engineering
Oregon State University, 233 Owen Hall, Corvallis, OR 97331-3212
(541)_602-6029
zubaidih@oregonstate.edu
hamsa.alzubaidi@qu.edu.iq

Jason C. Anderson, Ph.D.
Research Associate, Department of Civil and Environmental Engineering
Portland State University, 300 Suite, Engineering Building, Portland, OR 97201
(503) 725-4225
jason.c.anderson@pdx.edu

Salvador Hernandez, Ph.D.
Associate Professor, School of Civil and Construction Engineering
Oregon State University, 309 Owen Hall, Corvallis, OR 97331-3212
(541)-737-1757
sal.hernandez@oregonstate.edu

Abstract

Intersections present a significant safety concern; as such, in an effort to reduce the more serious injuries occurring at or near intersections, many jurisdictions have turned to implement roundabouts. Despite the advantages that roundabouts provide, crashes still occur, and less severe crashes are on the rise. The study presented in this paper investigates a crash-based analysis to better understand the factors that may influence less severe crashes to those of more severe crashes given various roundabout configurations and crash types. Using Oregon's crash database from 2011 to 2015, a series of log-likelihood ratio tests were conducted to validate that four separate random parameters binary probit models by configuration type were warranted. The outcome of each tested configuration (full, three & four leg, four leg, and three leg models) shows a major difference in both the combination and variables included in each model and the magnitude of the impact of those variables. These differences illustrate that various roundabout configurations (full, three & four leg, four leg, and three leg models) do, in fact, have different factors highlighting the need to examined crashes at roundabouts by configuration type. Variables related to driver error, weather, alcohol use, barrier conditions, vehicle movement, location of the crash, and restraint use were found as key differences between the various tested configurations.

Keywords: Roundabouts, Safety, Injury Severity, Unobserved Heterogeneity, Random Parameters Probit

2.1 Introduction

Intersections present a significant safety concern, accounting for roughly 2.21 million crashes and 6,770 fatal crashes in 2009, while during the period of 2011 to 2014, there were 48,733 (28%) drivers involved in fatal intersection crashes (AARP and WALC 2014; Gross et al. 2013; Lombardi et al. 2017; NHTSA 2009). Almost one in every four fatal crashes occur at or near an intersection (Haleem and Abdel-Aty 2010). In an effort to reduce the more serious injuries occurring at or near intersections, many jurisdictions have turned to implement roundabouts, a proven countermeasure (FHWA 2015; Gross et al. 2013; Nikitin et al. 2017). The construction of roundabouts as an alternative to signalized or stop sign-controlled intersections has increased over the years, with less than 100 in 1997 to as many as 3,200 in 2013 and growing (FHWA 2015; Montella 2011; Qin et al. 2011a). Many intersections have been converted to roundabouts to enhance traffic capacity and reduce crashes (Montella 2011). Compared to other types of intersections, roundabouts have some intrinsic properties favoring traffic safety; for example, they reduce speeds considerably and decrease the number of possible conflict points between road users (Daniels et al. 2011).

The Federal Highway Administration (FHWA) formally defines a conflict as an "*observable situation in which two or more road users approach each other in time and space to such an extent that there is a risk of collision if their movements remain unchanged*." Traffic at a roundabout is governed by the yield-at-entry rule, and the relatively lower levels of geometric design standards are intentionally applied to force vehicular trajectories at roundabouts into a very narrow space. International studies of intersections that have been converted to roundabouts indicate a steady reduction in injury crashes, particularly for crashes with fatal or severe injuries (Daniels et al. 2010a). These studies indicate that the crash frequencies (average annual crashes per roundabout) in the United States are still high compared to results from Australia, France, and the United Kingdom (Robinson et al. 2000). These same studies also report a considerable decrease in the number of severe crashes (fatalities and crashes involving severe injuries) compared to the reduction of the total number of injury crashes. However, the effects of property-damage-only (no injury) crashes are highly ambiguous (Daniels et al. 2008). Despite the advantages that roundabouts provide, crashes still occur.

Turning to driver injury severity analyses, various methodological and statistical modeling techniques have been developed and applied to identify various contributing factors to intersection-related crashes (Lombardi et al. 2017). Roadway geometric features, driver behavior, demographic information, traffic control elements, traffic compositions, and environmental characteristics are some examples of these factors (Liu et al. 2016; Lombardi et al. 2017; Mannering et al. 2016). To better understand the influences of these factors on roundabout crashes, it is essential to investigate their impacts on crash occurrences in order to develop effective countermeasures to reduce crash risk and severity.

Still, what is not clearly understood is the relationship between roundabout crash-related factors, crash types, injury severity, and roundabout configurations. Therefore, the objective of this study is to conduct a crash-based analysis to better understand the factors that may influence less severe crashes to those of more severe crashes given various configurations and crash types for roundabouts. This will be accomplished through exploring advanced econometric techniques applied to roundabout crash data that account for unobserved heterogeneity (unobservable in the data). These advanced econometric techniques have been shown to provide a more accurate understanding of contributing factors to overall safety issues (Mannering et al. 2016; Mannering and Bhat 2014). Specifically, this work utilizes the random parameters binary probit model. The random parameters binary probit model is used here to gain a better understanding of the complex interactions between factors found to be significant and those unobserved factors that may be influencing estimated outcomes. To accomplish this, Oregon crash data is used. The dataset consists of 1,006 crashes in seventeen counties in the State of Oregon for a five-year period (2011 to 2015). To the best of the authors' knowledge, this is the first attempt at modeling driver-injury severity for crashes occurring at roundabouts using a random parameter binary probit approach on two injury severity outcomes (injury or no injury) in Oregon.

2.2 Background

Given the sparsity of literature on roundabout injury severity modeling, this section presents studies related to methodological approaches that establish any links between crash characteristics, injury severity, road environments, and other factors related to roundabouts.

According to previous studies in Belgium, no studies for designing and improving road safety policy have ever been carried out in-depth. Considering this, De Brabander et al. (2005) studied the impact of roundabouts on the number of crashes and injury severity. Based on a classic negative binomial distribution, the results showed that roundabouts lead to a reduction of 34% in the number of injury crashes. After that study, many studies were completed in Belgium to evaluate the effectiveness of roundabouts. One such study found that the conversion of intersections into roundabouts led to a substantial rise (27%) in the number of injuries crashes involving bicyclists on or nearby roundabouts in Flanders, Belgium. This outcome was confirmed by Daniels *et al.* when attempting to ascertain if roundabouts have an impact on the safety of different types of road users; this was used to develop adequate decision-making criteria for roundabout design (Daniels et al. 2008). Then, in an expanded study in Flanders, Belgium, conducted by Daniels *et al.*, the authors looked into which factors might explain the severity of crashes or injuries at roundabouts constructed between 1990 and 2002 (Daniels et al. 2010b). To do this, Daniels *et al.* investigated the application of the Poisson and gamma modeling techniques to determine which variables might explain a structural part of the variation in crash rates at roundabouts in Flanders, Belgium (Daniels et al. 2010a).

Next, Daniels *et al.* extended the prediction models for crashes at roundabouts, in which regression models were fitted using available geometric and traffic variables (Daniels et al. 2011). The Poisson and gamma models were equipped with the resulting list of variables. Vulnerable road users (moped riders, motorcyclists, bicyclists, pedestrians) are more often involved in injury crashes at roundabouts. The overall number of crashes is more or less proportional to the number of motorized vehicles

(AADT). Three-leg roundabouts tend to perform worse than roundabouts with four or more legs. The larger the central island, the more single-vehicle crashes seem to occur. Substantial and highly significant crash reductions were observed by Montella following the conversion of signalized and stop-controlled intersections to roundabouts (Montella 2011). However, roundabout performances can degrade if precautions are not taken during either the design or the operation phase. Thus, the paper aimed to investigate the contributory crash factors at fifteen urban roundabouts located in Italy and to study the interdependencies between these factors. It was found that the most frequent category of contributory crash factors was geometric design. Markings were a contributory factor in more than half of the total crashes. The pavement was identified as a contributory factor in more than one-third of the total crashes, with the most common factor being low friction. Road environment factors were designated as a contributory factor in one-fifth of the crashes.

Furthermore, Kim and Choi investigated the significant factors that contribute to crashes at roundabouts in South Korea by comparing two conventional models: Poisson regression and negative binomial regression (Kim and Choi 2013). Based on their statistical analysis, it was found that the negative binomial regression approach performed best. In addition, the study provided a model with which to capture the relationship between geometric design elements and the occurrence of crashes at roundabouts.

Although roundabouts have gained popularity nationally and in Oregon, it is still not clearly understood what the relationship might be between crash types, injury severity, and roundabout configurations. As such, there is a need for further research to develop advanced crash prediction models that account for unobserved heterogeneity. In addition, from an injury severity analysis and configuration perspective, the lack of literature shows there is much more that can be done to capture the complex interactions of these variables. A reason for this may stem from the lack of available detailed crash-related data. Recent studies have illustrated the use of limited crash data sources to discover relationships between crash-related factors and injury severity through the use of advanced unobserved-heterogeneity-based econometric techniques (Al-Bdairi et al. 2018; Al-Bdairi and Hernandez 2017; Anderson and Hernandez 2017a;

Pahukula et al. 2015; Romo et al. 2014). Hence, the objective of this study is to conduct crash-based analyses to better understand the factors that may influence less severe crashes to those of more severe crashes given various configurations and crash types.

2.3 Empirical Setting

It is generally accepted that the number of crashes at roundabouts are fewer than those at signalized intersections. Therefore, obtaining detailed data that can capture the factors that contribute to crash severity is more complicated regarding the required sample size that accurately represents the population. As such, this research is based on crash data collected and compiled by Oregon's Department of Transportation (ODOT) Crash Analysis and Reporting Unit. The data includes crashes over a five-year period (2011 to 2015), in which 1,006 crashes occurred at roundabouts (shown in Table 2.1). Figure 2.1 illustrates the difference between the normalized crash data vs. the non-normalized crash data to compare the crash patterns during the period 2011 to 2015. The normalization was performed by calculating the average number of the crashes over five years and their distribution over this time period. These crashes occurred in seventeen counties at different types of roundabouts (unknown, three leg, four leg, and five leg roundabouts), as shown in Figure 2.2, with a different geometric design in both rural and urban areas.

Table 2.1: Crash Injury Severity at Roundabouts in Oregon from 2011-2015.

Year	Severe	Minor	No Injury	Sum	% of Total
2011	1	28	125	154	15.31
2012	1	41	179	221	21.97
2013	1	30	165	196	19.48
2014	1	34	160	195	19.38
2015	2	31	207	240	23.86
Total	6	164	836	1006	100

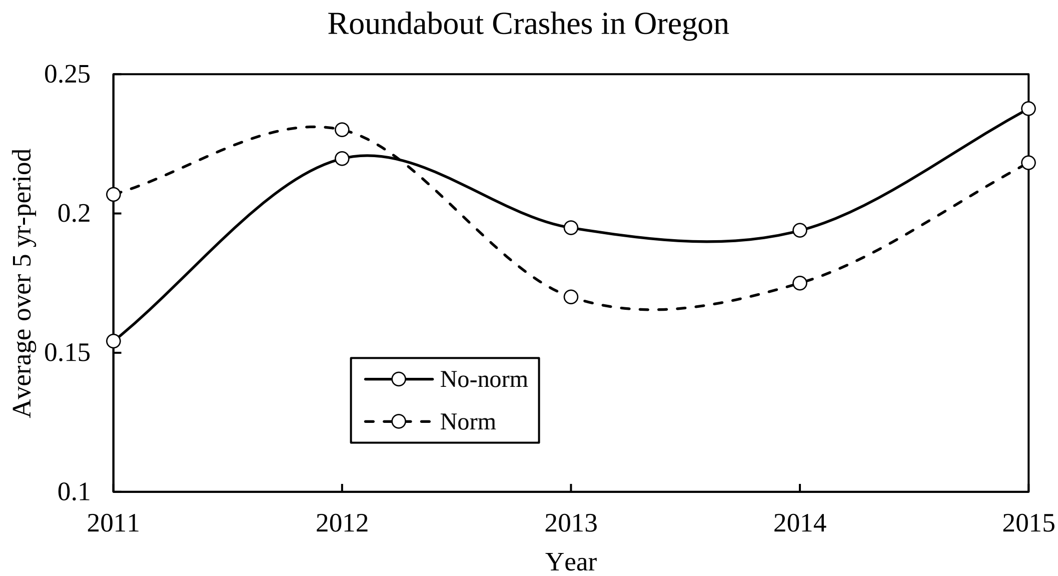


Figure 2.1: Normalized vs Non-normalized Data of Roundabout Crashes in Oregon from 2011 to 2015.

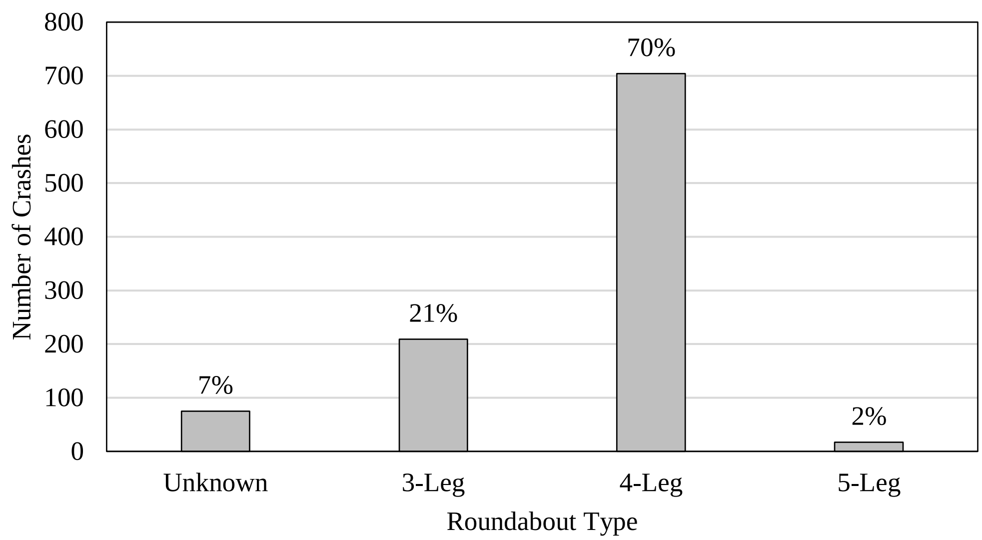


Figure 2.2: Number of crashes according to the type of roundabout.

As illustrated in Figure 2.2, roundabouts with four leg had the highest number of crashes at 704(70%), followed by roundabouts with three leg at 209 (21%) crashes, unknown roundabout type at 75(7%), and roundabouts with five leg at seventeen crashes. The distribution of the crash injury severity is comprised of six fatal and incapacitating injuries (severe), 164 minor injuries (non-incapacitating and possible injuries – complaint of pain), and 836 no injuries.

Table 2.2 provides descriptive statistics of variables that were used to model injury severity for all roundabout types (full model), three and four leg roundabouts (three and four leg), four leg roundabouts (four leg), and three leg roundabouts (three leg). These varibles consisted of factors related to gender, age of the driver, participant cause, crash level cause, safety equipment use, light condition, population, speed limit, weekdays, movement of the vehicle, weather condition, alcohol use ownership and the type of the vehicle, and barrier type and condition. The dependent variables in each of these models consisted of two specific outcomes: (1) no injury and (2) injury. The sample size characteristics and the frequency for each of these models are illustrate in Table 2.3.

Table 2.2: Descriptive Statistics of Key Variables in all Models.

Variable	Mean/Standard Deviation			
	Full Model	Three and Four leg Model	Four leg Model	Three leg Model
Vehicle Type (1 the vehicle weight <10,000 lb., 0 otherwise)	0.95/0.22	0.95/0.22	0.96/0.21	0.92/0.27
Gender (1 if male, 0 otherwise)	0.51/ 0.50	0.5/0.5	0.49/0.50	0.53/ 0.50
Age of Driver (1 if 21< age ≤35, 0 otherwise)	0.22 /0.42	0.22/0.41	0.21/0.41	0.97/0.17
Participant Safety Equipment Use (1 if seatbelt is used, 0 otherwise)	0.58/ 0.49	0.58/0.49	0.55/ 0.5	-
Participant Level Cause (1 if the driver followed too closely,0 otherwise)	0.13/0.33	0.13/0.34		
Vehicle Level Action (1 if the driver stopped in traffic not waiting to make a left turn, 0 otherwise)	0.16 /0.37	0.16/0.36		
Roadside (1 if the crash happened at the right roadside, 0 otherwise)	0.33/0.47	-	0.22/0.41	-
Participant level cause (1 if failed to avoid vehicle ahead, 0 otherwise)	0.04/ 0.21	-	0.07/0.26	-
Weekdays (1 if the crash happened during the weekdays, 0 otherwise)	-	-	0.79/0.41	0.78/0.41
Crash Level Cause (1 if the crash happened because careless driving, 0 otherwise)	-	-	0.03/0.18	0.04 /0.19

Variable (Continued)	Mean/Standard Deviation			
	Full Model	Three and Four leg Model	Four leg Model	Three leg Model
Movement of the Vehicle at the Time of the Crash (1 if stopped in traffic, 0 otherwise)	-	-	0.18/0.38	0.15/0.36
Participant Level Cause (1 if the driver followed too closely,0 otherwise)	-	-	0.14/0.35	-
Barrier Condition (1 if fair condition, 0 otherwise)	-	-	0.003/0.05	-
Posted Speed Limit (1 if the speed limit more than 35, 0 otherwise)	-	-	0.61/0.49	-
Crash Level Cause (1 if the driver disregarded other traffic control device, 0 otherwise)	-	-	-	0.04/ 0.19
Crash Level Cause (1 if failed to avoid vehicle ahead, 0 otherwise)	-	-	-	0.07/ 0.25
Barrier Type (1 if concrete type, 0 otherwise)	-	-	-	0.11 /0.31
Weather Condition (1 if cloudy, 0 otherwise)	-	-	-	0.08/0.27
Weather Condition (1 if rainy, 0 otherwise)	-	-	-	0.17/0.38
Alcohol Use (1 if that participant had been drinking, 0 otherwise)	-	-	-	0.06/0.23

* Population range is in thousand.

The vehicle type variable (passenger car, pickup, van, light delivery, and custom van) was found to be significant in all the models, additionally male driver and drivers older than 21 and but less than 36 years old was also found to be significant in all the models. Seatbelt being the safety equipment used by the driver was found to be significant in three models (full, three and four, and four leg models). Participant cause, such as the driver followed too closely and vehicle level action when driver stopped in traffic not waiting to make a left turn were found to be significant in the full and three & four leg models. Crash occurred at the right side of road and participant failed to avoid vehicle ahead were found to be significant in two models (full and four leg models). The variables related to the crash occurring during the weekdays, crash occurred because careless driving, and speed limit were found to be significant in the three and four leg models.

Driver disregarded other traffic control device, failed to avoid vehicle ahead, concrete barrier, cloudy and rainy weather, alcohol involved were all found to be significant in three leg model.

Table 2.3: Dependent Variable Frequency and percentage distribution in all the Models.

Model Number	No Injury	Injury
Full Model	836 (83.1%)	210 (16.9%)
Three and Four leg Model	761 (83.35%)	152 (16.65%)
Four leg Model	590 (83.81%)	114 (16.19%)
Three leg Model	171 (81.82%)	38 (18.18%)

2.4 Methodology

Many discrete choice modeling techniques have been used to formulate crash injury severity models. Such frameworks include multinomial logit models, ordered probit models, binary logit models, etc. For this research, fixed and random parameters binary probit models are used to model the probability of two possible crash severity outcomes. These outcomes represent the aggregation of: (1) injury-type crashes and fatal crashes, and (2) no injury crashes. This is done due to the substantial number of no injury crashes in comparison to crashes that result in injuries. Accordingly, the aggregated injury category consists of fatal, major, moderate, and minor injury outcomes, while the no injury category includes only no injury outcomes. The purpose for this aggregation is to increase the number of observations to reduce the variability caused by random effects when statistical methods are implemented (Chang and Mannering 1999). This is essential since the data that is used in this study has too few observations on incapacitating and fatal injuries to set apart their individual effects. Also, this research aims to discover what is influencing these no injury crashes.

To begin, the binary probit model takes on the form of a binary index response model (Wooldridge 2010):

$$\mathrm{P}(y = 1 \mid \boldsymbol{X}) = G(\boldsymbol{X\beta}) = p(\boldsymbol{X}) \tag{2.1}$$

where $\boldsymbol{X}$ is a $1 \times \boldsymbol{K}$, $\boldsymbol{\beta}$ is a $\boldsymbol{K} \times 1$, and the first element of $\boldsymbol{X}$ is taken to be unity. In the case of the probit model, in which $G(\cdot)$ is a cumulative distribution function (CDF), a more general form of the effect of an explanatory variable $\boldsymbol{X}$ on a binary outcome can be expressed as follows (Wooldridge 2010):

$$y^* = \alpha + \boldsymbol{\beta X} + \varepsilon \tag{2.2}$$

with:

$$y = 1[y^* > 0] \tag{2.3}$$

where $y = 1[y^* > 0]$ represents a crash in which an injury occurred ($y = 0$ otherwise). Considering these formulae, the probit model, which specifies the conditional probability, is then a special case of Eq. (1) (Cameron and Trivedi 2005; Wooldridge 2010):

$$\Phi(\boldsymbol{X}'\boldsymbol{\beta}) = \int_{-\infty}^{\boldsymbol{X}'\boldsymbol{\beta}} \phi(z)\, dz \tag{2.4}$$

Where $\Phi(\cdot)$ is the standard normal CDF, with derivative:

$$\phi(z) = \frac{\exp\left(-z^2/2\right)}{\sqrt{2\pi}} \tag{2.5}$$

where the probit model above is derived if $\mathcal{E}$ in the latent variable formulation has a standard normal distribution.

Using the presented probit model, the probability of being involved in an injury crash (i.e., y takes on the value 1) is computed. Referring to Eq. (1), $\boldsymbol{\beta}$ is a vector of estimable parameters and $\boldsymbol{X}$ represents a vector of explanatory variables (e.g., gender, age, safety equipment use, participant errors, residency of the participant, vehicle ownership, type of vehicle, intended movement, crash location, road surface condition, speed, pavement condition, and effect of striking vehicle), and ε is a disturbance term with a standard normal distribution.

2.4.1 Unobserved Heterogeneity

With the collected data, some of the many factors affecting the likelihood of a crash and the resulting injury severity are likely to be unavailable to the analyst. These unobservable factors, or unobserved heterogeneity, can introduce variation into the model impacting crash likelihood and injury severity (Mannering et al. 2016). For instance, consider gender as an observed human element that affects injury severity outcomes. However, there are clear physiological differences between men and women, as well as many variations across people of the same gender (for instance, differences in heigh, weight, bone density, etc.). These unobservables can result in unobserved heterogeneity, and if not accounted for, can result is biased parameter estimates. Examples of random parameters methods to account for unobserved heterogeneity can be found in Castro *et al.* (2013), Venkataraman *et al.* (2013), and Venkataraman *et al.* (2014). In an attempt to account for this data heterogeneity, a random parameters technique is applied as shown in Eq. (4) (Greene 2012):

$$\beta_i = \beta + u_i \tag{2.6}$$

Where u_i is a randomly distributed term. To estimate these random parameters, maximum likelihood estimation is performed through a simulation-based approach to address the computational complexity of computing the outcome probabilities. The chosen simulation approach utilizes Halton draws which have been shown to provide a more efficient distribution of the draws for numerical integration than purely random draws (Bhat 2003; Halton 1960; Pahukula et al. 2015). Lastly, marginal effects are computed to show the impact of a one-unit change of explanatory variable X on the injury outcome i as shown in Eq. (7) and referred to in Washington *et al.* (2011).

$$\frac{\partial Y}{\partial x_i} = \beta_i \, \phi \, (\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n \tag{2.7}$$

2.4.2 Log-likelihood Test

Maximum likelihood and simulation-based maximum likelihood methods are used to estimate the parameter vector. During analysis, normal, lognormal, triangular, and uniform distributions were considered for the random parameters' distribution; however, only the normal distribution was found to be statistically significant. In addition, the binary probit model is estimated using two hundred Halton draws, as it is stated in the literature that such number of Halton draws produces accurate estimates of the parameters (Bhat 2003; Gkritza and Mannering 2008; Hasan et al. 2011; Milton et al. 2008).

As mentioned before, there are different types of roundabouts (three leg, four leg, and five leg roundabouts). According to ODOT, there were about 1,006 roundabout crashes over a five-year period (2011 to 2015). Of the 1,006 crashes, most occurred on four-leg roundabouts (704) and three-leg roundabouts (209). Grouping the data for roundabouts for analysis may lead to erroneous inferences on the significance of particular explanatory variables. Subsequently, a log-likelihood ratio test is proposed to statistically test the overall significance of using a full model (all roundabout crashes regardless of configuration type) over separate models (a model with crashes on three and four leg roundabouts combined, another with crashes on four leg roundabouts only, and a model with crashes on three leg roundabouts). The first log-likelihood ratio test for transferability is as follows:

$$\chi^2 = -2[LL_{Full}(\beta^{Full}) - \sum LL_{Sep}(\beta^{Sep})] \tag{2.8}$$

where $LL_{Full}(\beta^{Full})$ is the log-likelihood at the convergence of the full model (-354.73), $LL_{Sep}(\beta^{Sep})$ is the log-likelihood at the convergence of a given subgroup (i.e., three and four leg, three leg, and four leg) using the same variables included in the full model, and Sep is the total number of subgroups (-610.36). Using Eq. (8) results in a chi-square statistic of -511.26 (statistic (x^2 = -511.26). The critical chi-square (χ^2) value associated with one-tailed probability level and degrees of freedom which equal to the summation of the number of the random estimated parameters in all separate models minus the number of the random estimated parameters in the full model provide a value much greater than 99.99% of confidence limit which the null hypothesis can be rejected. The null hypothesis states that there is no difference between the model parameters in the full model (all roundabout configuration) and the separate models (i.e., the parameters are the same).

For further validation, a more extensive transferability test was conducted to test if modeling crash severity at the roundabouts need to be modeled separately. This log-likelihood ratio test for transferability is as follows (Washington et al. 2011):

$$x^2 = -2\,LL\big(\beta_{M1_{M2}}\big) - \sum LL(\beta_{M1})] \tag{2.9}$$

Where $LL\big(\beta_{M1_{M2}}\big)$ is the log-likelihood at convergence for model M_1 using the data from model M_2 and is the log-likelihood at convergence for model M_1. As an illustration, in this equation M_1 refers to the model that utilizes the three and four legs data combination and M_2 refers to the model that can predict this data as shown in Table 2.4. Then, the variables and parameters estimate from the three leg best model were fixed and run with the three and four legs data combination. The corresponding log likelihood minus the log likelihood at convergence for three and four legs combination model, will show how well the three legs model (both variables and parameter estimates) can describe the three and four legs data combination.

Table 2.4: Chi-Square Statistics and Degrees of Freedom for Crash Severity related to the Roundabout Type Transferability Test.

M_1	M_2 Three and Four legs (Model)	Three legs (Model)	Four legs (Model)
Three and Four legs (Data)	0	254.44 (13)	146.32 (18)
Three legs (Data)	26.56 (15)	0	144.48 (18)
Four legs (Data)	**8.17*** (15)	287.25(13)	0

The results of the transferability test indicate with well over 99% confidence that injury severity analyses should be modeled according to the type of roundabout. The only exception being the four leg roundabout data, the chi-square 8.17 with fifteen degrees of freedom is less that the critical chi-square 24.996. This indicate that the estimated parameters from the three and four leg combination model are adequately describing the effects for four-legged roundabouts.

2.5 Discussion

Fixed and random parameters binary probit models were estimated based on two severity outcomes (no injury and injury) with 20 variables found to be statistically significant, where various variables were found to have estimated random parameters. The following sections illustrate the final estimation results of modeling crash data at roundabouts in Oregon.

2.5.1 Full Model

The estimation results for the original 1,006 crashes for fixed and random parameters binary probit models are summarized in Table 2.5. The marginal effects, which are illustrated in Table 2.5, provide additional insights on injury severity outcomes, their corresponding probabilities, and the magnitude of change. With regard to the interpretation of the marginal effects for roundabout crashes for example, such as the indicator variable representing drivers who are more than 21 and less than 36 years old, the marginal effects indicate that this age group has a 0.01 higher probability of sustaining an injury compared to other age groups.

Table 2.5: Fixed and Random Parameter Binary Probit Models of Injury Severity for All Roundabout Type.

Variable	Fixed Parameter		Random Parameter		
	Coefficient	t-statistic	Coefficient	t-statistic	Marginal Effects
Constant	-1.11	-1.87	-1.5	0.24	
Population (1 if the population is between 50-100, 0 otherwise)	-0.45	-3.77	-.90	-4.67	-0.01
Crash Level Cause (1 if the driver did not yield right-of-way, 0 otherwise)	-0.55	-4.10	-1.25	-5.26	-0.02
Vehicle ownership (1 if private, 0 otherwise)	1.75	2.84	3.86	2.91	0.05
Vehicle Type (1 vehicle weight <10,000 lb, 0 otherwise)	-1.64	-7.14	-3.51	-9.05	-0.05
Age of Driver (1 if 21< age ≤35, 0 otherwise)	0.3	2.45	0.65	3.53	0.01
Participant Level Cause (1 if the driver followed too closely,0 otherwise)	-0.72	-3.19	-1.56	-3.78	-0.02
Participant level cause (1 if failed to avoid vehicle ahead, 0 otherwise)	-1.05	-2.06	-2.19	-2.16	-0.03
Vehicle Level Action (1 if the driver stopped in traffic not waiting to make a left turn, 0 otherwise)	0.51	3.62	0.92	4.40	0.01
Standard Deviation of Parameter, Normally Distributed			0.56	3.54	

Variable (Continued)	Fixed Parameter		Random Parameter		
	Coefficient	t-statistic	Coefficient	t-statistic	Marginal Effects
Safety Equipment Use (1 if seatbelt is used, 0 otherwise)	0.94	6.97	0.87	4.15	0.01
Standard Deviation of Parameter, Normally Distributed			2.25	11.25	
Gender (1 if male, 0 otherwise)	-.052	-4.55	-1.32	-6.29	-0.02
Standard Deviation of Parameter, Normally Distributed			1.18	7.53	
Roadside (1 if the crash happened at the right roadside, 0 otherwise)	-0.49	-3.92	-1.32	-5.83	-0.02
Standard Deviation of Parameter, Normally Distributed			1.07	6.19	
Model Statistics					
Log-likelihood function	-361.08			-354.73	
McFadden Pseudo R-squared	0.21			0.22	
Number of Observation	1006			1006	
Log-likelihood function at Zero			-457.001		

Turning to the model, if a crash occurred where the driver was stopped in traffic and not waiting to make a left turn, there is an increase in the outcome probability of sustaining an injury. In addition, the estimated parameter for stopping in traffic and not waiting to make a left turn was found to be random and normally distributed with a mean of 0.92 and standard deviation of 0.56. This suggests that for 5% of crashes where the driver was stopped in traffic and not waiting to make a left turn the estimated parameter mean is less than zero, while 95% of them have an estimated parameter greater than zero. In other words, 5% of crashes involving drivers who stopped in traffic and not waiting to make a left turn are less likley to result in an injury, yet 95% are more likely to sustain an injury. This could possibly be attributed to unfamiliarity.

Seatbelt use by the driver was found to be significant and the estimated parameter was found to be random and normally distributed with a mean of 0.87 and standard deviation of 2.25. This implies that for roughly 35% (less than zero) of drivers, seatbelt decreased the likelihood of an injury while for 65.1 % of them it increased the likelihood of sustaining an injury. In spite of the benefits of the seatbelt in saving lives, there is a probability of getting injured due to unobserved factors. For example, body physiology differences and proper use of in-vehicle restraints as mentioned in Islam and Hernandez (2013). Anderson and Hernandez (2017) found a similar result on the use of seatbelts and the effects on injury severity. Comparably, other studies have suggested that seatbelt use is over-reported because of the legal implications of not wearing seatbelts (Amoros et al. 2006; Li et al. 1999; Malliaris et al. 1996; Stewart 1993; Streff and Wangenaar 1989). Li *et al.* (1999) showed that Australian police-reported seatbelt use overestimated actual use by 9% in crashes resulting in injuries (Li et al. 1999). Chen *et al.* illustrated that using seatbelts will significantly reduce the likelihood of drivers being fatally injured in rear-end collisions (Chen et al. 2015). However, Xie *et al.* concluded that wearing seatbelts could result in possible injuries to the participants, but was still critical for mitigating driver injury severity (Xie et al. 2012).

The indicator for males was also found to be statistically significant and negative. This may indicate that males are less likely to be involved in injury crashes, and this might be due to the physical differences between males and females as previously mentioned with seatbelt use (driver physiology). This indicator was also found to have a random and normally distributed estimated parameter with a mean of -1.32 and standard deviation of 1.18. This suggests that approximately 13.2% of observations have a mean of more than zero. That is to say, 13.2% of males are more likely to get injured in crashes, which follows findings of previous work (Al-Thaifani et al. 2016a; Leidman et al. 2016; Ulfarsson and Mannering 2004) . On the other hand, 86.8% of male drivers are less likely to sustain an injury.

Finally, for the roadside variable, results show that crashes which happened on the right roadside of the road have an estimated random parameter that is normally distributed with a mean of -1.32 and a standard deviation of 1.07. This implies that for roughly 11% of crashes that happened on the right side of the road increase the likelihood of sustaining an injury, while 89.1% of such crashes decrease the likelihood of sustaining an injury. This is most likely capturing driver inattentiveness.

2.5.2 Three and Four leg Combination Model

Table 2.6 illustrates the results of the three & four leg fixed and random parameter models.

Table 2.6: Fixed and Random Parameter Binary Probit Models Results for Three and Four leg Roundabout Combination.

Variable	Fixed Parameter		Random Parameter		
	Coefficient	t-statistic	Coefficient	t-statistic	Marginal Effects
Constant	-2.13	-2.93	-3.98	-2.61	
Population (1 if the population is between 10-25, 0 otherwise)	0.59	2.67	1.10	3.22	0.01
Crash Level Cause (1 if the driver did not yield right-of-way, 0 otherwise)	-0.43	-3.05	-1.001	-4.12	-0.01
Vehicle Movement (1 if turning right, 0 Otherwise)	-0.63	-2.66	-0.9	-2.87	-0.005
Vehicle Ownership (1 if private, 0 otherwise)	1.60	2.26	4.17	2.85	0.02
Vehicle Type (1 vehicle weight <10,000 lb, 0 otherwise)	-1.67	-6.84	-4.06	-8.55	-0.02
Age of driver (1 if 15< age ≤21, 0 otherwise)	0.52	2.72	2.21	3.24	0.01
Age of driver (1 if 21< age ≤35, 0 otherwise)	0.70	3.47	2.60	3.98	0.01
Age of driver (1 if 35< age ≤50, 0 otherwise)	-0.55	3.21	2.56	3.90	0.01
Age of driver (1 if 50< age ≤65, 0 otherwise)	1.06	2.39	1.84	2.82	0.01
Age of driver (1 if 65< age, 0 otherwise)	1.27	2.02	1.34	1.99	0.01
Participant Level Cause (1 if the driver followed too closely,0 otherwise)	1.18	0.05	-1.59	-3.63	-0.01
Safety Equipment Use (1 if seatbelt is not used, 0 otherwise)	0.9	4.93	0.8	3.44	0.004
Standard Deviation of Parameter, Normally Distributed			2.02	10.44	
Vehicle Movement (1 if the vehicle stopped in traffic, 0 otherwise)	0.78	3.38	0.8	3.16	0.004
Standard Deviation of Parameter, Normally Distributed			1.8	6.79	
Gender (1 if male, 0 otherwise)	-0.76	-4.55	-1.99	-7.07	-0.01
Standard Deviation of Parameter, Normally Distributed			1.87	8.48	
Model Statistics					
Log-likelihood function	-318.76		-314.30		
McFadden Pseudo R-squared	0.22		0.24		
Number of Observation	913		913		
Log-likelihood function at Zero			-411.09		

As seen from Table 2.6, three indicators were found to have random and normally distributed estimated parameters. As with the previous model (Table 2.5), males, seatbelt use, and being stopped in traffic were all found to have estimated random parameters. No seatbelt use was found to be random and normally distributed with a mean of 0.80 and a standard deviation of 2.02. This implies that for 34.6% of drivers who did not wear their seatbelt were less likely to be involved in an injury crashes and 65.4% were more likely. One possible explanation for this result may stem from the influence of speed at the time of the crash. Again, roundabouts are a known traffic calming countermeasure where lower speeds are generally observed.

Next, the parameter for vehicles being stopped in traffic was found to be random and normally distributed with a mean of 0.80 and a standard deviation of 1.80. This indicates that for 32.8% of crashes in which the vehicle was stopped in traffic the driver is less likely to sustain an injury and 67.2% of drivers are more likely. This might be due to the operational characteristics of roundabouts; specifically, as traffic approaches an entry point, drivers may have to stop to yield to traffic in the roundabout.

2.5.3 Four leg Model

Table 2.7 show the results of the four leg fixed and random parameter models.

Table 2.7: Random Parameter Binary Probit Model Results for Four leg Roundabout.

Variable	Fixed Parameter		Random Parameter		
	Coefficient	t-statistic	Coefficient	t-statistic	Marginal Effects
Constant	-0.58	-1.73	-0.70	-1.36	
Posted Speed Limit (1 if the speed limit more than 35, 0 otherwise)	0.30	2.09	0.85	3.34	0.001
Population (1 if the population is between 10-25, 0 otherwise)	0.62	2.34	1.32	2.82	0.002
Weekdays (1 if the crash happened during the weekdays, 0 otherwise)	0.34	1.94	0.99	3.18	0.002
Crash Level Cause (1 if the driver did not yield right-of-way, 0 otherwise)	-0.50	-2.90	-1.30	-4.01	-0.002
Participant level cause (1 if failed to avoid vehicle ahead, 0 otherwise)	-0.63	-2.07	-1.39	-2.72	-0.002
Crash Level Cause (1 if the crash happened because careless driving, 0 otherwise)	-0.92	-2.33	-2.46	-3.52	-0.004
Movement of the Vehicle at the Time of the Crash (1 if turning right, 0 otherwise)	-0.81	-2.52	-2.06	-3.43	-0.003
Movement of the Vehicle at the Time of the Crash (1 if stopped in traffic, 0 otherwise)	0.55	3.01	1.54	4.81	0.002
Age of driver (1 if $15 < \text{age} \leq 21$, 0 otherwise)	0.67	2.87	1.74	4.03	0.003

Variable (Continued)	Fixed Parameter		Random Parameter		
	Coefficient	t-statistic	Coefficient	t-statistic	Marginal Effects
Age of driver (1 if 21< age ≤35, 0 otherwise)	0.92	5.23	2.29	6.52	0.003
Participant Level Cause (1 if the driver followed too closely,0 otherwise)	-1.05	-3.50	-2.68	-4.36	-0.004
Roadside (1 if the crash happened at the right roadside, 0 otherwise)	-0.4	-2.30	-0.95	-3.35	-0.001
Condition of the Barrier (1 if fair condition, 0 otherwise)	2.62	2.27	6.88	3.13	0.01
Safety Equipment Use (1 if seatbelt is used, 0 otherwise)	0.99	5.89	1.58	5.12	0.002
Standard Deviation of Parameter, Normally Distributed			2.24	8.61	
Age of driver (1 if 35< age <51, 0 otherwise)	0.57	3.23	0.69	2.07	0.001
Standard Deviation of Parameter, Normally Distributed			2.37	6.57	
Vehicle Type (vehicle weight <10,000 lb, 0 otherwise)	-1.59	-5.49	-3.96	-6.82	-0.01
Standard Deviation of Parameter, Normally Distributed			0.53	4.28	
Gender (1 if male, 0 otherwise)	-0.49	-3.41	-1.6	-5.20	-0.002
Standard Deviation of Parameter, Normally Distributed			1.62	6.87	
Model Statistics					
Log-likelihood function	-228.66		-224.34		
McFadden Pseudo R-squared	0.27		0.28		
Number of Observation	704		704		
Log-likelihood function at Zero			-311.77		

For the four leg model, seventeen variables were found to be significant of which three of them were found to have estimated random parameters. Turning to the random parameters, the parameter for seatbelt use by the driver was found to be random and normally distributed with a mean of 1.58 and a standard deviation of 2.24. This suggests that for roughly 24.1% of drivers who wore a seatbelt were less likely to sustain an injury and 75.9% of drivers were more likely.

The indicator for drivers aged 35 years to 51 years was found to have a random and normally distributed estimated parameter with a mean of 0.69 and standard deviation of 2.37. This suggests that for 38.6% of drivers in this age group the likelihood of sustaining an injury decreases, while for 61.5% the opposite is true. In general, other studies have found similar results (Al-Thaifani et al. 2016b; Amoros et al. 2006; Bédard et al. 2002; Daniels et al. 2011; Ulfarsson and Mannering 2004). Mannering and Bhat,

(2014) found that for drivers in this age group the likelihood of getting injured decreases.

Vehicle type with weight less than 10,000 lb. was found also to be random and normally distribute with a mean of -3.96 and standard deviation of 0.53. This suggests that for roughly a small percent of the drivers who drive these types of vehicle have an increased probability of injury, whereas a larger proportion of them have the opposite effect. A possible explanation for this specific observation could be due to the increased aggressive driving behavior (e.g., entering the roundabout) a finding consistent with research that explored smaller to medium sized vehicle speeds to larger ones (Al-Thaifani et al. 2016b).

The indicator variable for male is statistically significant. The associated parameter was also found to be random and normally distributed with a mean of -1.60 and a standard deviation of 1.62. This suggests that for approximately 16.2% of male drivers there is an increase in the likelihood of sustaining an injury, while for 83.8% the opposite is true. Similar results are also found in Ulfarsson and Mannering, (2004), Al-Thaifani, Al-Rabeei and Dallak, (2016), Leidman *et al.*, (2016), and Grivna, Eid and Abu-zidan, (2017).

2.5.4 Three leg Model

Finally, for the three leg model, twelve variables were found to be significant and four of them were found to have random and normally distributed estimated parameters as shown in Table 2.8.

Table 2.8: Random Parameter Binary Probit Model Results for Roundabout with Three leg.

Variable	Fixed Parameter		Random Parameter		
	Coefficient	t-statistic	Coefficient	t-statistic	Marginal Effects
Constant	-0.82	-1.14	0.32	0.27	
Weather Condition (1 if cloudy, 0 otherwise)	0.65	1.67	1.16	1.93	-0.01
Weather Condition (1 if rainy, 0 otherwise)	-1.17	-2.32	-3.44	-3.47	-0.007
Weekdays (1 if the crash happened during the weekdays, 0 otherwise)	-0.77	-2.44	-1.88	-3.57	0.007
Crash Level Cause (1 if the driver disregarded other traffic control device, 0 otherwise)	0.95	1.73	1.93	2.05	-0.01

Variable (Continued)	Fixed Parameter		Random Parameter		
	Coefficient	t-statistic	Coefficient	t-statistic	Marginal Effects
Crash Level Cause (1 if the driver failed to avoid vehicle ahead, 0 otherwise)	-1.22	-1.77	-3.08	-2.16	-0.01
Vehicle ownership (1 if private, 0 otherwise)	1.96	2.34	2.92	1.99	0.01
Gender (1 if male, 0 otherwise)	-0.97	-3.38	-2.51	-4.04	-0.01
Barrier Type (1 if concrete type, 0 otherwise)	0.67	1.73	1.62	2.42	0.01
Vehicle Type (1 vehicle weight <10,000 lb, 0 otherwise)	-1.50	-3.49	-3.88	-4.50	-0.01
Standard Deviation of Parameter, Normally Distributed			2.01	5.06	
Crash Level Cause (1 if the crash happened because careless driving, 0 otherwise)	1.23	2.38	2.89	3.02	0.01
Standard Deviation of Parameter, Normally Distributed			1.86	1.83	
Alcohol Use (1 if that participant had been drinking, 0 otherwise)	1.06	2.16	1.12	1.03	0.004
Standard Deviation of Parameter, Normally Distributed			4.32	2.80	
Movement of the Vehicle at the Time of the Crash (1 if stopped in traffic, 0 otherwise)	1.15	3.55	2.66	4.03	0.01
Standard Deviation of Parameter, Normally Distributed			1.07	2.30	
Model Statistics					
Log-likelihood function	-72.64		-71.72		
McFadden Pseudo R-squared	0.27		0.28		
Number of observations	209		209		
Log-likelihood function at Zero		-99.1			

With respect to the random parameters, the estimated parameters for vehicle type, careless driving crash-level cause, driver alcohol use, and vehicles stopped in traffic were all found to be random and significant. In regard to vehicle type, the indicator for passenger car, pickup, van, light delivery, and custom van was found to have a random and normally distributed parameter with a mean of -3.88 and a standard deviation of 2.01. This suggests that roughly 3% of the drivers who drive these types of vehicle have an increased probability of sustaining an injury, whereas 97% of them are less likely. Again, as previously stated, a possible explanation for this specific observation could be due to the increased aggressive driving behavior (Al-Thaifani et al. 2016b).

The estimated parameter for careless driving was also found to random and normally distributed, with a mean of 2.89 and a standard deviation of 1.83. This suggests that for

6% of crashes where the crash-level cause was reported to be careless driving are less likely to result in an injury, while 94% of such crashes are more likely to result in an injury.

The next random parameter is associated with driver alcohol use. Specifically, the parameter for driver alcohol use was found to random and normally distributed with a mean of 1.12 and a standard deviation of 4.32. This suggests that for 39.8% of drivers who had been drinking alcohol, the probability of an injury decreased. On the other hand, 60.2% of drivers who had been drinking were more likely to sustain an injury. This random parameter may be capturing the varying degree of inebriation on driver performance and safety attitude around three legged roundabouts (Zhao et al. 2014).
The estimated parameter for vehicles that were stopped in traffic at the time of the crash was found to be random and normally distributed. A mean of 2.66 and a standard deviation of 1.07 suggest that for 0.06% crashes that occurred with a vehicle stopped in traffic were less likely to result in an injury, but 99.4% of them were more likely to result in an injury. Although stopping inside the roundabout is prohibited and dangerous, there is the possibility that in situations with dense traffic inside the roundabout or potential hazards, stopping may prevent a more serious crash from occurring.

2.6 Summary and Conclusions

This study involved the estimation of a random parameters binary probit model to capture the significant factors that contribute to specific levels of injury severity sustained by drivers involved in crashes at roundabouts in different locations in Oregon. Four models were estimated using five years (2011 to 2015) of Oregon crash data. For the current study, two injury severity outcomes were considered: no injury and injury. The four estimated models were based on the geometric design of the roundabout: full model (unknown, three leg, four leg, and five leg), three and four-leg combination model, three leg model, and four leg model.

A number of important factors were found to influence the level of injury severity at roundabouts. In each individual model, a number of variables are homogenous across crash observations (i.e., their estimated parameters are fixed across observations) and various variables are heterogeneous across crash observations (i.e., they have estimated random parameters). For example, vehicles stopped in traffic and not waiting to make a left turn, seatbelt usage, gender, type of vehicle, roadside crash characteristics, vehicle movement, age of the driver, careless driving, and alcohol use were found to have estimated random parameters.

This study provides useful insights and an increased understanding of the factors that contribute to either sustaining injury or not in in crashes at roundabouts through a random parameters approach. Although the results of this study are exploratory, they provide evidence that crashes are occurring at roundabouts and several factors lead to crashes that result in an injury. In addition, the modeling approach offers a methodology that can account for unobservable in the crash data.

This study aimed to analyze current and available databases to determine the most significant factors that contribute to injuries in crashes at roundabouts in Oregon. In future work, additional crash-specific variables are recommended to investigate roundabout injury severity, such as the specific location of the crash or additional geometric design details. In doing so, an injury severity picture with a higher resolution can be obtained, which in turn can offer more understanding of the design related factors that lead to severe crashes at roundabouts.

Acknowledgment

We would like to acknowledge the Oregon Department of Transportation, specifically the Transportation Data Department, for providing the crash dataset and for helping in our understanding of the data. Findings of this study do not necessarily reflect the views of the Oregon Department of Transportation.

Reference

AARP and WALC. (2014), “Modern Roundabouts”, No. August 2000.

Al-Bdairi, N.S.S. and Hernandez, S. (2017), “An empirical analysis of run-off-road injury severity crashes involving large trucks”, *Accident Analysis and Prevention*, Vol. 102, pp. 93–100.

Al-Bdairi, N.S.S., Hernandez, S. and Anderson, J. (2018), “Contributing Factors to Run-Off-Road Crashes Involving Large Trucks under Lighted and Dark Conditions”, *Journal of Transportation Engineering, Part A: Systems*, Vol. 144 No. 1, pp. 1–9.

Al-Thaifani, A.A., Al-Rabeei, N.A. and Dallak, A.M. (2016), “Study of the Injured Persons and the Injury Pattern in Road Traffic Accident in Sana’a City, Yemen”, *Advances in Public Health*, Vol. 2016, pp. 1–5.

Amoros, E., Martin, J.L. and Laumon, B. (2006), “Under-reporting of road crash casualties in France”, *Accident Analysis and Prevention*, Vol. 38 No. 4, pp. 627–635.

Anderson, J. and Hernandez, S. (2017), “Roadway classifications and the accident injury severities of heavy-vehicle drivers”, *Analytic Methods in Accident Research*, Elsevier Ltd, Vol. 15, pp. 17–28.

Bédard, M., Guyatt, G.H., Stones, M.J. and Hirdes, J.P. (2002), “The independent contribution of driver, crash, and vehicle characteristics to driver fatalities”, *Accident Analysis and Prevention*, Vol. 34 No. 6, pp. 717–727.

Bhat, C.R. (2003), “Simulation estimation of mixed discrete choice models using randomized and scrambled Halton sequences”, *Transportation Research Part B: Methodological*, Vol. 37 No. 9, pp. 837–855.

De Brabander, B., Nuyts, E. and Vereeck, L. (2005), “Road safety effects of roundabouts in Flanders”, *Journal of Safety Research*, Pergamon, Vol. 36 No. 3, pp. 289–296.

Cameron, A.C. and Trivedi, P.K. (2005), *Microeconometrics: Methods and Applications*, Cambridge University Press, New York, NY.

Castro, M., Paleti, R. and Bhat, C.R. (2013), “A spatial generalized ordered response model to examine highway crash injury severity”, *Accident Analysis and Prevention*, Elsevier Ltd, Vol. 52, pp. 188–203.

Chang, L.Y. and Mannering, F. (1999), “Analysis of injury severity and vehicle occupancy in truck- and non-truck-involved accidents”, *Accident Analysis and Prevention*, Vol. 31 No. 5, pp. 579–592.

Chen, C., Zhang, G., Tarefder, R., Ma, J., Wei, H. and Guan, H. (2015), “A multinomial logit model-Bayesian network hybrid approach for driver injury severity analyses in rear-end crashes”, *Accident Analysis and Prevention*, Vol. 80, pp. 76–88.

Daniels, S., Brijs, T., Nuyts, E. and Wets, G. (2010a), “Externality of risk and crash severity at roundabouts”, *Accident Analysis and Prevention*, Elsevier Ltd, Vol. 42 No. 6, pp. 1966–1973.

Daniels, S., Brijs, T., Nuyts, E. and Wets, G. (2010b), “Explaining variation in safety performance of roundabouts”, *Accident Analysis and Prevention*, Vol. 42 No. 2, pp. 393–402.

Daniels, S., Brijs, T., Nuyts, E. and Wets, G. (2011), “Extended prediction models for crashes at roundabouts”, *Safety Science*, Elsevier Ltd, Vol. 49 No. 2, pp. 198–207.

Daniels, S., Nuyts, E. and Wets, G. (2008), “The effects of roundabouts on traffic safety for bicyclists: An observational study”, *Accident Analysis and Prevention*, Vol. 40 No. 2, pp. 518–526.

FHWA. (2015), “A Review of Fatal and Severe Injury Crashes at Roundabouts”, Vol. IV No. September.

Gkritza, K. and Mannering, F.L. (2008), “Mixed logit analysis of safety-belt use in single- and multi-occupant vehicles.”, *Accident Analysis and Prevention*, Vol. 40 No. 2, pp. 443–51.

Greene, W.H. (2012), *Econometric Modeling Guide*, 10th ed., Econometric Software, Inc., Plainview, NY.

Grivna, M., Eid, H.O. and Abu-zidan, F.M. (2017), “Youth traffic-related injuries : a prospective study”, *World Journal of Emergency Surgery*, World Journal of Emergency Surgery, pp. 1–7.

Gross, F., Lyon, C., Persaud, B. and Srinivasan, R. (2013), “Safety effectiveness of converting signalized intersections to roundabouts”, *Accident Analysis and Prevention*, Elsevier Ltd, Vol. 50, pp. 234–241.

Haleem, K. and Abdel-Aty, M. (2010), “Examining traffic crash injury severity at unsignalized intersections”, *Journal of Safety Research*, Elsevier Ltd, Vol. 41 No. 4, pp. 347–357.

Halton, J.H. (1960), “On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals”, *Numerishe Mathematik*, Vol. 2, pp. 84–90.

Hasan, S., Ukkusuri, S., Gladwin, H. and Murray-Tuite, P. (2011), “Behavioral Model to Understand Household-Level Hurricane Evacuation Decision Making”, *Journal of Transportation Engineering*, Vol. 137 No. 5, pp. 341–348.

Islam, M. Bin and Hernandez, S. (2013), "Modeling Injury Outcomes of Crashes Involving Heavy Vehicles on Texas Highways", *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2388, pp. 28–36.

Kim, S. and Choi, J. (2013), "Safety analysis of roundabout designs based on geometric and speed characteristics", *KSCE Journal of Civil Engineering*, Vol. 17 No. 6, pp. 1446–1454.

Leidman, E., Maliniak, M., Sultan, A.-S.S., Hassan, A., Hussain, S.J. and Bilukha, O.O. (2016), "Road traffic fatalities in selected governorates of Iraq from 2010 to 2013: prospective surveillance.", *Conflict and Health*, Conflict and Health, Vol. 10, p. 2.

Li, L., Kim, K. and Nitz, L. (1999), "Predictors of safety belt use among crash-involved drivers and front seat passengers: Adjusting for over-reporting", *Accident Analysis and Prevention*, Vol. 31 No. 6, pp. 631–638.

Liu, Y., Li, Z., Liu, J. and Patel, H. (2016), "Data in Brief Vehicular crash data used to rank intersections by injury crash frequency and severity", *Data in Brief*, Elsevier, Vol. 8, pp. 930–933.

Lombardi, D.A., Horrey, W.J. and Courtney, T.K. (2017), "Age-related differences in fatal intersection crashes in the United States", *Accident Analysis and Prevention*, Elsevier Ltd, Vol. 99, pp. 20–29.

Malliaris, A., DeBlois, J. and Digges, K. (1996), "Light vehicle occupant ejections--a comprehensive investigation.", *Accident Analysis and Prevention*, Vol. 28 No. 1, pp. 1–14.

Mannering, F.L. and Bhat, C.R. (2014), "Analytic methods in accident research: Methodological frontier and future directions", *Analytic Methods in Accident Research*, Elsevier, Vol. 1, pp. 1–22.

Mannering, F.L., Shankar, V. and Bhat, C.R. (2016), "Unobserved heterogeneity and the statistical analysis of highway accident data", *Analytic Methods in Accident Research*, Elsevier, Vol. 11, pp. 1–16.

Milton, J.C., Shankar, V.N. and Mannering, F.L. (2008), "Highway accident severities and the mixed logit model: An exploratory empirical analysis", *Accident Analysis and Prevention*, Vol. 40 No. 1, pp. 260–266.

Montella, A. (2011), "Identifying crash contributory factors at urban roundabouts and using association rules to explore their relationships to different crash types", *Accident Analysis and Prevention*, Elsevier Ltd, Vol. 43 No. 4, pp. 1451–1463.

NHTSA. (2009), *TRAFFIC SAFETY FACTS 2009*, *National Highway Traffic Safety Administration*.

Nikitin, N., Patskan, V. and Savina, I. (2017), “Efficiency Analysis of Roundabout with Traffic Signals”, *Transportation Research Procedia*, available at:http://doi.org/10.1016/j.trpro.2017.01.072.

Pahukula, J., Hernandez, S. and Unnikrishnan, A. (2015), “A time of day analysis of crashes involving large trucks in urban areas”, *Accident Analysis & Prevention*, Elsevier Ltd, Vol. 75, pp. 155–163.

Qin, X., Khan, G., Bill, A. and Noyce, D.A. (2011), “Comprehensive Safety Evaluation of Roundabouts in Wisconsin”, *Journal of Transportation Safety & Security*, Vol. 3 No. 4, pp. 289–303.

Robinson, B., Rodegerdts, L.A., Troutbeck, R., Brilon, W., Bondizio, L., Courage, K., Kyte, M., et al. (2000), *Roundabouts : An Informational Guide*, *FHWA-RD-00-067*.

Romo, A., Hernandez, S. and Cheu, R.L. (2014), “Identifying Precrash Factors for Cars and Trucks on Interstate Highways: Mixed Logit Model Approach”, *Journal of Transportation Engineering*, Vol. 140 No. 3, p. 04013016.

Stewart, J.R. (1993), “Seat belt use and accident involvment: A comparison of driving behavior before and after a seat belt law”, *Accident Analysis & Prevention*, Vol. 25 No. 6, pp. 757–763.

Streff, M.F. and Wangenaar, C.A. (1989), “Are there really shortcut? estimating seat belt use with self-report measures”, *Accident Analysis & Prevention*, Vol. 21 No. 6, pp. 509–516.

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

Venkataraman, N., Shankar, V., Ulfarsson, G.F. and Deptuch, D. (2014), “A heterogeneity-in-means count model for evaluating the effects of interchange type on heterogeneous influences of interstate geometrics on crash frequencies”, *Analytic Methods in Accident Research*, Elsevier, Vol. 2, pp. 12–20.

Venkataraman, N., Ulfarsson, G.F. and Shankar, V.N. (2013), “Random parameter models of interstate crash frequencies by severity, number of vehicles involved, collision and location type”, *Accident Analysis and Prevention*, Elsevier Ltd, Vol. 59, pp. 309–318.

Washington, S.P., Karlaftis, M.G. and Mannering, F.L. (2010), *Statistical and Econometric Methods for Transportation Data Analysis, Second Edition*, CRC

Press, available at: http://books.google.com/books?id=-UcWqyWQc-IC&pgis=1 (accessed 22 July 2014).

Washington, S.P., Karlaftis, M.G. and Mannering, F.L. (2011), *Statistical and Econometric Methods for Transportation Data Analysis*, 2nd ed., CRC press, Boca Raton, FL.

Wooldridge, J.M. (2010), *Econometric Analysis of Cross Section and Panel Data*, 2nd ed., MIT Press, Cambridge, MA, available at:http://doi.org/10.1016/B978-0-323-05712-7.00031-3.

Xie, Y., Zhao, K. and Huynh, N. (2012), "Analysis of driver injury severity in rural single-vehicle crashes", *Accident Analysis and Prevention*, Elsevier Ltd, Vol. 47, pp. 36–44.

Zhao, X., Zhang, X. and Rong, J. (2014), "Study of the Effects of Alcohol on Drivers and Driving Performance on Straight Road", *Mathematical Problems in Engineering*, Hindawi, Vol. 2014, pp. 1–9.

Injury Severity Prediction at Roundabouts and Variable Selection with Econometric and Machine Learning Methods

By: Hamsa Abbas Zubaidi, Jason Anderson Ph.D., and Salvador Hernandez Ph.D.

Submitted to Journal: International Journal of Transportation Science and Technology
Address:
Issue:

Chapter 3: Injury Severity Prediction at Roundabouts and Variables Selection with Econometric and Machine Learning Methods

By:

Hamsa Abbas Zubaidi
*School of Civil and Construction Engineering
Oregon State University
Corvallis, Oregon 97331-3212.
zubaidih@oregonstate.edu
* Roads and Transport Department, College of Engineering, University of Al-Qadisiyah, Iraq.

Jason C. Anderson, PhD.
Research Faculty
Department of Civil and Environmental Engineering
Portland State University
Portland, OR 97201
(503) 725-4225
jason.c.anderson@pdx.edu

Salvador Hernandez, PhD.
Associate Professor
School of Civil and Construction Engineering
Oregon State University
Corvallis, OR 97331-3212
(541)-737-1757
sal.hernandez@oregonstate.edu

Abstract

The objective of this work is to assess the effects of variable selection methods on injury severity prediction at roundabouts using Oregon crash data. Variable selection was first determined using a random parameter binary probit model (econometric model), then through the use of a random forest. Based on marginal effects from the econometric model, careless driving, passenger cars, and male drivers have the highest effects on injury severity outcomes at three-leg roundabouts. For four-leg roundabouts, passenger cars, following too closely, and drivers aged 22-35 years have the highest effects on injury severity outcomes. From the random forest approach, time-of-day (afternoon), snowy weather, and drivers aged 36-50 years were found to be the most important injury severity predictors at three-leg roundabouts. At four-leg roundabouts, the most important injury severity predictors were poor pavement condition, lighting (dusk), and drivers losing control of the vehicle. Although these were deemed the most important, or impactful, a variety of additional variables were also considered during prediction comparison. Using the identified significant, or important, variables in each approach, injury severity predictions were compared between a traditional econometric model (binary probit) and a machine learning model, Support Vector Machine (SVM), where selected variables were used to predict in both models. For both prediction approaches, various training-and-testing proportions were considered, including 70-30, 80-20, and 90-10. In the end, regardless of the variable selection approach, the SVM models (regardless of kernel function) outperformed the econometric model in injury severity prediction. This work highlights that under pure prediction consideration, it is not necessary to consider significant explanatory factors and variable selection through a machine learning approach results in a higher accuracy. However, this work also highlights the necessity of a traditional econometric approach to account for data limitations and make inference on the likelihood of a respective injury severity outcome.

Keywords: Machine learning, Roundabout, Injury severity, Random parameters binary probit model, Support vector machine

3.1 Introduction

In recent years, many communities have resorted to either converting existing intersections to roundabouts or only using intersections as a calming traffic measure for improved safety. Although roundabouts decrease the probability of more severe crashes, the Federal Highway Administration claims that less severe crashes are on the rise (FHWA 2015). Understanding the contributing factors to such crashes is traditionally studied through the application and estimation of statistical and econometric models. Conventional models applied in this regard are those that take into account the ordinal nature of injury outcomes (e.g., no injury, non-fatal injury, and fatal injury) (Al-Bdairi and Hernandez 2017; Islam and Hernandez 2013b; Savolainen et al. 2011). In recent years, more advanced statistical models have been proposed for estimating crash injury severities that account for unobserved heterogeneity (Anderson and Hernandez 2017a; b; Mannering et al. 2016; Romo et al. 2014; Tay 2015b). Statistical models have been widely used for injury severity analysis, but they come with certain limitations.

Traditionally, statistical models (e.g., parametric models) require *a priori* assumptions about the data distribution and have predefined underlying relationships between response (dependent) and explanatory (independent) variables. The difficulty of validating such assumptions in some cases could lead to erroneous estimations of model parameters. Additionally, these models' prediction accuracy is often low; albeit, their strength is in explanatory power and addressing key data limitations (Ahmadi et al., 2018; Iranitalab and Khattak, 2017a; Abdel-Aty and Haleem, 2011).

Non-parametric methods and/or artificial intelligence models for analyzing injury severity have become popular due to their ability to outperform traditional statistical methods in predicting severity outcomes (Tixier et al., 2016). Specifically, machine learning provides a methodological approach that can account for nonlinearity and eliminates concerns of multicollinearity (Das et al. 2021; Goldstein et al. 2017; Storm et al. 2019; Wahab and Jiang 2019).

From a safety perspective, Abdel-Aty and Haleem (2011) explored the potential of applying a recently developed machine learning technique called multivariate adaptive regression spline (MARS) at unsignalized intersections; they then compared its results with those of an estimated negative binomial (NB) model. Li et al. (2012) developed a support vector machine (SVM) model for predicting crash injury severities at freeway diverge areas and compared their findings with those of an ordered probability model.

Zeng and Huang (2014) proposed a convex combination algorithm to train a neural network model for two-vehicle crash injury severity prediction; then, they compared it with the popular binary probit model. Chen et al. (2016) employed SVM models to investigate driver injury severity patterns in rollover crashes based on two years of crash data gathered in New Mexico. An artificial neural network was utilized by Alkheder et al. (2017) to predict the injury severity of traffic accidents based on 5,973 traffic accidents that were recorded in Abu Dhabi from 2008 to 2013. Iranitalab and Khattak (2017) used a multinomial logit model, nearest neighbor classification, SVM, and random forest (RF) prediction methods for classifying two-vehicle crash injury outcomes. A deep-learning model using a recurrent neural network was developed and employed by Sameen and Pradhan (2017) to predict the injury severity of traffic accidents based on 1130 accident records that occurred along Malaysia's North-South Expressway from 2009 to 2015. Ahmadi et al. (2018) applied a support vector machine, multinomial logit, and mixed logit for modeling the severity of rear-end crashes for five years of data from California. Zhang et al. (2018) compared the predictive performance, specifically prediction accuracy and estimation of variable importance, of various machine learning and statistical methods with distinct modeling logic in crash severity analysis for freeway diverge areas. Li et al. (2018) explored the process of significant factor identification from a multi-objective optimization standpoint at interstate highways. A two-layer stacking framework was proposed by Tang et al. (2019) to predict the crash injury severity at freeway diverge areas.

The application of non-parametric methods and/or artificial intelligence models to safety has become contemporary. A reason for this is that these methods have the potential to highlight important factors as they relate to injury severity, and they have more predictive power than do conventional econometric techniques (Abdel-Aty and Haleem 2011; Ahmadi et al. 2018; Wahab and Jiang 2019). With that in mind, although there have been a number of works related to safety, the application of machine learning in the context of roundabouts is limited. Further limited is the assessment of variable selection between these two approaches and how it can impact predictions.

Therefore, the objectives of this work are: (1) fill the gap in literature as it pertains to assessing variable selection in prediction accuracy by selecting variables in a traditional econometric model and through a random forest, (2) develop a machine learning model (using variables selected from both methods) to predict injury severity at roundabouts, (3) develop an econometric model (one to select variables and one using the variables identified in the random forest) to predict injury severity at roundabouts, and (4) compare prediction results. In short, this work will use a binary probit model and a random forest to select variables, then use these variables to predict injury severity outcomes using said binary probit model and SVM. For the SVM model, linear, nonlinear, polynomial, and sigmoid kernels are considered. The comparison will focus on predictions of the two injury outcomes (injury and no injury) for three- and four-leg roundabouts in Oregon. To the best of the authors' knowledge, this study is the first attempt at assessing variable selection within these two methods and applying it to roundabout safety.

3.2 Empirical Setting

This research is based on crash data collected and compiled by Oregon's Department of Transportation (ODOT) Crash Analysis and Reporting Unit. The data includes information on crashes over a five-year period (2011 to 2015) in which 1,006 crashes occurred at roundabouts (shown in Figure 3.1).

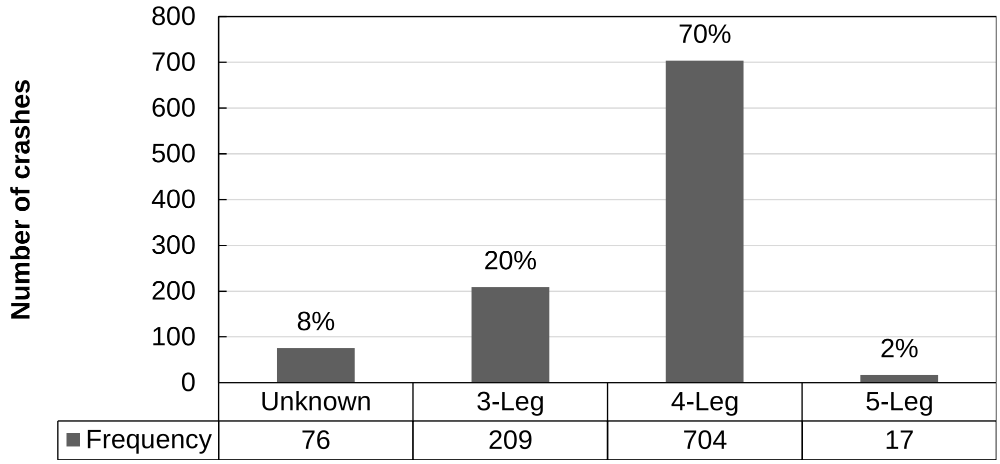
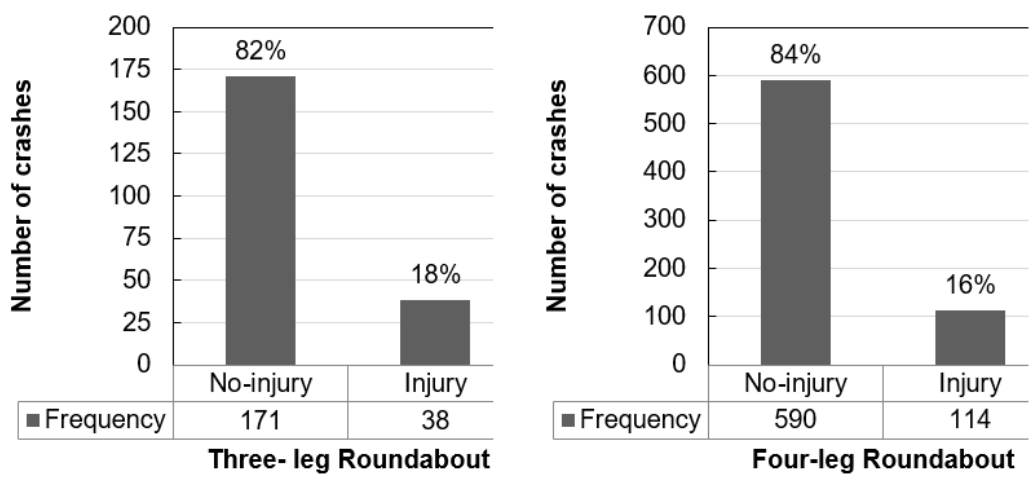


Figure 3.1: Number of Crashes at Different Roundabout Configurations.

As illustrated in Figure 3.1, roundabouts with four legs have the highest number of reported crashes at 704 (70%), followed by roundabouts with three legs at 209 (20%). Of the remaining configures, 76 crashes (8%) occurred at roundabouts in which the configuration was unknown or not specified, and 17 crashes (2%) occurred at five-leg roundabouts. Based on these statistics, this study considers only three- and four-leg roundabouts. The distributions of injury severity outcomes for both types of roundabout configurations are shown in Figure 3.2.



(a) (b)

Figure 3.2: Injury Severity Outcomes at (a) Three-leg Roundabouts and (b) Four-leg roundabouts ("Injury" includes all injury types, including fatalities).

3.3 Research Methodology

This section will outline each methodological approach used in the current study and Figure 3.3 illustrates the proposed methodology for both variable selection and prediction.

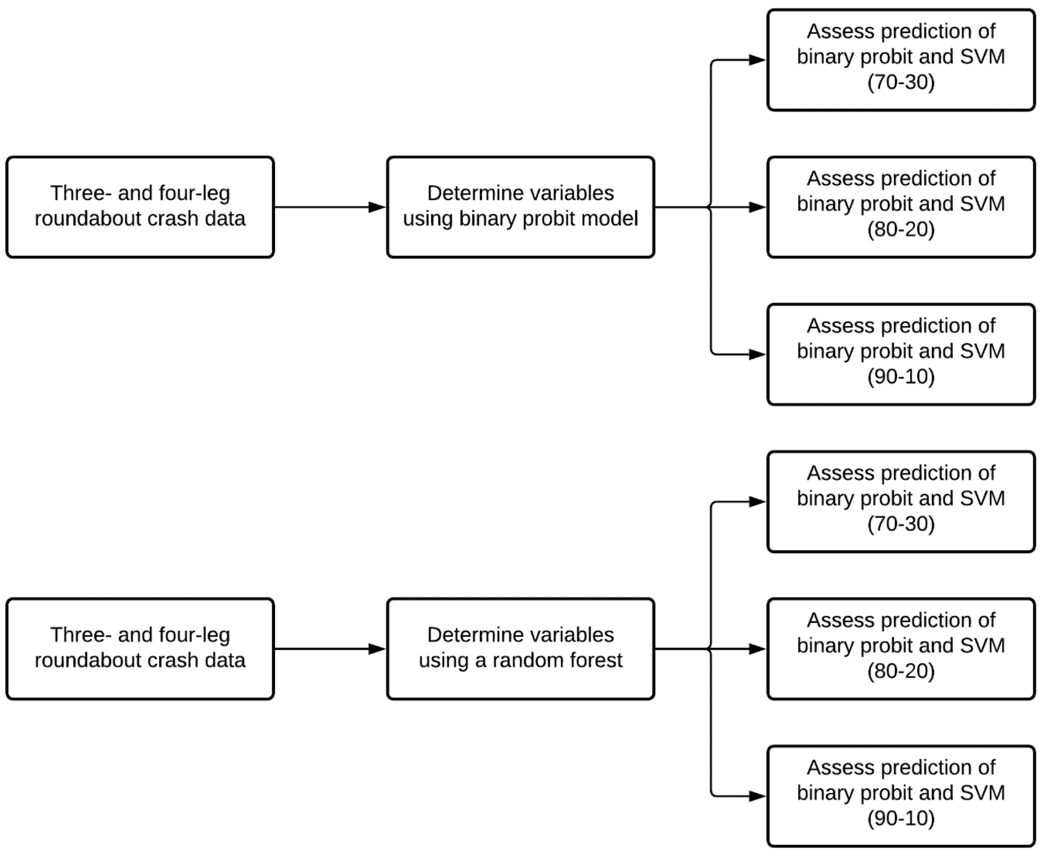


Figure 3.3: Methodological Process of Variable Selection and Model Prediction (numbers in parentheses refer to the split in training and test data, e.g., 80% training and 20% test).

3.3.1 Variable Selection

The first step consists of selecting variables to be used in the two prediction methods: binary probit and SVM. The two variable selection methods considered are a random forest and an econometric model (binary probit). As it pertains to the random forest, this was applied to identify variables that were determined to be important predictors for injury severity outcomes at roundabouts. Previous work has shown that random forests are helpful in this regard (Ahmed and Abdel-Aty 2012; Siddiqui et al. 2012; Yu and Abdel-Aty 2014).

The random forest method is a meta estimator that fits several decision tree classifiers on various sub-samples of the data and uses averaging to improve the predictive accuracy and control for over-fitting, unlike classification and regression tree models (Yu and Abdel-Aty 2014). The random forest classifier creates a set of decision trees (aggregating trees) from a randomly selected subset of the training data. It then aggregates the votes from different decision trees to decide the final class of the test object and helps with feature selection based on importance. Random forests, in general, deals with two free parameters: the number of trees (*ntree*) and the variables randomly sampled as candidates at each split (*mtry*). Random forests then work in three general steps, as shown in Figure 3.4.

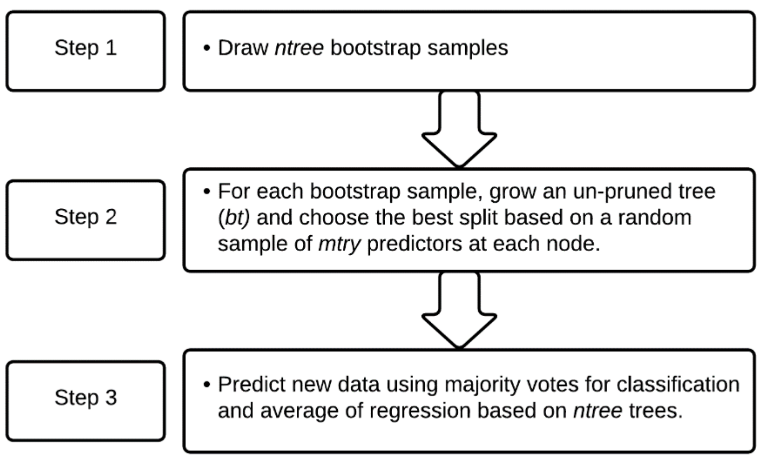


Figure 3.4: Random Forest Classifier

The second variable selection method consists of fitting a binary probit model, where unobserved heterogeneity was addressed through the estimation of random parameters. Final model specifications were obtained through a forward stepwise procedure, and variables in the final model were used for predicting injury severity outcomes (i.e., the selected variables using an econometric model). These same variables were then used in the SVM model. More detail on this model is given when prediction methods are discussed.

3.3.2 Prediction methods

As stated previously, after selecting variables using a random forest and an econometric model, an SVM model and a binary probit model were used to predict injury severity outcomes. The nature of the data, specifically the outcome to be predicted, drove the selection of both prediction methods. Referring to Figure 3.3 the outcome to be predicted is binary: "Injury" if the injury sustained was fatal, incapacitating, non-incapacitating, or possible, and "No Injury" if there was no injury sustained. Although this particular aggregation is unusual, due to the low number of crashes at roundabouts, it was necessary to arrive at an adequate number of observations for both outcomes. In doing so, variability caused by random effects is reduced (Chang and Mannering 1999). Therefore, considering this type of outcome, the econometric model chosen for the current study is the binary probit model, while the machine learning approach selected is an SVM model that utilizes different kernel functions.

3.3.2.1 Binary Probit Model

The specific model used for predicting was the binary probit model; however, random parameters were estimated in the model used for variable selection. This was done to address potential concerns related to unobserved heterogeneity. Of note, random parameters were not estimated when variables selected through a random forest were estimated in the binary probit model. In a random parameters model, some or all of the parameters are assumed to be random and will vary across observations. In this study, all random parameters were assumed to be normally distributed with a constant mean and variance. Since a normal distribution is symmetric and continuous, a coefficient for the same factor may be positive for some observations and negative for other observations regardless if the mean effect is positive (or negative). Also, if the variance or scale parameter is zero, then the parameter is not random, and the factor will have the same effect across all observations (Tay 2015b). For random parameters binary probit estimation model, see Zubaidi et al. (2020)

3.3.2.2 Support Vector Machine (SVM)

SVM is a powerful supervised machine learning technique developed by Boser et al. (1992) that can be utilized for linear and nonlinear classification and regression problems. This technique solves classification problems based on statistical learning theory, and it is best understood as approximating a target function f that maps input variables X to an output variable Y as follows:

$$Y = f(X) \tag{3.1}$$

This characterization describes the range of classification and prediction problems and the machine learning algorithms that can be used to address them.

A crucial component of SVM is choosing the right kernel function to succeed in the classification task and have the best SVM performance with a given dataset. The kernel is a way of computing the dot product of two vectors $\boldsymbol{X}$ and $\boldsymbol{Y}$ in some (possibly very high-dimensional) feature space, which is why a kernel function is sometimes called a "generalized dot product." An important consideration in learning the target function from the training data is how well the model generalizes to new data. Generalization is essential because the collected data is only a sample; it is incomplete and noisy. In general, SVM has four types of kernel functions: linear function, radial basis function (RBF), polynomial function, and sigmoid function (as shown in Table 3.1).

Table 3.1: SVM Kernel Functions

Type of Classifier	Kernel Function*
Linear kernel	$K = (x_i , x_j) = x_i^T x_j$
Radial basis kernel (RBF)	$K(x_i , x_j) = \exp(-\gamma \vert\vert x_i - x_j \vert\vert^2)$
Polynomial kernel	$K(x_i , x_j) = \gamma x_i^T x_j + r)^d$
Sigmoid kernel	$K(x_i , x_j) = \tanh(\gamma x_i^T x_j + r)$

* γ, r, and d are kernel parameters.

In this study, the linear, radial, polynomial, and sigmoid kernels were applied to take into consideration the linearity and nonlinearity of the data with different data splitting.

3.3.3 Evaluation metrics

Methodological advancement (including recent applications of advanced analytics and ensemble models) has substantially improved the understanding of the factors that affect crash frequencies and crash severities. It is perhaps the combination of evolving methodologies and assessment techniques that holds the greatest promise in advancing the analytical studies in this application domain (Lord and Mannering 2010). To assess these applications, metric evaluation should be considered, and to calculate these metrics, the confusion matrix (classification accuracy) is used:

$$\text{Classification Accuracy} = \frac{TP+TN}{TP+TN+FN+FP} \tag{3.2}$$

where classification accuracy is the ratio of the number of correct predictions to the total number of input samples. TP, TN, FP, and FN represent the number of classification cases that fall under true positive, true negative, false positive, and false negative counts, respectively. The overall accuracy, shown by Eq. (2), estimates the proportion of correctly classified test examples (i.e., the sum of all correctly classified samples divided by all of the samples) and therefore provides the overall ratio of correct classifications. Therefore, to provide a comparison of the analytical models in this study, various performance measures are assessed (Chinmoy et al. 2016):

$$\text{Sensitivity} = \frac{TP}{TP+FN} \tag{3.3}$$

$$\text{Specificity} = \frac{TN}{TN+FP} \tag{3.4}$$

$$\text{Precision} = \frac{TP}{TP+FP} \tag{3.5}$$

where sensitivity is the true positive rate (TPR) that is defined as $\frac{TP}{FN+TP}$. The TPR corresponds to the proportion of positive data points that are correctly considered as positive concerning all positive data points. Lastly, the false positive rate (FPR), or specificity, is defined as $\frac{FP}{FP+TN}$. The FPR corresponds to the proportion of negative data points that are mistakenly considered as positive for all negative data points.

3.4 Modeling Results and Discussion

A total of five prediction models (binary probit and SVM with four different kernel functions) were estimated using different training and test data ratios: (70-30), (80-20), and (90-10). The test dataset was used for the investigation of the methods' prediction accuracies, which were then compared. The details of the prediction processes are presented in this section. The coding and execution of the econometric and machine learning calculations were accomplished using NLOGIT6 and the R computer programming language development environment RStudio (The R Foundation for Statistical Computing).

3.4.1 Selected Variables Using the Binary Probit Model

Considering outcomes of injury and no injury, a binary probit model was estimated for 209 crashes at three-leg roundabouts and for 704 crashes at four-leg roundabouts from 2011-2015. Once more, this process was the first variable selection method, in which all significant variables in final model specifications were selected for use in the SVM predictions. Final model specifications, and therefore the first set of selected variables, are shown in Table 3.2 (descriptive statistics are shown in Table 3.3). As observed, several variables were included in the models, where six variables were found to be significant for the three-leg roundabout model and ten variables were found to be significant in the four-leg roundabout model. Additionally, one advantage of such a model is the ability to readily interpret the effects of significant variables on the likelihood of some outcome; hence, that discussion is provided here (see marginal effects in Table 3.2).

Results indicate that crashes due to careless driving and using a seat belt increased the likelihood of an injury for three-leg roundabouts. Careless driving significantly increases the risk of a crash, which can be lead to more severe injuries; this outcome was also found in Bener et al. (2017). Concerning the seat belt variable, a possible explanation is using a seat belt reduces the risk of death and severe injury, but at the same time, there is a chance of getting injured in the shoulder or the neck, especially with high-speed driving. The estimated parameters for the weather, movement of the vehicle, and gender were all found to be random and significant. Cloudy weather has a normally distributed parameter with a mean of 1.85 and a standard deviation of 1.97. This suggests that roughly 17% of those who drive in cloudy weather are less likely to sustain an injury, whereas 83% are more likely.

The indicators for safety equipment use, gender, drivers age 36-50 years, and crash cause in the four-leg roundabout model were found to be statistically significant. Additionally, the parameters for these indicators were random and normally distributed. The estimated parameter for a crash caused when the driver did not yield the right-of-way was found to be random and normally distributed with a mean of 2.74 and a standard deviation of 1.76. In other words, this suggests that for 6% of those drivers, the likelihood of sustaining an injury decreased, while for 94%, the opposite was true.

One issue when fitting a model is how well the generated model behaves when applied to new data. Generally, in explanatory econometric models, the whole of the dataset is used to arrive and model specifications, after which a full discussion on significant variables is given. However, for the current study, the primary focus is prediction. Therefore, to address this issue for predicting, the data was partitioned in two portions three times, each with a different ratio. The first portion is a training partition used to create the model, and the second portion is the test partition to evaluate the prediction performance for both the three and four-leg roundabout models. Using the significant variables shown in Table 3.2, four prediction models were assessed (binary probit, SVM-linear, SVM-radial, SVM- polynomial, and SVM- sigmoid) using three distinct partitioned datasets: (70-30), (80-20), and (90-10).

Table 3.2: Random Parameter Binary Probit Model Specifications

Variable	Three-leg Roundabout			Four-leg Roundabout		
	Coefficient	t-stat	Marginal Effect	Coefficient	t-stat	Marginal Effect
Constant	-0.23	-0.40	-	1.03	1.89	
Crash Level Cause (1 if the crash happened because of careless driving, 0 otherwise)	3.55	1.76	0.06	-	-	-
Weather (1 if cloudy, 0 otherwise)	1.15	1.85	0.02	-	-	-
(standard deviation, normally distributed)	(1.6)	(1.97)	-	-	-	-
Type of Vehicle (1 if passenger car, 0 otherwise)	-2.32	-3.30	-0.04	-3.67	-5.75	-0.019
Participant Safety Equipment Use (1 if seatbelt was used, 0 otherwise)	0.65	4.23	0.01	1.73	4.85)	0.009
(standard deviation, normally distributed)				(1.44)	6.95	
Movement of the Vehicle at the Time of the Crash (1 if stopped in traffic, 0 otherwise)	0.17	0.22	0.003	1.41	4.29	0.008
(standard deviation, normally distributed)	(3.93)	(2.96)	-	-	-	-
Gender (1 if male, 0 female)	-3.002	-3.33	-0.05	-1.48	-4.62	0.008
(standard deviation, normally distributed)	(2.79)	(3.65)	-	(1.6)	(5.87)	-
Age of Driver (1 if the driver age 22-35, 0 otherwise)	-	-	-	1.89	5.53	0.01
Participant Error (1 if following too close, 0 otherwise)	-	-	-	-2.87	-4.05	-0.02
Pavement Condition (1 if poor pavement, 0 otherwise)	-	-	-	-0.79	-2.92	-0.004
Location of the Crash (1 if at the right-hand side, 0 otherwise)	-	-	-	-0.57	-1.82	-0.003
Age of Driver (1 if driver age 36-50, 0 otherwise)	-	-	-	0.04	0.12	0.0002
(standard deviation, normally distributed)	-	-	-	(2.26)	(5.55)	-
Crash Cause (1 if the driver did not yield right-of-way, 0 otherwise)	-	-	-	0.86	2.74	0.005
(standard deviation, normally distributed)	-	-	-	(0.41)	(1.76)	-
Log Likelihood function		-56.18			-156.58	
Log Likelihood at zero		-72.78			-218.61	
AIC		132			343	
McFadden Pseudo R-squared		0.04			0.01	
No. of Observations		146			493	

Table 3.3: Descriptive Statistics of Significant Variables

Variable	Three-leg Roundabout		Four-leg Roundabout	
	Mean	Standard Deviation	Mean	Standard Deviation
Crash Level Cause (1 if the crash happened because of careless driving, 0 otherwise)	0.04	0.19	-	-
Weather (1 if cloudy, 0 otherwise)	0.10	0.27	-	-
Participant Safety Equipment Use (1 if seatbelt was used, 0 otherwise)	1.78	1.20	0.55	0.50
Movement of the Vehicle at the Time of the Crash (1 if stopped in traffic, 0 otherwise)	0.15	0.36	0.16	0.37
Gender (1 if male, 0 female)	0.53	0.50	0.49	0.50
Age of Driver (1 if the driver age 22-35, 0 otherwise)	-	-	0.21	0.41
Participant Error (1 if following too close, 0 otherwise)	-	-	0.14	0.35
Pavement Condition (1 if poor pavement, 0 otherwise)	-	-	0.34	0.47
Location of the Crash (1 if at the right-hand side, 0 otherwise)	-	-	0.22	0.41
Age of Driver (1 if driver age 36-50, 0 otherwise)	-	-	0.21	0.41
Crash Cause (1 if the driver did not yield right-of-way, 0 otherwise)	-	-	0.33	0.47

3.4.2 Selected Variables Using Random Forest

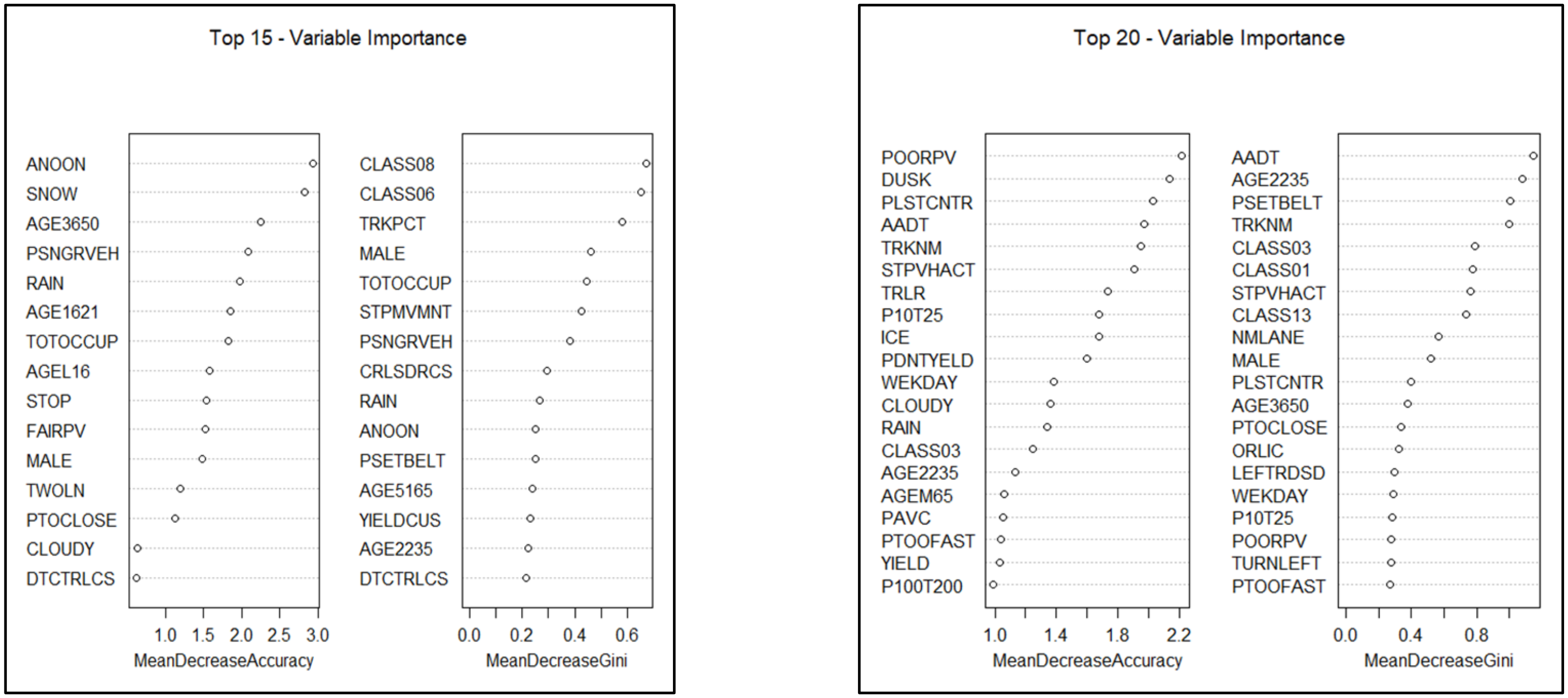
Variable importance was determined through a random forest algorithm, which served as the second variable selection method. Important variables in this framework are identified by monitoring how much the prediction error increases when the out-of-bag (OOB) data for that variable is permuted while all others were left unchanged (Yu and Abdel-Aty 2014). The R package ‘‘random forest’’ was used to generate the random forest and determine variable importance. In the three-leg model, $m = 1$ was specified so that one variable was randomly samples as a candidate for each split, and a total of 400 trees were constructed. The algorithm computes two measures of variable importance: the mean decrease in the Gini coefficient and the mean decrease in accuracy. For each bootstrap iteration and related tree, the prediction error using data, not in the bootstrap sample, called the out of bag (OOB) data, is estimated, and the accuracy is the measurement value for the classification data. In other words, the mean decrease in accuracy is usually described as the decrease in model accuracy from permuting the values in each feature. The mean decrease in the Gini coefficient is the average (mean) of a variable’s total decrease in node impurity, weighted by the proportion of samples reaching that node in each decision tree in the random forest. A higher mean decrease in the Gini coefficient indicates a higher in variable importance.

For the four-leg model, $m = 1$ was also used, and a total of 500 trees were constructed. Table 3.4 provides variable descriptions for the essential variables that were determined. Figure 3.5 shows the final results of the variable importance rankings, where the mean decrease in accuracy was the selection criteria for both the three- and four-leg roundabout models. These variables are the second set of variables used to assess prediction. Variables that are equal to or more than 0.5 mean accuracy have been included in the study. It can be drawn from the figure that the indicators “Time of Day: 12 p.m. - 6 p.m.” (ANOON in the figure), “Weather Condition: Snow” (SNOW), “Driver Age: 36-50” (AGE3650), “Vehicle Type: Passenger Car” (PSNGEVEH), and “Weather Condition: Rain” (RAIN) were identified as the most important variables in predicting injury severity at three-leg roundabouts. Figure 3.5 shows that “Pavement Condition: Poor” (POORV), “Light Condition: Dusk” (DUSK), “Participant (Driver)

Error: Speeding Too Fast for Conditions: Loss of Control" (PLSTCNTR), "Annual Average Daily Traffic" (AADT), "Number of Trucks" (TRKNM), and "Movement of the Vehicle at the Time of the Crash: Stopped" (STPVHACT) were identified as the most important variables in predicting injury severity at four-leg roundabouts.

Table 3.4: Variable Descriptions for Random Forest

Variable	Description
ANOON	Time of Day:12pm-6pm
WEKDAY	Weekday
SNOW	Weather Condition: Snow
RAIN	Weather Condition: Rain
CLOUDY	Weather Condition: Cloudy
ICE	Weather Condition: Ice
DUSK	Light Condition: Dusk
MALE	Gender of the Driver
AGEL16	Driver Age: 16
AGE1621	Driver Age: 16-21
AGE2235	Driver Age: 22-35
AGE3650	Driver Age: 36-50
AGEM65	Driver Age: More than 60
PSNGERVEH	Vehicle Type: Passenger Car
TOTOCCUP	Total Occupants in the Vehicle
PTOCLOSE	Participant (Driver) Error: Following Too Close
PLSTCNTR	Participant (Driver) Error: Speed Too Fast for Conditions: Loss of Control
PDNTYELD	Participant (Driver) Error: Did Not Yield Right-of-way
PTOOFAST	Participant (Driver) Error: Speed Too Fast for Conditions
DICTRLCS	Participant (Driver) Error: Disregarded Other Traffic Control Device
STPVHACT	Movement of the Vehicle at the Time of the Crash: Stopped
STOP	Traffic Control Device: Stop Sign
YIELD	Traffic Control Device: Yield Sign
FAIRPV	Pavement Condition: Fair
POORPV	Pavement Condition: Poor
PAVC	Type of Pavement: Concrete
P10T25	Population Range: 10,000 to 25,000
P100T200	Population Range: 100,000 to 200,000
TWOLAN	Number of Lanes
AADT	Annual Average Daily Traffic
TRKNM	Number of Trucks
CLASS03	Vehicle Classes: Four-tire, Single Unit
TRLR	Presence of Trailer



(a) (b)

Figure 3.5: Random Forest Variable Importance for (a) Three-leg Roundabouts and (b) Four-leg Roundabouts

3.4.3 Comparison of prediction performance

The primary objective of this study is to assess the prediction performance of the binary probit model and SVM model under different variable selection methods. For comparisons, the binary probit model and the SVM models were fitted using the same randomly separated training and testing datasets. The training was used to fit the model, while the testing dataset was used to evaluate the prediction performance of the model. To examine the accuracy of the prediction models, a validation dataset can be used, and the prediction results for each method can be summarized in a confusion matrix. Accuracy is a comprehensive indicator to reflect the number of observations that were predicted correctly. Comparing the situations of a no-injury sample being misclassified as an injury or an injury sample being misclassified as a no-injury, the latter will lead to more severe consequences (i.e., ignoring the impact of serious accidents). The parameter accuracy, which is the percent of correct predictions, was used for comparing the prediction performance of the models. It can be calculated using Eq. (2).

The SVM algorithm was applied to predict injury severity for the three-leg and four-leg roundabout crash data. In this study, four types of kernel functions were used to investigate the linear and nonlinear relationships between injury severity and the selected variables of both methods. SVM with a linear, radial, polynomial, and sigmoid kernels function were formalized in the RStudio environment.

Severity Prediction with Selected Variables from a Random Forest: For three-leg roundabouts, as shown in Table 3.5, the SVM model with a polynomial kernel function had the highest prediction accuracy at 86% (this was under the 70-30 split in the data). The other kernel functions also performed well, with the sigmoid kernel leading to an 84% prediction accuracy and the radial and linear kernels leading to 82% and 81% prediction accuracies, respectively. The binary probit model, using selected variables from a random forest, had the lowest accuracy at 76%.

For the 80-20 training-test split, the SVM model with polynomial and sigmoid kernels have the same accuracy at 83%. Accuracy also decreased for the radial kernel to 79%, while accuracy remained consistent with the linear kernel (81%). No predictions were made using the binary probit model in this scenario with the selected variables from the random forest.

The highest accuracy occurred in the 90-10 split for training and test data and was obtained using the polynomial kernel, specifically, with a prediction accuracy of 91%. This split also led to the highest prediction accuracy for the sigmoid kernel (86%), while it had varying effects on the linear and radial kernels. For example, in the 90-10 split, prediction accuracy was 81% (higher than the 80-20 split, but lower than the 70-30 split). Similarly, the linear kernel had the lowest accuracy with this split at 76%, which was 4% lower than the other splitting ratios (81% accuracy for each).

Overall prediction accuracies for four-leg roundabouts are shown in Table 3.6, Under the 70-30 split, prediction accuracies remained fairly consistent regardless of the kernel. The linear, polynomial, and sigmoid kernels all led to an 84% prediction accuracy, while the radial kernel led to an 83% accuracy. Each of these, however, outperformed the binary probit model, for which a prediction accuracy of 75% was obtained.

For the 80-20 split, similar results are observed, with all prediction accuracies but the radial kernel being marginally higher or the same. Specifically, using the polynomial and sigmoid kernels, an accuracy of 85% was obtained, while the binary probit accuracy increased to 77%. Accuracy using the radial kernel decreased to 82%, while prediction accuracy remained the same for the linear kernel at 84%.

Under the 90-10 split, all prediction accuracies but the radial kernel were the same as the 70-30 split, where the accuracy of the radial kernel decreased to 80%. For all other kernels (linear, polynomial, and sigmoid), there was accuracy of 84%.

Severity Prediction with Selected Variables from the Binary Probit Model: Table 3.7 and Table 3.8 show the prediction accuracies of all models using the variables selected through the development of the binary probit model. Referring to Table 3.7, all SVM models had higher prediction accuracies for three-leg roundabouts. Under the 70-30 split, the sigmoid kernel resulted in the highest accuracy at 86%. Accuracies using the radial and polynomial kernels were the same (82%), and prediction accuracy using the linear kernel was marginally lower at 81%. In regard to the binary probit model, this was the highest observed accuracy at 78%, which is likely linked to the selected variables being chosen based on the binary probit framework.

Under the 80-20 split, prediction accuracies of injury severity on three-leg roundabouts notably changed contingent on the SVM kernel function. For two kernels, accuracies decreased, while accuracies increased for the other two. Specifically, prediction accuracy using the linear kernel decreased to 79% (from 81% in the 70-30 split), and prediction accuracy using the sigmoid kernel decreased to 83% (from 86% in the 70-30 split). On the other hand, prediction accuracies using the radial kernel increased to 86% (from 82% in the 70-30 split), and prediction accuracies using the polynomial kernel increased to 83% (from 82% in the 70-30 split). For the 90-10 split, all accuracies were the same regardless of the kernel.

In regard to four-leg roundabouts (Table 3.8) under the 70-30 split, the highest accuracy was again obtained with the sigmoid kernel function at 87%. Accuracies for the other three kernels (linear, radial, and polynomial) were identical at 84%, while the accuracy of the binary probit model was notable lower at 75%.

For the 80-20 split on four-leg roundabouts, the highest prediction accuracy was also observed using the sigmoid kernel (86%). Accuracies for the other three kernels were nearly identical to the 70-30 split, where the accuracy of the linear and polynomial kernels remained at 84%, and the accuracy of the radial kernel decreased marginally to 83% (from 84% in the 70-30 split).

Lastly, under the 90-10 split on four-leg roundabouts, the radial, and polynomial kernel functions resulted in the highest prediction accuracy at 87%, while the sigmoid kernel resulted in a comparable accuracy at 86%. As for the linear kernel, as was the case with all other splits, prediction accuracy remained at 84%.

A visual comparison of prediction accuracies is shown in Figure 3.6, and a tabulated comparison is shown in Table 3.9. Overall, the SVM models outperformed the binary probit model regardless of training-test ratios, kernel functions, and variable selection methods. In addition, the prediction rate remained fairly consistent regardless of split ratios, roundabout configuration, and variable selection.

The highest prediction accuracy of the binary probit model (78%) corresponds to variables selected by said model, a 70-30 split, and for three-leg roundabouts. This scenario also resulted in the most comparable predictions across all models, with the exception of SVM-sigmoid.

Additional noteworthy factors include the highest rate of prediction, which was observed for three-leg roundabouts, a 90-10 split, and predicted by the SVM-polynomial model. This prediction accuracy was 91% (no other accuracy was greater than 87%). All SVM models using selected variables from the binary probit model had the same prediction rates in a 90-10 split for three-leg roundabouts (81%). This was the only scenario in which this was observed. Ultimately, regardless of variable selection methods, kernel functions, and split ratios, the SVM-based models resulted in higher prediction accuracy in this context. In regard to SVM alone, while considering variable selection methods, results varied (in some cases, prediction rates were better with variables selected via a random forest, and in some cases, the opposite).

Table 3.5: Prediction models Results for Three-leg Roundabouts Using Selected Variables from a Random Forest.

Model Type	Three-leg Roundabout Model									
		Confusion Matrix (70-30)		Model Accuracy (%)	Confusion Matrix (80-20)		Model Accuracy (%)	Confusion Matrix (90-10)		Model Accuracy (%)
		No-injury	Injury		No-injury	Injury		No-injury	Injury	
Binary Probit Model	No-injury	40	7	76	-	-	-	-	-	-
	Injury	6	2		-	-		-	-	
Support Vector Machine (SVM-linear)	No-injury	48	3	81	34	0	81	16	1	76
	Injury	9	2		8	0		4	0	
Support Vector Machine (SVM- radial)	No-injury	51	0	82	33	1	79	17	0	81
	Injury	11	0		8	0		4	0	
Support Vector Machine (SVM- polynomial)	No-injury	50	1	86	33	1	83	17	1	91
	Injury	8	3		6	2		3	0	
Support Vector Machine (SVM- sigmoid)	No-injury	51	0	84	34	6	83	17	0	86
	Injury	10	1		7	1		3	1	

- The sign (-) means that the model couldn’t predict the outcomes for this ratio

Table 3.6: Prediction Results for Four-leg Roundabouts Using Selected Variables from a Random Forest.

Model Type	Four-leg Roundabout Model									
		Confusion Matrix (70-30)		Model Accuracy (%)	Confusion Matrix (80-20)		Model Accuracy (%)	Confusion Matrix (90-10)		Model Accuracy (%)
		No-injury	Injury		No-injury	Injury		No-injury	Injury	
Binary Probit Model	No-injury	148	26	75	103	15	77	-	-	-
	Injury	25	8		17	5		-	-	
Support Vector Machine (SVM-linear)	No-injury	177	0	84	118	0	84	59	0	84
	Injury	34	0		23	0		11	0	
Support Vector Machine (SVM-radial)	No-injury	175	2	83	115	3	82	56	3	80
	Injury	34	0		23	0		11	0	
Support Vector Machine (SVM-polynomial)	No-injury	177	0	84	118	0	85	59	0	84
	Injury	33	1		22	1		11	0	
Support Vector Machine (SVM-sigmoid)	No-injury	177	0	84	117	1	85	59	0	84
	Injury	33	1		20	3		11	0	

Table 3.7: Prediction Results for Three-leg Roundabouts Using Selected Variables from Binary Probit Model

Model Type		Three-leg Roundabout Model								
		Confusion Matrix (70-30)		Model Accuracy (%)	Confusion Matrix (80-20)		Model Accuracy (%)	Confusion Matrix (90-10)		Model Accuracy (%)
		No-injury	Injury		No-injury	Injury		No-injury	Injury	
Binary Probit Model	No-injury	43	8	78	-	-	-	-	-	-
	Injury	5	4		-	-		-	-	
Support Vector Machine (SVM-linear)	No-injury	50	1	81	33	1	79	17	0	81
	Injury	11	0		8	0		4	0	
Support Vector Machine (SVM- radial)	No-injury	49	2	82	34	0	86	17	0	81
	Injury	9	2		6	2		4	0	
Support Vector Machine (SVM-polynomial)	No-injury	51	0	82	33	1	83	17	0	81
	Injury	11	0		6	2		4	0	
Support Vector Machine (SVM- sigmoid)	No-injury	51	0	86	33	1	83	17	0	81
	Injury	9	2		6	2		4	0	

Table 3.8: Prediction Results for Three-leg Roundabouts Using Selected Variables from Binary Probit Model.

Model Type	Four-leg Roundabout Model									
	Confusion Matrices (70-30)			Model Accuracy (%)	Confusion Matrices (80-20)		Model Accuracy (%)	Confusion Matrices (90-10)		Model Accuracy (%)
	#	No-injury	Injury		No-injury	Injury		No-injury	Injury	
Binary Probit Model	No-injury	145	26	75	97	21	72	-	-	-
	Injury	25	12		17	3		-	-	
Support Vector Machine (SVM-linear)	No-injury	177	0	84	118	0	84	59	0	84
	Injury	34	0		23	0		11	0	
Support Vector Machine (SVM- radial)	No-injury	172	5	84	113	5	83	58	1	87
	Injury	28	6		19	4		8	3	
Support Vector Machine (SVM-polynomial)	No-injury	170	7	84	118	0	84	57	2	87
	Injury	26	8		23	0		7	4	
Support Vector Machine (SVM-sigmoid)	No-injury	176	1	87	115	3	86	56	3	86
	Injury	27	7		17	6		7	4	

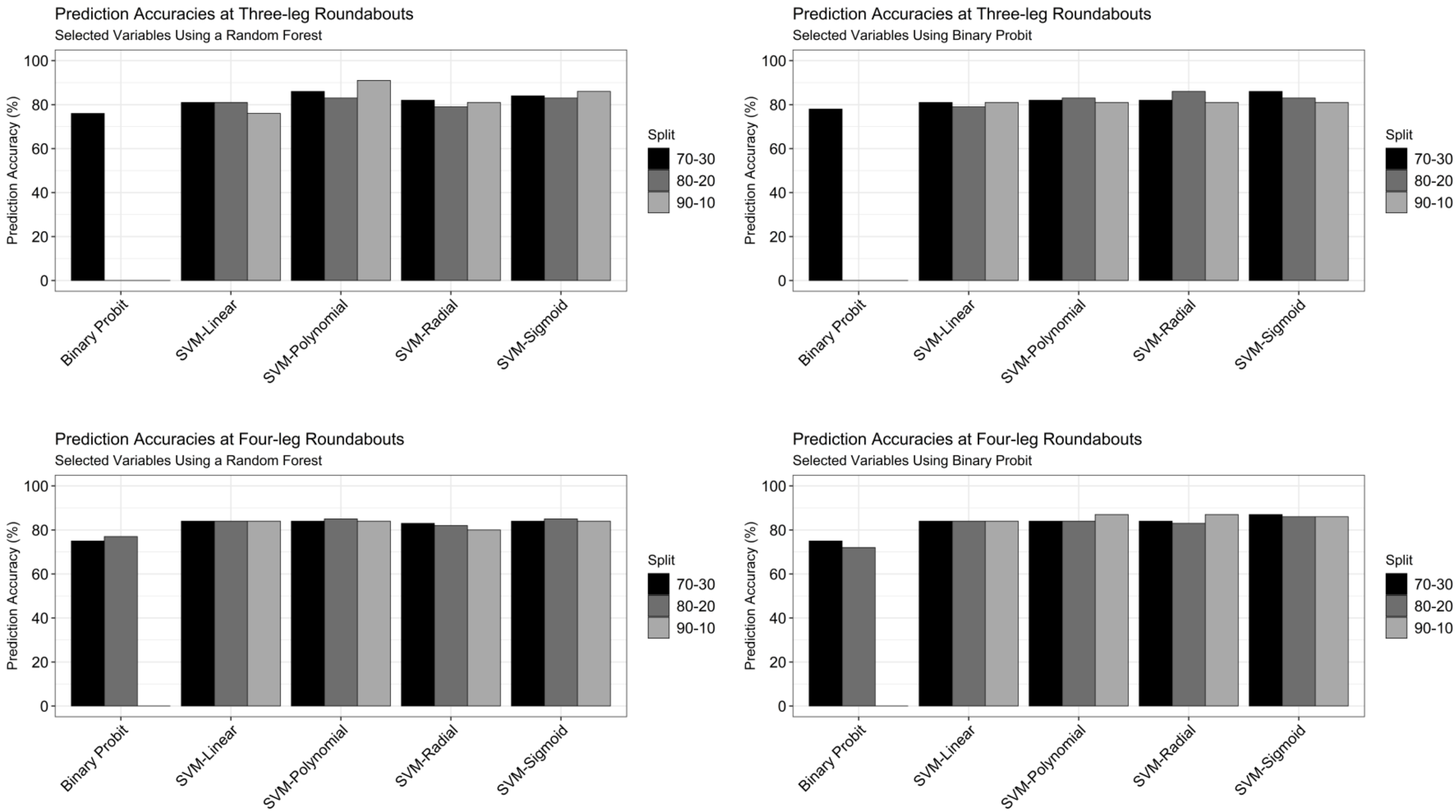


Figure 3.6: Visual Comparison of Prediction Accuracies by Model and Variable Selection Method

Table 3.9: Comparison of Prediction Accuracy Across Models and Variable Selection

Model	Three-leg Roundabout Accuracy (%)					
	Random Forest Variables			Binary Probit Variables		
	70-30	80-20	90-10	70-30	80-20	90-10
Binary Probit	76	-	-	78	-	-
SVM-Linear	81	81	76	81	79	81
SVM-Radial	82	79	81	82	86	81
SVM-Polynomial	86	83	91	82	83	81
SVM-Sigmoid	84	83	86	86	83	81
Model	Four-leg Roundabout Accuracy (%)					
	Random Forest Variables			Binary Probit Variables		
Binary Probit	75	77	-	75	72	-
SVM-Linear	84	84	84	84	84	84
SVM-Radial	83	82	80	84	83	87
SVM-Polynomial	84	85	84	84	84	87
SVM-Sigmoid	84	85	84	87	86	86

3.5 Conclusion

This study compared the predictive performance of injury severity between various machine-learning and econometric techniques based on three-leg and four-leg roundabout crash data from 2011 to 2015 in Oregon. A key component of this analysis was to assess the impact of variable selection on injury severity prediction. Variable selection was conducted using a random forest and fitting a binary probit model, after which a series of SVM-based models and a binary probit model were used to predict injury severity outcomes. In addition to assessing variable selection on prediction accuracy, three different training-test data ratios were considered. Results showed that regardless of the variable selection method and training-test ratios, the SVM models consistently outperformed the traditional econometric approach.

The binary model performed best when predicting injury severity at three-leg roundabouts and under a 70-30 split in training-test ratio. Specifically, 76% accuracy using variables selected by a random forest and 78% accuracy using variables selected by the binary probit model. Prediction rates for the binary probit model were lower when considering four-leg roundabouts, but a 77% accuracy was observed when considered an 80-20 split and variables selected by a random forest.

The SVM-linear model has comparable predictions for both three-leg and four-leg models, under both variable selection methods, with 81% and 84%, respectively. SVM-radial had a higher prediction, specifically for the four-leg model and using variables selected by the binary probit model (the highest prediction accuracy was 87% with a 90-10 split). SVM-polynomial performed best in the three-leg model using variables selected by a random forest (91% accuracy under a 90-10 split, also the highest observed accuracy) and in the four-leg model using variables selected by the binary probit model (87% accuracy under a 90-10 split). Lastly, SVM-sigmoid was the most consistent of the SVM models across all training-test ratios and variable selection. Specifically, SVM-sigmoid performed best for four-leg models using variables selected by the binary probit model and had the lowest prediction rate in the three-leg model using the same variables.

In regards to variable importance, a random forest analysis indicated that afternoon, snowy weather, and drivers aged 36-50 years were the most important injury severity predictors for three-leg roundabouts. For four-leg roundabouts, poor pavement condition, dusk lighting, and losing control of the vehicle were identified as the most important predictors for injury severity at four-leg roundabouts. For the econometric model, based on marginal effects, careless driving, passenger cars, and male drivers have the largest effect on injury severity outcomes at three-leg roundabouts. For four-leg roundabouts, passenger cars, following too closely, and drivers aged 22-35 years have the largest effect on injury severity outcomes, according to marginal effects.

In summary, when accurately predicting outcomes is a primary goal, machine learning (SVM in the current study) is advantageous over traditional econometric methods. Such methods can be used to help confront issues of multiple and correlated predictors and non-linear relationships. However, when using machine learning methods, extra care is needed in the form of model validation. Although this study thoroughly investigated injury severity at roundabouts with three- and four-legs, there were some limitations. For example, due to the data limitations, injury severity was categorized into two levels, injury and no-injury. Also, there was information missing for several factors that could have been important, such as the geometric design of the roundabout, the exact location of the crashes, the presence of a work zone, and route numbers. With more crash data in the future, outlook studies could classify the outcomes into more levels and may focus on identifying new significant factors that may lead to more detailed classifications of injury severity. Until then, the application of different machine-learning techniques can handle the small ratio of specific outcomes with the existing data.

Acknowledgment

The authors want to acknowledge the Oregon Department of Transportation, specifically the Transportation Data Department, for providing the crash dataset and for helping to understand the data. The findings of this study do not necessarily reflect the views of the Oregon Department of Transportation.

References

Abdel-Aty, M., Haleem, K., 2011. Analyzing angle crashes at unsignalized intersections using machine learning techniques. Accid. Anal. Prev. 43, 461–470.

Ahmadi, A., Jahangiri, A., Berardi, V., Machiani, S.G., 2018. Crash severity analysis of rear-end crashes in California using statistical and machine learning classification methods. J. Transp. Saf. Secur.

Ahmed, M.M., Abdel-Aty, M.A., 2012. The viability of using automatic vehicle identification data for real-time crash prediction. IEEE Trans. Intell. Transp. Syst. 13, 459–468.

Al-Bdairi, N.S.S., Hernandez, S., 2017. An empirical analysis of run-off-road injury severity crashes involving large trucks. Accid. Anal. Prev. 102, 93–100.

Alkheder, S., Taamneh, M., Taamneh, S., 2017. Severity Prediction of Traffic Accident Using an Artificial Neural Network. J. Forecast. 36, 100–108.

Anderson, J., Hernandez, S., 2017a. Roadway classifications and the accident injury severities of heavy-vehicle drivers. Anal. Methods Accid. Res. 15, 17–28.

Anderson, J., Hernandez, S., 2017b. Analytic Methods in Accident Research Roadway classifications and the accident injury severities of heavy-vehicle drivers. Anal. Methods Accid. Res. 15, 17–28.

Bener, A., Yildirim, E., Özkan, T., Lajunen, T., 2017. Driver sleepiness, fatigue, careless behavior and risk of motor vehicle crash and injury: Population based case and control study. J. Traffic Transp. Eng. (English Ed. 4, 496–502.

Boser, E., Vapnik, N., Guyon, I.M., Laboratories, T.B., 1992. A training algorithm for optimal margin classifiers. ACM New York 144–152.

Chang, L.Y., Mannering, F., 1999. Analysis of injury severity and vehicle occupancy in truck- and non-truck-involved accidents. Accid. Anal. Prev. 31, 579–592.

Chen, C., Zhang, G., Qian, Z., Tarefder, R.A., Tian, Z., 2016. Investigating driver injury severity patterns in rollover crashes using support vector machine models. Accid. Anal. Prev. 90, 128–139.

Chinmoy, P., Tomosaburo, O., Vimalathithan, K., 2016. Factors Influencing Specificity and Sensitivity of Injury Severity Prediction (ISP) Algorithm for AACN. Int. J. Automot. Eng. 7, 15–22.

Das, S., Datta, S., Zubaidi, H.A., Obaid, I.A., 2021. Applying interpretable machine learning to classify tree and utility pole related crash injury types. IATSS Res.

FHWA, 2015. A Review of Fatal and Severe Injury Crashes at Roundabouts IV.

Goldstein, B.A., Navar, A.M., Carter, R.E., 2017. Moving beyond regression techniques in cardiovascular risk prediction: Applying machine learning to address analytic challenges. Eur. Heart J. 38, 1805–1814.

Iranitalab, A., Khattak, A., 2017. Comparison of four statistical and machine learning methods for crash severity prediction. Accid. Anal. Prev. 108, 27–36.

Islam, M., Hernandez, S., 2013. Large Truck–Involved Crashes: Exploratory Injury Severity Analysis. J. Transp. Eng. 139, 596–604.

Li, Y., Ma, D., Zhu, M., Zeng, Z., Wang, Y., 2018. Identification of significant factors in fatal-injury highway crashes using genetic algorithm and neural network. Accid. Anal. Prev. 111, 354–363.

Li, Z., Liu, P., Wang, W., Xu, C., 2012. Using support vector machine models for crash injury severity analysis. Accid. Anal. Prev. 45, 478–486.

Lord, D., Mannering, F., 2010. The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives. Transp. Res. Part A Policy Pract. 44, 291–305.

Mannering, F.L., Shankar, V., Bhat, C.R., 2016. Unobserved heterogeneity and the statistical analysis of highway accident data. Anal. Methods Accid. Res. 11, 1–16.

Romo, A., Hernandez, S., Cheu, R.L., 2014. Identifying Precrash Factors for Cars and Trucks on Interstate Highways: Mixed Logit Model Approach. J. Transp. Eng. 140, 04013016.

Sameen, M., Pradhan, B., 2017. Severity Prediction of Traffic Accidents with Recurrent Neural Networks. Appl. Sci. 7, 476.

Savolainen, P.T., Mannering, F.L., Lord, D., Quddus, M.A., 2011. The statistical analysis of highway crash-injury severities : A review and assessment of methodological alternatives. Accid. Anal. Prev. 43, 1666–1676.

Siddiqui, C., Abdel-Aty, M., Huang, H., 2012. Aggregate nonparametric safety analysis of traffic zones. Accid. Anal. Prev. 45, 317–325.

Storm, H., Baylis, K., Heckelei, T., 2019. Machine learning in agricultural and applied economics. Eur. Rev. Agric. Econ. 1–44.

Tang, J., Liang, J., Han, C., Li, Z., Huang, H., 2019. Crash injury severity analysis using a two-layer Stacking framework. Accid. Anal. Prev. 122, 226–238.

Tay, R., 2015. A random parameters probit model of urban and rural intersection crashes. Accid. Anal. Prev. 84, 38–40.

Tixier, A.J.P., Hallowell, M.R., Rajagopalan, B., Bowman, D., 2016. Application of machine learning to construction injury prediction. Autom. Constr. 69, 102–114.

Wahab, L., Jiang, H., 2019. A comparative study on machine learning based algorithms for prediction of motorcycle crash severity. PLoS One 14, 1–17.

Yu, R., Abdel-Aty, M., 2014. Analyzing crash injury severity for a mountainous freeway incorporating real-time traffic and weather data. Saf. Sci. 63, 50–56.

Zeng, Q., Huang, H., 2014. A stable and optimized neural network model for crash injury severity prediction. Accid. Anal. Prev. 73, 351–358.

Zhang, J., Li, Z., Pu, Z., Xu, C., 2018. Comparing prediction performance for crash injury severity among various machine learning and statistical methods. IEEE Access 6, 60079–60087.

Zubaidi, H.A., Anderson, J.C., Hernandez, S., 2020. Understanding roundabout safety through the application of advanced econometric techniques. Int. J. Transp. Sci. Technol.

The effect of driver age and gender on driver-injury severities at roundabouts: A random parameters binary probit model with heterogeneity-in-means approach

By Hamsa Abbas Zubaidi and Salvador Hernandez, Ph.D.

Journal: (TBD)
Address:
Issue:

Chapter 4: The effect of driver age and gender on driver-injury severities at roundabouts: A random parameters binary probit model with a heterogeneity-in-means approach

By:

Hamsa Abbas Zubaidi
*School of Civil and Construction Engineering
Oregon State University, Corvallis, Oregon 97331-3212.
* Roads and Transport Department, College of Engineering, University of Al-Qadisiyah, Iraq.
zubaidih@oregonstate.edu

Salvador Hernandez, PhD.
Associate Professor
School of Civil and Construction Engineering
Oregon State University, Corvallis, OR 97331-3212
(541)-737-1757
sal.hernandez@oregonstate.edu

.

Abstract

This paper investigates the risk factors that significantly affect the severity of drivers' injuries in roundabout-related accidents while systematically accounting for unobserved heterogeneity and variance in terms of the random parameter within the crash data. It uses the data collected by the Washington State Department of Transportation (WSDOT) over a six-year period (2013 to 2018), during which 8548 crashes occurred at roundabouts. A random parameter binary probit model with heterogeneity in the means of random parameters employed to explore the effects of a wide range of variables on driver injury severity-related outcomes. The dataset was separated into four groups based on the age and gender of the drivers: young female, adult female, young male, and adult male. A log-likelihood ratio test and extensive transferability test were conducted to verify whether the modeling of crash severity at roundabouts needed to be carried out separately. The results indicate, with 99% confidence, that such accidents need to be modeled separately according to the drivers' age and gender. The model estimation results show that using the random parameter with heterogeneity in means improves overall model fit and yields essential new insights. Many factors potentially affect the likelihood of the driver injury severity estimation results for roundabout crashes outcomes. This includes crashes that occur during the weekdays and at two-lane roundabouts and those that involve sideswipes, driving under the influence of alcohol, collisions with pedal cycles, collisions with motorcycles, vehicles entering at an angle and so on. The findings of this research highlight the need to further study the factors that contribute to driver injury severity in roundabout-related accidents.

Key Words: Driver injury severity; Roundabout; Random parameter; Heterogeneity in the mean; Driver age and gender

4.1 Introduction

The construction of roundabouts continues to increase across the United States, as they serve as a favorable alternative to signalized or stop sign-controlled intersections (FHWA, 2015; Montella, 2011; Rodegerdts et al., 2015; WIDOT, 2020). A significant amount of research has indicated that drivers might not be as skilled at navigating roundabouts as they believe themselves to be. In a recent survey of 1,200 Washington and Oregon residents by the Washington State Department of Transportation WSDOT, three-quarters of the respondents claimed that they drive around roundabouts correctly, while two-thirds said that they see others making mistakes. In another study conducted by Day et al. (2018), it was found that some new drivers described being unsure about where to look at junctions and roundabouts while driving. Further, de Winter et al. (2009) indicated that drivers who learn to negotiate crossroads well do not automatically learn how to do the same for roundabouts as well.

Furthermore, in a study conducted by Al-Saleh and Bendak (2012) on the drivers' behavior at roundabouts, it was found that two-thirds of the drivers left the roundabout without indicating, which was the most prominent violation type observed. In addition, more than one-third of the drivers were found to be entering the roundabout without giving way or taking into consideration other cars that were already in the roundabout. Changing lanes unnecessarily was the third most frequent violation, followed by not slowing down when approaching roundabouts, and, finally, tooting, which was the least frequent and least severe. Moreover, Ziolkowski (2014) found that drivers are more likely to meet in an accident when navigating large roundabouts due to the associated conditions when driving at high speeds.

Regarding older drivers, Payyanadan et al. (2018) found that drivers who were 65 years and older were also more likely to report having issues with the complicated driving maneuvers involved in negotiating a roundabout. Burdett et al. (2017) found that younger drivers, under the age of 25, were engaged in 29% of all single-vehicle roundabout crashes, drivers aged 18–24 were involved in 24% of such crashes. Further, drivers aged 45–64 were involved in 22% of roundabout crashes. Regarding the exploratory factors, younger drivers were involved in 35% of all weather-related crashes and 61.9% of speed-related crashes at roundabouts. Several studies have been recently carried out to improve safety at roundabouts. Nevertheless, only a limited number of them have explored the factors that influence the severity of driver injury in roundabout-related accidents. Thus, it can be assumed that studies that explore this aspect are limited and scattered in terms of their varying objectives.

In addition to the previously mentioned studies, which evaluated the influence of the driver's characteristics on the frequency of roundabout crashes, another crucial research area is to identify the factors that lead to specific levels of injury severity for the drivers involved in these accidents. In an attempt to determine the impacts of such components, many studies have designed various sorts of discrete outcomes models. Driver injury severity have been statistically modeled using a wide variety of ordered and unordered discrete outcome approaches, such as the binary probit/logit models, ordered probit/logit models, multinomial logit model, and nested logit models. These models treat parameters as a constant across the observations (Ahmadi et al., 2018; Mannering and Bhat, 2014; Savolainen et al., 2011; Xie et al., 2012). A profusion of recent research has emphasized the importance of accounting for unobserved heterogeneity (factors that affect crash severity but are unobserved by the analyst) in the analysis of vehicle crash data (Mannering et al., 2016). Unobserved heterogeneity can arise from a number of sources, including unobserved environmental effects, interactions between the driver and vehicle, interactions between vehicles, and so on. Therefore, many studies accounted for unobserved heterogeneity through the inclusion of the random parameter approaches by assuming the estimated parameters vary across the observation according to some pre-specified distribution such as the random parameter multinomial logit model (mixed logit model), random parameter order probit model, random parameter binary probit model (Al-Bdairi et al., 2018; Anderson and Hernandez, 2017; Cerwick et al., 2014; Eluru and Yasmin, 2015; Haleem and Gan, 2013; Kim et al., 2013; Milton et al., 2008; Moore et al., 2011; Wu et al., 2014, 2016; Yasmin et al., 2015; Ye and Lord, 2014; Zubaidi et al., 2020).

In random parameter models, the distributions derived from the estimated random parameters are defined in terms of the full sample, and each respondent is randomly assigned an estimate drawn from the full distribution (Greene, 2012). However, this would make it impossible to verify whether the unobserved heterogeneity is a function of the explanatory variables. Therefore, to assess whether such effects exist, many recent studies introduced heterogeneity around the mean and/or invariance of the random parameters (Al-Bdairi et al., 2020; Anastasopoulos, 2016; Behnood and Mannering, 2017a, 2017b; Seraneeprakarn et al., 2017).

Despite the many studies conducted regarding roundabout safety (Al-Ghandour et al., 2011; Al-Marafi et al., 2019; AlKheder et al., 2020; Bahmankhah et al., 2019; Baker, 2020; Balado et al., 2019; Bastos et al., 2006; Burdett et al., 2017, 2016; Campisi et al., 2020; Chen et al., 2020; Echab et al., 2016; Ghanim et al., 2020; Landolsi et al., 2015a, 2015b; Nikitin et al., 2017; Patnaik et al., 2020; Pratelli et al., 2020; Pratic et al., 2015; Qin et al., 2013; Sadeq and Sayed, 2016; Shaaban and Hamad, 2020; Shen et al., 2020; Sisiopiku and Oh, 2001a, 2001b; Tollazzi, 2015; Vasconcelos et al., 2012; Yap et al., 2013; Zohdy and Rakha, 2013), the effect of the driver characteristics remains vague.

In other words, the relationship between the specific levels of driver injury severity and the gender and age of the driver remains unclear for roundabout-related accidents. Thereby, the majority of this research focuses on and intends to contribute to a better understanding of the influence of driver characteristics on the specific outcome of injury severity in roundabout crashes, with the heterogeneous mean being specified as a function of drivers' age and gender. This will be carried out by conducting crash-based analyses in which the unobserved heterogeneity and variance in the means of the random parameter are taken into consideration.

4.2 Methodology

4.2.1 The statistical model

In an attempt to better understand the effect of the driver's age and gender on the driver injury severity outcome for roundabout accidents, a random parameter binary probit model with heterogeneity in the means was developed. An additional layer of heterogeneity has been added, which is associated with the mean of the distribution of the estimated random parameter—in other words, allowing the random parameter to vary as per the explanatory variables. The dataset was divided into four groups depending on the age and gender of the drivers: young female, adult female, young male, and adult male. To begin, the effect of an explanatory variable $\boldsymbol{X}$ on a binary outcome can be expressed by assuming the disturbance term Ɛ to be normally distributed (Wooldridge, 2010) as follows:

$$y^* = \alpha + \boldsymbol{\beta X} + \varepsilon \tag{4.1}$$

Where,

$$y = 1[y^* > 0] \tag{4.2}$$

Where, $y = 1[y^* > 0]$ represents a crash in which an injury occurred (otherwise, $y = 0$).

To account for the unobserved heterogeneity—which can introduce variation into the model and, as a result, can affect the likelihood function of the driver injury severity outcome (Mannering et al., 2016)—a random parameter technique is applied as shown in Eq. (4.3) (Greene, 2012):

$$\beta_i = \beta + u_i \tag{4.3}$$

Where, β is the mean parameter, and u_i is a randomly distributed term that captures unobserved heterogeneity across crashes. Maximum log-likelihood estimation is performed to estimate the random parameter by utilizing 200 Halton draws, which provide an efficient distribution of the draws for numerical integration (Bhat, 2003; Pahukula et al., 2015). Then, to account for the impact of the one-unit change in the features (explanatory variable X) on the injury severity outcomes, marginal effects are computed as shown in Eq. (4.4) and referred to in Washington et al. (2011).

$$\frac{\partial Y}{\partial x_i} = \beta_i \, \phi \; (\beta_0 \; + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n \tag{4.4}$$

Finally, by assuming that the degree of unobserved heterogeneity could have some function of explanatory variables, the mean of the estimated parameter enabled the function of the explanatory variables, and Eq. (4.3) gets converted into Eq. (4.5).

$$\beta_i = \beta + \Theta Z_i + u_i \tag{4.5}$$

Where, $Z_i$is a vector of the explanatory variables that influence the mean of β_i , and Θ is a vector of the estimable parameters.

The maximum likelihood and simulation-based maximum likelihood methods are used to estimate the parameter vector. The normal distribution was found to be statistically significant among different distributions. In addition, the binary probit model is estimated using two hundred Halton draws, as it is stated in the literature that this number of Halton draws produces accurate estimates of the parameters (Bhat, 2003; Gkritza and Mannering, 2008; Hasan et al., 2011; Milton et al., 2008). Subsequently, a log-likelihood ratio test is proposed to statistically test the overall significance of using all the data models overusing separate models (young female, adult female, young male, and adult male). The first log-likelihood ratio test for transferability is as follows:

$$\chi^2 = -2[LL_{All\ Data}(\beta^{All\ Data}) - \sum LL_{Sep}(\beta^{Sep})] \tag{4.6}$$

Where, $LL_{All\ Data}(\beta^{All\ Data})$ is the log-likelihood at the convergence of the model with all the data, and $LL_{Sep}(\beta^{Sep})$ is the log-likelihood at the convergence of the subgroups mentioned above.

In addition to further validation, a more extensive transferability test was conducted to check if the modeling of crash severity at roundabouts needs to be carried out separately. This log-likelihood ratio test for transferability is as follows (Washington et al., 2011):

$$x^2 = -2\left[LL\left(\beta_{M1_{M2}}\right) - LL(\beta_{M1})\right] \tag{4.7}$$

Where, $LL(\beta_{M1_{M2}})$ is the log-likelihood at convergence for Model M_1 using the data from Model M_2, and$LL(\beta_{M1})$ is the log-likelihood at convergence for Model M_1.

Data Description

This study was based on crash data collected and compiled by the WSDOT. The data included information about all the accidents that occurred over a six-year period (2013 to 2018), during which time 8548 crashes occurred at roundabouts. To test the significance of the driver's gender and age in the context of these crashes, the data was split into four categories: young female driver (under the age of 25, 831 crashes), adult female driver (25 years or older, 5020 crashes), young male driver (under the age of 25, 910 crashes), and adult male driver (25years or older, 3491 crashes).

Due to the limited number of accidents that resulted in a disabling injury or fatality, it was not statistically possible to estimate all the five injury-level categories on KABCO scale (fatal, incapacitating, moderate, possible, and possible damage only). Thus, only two categories have been considered in this study: no-injury and injury. Table 4.1 presents the distribution level of injury severity across the four categories. Young males constitute the highest percentage of the no injury outcome with 86%, followed by both young females and adult males, who constitute have 85% of the total cases, and, finally, by adult females, who contribute to 55% of the cases. On the other hand, the adult females account for the highest number of accidents (55%), followed by both young females and adult males (15%), and, last, by the young males (14%).

Table 4.2 illustrates the descriptive statistics for the significant variables in the conducted models. The variables of using the lap and shoulder belts and collision with another vehicle present high percentages (86% and 84%, respectively) for adult male drivers. In contrast, 72% and 56% of crashes that occur during the weekdays and when entering roundabouts, respectively, can be attributed to adult females, whereas 49% of the accidents that occur when entering roundabouts are contributed by young males. Finally, young female drivers account for 23% of the crashes that occur when exiting roundabouts.

Table 4.1: Injury severity frequency and percentage distribution by different categories

Driver Characteristics	Injury Severity	Observation	Percentage (%)
Young female < 25	No-Injury	709	85
	Injury	122	15
	Total	831	100
Adult female ≥ 25	No-Injury	2263	45
	Injury	2757	55
	Total	5020	100
Young male < 25	No-Injury	785	86
	Injury	125	14
	Total	910	100
Adult male ≥ 25	No-Injury	2952	85
	Injury	539	15
	Total	3491	100

Table 4.2: Descriptive statistics of the significant variables in the injury severity models

Variables	Descriptive Statistics							
	Adult Female		Young Female		Adult Male		Young Male	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Temporal Characteristics								
Weekday (1 if the crash occurred during weekdays, otherwise 0)	0.72	0.45	-	-	-	-	-	-
Season (1 if winter (Dec–Feb), otherwise 0)	-	-	-	-	0.24	0.43	-	-
Spatial Characteristics								
County (1 if Snohomish, otherwise 0)	0.23	0.42	-	-	-	-	-	-
County (1 if Stevens, otherwise 0)	0.08	0.26	-	-	-	-	-	-
County (1 if Pacific, otherwise 0)	-	-	0.06	0.24	-	-	-	
County (1 if Island, otherwise 0)	-	-	-	-	-	-	0.17	0.38
Roadway Characteristics								
Roadway characteristics (1 if the road is curve and graded, otherwise 0)	0.05	0.23	-	-	-	-	-	-
Roadway characteristics (1 if the road is straight and graded, otherwise 0)	0.05	0.22	-	-	-	-	-	-
Posted speed limit (1 if speed limit is ≤ 25 m/hr, otherwise 0)	0.07	0.26	-	-	-	-	-	-
Number of lanes (1 if two-lane roundabout, otherwise 0)	-	-	-	-	0.12	0.32	-	-
Posted speed limit (1 if 20m/hr under a speed limit of <50 m/hr, otherwise 0)	-	-	-	-	0.66	0.48	-	-
Driver Action and Contribution								
Vehicle action (1 if stopped on the road, otherwise 0)	0.11	0.31	-	-	0.1	0.25	-	-
Vehicle action (1 if making a right, otherwise 0)	0.07	0.26	-	-	-	-	-	-
Vehicle action (1 if making a left, otherwise 0)	-	-	0.07	0.26	0.07	0.26	-	-
Vehicle action (1 if slowing down, otherwise 0)	-	-	-	-	0.03	0.17	-	-
Driver contribution (1 if passing incorrectly, otherwise 0)	0.03	0.17	0.03	0.16	-	-	-	-
Driver contribution (1 if under the influence of alcohol, otherwise 0)	-	-	-	-	0.07	0.26	0.06	0.23
Contributing circumstance (1 if inattentive, otherwise 0)	-	-	-	-	0.14	0.34	-	-
Speed condition (1 if reasonable speed is exceeded, otherwise 0)	0.03	0.14	0.07	0.25	-	-	-	-
Collision Types								
Collision type (1 if entering at angle, otherwise 0)	0.08	0.27	-	-	-	-	-	-
Collision type (1 if sideswiped, otherwise 0)	0.23	0.42	-	-	0.21	0.41	0.19	0.39
Collision type (1 if collided head-on, otherwise 0)	-	-	0.11	0.32	0.14	0.35	-	-
Collision type (1 if a transport vehicle is involved, otherwise 0)	-	-	-	-	0.84	0.37	-	-
Collision type (1 if collided with fixed object, otherwise 0)	-	-	-	-	0.1	0.29	-	-
Crash type (1 if vehicle goes off the road, otherwise 0)	-	-	-	-	0.05	0.21	-	-
Junction Relationship								
Junction relationship (1 if entering the roundabout, otherwise 0)	0.56	0.5	-	-	-	-	0.49	0.50

Variables (Continued)	Descriptive Statistics							
	Adult Female		Young Female		Adult Male		Young Male	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Junction relationship (1 if circling the roundabout, otherwise 0)	-	-	0.17	0.37	0.19	0.39	0.20	0.40
Junction relationship (1 if exiting the roundabout, otherwise 0)	-	-	0.23	0.42	0.22	0.41	0.23	0.42
Airbag type (1 if deployed, otherwise 0)	0.04	0.15	-	-	0.03	0.17	-	-
Road User Involvement								
Road user indicator (1 if a motorcycle is involved, otherwise 0)	0.03	0.12	0.03	0.11	0.04	0.19	0.03	0.16
Road user indicator (1 if a pedal cycle is involved, otherwise 0)	0.03	0.18	-	-	0.03	0.08	-	-
Road user indicator (1 if a pedestrian is involved, otherwise 0)	0.04	0.16	-	-	-	-	-	-
Truck involvement (1 if the truck weighs <10,000 lbs, otherwise 0)	0.47	0.5	-	-	-	-	-	-
Truck involvement (1 if the truck weighs >10,000 lbs, otherwise 0)	-	-	-	-	0.12	0.33	-	-
Restraint System Type								
Airbag type (1 if no airbag equipped, otherwise 0)	0.04	0.16	-	-	-	-	-	-
Airbag type (1 if deployed, otherwise 0)	0.03	0.15	-	-	-	-	0.05	0.21
Restraining system type (1 if lap and shoulder belts are used, otherwise 0)	-	-	-	-	0.86	0.34	-	-
Weather Condition								
Weather (1 if raining, otherwise 0)	-	-	-	-	0.16	0.37	-	-

4.3 Model Estimation Results

4.3.1 Likelihood ratio test results

Using Eq. (4.6), with regard to the model separation, a chi-square statistic of 236.56 was determined. Further, it was found that the total number of estimated random parameters in the four age and gender models had 15 degrees of freedom minus the number of the random estimated parameters in all the data models. The results indicate, with 99% confidence, that the crashes that occur at roundabouts need be modeled separately as per the drivers' age and gender. Therefore, only the separated models will be presented and discussed in this study (adult female, young female, adult male, and young male).

The transferability test results obtained by applying Eq. (4.7) are shown in Table 4.3. The results have been presented according to the estimated chi-squares (with the specific degrees of freedom values given in parentheses) with 99.99% confidence level using separated models based on driver age and gender.

Table 4.3: Chi-square statistics and degrees of freedom for driver injury severity regarding driver age and gender transferability test.

M_2	M_1			
	Adult Female (model)	Young Female (model)	Adult Male (model)	Young Male (model)
Adult Female (Data)	-	584(9) [> 99.99%]	2394.58 (22) [> 99.99%]	765.48 (9) [> 99.99%]
Young Female (Data)	745.08 (20) [> 99.99%]	-	762 (22) [> 99.99%]	210 (9) [> 99.99%]
Adult Male (Data)	1441.07 (20) [> 99.99%]	290 (9) [> 99.99%]	-	536.58 (9) [> 99.99%]
Young Male (Data)	361.14 (20) [> 99.99%]	53 (9) [> 99.99%]	442.46 (22) [> 99.99%]	-

4.4 Models Result Discussion

4.4.1 Temporal Characteristics

Many different temporal characteristics were found to be significant and to lead to different driver injury severity outcomes. In the current study, the "crashes that occur during weekdays" indicator variables were found to be statistically significant with a random parameter that is normally distributed, with a mean of 0.01 and a standard deviation of 0.73 for adult females. This result indicates that half of the crashes (more than zero) are more likely to involve injury, whereas the other half have less likelihood in the adult female model. This was not in line with the result obtained by Behnood and Mannering (2019) and Islam and Hernandez (2013). With regard to weather conditions (1 if winter (Dec–Feb), otherwise 0), the parameter was found to be statistically significant with a random parameter that is normally distributed using a statistically significant random parameter binary probit model with heterogeneity in the mean. The means of -1.07 and -1.03 and standard deviations of 1.5 and 1.49 were obtained for adult males using the binary probit model with and without heterogeneity in the mean, respectively. The result implied that 76% of the cases (for both models) involved vehicles that are less likely to result in injury, whereas 24% (for both models) are more likely to cause the same. This might be due to the snow season and the adverse weather conditions that occur during this period of time that could lead to driver impairment. Several studies have been conducted to investigate the effects of the impact factors on the occurrence of injury during the snow season (Heqimi et al., 2018; Seeherman and Liu, 2015).

4.4.2 Spatial Characteristics

The probability of injury in Snohomish County was found to decrease by -0.020 in the adult female model only, as shown in Table 4.5 that displays the marginal effects. Another notable finding is that the county variable (1 if Stevens) resulted in a -0.048 decrease in the probability of injury in the adult female model. Further, accidents that occur in Pacific County are significantly less likely to result in a disabling injury (-0.067) in the young female model. The indicator of the crashes that happen in Island County were found to have a random parameter that is normally distributed in the binary probit model with and without heterogeneity in the mean. The indicator variable for this variable as found to be statistically significant with a random parameter that is normally distributed, with a mean of -0.67 and standard deviation of 1.55 in the young model. This suggests that about 67% of the crashes involving multiple vehicles have a mean that is less than zero, while about 33% of them have a mean that is more than zero in the random parameter binary probit model with heterogeneity in the mean. In fact, there are many reasons that cause the accidents, and the explanations for each one are varied. Furthermore, the causes for the accidents differed according to the circumstances, times, and places associated with the occurrence of the crashes.

4.4.3 Roadway Characteristics

Crashes that occurred on curved roads with a gradient were found to involve less minor injuries; the parameter decreased the probability of minor injuries by -0.032 in the adult female model, as shown in Table 4.5. This might be due to adult females being more aware and also having more experience, in general, with these types of roads, which could tend to encourage them to drive at low speeds. With regard to the effect of roadway characteristics, the analysis indicated that the estimated parameter for a straight road with a grade was found to be statistically significant and random, with a mean of 0.14 and a standard deviation of 1.51 for adult females only. This indicates that 46% of drivers present a value of less than zero, which means that they are less likely to cause injury to themselves; this might be related to the slow speed at which adult females drive when navigating roundabouts. In contrast, 54% of them present a value greater than zero, which means that there is an increased probability of them experiencing an injury.

Moreover, the study found that the crashes that occur at roundabouts that have two lanes present a higher probability of possible injury in the full model. The possibilities involving the number of lanes (two lanes) resulted in an increase in the injury outcome by 0.004 in the adult male model. The one possible explanation for this is that the drivers might have been driving in the wrong lane. This may have caused them to change lanes, which is prohibited when driving around a roundabout and may lead to crashes. This finding in line with that of Isebrands and Hallmark (2012), Persaud et al. (2001), and Robinson et al. (2004).

With regard to the roadway characteristic, the speed of less than or equal to 20 mph was found to be normally distributed, with a mean of -1.32 and a standard deviation of 1.48. The distribution is over zero for almost 29%, suggesting that the speed of less than or equal to 20 mph almost always result in injury. This might be due to the traffic conditions at the time of the accident, and this result is not in line with that of Martin (2002). Finally, the posted speed limit (1 if 20 mph less that a speed limit of less than 50 mph) was found to increase the likelihood of severe injuries being sustained in the adult male model.

4.4.4 Driver Action and Contribution

Crashed related to driver action (slowing down) were found to have an effect on the injury outcome and resulted in a normally distributed random parameter for both models (adult only). A mean of 1.25 and a standard deviation of 0.35 were obtained in the adult male model when using the binary probit model with heterogeneity in the mean. Thus, in 0.2% of the observed accidents, the crashes that occurred due to driver action (if slowing down) are associated with a lower probability of injury outcome.

Furthermore, the driver's action (making a right) was also found to be significant, with negative coefficients indicating that the likelihood of injury is increased by 0.026 in the adult female model. This might be related to the lack of speed reduction when entering the roundabout. When examining the effect of driver action (making a left turn), the variable is found to have a normally distributed random parameter, with a mean of -2.5 and a standard deviation of 1.78 in the young female model. This implies that driver actions (making a left turn) that cause crashes increase the likelihood of injury for 8% of the observed cases and decrease the likelihood of injury for 92% of them. As shown in Table 4.4, driver contribution (if inattentive) was found to be normally distributed and random in the adult male model, with a mean of 0.26 and a standard deviation of 1.2 in the binary probit model with heterogeneity in the mean. This result supports the hypothesis that the drivers involved in such crashes are more likely to sustain injuries.

Moreover, for the driver contribution (if under the influence of alcohol) that affects the likelihood of injury, the marginal effects indicated that the driver contribution (if under the influence of alcohol) increases the probability of the occurrence of injury by 0.002 and 0.105 for the adult and young male models, respectively. This finding is in line with that of Behnood et al. (2014) and Behnood and Mannering (2017c). Among these variables, the crashes related to the driver contribution (if driving incorrectly) were found to be statistically significant with a random and normally distributed parameter. A mean of -6.1 and a standard deviation of 11.5 indicate that 70% (less than zero) of the crashes present less likelihood of causing injury, whereas 30% of them present more likelihood of this in the adult female model. Furthermore, the indicator variable for driver contribution (if stopped on the road) was found to be a statistically significant random parameter. The obtained mean (standard deviation) of the indictor variable were -0.82 (3.60) and -0.30 (3.7) in the adult female and male models, respectively. The results indicate that 59% and 53% of the distribution is less than 0, and the remaining 41% and 47% is greater than 0, for adult females and males, respectively.

Last, speed conditions (if the reasonable speed is exceeded) are also normally distributed, with an obtained mean (standard deviation) of 0.28 (1.12) and -0.08 (1.55) for adult and young females, respectively. For adult females, 40% of the distribution is less than 0, and the remaining 60% is greater than 0. Based on this, in 60% of the accident observations associated with speed conditions (if the reasonable speed is exceeded), adult females are more likely to suffer injuries. On the other hand, for young females, 52% of the distribution is less than 0, and the remaining 48% is greater than

0. The high-speed crashes related to traffic lights might be one acceptable reason for this result.

4.4.5 Collision Types

In the collision-type variables, the random parameter for the variable "involving a transport vehicle" has a mean of -3.48 and a standard deviation of 2.09 for the adult male model using the random parameter binary probit model with heterogeneity in the mean. Thus, 5% of the crash observations involving transport vehicles are associated with a higher probability of injury. This might be due to the adult male drivers undertaking relatively longer journeys. Therefore, they are more susceptible to getting into crashes with other road users (Adebisi et al., 2019).

Moving on to other parameters, the collision type (1 if entering at an angle, otherwise 0) was found to be statistically significant with a random and normally distributed parameter in the adult female model. A mean of 0.61 and a standard deviation of 0.22 indicate that 0.28% (less than zero) of the crashes are less likely to involve injuries, whereas 99.72% of them have more likelihood of the same occurring. One possible reason for this could be that the drivers do not heed to the yield sign at roundabouts. In addition, the collision type (1 if sideswiped, otherwise 0) is also normally distributed, with an obtained mean (standard deviation) of 0.36 (0.59) and -0.26 (2.23) for adult females and adult males, respectively. For adult females, 27% of the distribution is less than zero, while the remaining 73% is greater than zero. Based on this, 27% of sideswipe-related crash observations for adult females are associated with fewer injury outcomes. This result might be related to the slow speed at which they navigate roundabouts. On the other hand, for the adult males (random parameter binary probit model with heterogeneity in mean), 55% of the distribution is less than zero, while the remaining 45% is greater than zero. This may be explained by the adult male drivers being more experienced and becoming more responsible while driving. Mandavilli et al. (2009) found that sideswipe-related crashes are one of the significant accident types.

With regard to the collision types, head-on collisions were found to present a decreased likelihood of injury. The marginal effects show that head-on crashes decrease the probability of the occurrence of injury by -0.049 and -0.003 in the young female and adult male models, respectively. These results are expected and might be related to the fact that vehicles do not face each other at roundabouts. Moreover, the collision type (1 if involving a collision with a fixed object, otherwise 0) was found to decrease the likelihood of injuries occurring. The average marginal effect (Table 4.5) shows that the "collision with a fixed object" variable decreases the probability of injury by -0.03 in the adult male model. Several studies have shown that single-vehicle fixed-object collisions are frequent at roundabouts (Burdett et al., 2017; Mandavilli et al., 2009). The collision type (1 if the vehicle goes off-road, otherwise 0) resulted in a significant reduction in the probability of injury (with an average marginal effect of -0.009) for adult males.

4.4.6 Junction Relationship

The mean of the junction relationship (1 if entering a roundabout, otherwise 0) indicator increased if the adult driver was female. This variable has a mean of 0.23 and a standard deviation of 0.29, suggesting, with a normal distribution, that this variable is negative for 79% of the observations (increasing the likelihood of injury) and positive for 21% of the observations (decreasing the likelihood of injury). This might be attributed to high speeds and the driver not heeding to the yield sign when entering the roundabout.

With respect to the junction relationships displayed in Table 4.4, the analysis indicated that the estimated parameter of "if circling the roundabout" was found to be statistically significant using the random parameter binary probit model with and without heterogeneity in the mean. For this, the obtained mean (standard deviation) was -0.88 (1.27) and -1.61 (1.17) for the adult and young male models (random parameter binary probit model with heterogeneity in the mean), respectively. This indicates that 24% of the adult male drivers present a value of more than zero, while

8% of the young drivers present a value of more than zero, which means that they are less likely to bring about injury-related outcomes. This might be related to high speeds and the young drivers not having enough driving experience.

Finally, the junction relationship (1 if exiting the roundabout, otherwise 0) was found to be random and normally distributed, with a mean of -1.34 and standard deviation of 1.46 for young female drivers. Furthermore, 18% of the young female drivers were found to be more likely to sustain injuries, whereas 82% of the other drivers are less likely to yield the same outcome. This might be attributed to young female drivers who do not have enough driving experience or who drove at high speeds.

4.4.7 Road User Involvement

In Table 4.4, the parameter for the road user indicator (1 if a pedal cyclist is involved, otherwise 0) is found to be statistically significant with random distribution for both models (adult only). The parameters of the distribution are estimated to have a mean of 6.87 and a standard deviation of 8.44 for adult males using a random parameter binary probit model with heterogeneity in the mean. This revealed that almost 79% of the distribution is above zero, and only 21% of the drivers are less likely to be involved with injury-related outcomes.

Moving on to another parameter, the road user indicator (1 if a motorcycle is involved, otherwise 0) was found to be statistically significant with a random and normally distributed parameter using a random parameter binary probit model with heterogeneity in the mean in the adult male model. A mean of 5.20 and a standard deviation of 4.59 indicate that 87% (more than zero) of the crashes are more likely to involve injuries, whereas 13% of them present less likelihood for the same in the random parameter binary probit model with heterogeneity in the mean. This might be due to the adult males having sufficient driving experience and them driving carefully. This finding not in line with that obtained by Somasundaraswaran and Richardson (2019).

Furthermore, the road user indicator (1 if a pedestrian involved, otherwise 0) produced a statistically significant result. The marginal effects in Table 4.5 show that this indicator variable has the effect of increasing the likelihood of possible injury by 0.257 in the adult female model. This might be attributed to old adult females not focusing on the road at roundabouts.

For truck involvement (1 if the truck weighs more than 10,000 lbs, otherwise 0), adult male drivers were found to present less likelihood of sustaining injuries. The marginal effects indicate that adult male drivers decrease the probability of injury-related outcomes by -0.008. Last, truck involvement (1 if the truck weighs less than 10,000 lbs, otherwise 0) was found to be a normally distributed parameter, with a mean of -0.46 and a standard deviation of 0.70. This results in 26% of the distribution being more than 0, and 74% of the distribution being less than 0. Thus, for almost 26% of the roadway segments, the likelihood of injury-related outcomes is increased.

4.4.8 Restraint System Type

In Table 4.4, the airbag type (1 if no airbag equipped, otherwise 0) was found to be a fixed parameter that was significant and to have a positive effect on injury-related outcomes. The marginal effects show that the airbag type (1 if no airbag equipped, otherwise 0) increases the probability of the occurrence of injury by 0.032 in the adult female model.

Moving to the other airbag type (1 if deployed, otherwise 0), this was determined using a normally distributed parameter binary probit model and a random parameter binary probit model with heterogeneity in the mean, with obtained means (standard deviations) of -1.03 (4.29) and 0.83 (2.36) for adult and young males (random parameter binary probit model with heterogeneity in the mean), respectively. With regard to the estimated parameters, 41% of the distribution is more than 0, and 59% is less than 0, for an adult male, while 64% of the distribution is more than 0, and 36% is less than 0, for young males. This implies that 38% of the airbag type (1 if deployed, otherwise 0) increases the likelihood of accidents that cause injuries for adult males, while 19% of the airbag type (1 if deployed, otherwise 0) reduces the likelihood of injury-related outcomes.

Finally, the restraining system type (1 if lap and shoulder belts are used, otherwise 0) was found to be a fixed parameter that significantly reduced the likelihood of injury. The marginal effects show that the restraining system type (1 if lap and shoulder belts are used, otherwise 0) decreases the probability of the occurrence of injuries by -0.003 in the adult male model.

4.4.9 Weather Conditions

Weather (1 if raining, otherwise 0) was found to decrease the likelihood of injury-related outcome. The marginal effects (Table 4.5) show that the weather (1 if raining, otherwise 0) variable decreased the probability of injury by -0.003 in the adult male model. This might be due to adult male drivers preferring to reduce their speed during bad weather and them being more aware.

4.4.10 Heterogeneity in the Mean of the Random Parameters

The explanatory variables in all the models were tested for the possibility of significantly affecting the means and variances of the random parameters. The only models that produced significant heterogeneity in the means of random parameters were the adult and young male models, as shown in Table 4.4. Using the adult male data in the mixed logit model with heterogeneity in means, involvement of alcohol use was found to significantly affect the mean of the five random parameters.

The indicator variable for alcohol involvement was found to increase the mean of the variable "involving a transport vehicle," which indicates an increase in injury-related outcomes when a driver operating their vehicle under the influence of alcohol crashes into a transport vehicle. Further, alcohol involvement was found to decrease the mean of accidents involving inattentive drivers, making injuries less likely. In addition, a driver under the influence of alcohol was found to decrease the mean of the deployed airbag variable. This means that driving after consuming alcohol is less likely to result in injuries if the airbags are deployed.

Finally, for the adult male model, alcohol involvement was found to decrease the mean for collisions involving sideswiping or slowing down, making injury less likely for adult males. With regard to using the young male data, in the mixed logit model with heterogeneity in the mean, sideswiping collisions were found to increase the mean of the accidents in Island County and make injury-related outcomes.

4.5 Summary

This study was based on crash data collected and compiled by the WSDOT. The data included information about the crashes that occurred over a six-year period (2013 to 2018), during which 8548 crashes occurred at roundabouts. To determine the effect of the drivers' gender and age on their driving ability to negotiate the roundabouts, the data was divided into four categories: young female drivers (under the age of 25, 831 cases, adult female drivers (25 years or older, 5020 cases), young male drivers (under the age of 25, 910 cases), and adult male drivers (25 years or older, 3491 cases).

In an attempt to better understand the effect of the drivers' age and gender on injury severity outcomes for roundabout-related accidents, a random parameter binary probit model was used to account for the unobserved heterogeneity, which can introduce variation into the model and, as a result, affect the likelihood of the driver injury severity outcomes. An additional layer of heterogeneity has been added, which is associated with the mean of the distribution of the estimated random parameters. In other words, this allows the random parameter with heterogeneity in the means of the developed parameter to vary as per the explanatory variables, which improves the overall model fit and yields critical new insights. Due to the limited number of accidents that resulted in a disabling injury and fatality, only two categories have been considered: no-injury and injury. A log-likelihood ratio test and extensive transferability test were conducted to check whether the modeling of the crash severity at roundabouts needs to be carried out separately.

The results indicate, with 99% confidence, that the crashes that occur at roundabouts need to be modeled separately according to the drivers' age and gender. This data has been classified into different groups, namely temporal characteristics, spatial characteristics, roadway characteristics, driver actions and contributions, collision types, junction relationships, and restraint system types. Overall, there were 21 variables that presented a random parameter that is normally distributed. Both the adult male and the young male models have a parameter with heterogeneity in the mean of the random parameters. In addition, allowing for heterogeneity in the means of the random parameters empirically provides much more flexibility when tracking the unobserved heterogeneity in the data with any of the given distributional assumptions.

With regard to the estimation results for driver injury severity, a wide range of variables were found to increase the likelihood of drivers getting injured in roundabout-related accidents, including those that occur during the weekdays and at two-lane roundabouts and those that involve sideswipes, driving under the influence of alcohol, collisions with pedal cycles, collisions with motorcycles, entering an angle, and so on.

The findings of this research underscore the importance of fully accounting for unobserved heterogeneity by considering the possible heterogeneity in the means of the parameters. With the growing importance of roundabout safety, this paper provides not only certain essential initial findings using the WSDOT data but also, hopefully, some guidance for the analysis of other roundabout-accident databases from other geographic locations and time periods.

Table 4.4: Estimation results for random parameter binary probit model with/without heterogeneity in the mean.

Variable	Random parameter binary probit model								Random parameter binary probit model with heterogeneity in mean			
	Adult Female		Young Female		Adult Male		Young male		Adult Male		Young Male	
	Coeff	t-stat	Coeff	t-stat	Coeff	t-stat	Coeff	t-stat	Coeff	t-stat	Coeff	t-stat
Constant	-1.51	-14.31	-0.95	-12.78	0.69	2.20	-0.83	-4.74	0.92	2.83	-0.82	-4.51
Standard deviation (normally distributed)	0.29	7.56	-	-	0.38	7.93	-	-	0.49	9.55	-	-
Temporal Characteristics												
Weekday (1 if the crash occurs during weekdays, otherwise 0)	0.01	0.06	-	-	-	-	-	-	-	-	-	-
Standard deviation (normally distributed)	0.73	13.99	-	-	-	-	-	-	-	-	-	-
Season (1 if winter (Dec–Feb), otherwise 0)	-	-	-	-	-1.03	-6.82	-	-	-1.07	-6.67	-	-
Standard deviation (normally distributed)	-	-	-	-	1.49	10.97	-	-	1.5	10.86	-	-
Spatial Characteristics												
County (1 if Snohomish, otherwise 0)	-0.26	-2.76	-	-	-	-	-	-	-	-	-	-
County (1 if Stevens, otherwise 0)	-0.63	-3.51	-	-	-	-	-	-	-	-	-	-
County (1 if Pacific, otherwise 0)	-	-	-0.55	-1.73	-	-	-	-	-	-	-	-
County (1 if Island, otherwise 0)	-	-	-	-	-	-	-0.21	-0.99	-	-	-0.67	-2.21
Standard deviation (normally distributed)	-	-	-	-	-	-	1.20	5.53	-	-	1.55	5.71
Roadway Characteristics												
Roadway characteristics (1 if the road is curved and graded, otherwise 0)	-0.41	-2.24	-	-	-	-	-	-	-	-	-	-
Roadway characteristics (1 if the road is straight and graded, otherwise 0)	0.14	.73	-	-	-	-	-	-	-	-	-	-
Standard deviation (normally distributed)	1.51	6.59	-	-	-	-	-	-	-	-	-	-
Number of lanes (1 if two-lane roundabout, otherwise 0)	-	-	-	-	0.42	3.17	-	-	0.46	3.39	-	-
Posted speed limit (1 if the speed ≤ 20 mph)	-1.32	-4.27	-	-	-	-	-	-	-	-	-	-
Standard deviation (normally distributed)	1.48	5.77	-	-	-	-	-	-	-	-	-	-
Posted speed limit (1 if 20 mph less than a speed limit < 50 mph, otherwise 0)	-	-	-	-	0.39	3.91	-	-	0.43	3.99	-	-

Variable (Continued)	Random parameter binary probit model								Random parameter binary probit model with heterogeneity in mean			
	Adult Female		Young Female		Adult Male		Young male		Adult Male		Young Male	
	Coeff	t-stat	Coeff	t-stat	Coeff	t-stat	Coeff	t-stat	Coeff	t-stat	Coeff	t-stat
Driver Action and Contribution												
Driver action (1 if slowing down, otherwise 0)	-	-	-	-	1.09	4.73	-	-	1.25	5.01	-	-
Standard deviation (normally distributed)	-	-	-	-	0.41	1.83	-	-	0.35	1.54	-	-
Driver action (1 if making a right, otherwise 0)	0.33	2.26	-	-	-	-	-	-	-	-	-	-
Driver action (1 if making a left, otherwise 0)	-	-	-2.5	-2.49	-0.83	-3.73	-	-	-0.98	-4.15	-	-
Standard deviation (normally distributed)	-	-	1.78	0.72	-	-	-	-	-	-	-	-
Driver contribution (1 if inattentive, otherwise 0)	-	-	-	-	0.22	1.55	-	-	0.26	1.78	-	-
Standard deviation (normally distributed)	-	-	-	-	1.09	8.08	-	-	1.2	8.45	-	-
Driver contribution (1 if under the influence of alcohol, otherwise 0)	-	-	-	-	0.52	3.34	0.59	2.52	0.24	0.00	0.61	2.48
Driver contribution (1 if driving incorrectly, otherwise 0)	-6.1	-2.85	1.03	3.62	-	-	-	-	-	-	-	
Standard deviation (normally distributed)	11.50	3.48	-	-	-	-	-	-	-	-	-	-
Driver contribution (1 if stopped on the road, otherwise 0)	-0.82	-3.25	-	-	-0.30	-1.10	-	-	-0.27	0.33	-	-
Standard deviation (normally distributed)	3.60	9.66	-	-	3.7	10.91	-	-	4.12	11.17	-	-
Speed condition (1 if the reasonable speed is exceeded, otherwise 0)	0.28	1.08	-0.08	-.25	-	-	-	-	-	-	-	-
Standard deviation (normally distributed)	1.12	3.94	1.55	3.69	-	-	-	-	-	-	-	-
Collision Types												
Collision type (1 if involving a transport vehicle, otherwise 0)	-	-	-	-	-2.93	-9.79	-	-	-3.48	-10.79	-	-
Standard deviation (normally distributed)	-	-	-	-	1.81	19.33	-	-	2.09	19.13	-	-
Collision type (1 if entering at angle, otherwise 0)	0.61	4.62	-	-	-	-	-	-	-	-	-	-
Standard deviation (normally distributed)	0.22	1.83	-	-	-	-	-	-	-	-	-	-
Collision type (1 if sideswiped, otherwise 0)	0.36	3.31	-	-	-0.39	-2.48	0.37	2.48	-0.26	-1.55	0.29	1.76
Standard deviation (normally distributed)	0.59	6.73	-	-	2.11	13.3	-	-	2.23	13.42	-	-
Collision type (1 if collided head-on, otherwise 0)	-	-	-0.40	-1.67	-0.76	-4.15	-	-	-0.89	-4.49	-	-

Variable (Continued)	Random parameter binary probit model								Random parameter binary probit model with heterogeneity in mean			
	Adult Female		Young Female		Adult Male		Young male		Adult Male		Young Male	
	Coeff	t-stat	Coeff	t-stat	Coeff	t-stat	Coeff	t-stat	Coeff	t-stat	Coeff	t-stat
Collision type (1 if collided with a fixed object, otherwise 0)	-	-	-	-	-2.40	-7.60	-	-	-2.56	-7.68	-	-
Collision type (1 if vehicle goes off-road, otherwise 0)	-	-	-	-	-2.04	-6.21	-	-	-2.17	-6.31	-	-
Collision type (1 if involving a transport vehicle, otherwise 0)	-	-	-	-	-2.93	-9.79	-	-	-3.48	- 10.79	-	-
Junction Relationship												
Junction relationship (1 if entering roundabout, otherwise 0)	0.23	2.82	-	-	-	-	-0.61	-3.29	-	-	-0.61	-3.21
Standard deviation (normally distributed)	0.29	5.61	-	-	-	-	-	-	-	-	-	-
Junction relationship (1 if circling roundabout, otherwise 0)	-	-	-0.35	-1.94	-0.76	-4.75	-1.56	-4.94	-0.88	-4.82	-1.61	-4.71
Standard deviation (normally distributed)	-	-	-	-	1.003	7.47	1.09	4.87	1.27	8.48	1.17	4.90
Junction relationship (1 if exiting roundabout, otherwise 0)	-	-	-1.34	-4.03	-0.65	-5.05	-0.78	-3.72	-0.77	-5.64	-0.79	-3.69
Standard deviation (normally distributed)	-	-	1.46	5.22	-	-	-	-	-	-	-	-
Road User Involvement												
Road user indicator (1 if a pedal cycle is involved, otherwise 0)	4.1	6.52	-	-	6.19	1.65	-	-	6.87	1.66	-	-
Standard deviation (normally distributed)	-	-	-	-	7.43	1.72	-	-	8.44	1.76	-	-
Road user indicator (1 if a motorcycle is involved, otherwise 0)	2.10	8.12	2.03	4.19	3.96	11.95	2.19	6.85	5.20	11.54	2.20	6.81
Standard deviation (normally distributed)	-	-	-	-	3.06	8.33	-	-	4.59	8.58	-	-
Road user indicator (1 if a pedestrian is involved, otherwise 0)	3.34	5.98	-	-	-	-	-	-	-	-	-	-
Truck involvement (1 if the truck weighs > 10,000 lbs, otherwise 0)	-	-	-	-	-0.79	-4.42	-	-	-0.96	-5.01	-	-
Truck involvement (1 if the truck weighs < 10,000 lbs, otherwise 0)	-0.46	-5.44	-	-	-	-	-	-	-	-	-	-
Standard deviation (normally distributed)	0.70	10.31	-	-	-	-	-	-	-	-	-	-
Restraint System Type												
Airbag type (1 if no airbag equipped, otherwise 0)	0.41	1.92	-	-	-	-	-	-	-	-	-	-

Variable (Continued)	Random parameter binary probit model								Random parameter binary probit model with heterogeneity in mean			
	Adult Female		Young Female		Adult Male		Young male		Adult Male		Young Male	
	Coeff	t-stat	Coeff	t-stat	Coeff	t-stat	Coeff	t-stat	Coeff	t-stat	Coeff	t-stat
Airbag type (1 if deployed, otherwise 0)	1.62	7.34	-	-	-8.03	-2.62	1.12	4.18	-1.03	-3.01	0.83	2.33
Standard deviation (normally distributed)	-	-	-	-	25.67	3.27	1.26	3.66	4.29	6.60	2.36	3.68
Restraining system type (1 if lap and shoulder belts are used, otherwise 0)	-	-	-	-	-0.34	-2.59	-	-	-0.40	-2.85	-	-
Weather Condition												
Weather (1 if raining, otherwise 0)	-	-	-	-	-0.38	-2.92	-	-	-0.39	-2.84	-	-
Heterogeneity in the Means of Random Parameters												
Collision with vehicle in transit: Driver contribution (1 if under the influence of alcohol, otherwise 0)	-	-	-	-	-	-	-	-	2.45	6.39	-	-
Inattention during driving: Driver contribution (1 if under the influence of alcohol, otherwise 0)	-	-	-	-	-	-	-	-	-3.35	-1.88	-	-
Deployed airbag: Driver contribution (1 if under the influence of alcohol, otherwise 0)	-	-	-	-	-	-	-	-	-2.86	-3.93	-	-
Sideswipe collision: Driver contribution (1 if under the influence of alcohol, otherwise 0)	-	-	-	-	-	-	-	-	-3.18	-3.68	-	-
Slowing down: Driver contribution (1 if under the influence of alcohol, otherwise 0)	-	-	-	-	-	-	-	-	-2.53	-2.80	-	-
Island County: Collision type (1 if sideswiped, otherwise 0)	-	-	-	-	-	-	-	-	-	-	0.97	2.16
Model Statistics												
Log-likelihood function	-1106.58		-320.74		-1259.58		-312.65		-1242.83		-309.51	
Log-likelihood function at zero	-1259.97		-348.22		-1509.04		-366.04		-		-	
AIC	2275.2		665.5		2585.2		649.3		2573.7		649.0	
McFadden's pseudo r-squared	0.007		0.004		0.011		0.005		0.024		0.015	
No. of observations	2754		841		3533		923		3533		923	

Table 4.5: Averaged marginal effects for the random parameter binary probit model with /without heterogeneity in the mean.

Variable	Marginal effect for random parameter binary probit model				Marginal effect for random parameter binary probit model with heterogeneity in mean	
	Adult Female	Youth Female	Adult Male	Youth Male	Adult Male	Youth Male
Temporal Characteristics						
Weekday (1 if the crash occurred during weekdays, otherwise 0)	0.0004	-	-	-	-	-
Season (1 if winter (Dec–Feb), otherwise 0)	-	-	-.0524	-	-0.009	-
Spatial Characteristics						
County (1 if Snohomish, otherwise 0)	-0.020	-	-	-	-	-
County (1 if Stevens, otherwise 0)	-0.048	-	-	-	-	-
County (1 if Island, otherwise 0)	-	-	-	-0.029	-	0.116
County (1 if Pacific, otherwise 0)	-	-0.067	-	-	-	-
Roadway Characteristics						
Roadway characteristics (1 if the road is curved and graded, otherwise 0)	-0.032	-	-	-	-	-
Roadway characteristics (1 if the road is straight and graded, otherwise 0)	0.011	-	-	-	-	-
Posted speed limit (1 if speed limit ≤ 25 mph, otherwise 0)	-0.101	-	-	-	-	-
Number of lanes (1 if two-lane roundabout, otherwise 0)	-	-	0.030	-	0.004	-
Posted speed limit (1 if 20 mph less than a speed limit < 50 mph, otherwise 0)	-	-	0.025	-	0.004	-
Driver Action and Contribution						
Vehicle action (1 if stopped on the road, otherwise 0)	-0.063	-	0.017	-	-0.002	-
Vehicle action (1 if making a right, otherwise 0)	0.026	-	-	-	-	-
Vehicle action (1 if making a left, otherwise 0)	-	-0.304	-0.053	-	-0.008	-
Vehicle action (1 if slowing down, otherwise 0)	-	-	0.067	-	0.01	-
Driver contribution (1 if passing incorrectly, otherwise 0)	-0.469	0.126	-	-	-	-
Driver contribution (1 if under the influence of alcohol, otherwise 0)	-	-	0.048	0.082	0.002	0.105
Contributing circumstance (1 if inattentive, otherwise 0)	-	-	0.007	-	0.002	-
Speed condition (1 if the reasonable speed is exceeded, otherwise 0)	0.021	-.0103	-	-	-	-
Collision Types						
Collision type (1 if entering at an angle, otherwise 0)	0.047	-	-	-	-	-
Collision type (1 if sideswiped, otherwise 0)	0.028	-	-0.011	0.051	-0.002	0.049

Variable (Continued)	Marginal effect for random parameter binary probit model				Marginal effect for random parameter binary probit model with heterogeneity in mean	
	Adult Female	Youth Female	Adult Male	Youth Male	Adult Male	Youth Male
Collision type (1 if collided head-on, otherwise 0)	-	-0.049	-0.048	-	-0.003	-
Collision type (1 if involving a transport vehicle, otherwise 0)	-	-	-0.188	-	-0.03	-
Collision type (1 if collided with a fixed object, otherwise 0)	-	-	-0.203	-	-0.022	-
Crash type (1 if the vehicle goes off-road, otherwise 0)	-	-	-0.178	-	-0.019	-
Junction Relationship						
Junction relationship (1 if entering roundabout, otherwise 0)	0.017	-	-	-0.085	-	-0.104
Junction relationship (1 if circling roundabout, otherwise 0)	-	-0.043	-0.042	-0.218	-0.008	-0.277
Junction relationship (1 if exiting roundabout, otherwise 0)	-	-0.164	-0.044	-0.109	-0.007	-0.136
Road User Involvement						
Road user indicator (1 if a pedal cycle is involved, otherwise 0)	0.315	-	0.223	-	0.059	-
Road user indicator (1 if a motorcycle is involved, otherwise 0)	0.162	0.247	0.284	0.306	0.044	0.379
Road user indicator (1 if a pedestrian is involved, otherwise 0)	0.257	-	-	-	-	-
Truck involvement (1 if the truck weighs > 10,000 lbs, otherwise 0)	-	-	-0.05	-	-0.008	-
Truck involvement (1 if the truck weighs < 10,000 lbs, otherwise 0)	-0.036	-	-	-	-	-
Restraint System Type						
Airbag type (1 if no airbag equipped, otherwise 0)	0.032	-	-	-	-	-
Airbag type (1 if deployed, otherwise 0)	0.125	-	-0.051	0.157	-0.009	0.143
Restraining system type (1 if lap and shoulder belts are used, otherwise 0)	-	-	-0.027	-	-0.003	-
Weather Condition						
Weather (1 if raining, otherwise 0)	-	-	-0.024	-	-0.003	-

References

Adebisi, A., Ma, J., Masaki, J. and Sobanjo, J. (2019), "safety Age-Related Di ff erences in Motor-Vehicle Crash Severity in California", pp. 1–17.

Ahmadi, A., Jahangiri, A., Berardi, V. and Machiani, S.G. (2018), "Crash severity analysis of rear-end crashes in California using statistical and machine learning classification methods", *Journal of Transportation Safety and Security*, No. February 2019, available at:https://doi.org/10.1080/19439962.2018.1505793.

Al-Bdairi, N.S.S., Behnood, A. and Hernandez, S. (2020), "Temporal stability of driver injury severities in animal-vehicle collisions: A random parameters with heterogeneity in means (and variances) approach", *Analytic Methods in Accident Research*, Vol. 26 No. September 2019, pp. 1–52.

Al-Bdairi, N.S.S., Hernandez, S. and Anderson, J. (2018), "Contributing Factors to Run-Off-Road Crashes Involving Large Trucks under Lighted and Dark Conditions", *Journal of Transportation Engineering, Part A: Systems*, Vol. 144 No. 1, pp. 1–9.

Al-Ghandour, M.N., Schroeder, B.J., Williams, B.M. and Rasdorf, W.J. (2011), "Conflict Models for Single-Lane Roundabout Slip Lanes from Microsimulation", *Transportation Research Record: Journal of the Transportation Research Board*, available at:https://doi.org/10.3141/2236-11.

Al-Marafi, M.N., Somasundaraswaran, K. and Ayers, R. (2019), "Developing crash modification factors for roundabouts using a cross-sectional method", *Journal of Traffic and Transportation Engineering (English Edition)*, Elsevier Ltd, No. xxx, available at:https://doi.org/10.1016/j.jtte.2018.10.012.

Al-Saleh, K. and Bendak, S. (2012), "Drivers' behaviour at roundabouts in Riyadh", *International Journal of Injury Control and Safety Promotion*, Vol. 19 No. 1, pp. 19–25.

AlKheder, S., Al-Rukaibi, F. and Al-Faresi, A. (2020), "Driver behavior at Kuwait roundabouts and its performance evaluation", *IATSS Research*, available at:https://doi.org/https://doi.org/10.1016/j.iatssr.2020.03.004.

Anastasopoulos, P.C. (2016), "Random parameters multivariate tobit and zero-inflated count data models: Addressing unobserved and zero-state heterogeneity in accident injury-severity rate and frequency analysis", *Analytic Methods in Accident Research*, Elsevier, Vol. 11, pp. 17–32.

Anderson, J. and Hernandez, S. (2017), "Roadway classifications and the accident injury severities of heavy-vehicle drivers", *Analytic Methods in Accident Research*, Elsevier Ltd, Vol. 15, pp. 17–28.

Bahmankhah, B., Fernandes, P., Teixeira, J. and Coelho, M.C. (2019), "Interaction between motor vehicles and bicycles at two-lane roundabouts: a driving volatility-based analysis", *International Journal of Injury Control and Safety Promotion*, Taylor & Francis, Vol. 26 No. 3, pp. 205–215.

Baker, B.D. (2020), *The Future of Roundabouts in Northeast Ohio*.

Balado, J., Díaz-Vilariño, L., Arias, P. and Novo, A. (2019), "A safety analysis of roundabouts and turbo roundabouts based on Petri nets", *Traffic Injury Prevention*, Taylor & Francis, Vol. 20 No. 4, pp. 400–405.

Bastos, S.A., DA Maia, S.Á.J. and DA Silva, J.P.C. (2006), "Characterization of Trajectories Adopted at Roundabout Crossings", *Proceedings of the Eurpean Transport Conference (ETC)*.

Behnood, A. and Mannering, F. (2017a), "The effect of passengers on driver-injury severities in single-vehicle crashes: A random parameters heterogeneity-in-means approach", *Analytic Methods in Accident Research*, Elsevier Ltd, Vol. 14, pp. 41–53.

Behnood, A. and Mannering, F. (2017b), "Determinants of bicyclist injury severities in bicycle-vehicle crashes: A random parameters approach with heterogeneity in means and variances", *Analytic Methods in Accident Research*, Elsevier Ltd, Vol. 16, pp. 35–47.

Behnood, A. and Mannering, F. (2019), "Time-of-day variations and temporal instability of factors affecting injury severities in large-truck crashes", *Analytic Methods in Accident Research*, Elsevier Ltd, Vol. 23, p. 100102.

Behnood, A. and Mannering, F.L. (2017c), "The effects of drug and alcohol consumption on driver injury severities in single-vehicle crashes".

Behnood, A., Roshandeh, A.M. and Mannering, F.L. (2014), "Latent class analysis of the effects of age, gender, and alcohol consumption on driver-injury severities", *Analytic Methods in Accident Research*, Elsevier, Vol. 3–4, pp. 56–91.

Bhat, C.R. (2003), "Simulation estimation of mixed discrete choice models using randomized and scrambled Halton sequences", *Transportation Research Part B: Methodological*, Vol. 37 No. 9, pp. 837–855.

Burdett, B., Alsghan, I., Chiu, L.-H., Bill, A.R. and Noyce, D.A. (2016), "Analysis of Rear-End Collisions at Roundabout Approaches", *Transportation Research Record: Journal of the Transportation Research Board*, available at:https://doi.org/10.3141/2585-04.

Burdett, B., Bill, A.R. and Noyce, D.A. (2017), "Evaluation of Roundabout-Related Single-Vehicle Crashes", *Transportation Research Record: Journal of the Transportation Research Board*, available at:https://doi.org/10.3141/2637-03.

Campisi, T., Tibljaš, A.D.-, Tesoriere, G., Canale, A., Rencelj, M. and Šurdonja, S. (2020), "Cycling traffic at turbo roundabouts: some considerations related to cyclist mobility and safety", *Transportation Research Procedia*, Vol. 45, pp. 627–634.

Cerwick, D.M., Gkritza, K., Shaheed, M.S. and Hans, Z. (2014), "A comparison of the mixed logit and latent class methods for crash severity analysis", *Analytic Methods in Accident Research*, Elsevier, Vol. 3–4, pp. 11–27.

Chen, Y., Zheng, X., Peterson, A. and Ahmadian, M. (2020), *Simulation Evaluation on the Rollover Propensity of Multi-Trailer Trucks at Roundabouts*, SAE Technical Paper.

Day, M.R., Thompson, A.R., Poulter, D.R., Stride, C.B. and Rowe, R. (2018), "Why do drivers become safer over the first three months of driving? A longitudinal qualitative study", *Accident Analysis and Prevention*, Elsevier, Vol. 117 No. July 2017, pp. 225–231.

Echab, H., Lakouari, N., Ez-Zahraouy, H. and Benyoussef, A. (2016), "Simulation study of traffic car accidents at a single lane roundabout", *International Journal of Modern Physics C*, available at:https://doi.org/10.1142/S0129183116500091.

Eluru, N. and Yasmin, S. (2015), "A note on generalized ordered outcome models", *Analytic Methods in Accident Research*, Elsevier, Vol. 8, pp. 1–6.

FHWA. (2015), "A Review of Fatal and Severe Injury Crashes at Roundabouts", Vol. IV No. September.

Ghanim, M., Kharbeche, M., Hannun, J., Hannun, J. and Shamiyeh, K. (2020), "Safety and Operational Performance of Signalized Roundabouts: A Case Study in Doha", *Procedia Computer Science*, Elsevier B.V., Vol. 170, pp. 427–433.

Gkritza, K. and Mannering, F.L. (2008), "Mixed logit analysis of safety-belt use in single- and multi-occupant vehicles", *Accident Analysis and Prevention*, Vol. 40 No. 2, pp. 443–451.

Greene, W.H. (2012), *Econometric Modeling Guide*.

Haleem, K. and Gan, A. (2013), "Effect of driver's age and side of impact on crash severity along urban freeways: A mixed logit approach", *Journal of Safety Research*, National Safety Council and Elsevier Ltd, Vol. 46, pp. 67–76.

Hasan, S., Ukkusuri, S., Gladwin, H. and Murray-Tuite, P. (2011), "Behavioral Model to Understand Household-Level Hurricane Evacuation Decision Making", *Journal of Transportation Engineering*, Vol. 137 No. 5, pp. 341–348.

Heqimi, G., Gates, T.J. and Kay, J.J. (2018), "Using spatial interpolation to determine impacts of annual snowfall on traffic crashes for limited access freeway segments", *Accident Analysis and Prevention*, Elsevier, Vol. 121 No. February, pp. 202–212.

Isebrands, H. and Hallmark, S. (2012), "Statistical Analysis and Development of Crash Prediction Model for Roundabouts on High-Speed Rural Roadways", No. 2312, pp. 3–13.

Islam, M. and Hernandez, S. (2013), "Large truck-involved crashes: Exploratory injury severity analysis", *Journal of Transportation Engineering*, Vol. 139 No. 6, pp. 596–604.

Kim, J.-K., Ulfarsson, G.F., Kim, S. and Shankar, V.N. (2013), "Driver-injury severity in single-vehicle crashes in California: A mixed logit analysis of heterogeneity due to age and gender.", *Accident; Analysis and Prevention*, Elsevier Ltd, Vol. 50, pp. 1073–81.

Landolsi, J., Rehimi, F. and Kalboussi, A. (2015), "Macroscopic traffic modeling on a network: Application to a Roundabout", *GSCIT 2015 - Global Summit on Computer and Information Technology - Proceedings*, available at:https://doi.org/10.1109/GSCIT.2015.7353327.

Mandavilli, S., McCartt, A.T. and Retting, R.A. (2009), "Crash patterns and potential engineering countermeasures at maryland roundabouts", *Traffic Injury Prevention*, Vol. 10 No. 1, pp. 44–50.

Mannering, F.L. and Bhat, C.R. (2014), "Analytic methods in accident research: Methodological frontier and future directions", *Analytic Methods in Accident Research*, Elsevier, Vol. 1, pp. 1–22.

Mannering, F.L., Shankar, V. and Bhat, C.R. (2016), "Unobserved heterogeneity and the statistical analysis of highway accident data", *Analytic Methods in Accident Research*, Elsevier, Vol. 11, pp. 1–16.

Martin, J. (2002), "Relationship between crash rate and hourly traffic flow on interurban motorways", Vol. 34, pp. 619–629.

Milton, J.C., Shankar, V.N. and Mannering, F.L. (2008), "Highway accident severities and the mixed logit model: An exploratory empirical analysis", *Accident Analysis and Prevention*, Vol. 40 No. 1, pp. 260–266.

Montella, A. (2011), "Identifying crash contributory factors at urban roundabouts and using association rules to explore their relationships to different crash types", *Accident Analysis and Prevention*, Elsevier Ltd, Vol. 43 No. 4, pp. 1451–1463.

Moore, D.N., Schneider IV, W.H., Savolainen, P.T. and Farzaneh, M. (2011), "Mixed logit analysis of bicyclist injury severity resulting from motor vehicle crashes at intersection and non-intersection locations", *Accident Analysis and Prevention*, Elsevier Ltd, Vol. 43 No. 3, pp. 621–630.

Nikitin, N., Patskan, V. and Savina, I. (2017), "Efficiency Analysis of Roundabout with Traffic Signals", *Transportation Research Procedia*, available at:https://doi.org/10.1016/j.trpro.2017.01.072.

Pahukula, J., Hernandez, S. and Unnikrishnan, A. (2015), "A time of day analysis of crashes involving large trucks in urban areas", *Accident Analysis and Prevention*, Elsevier Ltd, Vol. 75, pp. 155–163.

Patnaik, A.K., Agarwal, L.A., Panda, M. and Bhuyan, P.K. (2020), "Entry capacity modelling of signalized roundabouts under heterogeneous traffic conditions", *Transportation Letters*, Taylor & Francis, Vol. 12 No. 2, pp. 100–112.

Payyanadan, R.P., Lee, J.D. and Grepo, L.C. (2018), "Challenges for older drivers in Urban, Suburban, and rural settings", *Geriatrics (Switzerland)*, Vol. 3 No. 2, available at:https://doi.org/10.3390/geriatrics3020014.

Persaud, B., Retting, R., Garder, P. and Lord, D. (2001), "Safety Effect of Roundabout Conversions in the United States: Empirical Bayes Observational Before-After Study", *Transportation Research Record: Journal of the Transportation Research Board*, Transportation Research Board of the National Academies , Vol. 1751, pp. 1–8.

Pratelli, A., Lupi, M., Pratelli, C. and Farina, A. (2020), "Mini-roundabouts for Improving Urban Accessibility", *Modelling of the Interaction of the Different Vehicles and Various Transport Modes*, Springer, pp. 333–382.

Pratic??, F.G., Vaiana, R. and Gallelli, V. (2015), "Micro-simulation effectiveness in predicting operating speed profiles in a roundabout", *Advances in Transportation Studies*.

Qin, X., Bill, A., Chitturi, M. and Noyce, D.A. (2013), "Evaluation of roundabout safety", *Transportation Research Board 92nd Annual Meeting*.

Robinson, B., Blogg, M. and Waltman, E. (2004), "MODERN ROUNDABOUTS AND OTHER INTERSECTION TYPES", pp. 1–55.

Rodegerdts, L.A., Malinge, A., Marnell, P.S., Beaird, S.G., Kittelson, M.J. and Mereszczak, Y.S. (2015), "Assessment of Roundabout Capacity Models for the Highway Capacity Manual", Vol. II No. September, p. 94.

Sadeq, H. and Sayed, T. (2016), "Automated Roundabout Safety Analysis: Diagnosis and Remedy of Safety Problems", *Journal of Transportation Engineering*, available at:https://doi.org/10.1061/(ASCE)TE.1943-5436.0000887.

Savolainen, P.T., Mannering, F.L., Lord, D. and Quddus, M.A. (2011), "The statistical analysis of highway crash-injury severities : A review and assessment of methodological alternatives", *Accident Analysis and Prevention*, Elsevier Ltd, Vol. 43 No. 5, pp. 1666–1676.

Seeherman, J. and Liu, Y. (2015), "Effects of extraordinary snowfall on traffic safety", *Accident Analysis and Prevention*, Vol. 81, pp. 194–203.

Seraneeprakarn, P., Huang, S., Shankar, V., Mannering, F., Venkataraman, N. and Milton, J. (2017), "Occupant injury severities in hybrid-vehicle involved crashes: A random parameters approach with heterogeneity in means and variances", *Analytic Methods in Accident Research*, Elsevier Ltd, Vol. 15, pp. 41–55.

Shaaban, K. and Hamad, H. (2020), "Critical gap comparison between one-, two-, and three-lane roundabouts in Qatar", *Sustainability (Switzerland)*, Vol. 12 No. 10, pp. 1–14.

Shen, J., Wang, T., Zheng, C. and Yu, M. (2020), “Determinants of Bicyclist Injury Severity Resulting from Crashes at Roundabouts, Crossroads, and T-Junctions”, *Journal of Advanced Transportation*, Hindawi, Vol. 2020.

Sisiopiku, V.P. and Oh, H.-U. (2001), “Evaluation of roundabout performance using SIDRA”, *Journal of Transportation Engineering*, available at:https://doi.org/10.1061/(ASCE)0733-947X(2001)127:2(143).

Somasundaraswaran, K. and Richardson, M. (2019), “Investigating the Possibilities of Specific Crash Type at Roundabouts”, *Journal of Traffic and Transportation Engineering*, Vol. 7 No. 5, pp. 206–211.

Tollazzi, T. (2015), “First developments of different roundabout types”, *Springer Tracts on Transportation and Traffic*, available at:https://doi.org/10.1007/978-3-319-09084-9_2.

Vasconcelos, L., Seco, Á. and Silva, A. (2012), “A Comparison of Roundabout Capacity Models”, *IASTED - International Association of Science and Technology for Development*, available at:https://doi.org/10.2316/P.2012.783-023.

Washington, S.P., Karlaftis, M.G. and Mannering, F.L. (2010), *Statistical and Econometric Methods for Transportation Data Analysis, Second Edition*, CRC Press.

Washington, S.P., Karlaftis, M.G. and Mannering, F.L. (2011), *Statistical and Econometric Methods for Transportation Data Analysis, Chapman & Hall/CRC*, 2nd ed., Chapman & Hall/CRC, available at:https://doi.org/10.1198/tech.2004.s238.

WIDOT. (2020), “Motorists’ handbook”.

de Winter, J.C.F., de Groot, S., Mulder, M., Wieringa, P.A., Dankelman, J. and Mulder, J.A. (2009), “Relationships between driving simulator performance and driving test results”, *Ergonomics*, Vol. 52 No. 2, pp. 137–153.

Wooldridge, J.M. (2010), *Econometric Analysis of Cross Section and Panel Data*, 2nd ed., MIT Press, Cambridge, MA, available at:https://doi.org/10.1016/B978-0-323-05712-7.00031-3.

Wu, Q., Chen, F., Zhang, G., Liu, X.C., Wang, H. and Bogus, S.M. (2014), “Mixed logit model-based driver injury severity investigations in single- and multi-vehicle crashes on rural two-lane highways”, *Accident Analysis and Prevention*, Elsevier Ltd, Vol. 72, pp. 105–115.

Wu, Q., Zhang, G., Zhu, X., Liu, X.C. and Tarefder, R. (2016), “Analysis of driver injury severity in single-vehicle crashes on rural and urban roadways”, *Accident Analysis and Prevention*, Elsevier Ltd, Vol. 94, pp. 35–45.

Xie, Y., Zhao, K. and Huynh, N. (2012), “Analysis of driver injury severity in rural single-vehicle crashes”, *Accident Analysis and Prevention*, Elsevier Ltd, Vol. 47, pp. 36–44.

Yap, Y.H., Gibson, H.M. and Waterson, B.J. (2013), "An International Review of Roundabout Capacity Modelling", *Transport Reviews*, available at:https://doi.org/10.1080/01441647.2013.830160.

Yasmin, S., Eluru, N. and Pinjari, A.R. (2015), "Analyzing the continuum of fatal crashes: A generalized ordered approach", *Analytic Methods in Accident Research*, Elsevier, Vol. 7, pp. 1–15.

Ye, F. and Lord, D. (2014), "Comparing three commonly used crash severity models on sample size requirements: Multinomial logit, ordered probit and mixed logit models", *Analytic Methods in Accident Research*, Elsevier, Vol. 1, pp. 72–85.

Ziolkowski, R. (2014), "The influence of roundabouts on drivers' speed and behaviour", *Advanced Materials Research*, Vol. 1020 No. Table 1, pp. 674–679.

Zohdy, I. and Rakha, H. (2013), *Enhancing Roundabout Operations via Vehicle Connectivity, Transportation Research Record,* available at:https://doi.org/10.3141/2381-11.

Zubaidi, H.A., Anderson, J.C. and Hernandez, S. (2020), "Understanding roundabout safety through the application of advanced econometric techniques", *International Journal of Transportation Science and Technology*, Tongji University and Tongji University Press, No. xxxx, available at:https://doi.org/10.1016/j.ijtst.2020.03.001.

Chapter 5: Summary and Conclusion

5.1 Summary

The construction of roundabouts as alternatives to signalized or STOP-controlled intersections has increased in the united states due to their safety performance characteristics (Montella 2011). The rapid growth in the number of roundabouts raises some significant research challenges, especially in the area of safety performance and evaluation for U.S. specific data under varying conditions. Despite the advantages in the roundabout geometric design, crashes still occur. While roundabouts are great at preventing severe crashes, they may bring on more non-fatal wrecks. This will be accomplished through exploring relatively new econometric techniques and machine learning algorithms applied to roundabout crash data. With this in mind, what is still not clearly understood is the relationship between roundabout crash-related factors, crash types, injury severity, and roundabout configurations. A reason for this may stem from the lack of available detailed crash-related data. The current econometric models were good 10 to 15 years ago; however, since that time, there has been considerable advancement in econometrics, especially econometrics methods that account for unobserved heterogeneity. These advancements have been shown to provide a more reasonable understanding of contributing factors to overall safety issues. It is envisioned that through this research, additional variables will be included that account for the human element compared to simply using exposure-based variables that current methods rely on (Mannering and Bhat 2014). This is important since most crashes are a result of driver behaviors and other environmental factors that play a big role in crash outcomes.

For that, the aim of this dissertation is to conduct crash-based analyses to better understand the factors that may influence less severe crashes to those of more severe crashes given various configurations and crash types utilizing Oregon crash data in the first and second manuscripts and Washington crash data for the third manuscript.

The first manuscript investigated a crash-based analysis to better understand the factors that may influence less severe crashes to those of more severe crashes given various roundabout configurations and crash types. Using Oregon's crash database from 2011 to 2015 in which 1,006 crashes occurred at roundabouts. A series of log-likelihood ratio tests were conducted to validate that four separate random parameters binary probit models by configuration type were warranted.

The objective of the second manuscript is to develop a machine learning methodology that evaluates crash injury severity at roundabouts and compares this method with traditional econometric techniques. Precisely, this work will estimate a Random Parameter Binary Probit model (RPBP) and compare its predictions with those rendered from machine learning techniques, namely, Support Vector Machine (SVM) with linear, radial, polynomial, and sigmoid kernels. This is accomplished by utilizing Oregon crash data from 2011 to 2015 and focuses on both three- and four-leg roundabouts. Two significant variable sets have been conducted by utilizing random forest and binary models.

Finally, the third manuscript investigated risk factors that significantly contribute to the driver injury severity of roundabout crashes while systematically accounting for unobserved heterogeneity and the variance in means of the random parameter within the crash data. Using the data from the Washington State Department of Transportation (WSDOT) over a six-year period (2013 to 2018) in which 8548 crashes occurred at roundabouts. A random parameter binary probit model with heterogeneity in random parameter means is estimated to explore the effects of a wide range of variables on driver injury severity outcomes.

5.2 Study Findings

It is generally accepted that the number of crashes at roundabouts are fewer than those at signalized intersections. Therefore, obtaining detailed data that can capture the factors that contribute to crash severity is more complicated regarding the required sample size that accurately represents the population. Accordingly, the aggregated injury category consists of fatal, major, moderate, and minor injury outcomes, while the no injury category includes only no injury outcomes. The purpose of this aggregation is to increase the number of observations to reduce the variability caused by random effects when statistical methods are implemented (Chang and Mannering

1999). This is essential since the data that is used in this study has too few observations on incapacitating and fatal injuries to set apart their individual effects. As a result, The dependent variables in each of these manuscripts consisted of two specific outcomes: (1) no injury and (2) injury, and the findings of this manuscript are summarized.

First Manuscript: Four estimated models were developed based on the geometric design of the roundabout: full model (unknown, three-leg, four legs, and five legs), three and four-leg combination model, three-leg model, and four-leg model. A number of important factors were found to influence the level of injury severity at roundabouts. In each individual model, a number of variables are homogenous across crash observations (i.e., their estimated parameters are fixed across observations), and various variables are heterogeneous across crash observations (i.e., they have estimated random parameters). For example, vehicles stopped in traffic and not waiting to make a left turn, seatbelt usage, gender, type of vehicle, roadside crash characteristics, vehicle movement, age of the driver, careless driving, and alcohol use were found to have estimated random parameters.

Second Manuscript: The results demonstrated that the binary model performed best when predicting injury severity at three-leg roundabouts and under a 70-30 split in training-test ratio. Specifically, 76% accuracy using variables selected by a random forest and 78% accuracy using variables selected by the binary probit model. Prediction rates for the binary probit model were lower when considering four-leg roundabouts, but a 77% accuracy was observed when considered an 80-20 split and variables selected by a random forest.

The SVM-linear model has comparable predictions for both three-leg and four-leg models, under both variable selection methods, with 81% and 84%, respectively. SVM-radial had a higher prediction, specifically for the four-leg model and using variables selected by the binary probit model (the highest prediction accuracy was 87% with a 90-10 split). SVM-polynomial performed best in the three-leg model using variables selected by a random forest (91% accuracy under a 90-10 split, also the highest observed accuracy) and in the four-leg model using variables selected by the binary probit model (87% accuracy under a 90-10 split). Lastly, SVM-sigmoid was the most consistent of the SVM models across all training-test ratios and variable selection. Specifically, SVM-sigmoid performed best for four-leg models using variables selected by the binary probit model and had the lowest prediction rate in the three-leg model using the same variables.

In regards to variable importance, a random forest analysis indicated that afternoon, snowy weather, and drivers aged 36-50 years were the most important injury severity predictors for three-leg roundabouts. For four-leg roundabouts, poor pavement condition, dusk lighting, and losing control of the vehicle were identified as the most important predictors for injury severity at four-leg roundabouts. For the econometric model, based on marginal effects, careless driving, passenger cars, and male drivers have the largest effect on injury severity outcomes at three-leg roundabouts. For four-leg roundabouts, passenger cars, following too closely, and drivers aged 22-35 years have the largest effect on injury severity outcomes, according to marginal effects.

In summary, when accurately predicting outcomes is a primary goal, machine learning (SVM in the current study) is advantageous over traditional econometric methods.

Third Manuscript: To test the effect of the driver gender and age, the data was split into four categories: youth female driver <25 with 831 observations, adult female driver ≥ 25 with 5020 observations, youth male driver <25 with 910 crashes, and 3491 crashes for adult male driver ≥ 25. The results indicate that with 99% confidence suggests that crashes at the roundabouts need to be modeled separately according to the driving age and gender. That data has been categorized into different groups like temporal characteristics, spatial characteristics, roadway characteristics, driver action and contribution, collision types, junction relationship, and restraint system type. Overall, there were 21 variables that have a random parameter that is normally distributed. With regard to driver injury severity estimation results, wide ranges of variables were found to increase the likelihood of getting injured in roundabout crashes, including crashes during the weekday, two-lane roundabout, sideswipe, drivers under the influence of alcohol, hit ped cycle, hit a motorcycle, and entering an angle and so on.

5.3 Discussion of Study Findings

Random parameters binary probit models were estimated based on two severity outcomes (no injury and injury) with many variables found to be statistically significant, where various variables were found to have estimated random parameters. In each individual model, a number of variables are homogenous across crash observations (i.e., their estimated parameters are fixed across observations), and various variables are heterogeneous across crash observations (i.e., they have estimated random parameters). With the collected data, some of the many factors affecting the likelihood of a crash and the resulting injury severity are likely to be unavailable to the analyst. These unobservable factors, or unobserved heterogeneity, can introduce variation into the model, impacting crash likelihood and injury severity (Mannering et al. 2016). For instance, consider gender as an observed human element that affects injury severity outcomes. However, there are clear physiological differences between men and women, as well as many variations across people of the same gender (for instance, differences in height, weight, bone density, etc.). These unobservable can result in unobserved heterogeneity, and if not accounted for, can result in biased parameter estimates Castro *et al.* (2013); Venkataraman *et al.* (2013); and Venkataraman *et al.* (2014). An additional layer of heterogeneity has been added that is associated with the mean of the distribution of the estimated random parameter, in other words, allowing the random with heterogeneity in means developed parameter to vary by the explanatory variables, which improves overall model fit and allow critical new insights.

Regarding the crash prediction, creating a model that most accurately predicts an outcome is the primary goal; machine-learning algorithms can be advantageous over traditional econometric methods. Such methods can be employed to help confront issues of multiple and correlated predictors and non-linear relationships. However, when using machine-learning methods, extra care is needed in the form of model validation.

5.4 Study Application

This study provides useful insights and an increased understanding of the factors that contribute to either sustaining injury or not in crashes at roundabouts through a random parameters approach. Although the results of this study are exploratory, they provide evidence that crashes are occurring at roundabouts and several factors lead to crashes that result in an injury. In addition, the modeling approach offers a methodology that can account for unobservable crash data. The findings of this research underscore the importance of fully accounting for unobserved heterogeneity by considering possible heterogeneity in the means of parameters. With the growing and importance relating to roundabout safety, this paper provides some essential initial findings with Washington data, but also hopefully can provide some guidance for the analysis of other roundabouts-crash databases from other geographic locations and time periods.

Several low-cost mitigation measures can reduce the number of crashes at roundabouts. First, improving pavement marking and signage to guide the motorist better and enhance driver expectancy. Furthermore, educating the public, including public-private partnerships between law enforcement agencies, driver's education instructors, transportation engineering groups, and insurance companies.

5.5 Limitation and Future work

Although this study thoroughly investigated crash injury severity for roundabouts with a different configuration, there were some limitations. For example, due to the data limitations, crash injury severity was categorized into only two levels, injury and no-injury. Also, there was information missing for several factors that could have been important, such as the geometric design of the roundabout, the exact location of the crashes, the presence of a work zone or not, and route numbers. In future work, additional crash-specific variables are recommended to investigate roundabout injury severity, such as the specific location of the crash or additional geometric design details. In doing so, an injury severity picture with a higher resolution can be obtained, which in turn can offer more understanding of the design-related factors that lead to severe crashes at roundabouts. Also, with more crash data in the future, outlook studies could classify the outcomes into more levels and may focus on identifying new significant factors that may lead to more detailed classifications of injury severity. Until then, the application of different machine-learning techniques can handle the small ratio of specific outcomes with the existing data.

References

Castro, M., Paleti, R., Bhat, C.R., 2013. A spatial generalized ordered response model to examine highway crash injury severity. Accid. Anal. Prev. 52, 188–203.

Chang, L.Y., Mannering, F., 1999. Analysis of injury severity and vehicle occupancy in truck- and non-truck-involved accidents. Accid. Anal. Prev. 31, 579–592.

Mannering, F.L., Bhat, C.R., 2014. Analytic methods in accident research: Methodological frontier and future directions. Anal. Methods Accid. Res. 1, 1–22.

Mannering, F.L., Shankar, V., Bhat, C.R., 2016. Unobserved heterogeneity and the statistical analysis of highway accident data. Anal. Methods Accid. Res. 11, 1–16.

Montella, A., 2011. Identifying crash contributory factors at urban roundabouts and using association rules to explore their relationships to different crash types. Accid. Anal. Prev. 43, 1451–1463.

Venkataraman, N., Shankar, V., Ulfarsson, G.F., Deptuch, D., 2014. A heterogeneity-in-means count model for evaluating the effects of interchange type on heterogeneous influences of interstate geometrics on crash frequencies. Anal. Methods Accid. Res. 2, 12–20.

Venkataraman, N., Ulfarsson, G.F., Shankar, V.N., 2013. Random parameter models of interstate crash frequencies by severity, number of vehicles involved, collision and location type. Accid. Anal. Prev. 59, 309–318.

Chapter 6: References

AARP, and WALC. (2014). "Modern Roundabouts." (August 2000).

Abdel-Aty, M., and Haleem, K. (2011). "Analyzing angle crashes at unsignalized intersections using machine learning techniques." *Accident Analysis and Prevention*, Elsevier Ltd, 43(1), 461–470.

Adebisi, A., Ma, J., Masaki, J., and Sobanjo, J. (2019). "safety Age-Related Di ff erences in Motor-Vehicle Crash Severity in California." 1–17.

Aghayan, I., Hosseinlou, M. H., and Kunt, M. M. (2015). "Application of Support Vector Machine for Crash Injury Severity Prediction : A Model Comparison Approach." *Journal of Civil Engineering and Urbanism*, 5(5), 193–199.

Ahmadi, A., Jahangiri, A., Berardi, V., and Machiani, S. G. (2018). "Crash severity analysis of rear-end crashes in California using statistical and machine learning classification methods." *Journal of Transportation Safety and Security*, (February 2019).

Ahmadi, A., Jahangiri, A., Berardi, V., and Machiani, S. G. (2020). "Crash severity analysis of rear-end crashes in California using statistical and machine learning classification methods." *Journal of Transportation Safety and Security*, 12(4), 522–546.

Ahmed, M. M., and Abdel-Aty, M. A. (2012). "The viability of using automatic vehicle identification data for real-time crash prediction." *IEEE Transactions on Intelligent Transportation Systems*, 13(2), 459–468.

Al-Bdairi, N. S. S., Behnood, A., and Hernandez, S. (2020). "Temporal stability of driver injury severities in animal-vehicle collisions: A random parameters with heterogeneity in means (and variances) approach." *Analytic Methods in Accident Research*, 26(September 2019), 1–52.

Al-Bdairi, N. S. S., and Hernandez, S. (2017). "An empirical analysis of run-off-road injury severity crashes involving large trucks." *Accident Analysis and Prevention*, 102, 93–100.

Al-Bdairi, N. S. S., Hernandez, S., and Anderson, J. (2018). "Contributing Factors to Run-Off-Road Crashes Involving Large Trucks under Lighted and Dark Conditions." *Journal of Transportation Engineering, Part A: Systems*, 144(1), 1–9.

Al-Ghandour, M. N., Schroeder, B. J., Williams, B. M., and Rasdorf, W. J. (2011). "Conflict Models for Single-Lane Roundabout Slip Lanes from Microsimulation." *Transportation Research Record: Journal of the Transportation Research Board*.

Al-Marafi, M. N., Somasundaraswaran, K., and Ayers, R. (2019). "Developing crash modification factors for roundabouts using a cross-sectional method." *Journal of Traffic and Transportation Engineering (English Edition)*, Elsevier Ltd, (xxx).

Al-Saleh, K., and Bendak, S. (2012). "Drivers' behaviour at roundabouts in Riyadh." *International Journal of Injury Control and Safety Promotion*, 19(1), 19–25.

Al-Thaifani, A. A., Al-Rabeei, N. A., and Dallak, A. M. (2016a). "Study of the Injured Persons and the Injury Pattern in Road Traffic Accident in Sana'a City, Yemen." *Advances in Public Health*, 2016, 1–5.

Al-Thaifani, A. A., Al-Rabeei, N. A., and Dallak, A. M. (2016b). "Study of the Injured Persons and the Injury Pattern in Road Traffic Accident in Sana'a City, Yemen." *Advances in Public Health*, 2016, 1–5.

AlKheder, S., Al-Rukaibi, F., and Al-Faresi, A. (2020). "Driver behavior at Kuwait roundabouts and its performance evaluation." *IATSS Research*.

Alkheder, S., Taamneh, M., and Taamneh, S. (2017). "Severity Prediction of Traffic Accident Using an Artificial Neural Network." *Journal of forecasting*, 36(1), 100–108.

Ambros, J., Novak, J., Borsos, A., Hoz, E., Kiec, M., MacHcinik, ??tefan, and Ondrejka, R. (2016). "Central European Comparative Study of Traffic Safety on Roundabouts." *Transportation Research Procedia*, 14, 4200–4208.

Amoros, E., Martin, J. L., and Laumon, B. (2006). "Under-reporting of road crash casualties in France." *Accident Analysis and Prevention*, 38(4), 627–635.

Anastasopoulos, P. C. (2016). "Random parameters multivariate tobit and zero-inflated count data models: Addressing unobserved and zero-state heterogeneity in accident injury-severity rate and frequency analysis." *Analytic Methods in Accident Research*, Elsevier, 11, 17–32.

Anderson, J., and Hernandez, S. (2017a). "Roadway classifications and the accident injury severities of heavy-vehicle drivers." *Analytic Methods in Accident Research*, Elsevier Ltd, 15, 17–28.

Anderson, J., and Hernandez, S. (2017b). "Analytic Methods in Accident Research Roadway classifications and the accident injury severities of heavy-vehicle drivers." *Analytic Methods in Accident Research*, Elsevier Ltd, 15, 17–28.

Anjana, S., and L. R. Anjaneyulu, M. V. (2015). "Development of Safety Performance Measures for Urban Roundabouts in India." *Journal of Transportation Engineering*, 141(1), 04014066.

Bahmankhah, B., Fernandes, P., Teixeira, J., and Coelho, M. C. (2019). "Interaction between motor vehicles and bicycles at two-lane roundabouts: a driving volatility-based analysis." *International Journal of Injury Control and Safety Promotion*, Taylor & Francis, 26(3), 205–215.

Baker, B. D. (2020). *The Future of Roundabouts in Northeast Ohio*.

Balado, J., Díaz-Vilariño, L., Arias, P., and Novo, A. (2019). "A safety analysis of roundabouts and turbo roundabouts based on Petri nets." *Traffic Injury Prevention*, Taylor & Francis, 20(4), 400–405.

Bastos, S. A., DA Maia, S. Á. J., and DA Silva, J. P. C. (2006). "Characterization of Trajectories Adopted at Roundabout Crossings." *Proceedings of the Eurpean Transport Conference (ETC)*.

Bédard, M., Guyatt, G. H., Stones, M. J., and Hirdes, J. P. (2002). "The independent contribution of driver, crash, and vehicle characteristics to driver fatalities." *Accident Analysis and Prevention*, 34(6), 717–727.

Behnood, A., and Mannering, F. (2017a). "The effect of passengers on driver-injury severities in single-vehicle crashes: A random parameters heterogeneity-in-means approach." *Analytic Methods in Accident Research*, Elsevier Ltd, 14, 41–53.

Behnood, A., and Mannering, F. (2017b). "Determinants of bicyclist injury severities in bicycle-vehicle crashes: A random parameters approach with heterogeneity in means and variances." *Analytic Methods in Accident Research*, Elsevier Ltd, 16, 35–47.

Behnood, A., and Mannering, F. (2019). "Time-of-day variations and temporal instability of factors affecting injury severities in large-truck crashes." *Analytic Methods in Accident Research*, Elsevier Ltd, 23, 100102.

Behnood, A., and Mannering, F. L. (2017c). "The effects of drug and alcohol consumption on driver injury severities in single-vehicle crashes."

Behnood, A., Roshandeh, A. M., and Mannering, F. L. (2014). "Latent class analysis of the effects of age, gender, and alcohol consumption on driver-injury severities." *Analytic Methods in Accident Research*, Elsevier, 3–4, 56–91.

Bener, A., Yildirim, E., Özkan, T., and Lajunen, T. (2017). "Driver sleepiness, fatigue, careless behavior and risk of motor vehicle crash and injury: Population based case and control study." *Journal of Traffic and Transportation Engineering (English Edition)*, 4(5), 496–502.

Bhat, C. R. (2003). "Simulation estimation of mixed discrete choice models using randomized and scrambled Halton sequences." *Transportation Research Part B: Methodological*, 37(9), 837–855.

Boser, E., Vapnik, N., Guyon, I. M., and Laboratories, T. B. (1992). "A training algorithm for optimal margin classifiers." *ACM New York*, 144–152.

De Brabander, B., Nuyts, E., and Vereeck, L. (2005). "Road safety effects of roundabouts in Flanders." *Journal of Safety Research*, 36(3), 289–296.

Burdett, B., Alsghan, I., Chiu, L.-H., Bill, A. R., and Noyce, D. A. (2016). "Analysis of Rear-End Collisions at Roundabout Approaches." *Transportation Research Record: Journal of the Transportation Research Board*.

Burdett, B., Bill, A. R., and Noyce, D. A. (2017). "Evaluation of Roundabout-Related Single-Vehicle Crashes." *Transportation Research Record: Journal of the Transportation Research Board*.

Cameron, A. C., and Trivedi, P. K. (2005). *Microeconometrics: Methods and Applications*. Cambridge University Press, New York, NY.

Campisi, T., Tibljaš, A. D.-, Tesoriere, G., Canale, A., Rencelj, M., and Šurdonja, S. (2020). "Cycling traffic at turbo roundabouts: some considerations related to cyclist mobility and safety." *Transportation Research Procedia*, 45, 627–634.

Castro, M., Paleti, R., and Bhat, C. R. (2013). "A spatial generalized ordered response model to examine highway crash injury severity." *Accident Analysis and Prevention*, Elsevier Ltd, 52, 188–203.

Cerwick, D. M., Gkritza, K., Shaheed, M. S., and Hans, Z. (2014). "A comparison of the mixed logit and latent class methods for crash severity analysis." *Analytic Methods in Accident Research*, Elsevier, 3–4, 11–27.

Chang, L. Y., and Mannering, F. (1999). "Analysis of injury severity and vehicle occupancy in truck- and non-truck-involved accidents." *Accident Analysis and Prevention*, 31(5), 579–592.

Chen, C., Zhang, G., Qian, Z., Tarefder, R. A., and Tian, Z. (2016). "Investigating driver injury severity patterns in rollover crashes using support vector machine models." *Accident Analysis and Prevention*, Elsevier Ltd, 90, 128–139.

Chen, C., Zhang, G., Tarefder, R., Ma, J., Wei, H., and Guan, H. (2015). "A multinomial logit model-Bayesian network hybrid approach for driver injury severity analyses in rear-end crashes." *Accident Analysis and Prevention*, 80, 76–88.

Chen, Y., Persaud, B., Sacchi, E., and Bassani, M. (2013). "Investigation of models for relating roundabout safety to predicted speed." *Accident Analysis and Prevention*, Elsevier Ltd, 50, 196–203.

Chen, Y., Zheng, X., Peterson, A., and Ahmadian, M. (2020). *Simulation Evaluation on the Rollover Propensity of Multi-Trailer Trucks at Roundabouts*. SAE

Technical Paper.

Chinmoy, P., Tomosaburo, O., and Vimalathithan, K. (2016). “Factors Influencing Specificity and Sensitivity of Injury Severity Prediction (ISP) Algorithm for AACN.” *International Journal of Automotive Engineering*, 7, 15–22.

Chiu, L. (2014). “the Development of Safety Performance.”

Coelho, M. C., Farias, T. L., and Rouphail, N. M. (2006). “Effect of roundabout operations on pollutant emissions.” *Transportation Research Part D: Transport and Environment*, Pergamon, 11(5), 333–343.

Daniels, S., Brijs, T., Nuyts, E., and Wets, G. (2010a). “Externality of risk and crash severity at roundabouts.” *Accident Analysis and Prevention*, Elsevier Ltd, 42(6), 1966–1973.

Daniels, S., Brijs, T., Nuyts, E., and Wets, G. (2010b). “Explaining variation in safety performance of roundabouts.” *Accident Analysis and Prevention*, 42(2), 393–402.

Daniels, S., Brijs, T., Nuyts, E., and Wets, G. (2011). “Extended prediction models for crashes at roundabouts.” *Safety Science*, Elsevier Ltd, 49(2), 198–207.

Daniels, S., Nuyts, E., and Wets, G. (2008). “The effects of roundabouts on traffic safety for bicyclists: An observational study.” *Accident Analysis and Prevention*, 40(2), 518–526.

Das, S., Datta, S., Zubaidi, H. A., and Obaid, I. A. (2021). “Applying interpretable machine learning to classify tree and utility pole related crash injury types.” *IATSS Research*.

Day, M. R., Thompson, A. R., Poulter, D. R., Stride, C. B., and Rowe, R. (2018). “Why do drivers become safer over the first three months of driving? A longitudinal qualitative study.” *Accident Analysis and Prevention*, Elsevier, 117(July 2017), 225–231.

Delen, D., Tomak, L., Topuz, K., and Eryarsoy, E. (2017). “Investigating injury severity risk factors in automobile crashes with predictive analytics and sensitivity analysis methods.” *Journal of Transport and Health*, Elsevier Ltd, 4, 118–131.

Dixon, K. K. (2012). “Developing Safety Performance Functions for 4-leg Single-lane Roundabouts Based on Oregon Data: A Case Study.”

Dixon, K. K., and Zheng, J. (2013). *Developing safety performance measures for roundabout applications in the State of Oregon*.

Echab, H., Lakouari, N., Ez-Zahraouy, H., and Benyoussef, A. (2016). “Simulation study of traffic car accidents at a single lane roundabout.” *International Journal of Modern Physics C*.

Eluru, N., and Yasmin, S. (2015). "A note on generalized ordered outcome models." *Analytic Methods in Accident Research,* Elsevier, 8, 1–6.

FHWA. (2015). "A Review of Fatal and Severe Injury Crashes at Roundabouts." IV(September).

Flannery, A. (2001). "Geometric Design and Safety Aspects of Roundabouts." *Transportation Research Record: Journal of the Transportation Research Board*, Transportation Research Board of the National Academies , 1751, 76–81.

Ghanim, M., Kharbeche, M., Hannun, J., Hannun, J., and Shamiyeh, K. (2020). "Safety and Operational Performance of Signalized Roundabouts: A Case Study in Doha." *Procedia Computer Science,* Elsevier B.V., 170, 427–433.

Gkritza, K., and Mannering, F. L. (2008). "Mixed logit analysis of safety-belt use in single- and multi-occupant vehicles." *Accident Analysis and Prevention,* 40(2), 443–451.

Goldstein, B. A., Navar, A. M., and Carter, R. E. (2017). "Moving beyond regression techniques in cardiovascular risk prediction: Applying machine learning to address analytic challenges." *European Heart Journal*, 38(23), 1805–1814.

Greene, W. H. (2012). *Econometric Modeling Guide*.

Grivna, M., Eid, H. O., and Abu-zidan, F. M. (2017). "Youth traffic-related injuries : a prospective study." *World Journal of Emergency Surgery,* World Journal of Emergency Surgery, 1–7.

Gross, F., Lyon, C., Persaud, B., and Srinivasan, R. (2013). "Safety effectiveness of converting signalized intersections to roundabouts." *Accident Analysis and Prevention*, Elsevier Ltd, 50, 234–241.

Haleem, K., and Abdel-Aty, M. (2010). "Examining traffic crash injury severity at unsignalized intersections." *Journal of Safety Research*, Elsevier Ltd, 41(4), 347–357.

Haleem, K., and Gan, A. (2013). "Effect of driver's age and side of impact on crash severity along urban freeways: A mixed logit approach." *Journal of Safety Research*, National Safety Council and Elsevier Ltd, 46, 67–76.

Halton, J. H. (1960). "On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals." *Numerishe Mathematik*, 2, 84–90.

Hasan, S., Ukkusuri, S., Gladwin, H., and Murray-Tuite, P. (2011). "Behavioral Model to Understand Household-Level Hurricane Evacuation Decision Making." *Journal of Transportation Engineering*, 137(5), 341–348.

Heqimi, G., Gates, T. J., and Kay, J. J. (2018). "Using spatial interpolation to determine impacts of annual snowfall on traffic crashes for limited access freeway segments." *Accident Analysis and Prevention*, Elsevier, 121(February), 202–212.

Iranitalab, A., and Khattak, A. (2017a). "Comparison of four statistical and machine learning methods for crash severity prediction." *Accident Analysis and Prevention*, Elsevier, 108(August), 27–36.

Iranitalab, A., and Khattak, A. (2017b). "Comparison of four statistical and machine learning methods for crash severity prediction." *Accident Analysis and Prevention*, Elsevier, 108(February), 27–36.

Isebrands, H., and Hallmark, S. (2012). "Statistical Analysis and Development of Crash Prediction Model for Roundabouts on High-Speed Rural Roadways." (2312), 3–13.

Islam, M. Bin, and Hernandez, S. (2013a). "Modeling Injury Outcomes of Crashes Involving Heavy Vehicles on Texas Highways." *Transportation Research Record: Journal of the Transportation Research Board*, 2388, 28–36.

Islam, M., and Hernandez, S. (2013b). "Large Truck–Involved Crashes: Exploratory Injury Severity Analysis." *Journal of Transportation Engineering*, 139(6), 596–604.

Islam, M., and Hernandez, S. (2013c). "Large truck-involved crashes: Exploratory injury severity analysis." *Journal of Transportation Engineering*, 139(6), 596–604.

Jeong, H., Jang, Y., Bowman, P. J., and Masoud, N. (2018). "Classification of motor vehicle crash injury severity: A hybrid approach for imbalanced data." *Accident Analysis and Prevention*, 120(August), 250–261.

Kamla, J., Parry, T., and Dawson, A. (2016). "Roundabout Accident Prediction Model." *Transportation Research Record: Journal of the Transportation Research Board*, 2585(2), 11–19.

Kashani, A. T., and Mohaymany, A. S. (2011). "Analysis of the traffic injury severity on two-lane, two-way rural roads based on classification tree models." *Safety Science*, Elsevier Ltd, 49(10), 1314–1320.

Kim, J.-K., Ulfarsson, G. F., Kim, S., and Shankar, V. N. (2013). "Driver-injury severity in single-vehicle crashes in California: A mixed logit analysis of heterogeneity due to age and gender." *Accident; analysis and prevention*, Elsevier Ltd, 50, 1073–81.

Kim, S., and Choi, J. (2013). "Safety analysis of roundabout designs based on geometric and speed characteristics." *KSCE Journal of Civil Engineering*, 17(6), 1446–1454.

Landolsi, J., Rehimi, F., and Kalboussi, A. (2015). "Macroscopic traffic modeling on a network: Application to a Roundabout." *GSCIT 2015 - Global Summit on Computer and Information Technology - Proceedings*.

Leidman, E., Maliniak, M., Sultan, A.-S. S., Hassan, A., Hussain, S. J., and Bilukha, O. O. (2016). "Road traffic fatalities in selected governorates of Iraq from 2010 to 2013: prospective surveillance." *Conflict and health*, Conflict and Health, 10, 2.

Li, L., Kim, K., and Nitz, L. (1999). "Predictors of safety belt use among crash-involved drivers and front seat passengers: Adjusting for over-reporting." *Accident Analysis and Prevention*, 31(6), 631–638.

Li, Y., Ma, D., Zhu, M., Zeng, Z., and Wang, Y. (2018a). "Identification of significant factors in fatal-injury highway crashes using genetic algorithm and neural network." *Accident Analysis and Prevention*, Elsevier, 111(December 2017), 354–363.

Li, Y., Ma, D., Zhu, M., Zeng, Z., and Wang, Y. (2018b). "Identification of significant factors in fatal-injury highway crashes using genetic algorithm and neural network." *Accident Analysis and Prevention*, Elsevier, 111(September 2017), 354–363.

Li, Z., Liu, P., Wang, W., and Xu, C. (2012). "Using support vector machine models for crash injury severity analysis." *Accident Analysis and Prevention*, Elsevier Ltd, 45, 478–486.

Liu, Y., Li, Z., Liu, J., and Patel, H. (2016). "Data in Brief Vehicular crash data used to rank intersections by injury crash frequency and severity." *Data in Brief*, Elsevier, 8, 930–933.

Lombardi, D. A., Horrey, W. J., and Courtney, T. K. (2017). "Age-related differences in fatal intersection crashes in the United States." *Accident Analysis and Prevention*, Elsevier Ltd, 99, 20–29.

Lord, D., and Mannering, F. (2010). "The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives." *Transportation Research Part A: Policy and Practice*, Elsevier Ltd, 44(5), 291–305.

Mafi, S., AbdelRazig, Y., and Doczy, R. (2018). "Machine Learning Methods to Analyze Injury Severity of Drivers from Different Age and Gender Groups." *Transportation Research Record*, 2672(38), 171–183.

Malliaris, A., DeBlois, J., and Digges, K. (1996). "Light vehicle occupant ejections--a comprehensive investigation." *Accident analysis and prevention*, 28(1), 1–14.

Mandavilli, S., McCartt, A. T., and Retting, R. A. (2009). "Crash patterns and potential engineering countermeasures at maryland roundabouts." *Traffic Injury Prevention*, 10(1), 44–50.

Mannering, F. L., and Bhat, C. R. (2014). "Analytic methods in accident research: Methodological frontier and future directions." *Analytic Methods in Accident Research*, Elsevier, 1, 1–22.

Mannering, F. L., Shankar, V., and Bhat, C. R. (2016). "Unobserved heterogeneity and the statistical analysis of highway accident data." *Analytic Methods in Accident Research*, Elsevier, 11, 1–16.

Martin, J. (2002). "Relationship between crash rate and hourly traffic flow on interurban motorways." 34, 619–629.

Milton, J. C., Shankar, V. N., and Mannering, F. L. (2008). "Highway accident severities and the mixed logit model: An exploratory empirical analysis." *Accident Analysis and Prevention*, 40(1), 260–266.

Montella, A. (2007). "Roundabout In-Service Safety Reviews: Safety Assessment Procedure." *Transportation Research Record: Journal of the Transportation Research Board*, Transportation Research Board of the National Academies , 2019, 40–50.

Montella, A. (2011). "Identifying crash contributory factors at urban roundabouts and using association rules to explore their relationships to different crash types." *Accident Analysis and Prevention*, Elsevier Ltd, 43(4), 1451–1463.

Moore, D. N., Schneider IV, W. H., Savolainen, P. T., and Farzaneh, M. (2011). "Mixed logit analysis of bicyclist injury severity resulting from motor vehicle crashes at intersection and non-intersection locations." *Accident Analysis and Prevention*, Elsevier Ltd, 43(3), 621–630.

NHTSA. (2009). *TRAFFIC SAFETY FACTS 2009*. *National Highway Traffic Safety Administration*.

Nikitin, N., Patskan, V., and Savina, I. (2017). "Efficiency Analysis of Roundabout with Traffic Signals." *Transportation Research Procedia*.

Pahukula, J., Hernandez, S., and Unnikrishnan, A. (2015). "A time of day analysis of crashes involving large trucks in urban areas." *Accident Analysis and Prevention*, Elsevier Ltd, 75, 155–163.

Pal, C., Okabe, T., Kulothungan, V., Sangolla, N., Manoharan, J., Stewart, W., and Combest, J. (2016). "Factors influencing specificity and sensitivity of injury severity prediction (ISP) algorithm for AACN." *International Journal of Automotive Engineering*, 7(1), 15–22.

Patnaik, A. K., Agarwal, L. A., Panda, M., and Bhuyan, P. K. (2020). “Entry capacity modelling of signalized roundabouts under heterogeneous traffic conditions.” *Transportation Letters,* Taylor & Francis, 12(2), 100–112.

Payyanadan, R. P., Lee, J. D., and Grepo, L. C. (2018). “Challenges for older drivers in Urban, Suburban, and rural settings.” *Geriatrics (Switzerland),* 3(2).

Persaud, B., Retting, R., Garder, P., and Lord, D. (2001). “Safety Effect of Roundabout Conversions in the United States: Empirical Bayes Observational Before-After Study.” *Transportation Research Record: Journal of the Transportation Research Board*, Transportation Research Board of the National Academies , 1751, 1–8.

Polders, E., Daniels, S., Casters, W., and Brijs, T. (2015). “Identifying crash patterns on roundabouts.” *Traffic injury prevention*, 16(2), 202–7.

Pratelli, A. (2006). “Design of modern roundabouts in urban traffic systems.” *Urban Transport XII: Urban Transport and the Environment in the 21st Century*, WIT Transactions on The Built Environment, Vol 89, WIT Press, Southampton, UK, 83–93.

Pratelli, A., Lupi, M., Pratelli, C., and Farina, A. (2020). “Mini-roundabouts for Improving Urban Accessibility.” *Modelling of the Interaction of the Different Vehicles and Various Transport Modes,* Springer, 333–382.

Pratic??, F. G., Vaiana, R., and Gallelli, V. (2015). “Micro-simulation effectiveness in predicting operating speed profiles in a roundabout.” *Advances in Transportation Studies*.

Qin, X., Bill, A., Chitturi, M., and Noyce, D. A. (2013). “Evaluation of roundabout safety.” *Transportation Research Board 92nd Annual Meeting*.

Qin, X., Khan, G., Bill, A., and Noyce, D. A. (2011a). “Comprehensive Safety Evaluation of Roundabouts in Wisconsin.” *Journal of Transportation Safety & Security*, 3(4), 289–303.

Qin, X., Khan, G., Bill, A., and Noyce, D. A. (2011b). “Comprehensive Safety Evaluation of Roundabouts in Wisconsin.” *Journal of Transportation Safety & Security*, Taylor & Francis Group , 3(4), 289–303.

Rezapour, M., Mehrara Molan, A., and Ksaibati, K. (2020). “Analyzing injury severity of motorcycle at-fault crashes using machine learning techniques, decision tree and logistic regression models.” *International Journal of Transportation Science and Technology*, Tongji University and Tongji University Press, 9(2), 89–99.

Robinson, B., Blogg, M., and Waltman, E. (2004). “MODERN ROUNDABOUTS AND OTHER INTERSECTION TYPES.” 1–55.

Robinson, B., Rodegerdts, L. A., Troutbeck, R., Brilon, W., Bondizio, L., Courage, K., Kyte, M., Mason, J., Flannery, A., Myers, E., Bunker, J., and Jacquemart, G. (2000). *Roundabouts : An Informational Guide*. *FHWA-RD-00-067*.

Rodegerdts, L. A., Malinge, A., Marnell, P. S., Beaird, S. G., Kittelson, M. J., and Mereszczak, Y. S. (2015). “Assessment of Roundabout Capacity Models for the Highway Capacity Manual.” II(September), 94.

Romo, A., Hernandez, S., and Cheu, R. L. (2014). “Identifying Precrash Factors for Cars and Trucks on Interstate Highways: Mixed Logit Model Approach.” *Journal of Transportation Engineering*, 140(3), 04013016.

Sadeq, H., and Sayed, T. (2016). “Automated Roundabout Safety Analysis: Diagnosis and Remedy of Safety Problems.” *Journal of Transportation Engineering*.

Sameen, M. I., and Pradhan, B. (2017a). “Severity prediction of traffic accidents with recurrent neural networks.” *Applied Sciences (Switzerland)*, 7(6).

Sameen, M., and Pradhan, B. (2017b). “Severity Prediction of Traffic Accidents with Recurrent Neural Networks.” *Applied Sciences*, 7(6), 476.

Savolainen, P. T., Mannering, F. L., Lord, D., and Quddus, M. A. (2011). “The statistical analysis of highway crash-injury severities : A review and assessment of methodological alternatives.” *Accident Analysis and Prevention*, Elsevier Ltd, 43(5), 1666–1676.

Seeherman, J., and Liu, Y. (2015). “Effects of extraordinary snowfall on traffic safety.” *Accident Analysis and Prevention*, 81, 194–203.

Seraneeprakarn, P., Huang, S., Shankar, V., Mannering, F., Venkataraman, N., and Milton, J. (2017). “Occupant injury severities in hybrid-vehicle involved crashes: A random parameters approach with heterogeneity in means and variances.” *Analytic Methods in Accident Research*, Elsevier Ltd, 15, 41–55.

Shaaban, K., and Hamad, H. (2020). “Critical gap comparison between one-, two-, and three-lane roundabouts in Qatar.” *Sustainability (Switzerland)*, 12(10), 1–14.

Shen, J., Wang, T., Zheng, C., and Yu, M. (2020). “Determinants of Bicyclist Injury Severity Resulting from Crashes at Roundabouts, Crossroads, and T-Junctions.” *Journal of Advanced Transportation*, Hindawi, 2020.

Siddiqui, C., Abdel-Aty, M., and Huang, H. (2012). “Aggregate nonparametric safety analysis of traffic zones.” *Accident Analysis and Prevention*, Elsevier Ltd, 45, 317–325.

Sisiopiku, V. P., and Oh, H.-U. (2001). “Evaluation of roundabout performance using SIDRA.” *Journal of Transportation Engineering*.

Somasundaraswaran, K., and Richardson, M. (2019). “Investigating the Possibilities of Specific Crash Type at Roundabouts.” *Journal of Traffic and Transportation Engineering*, 7(5), 206–211.

Stewart, J. R. (1993). “Seat belt use and accident involvment: A comparison of driving behavior before and after a seat belt law.” *Accident Analysis & Prevention*, 25(6), 757–763.

Storm, H., Baylis, K., and Heckelei, T. (2019). “Machine learning in agricultural and applied economics.” *European Review of Agricultural Economics*, 1–44.

Streff, M. F., and Wangenaar, C. A. (1989). “Are there really shortcut? estimating seat belt use with self-report measures.” *Accident Analysis & Prevention*, 21(6), 509–516.

Taekratok, T. (1998). “Modern roundabouts for Oregon.” *Oregon Department of Transportation Research Unit*, (June).

Tang, J., Liang, J., Han, C., Li, Z., and Huang, H. (2019a). “Crash injury severity analysis using a two-layer Stacking framework.” *Accident Analysis and Prevention*, Elsevier, 122(November 2018), 226–238.

Tang, J., Liang, J., Han, C., Li, Z., and Huang, H. (2019b). “Crash injury severity analysis using a two-layer Stacking framework.” *Accident Analysis and Prevention*, Elsevier, 122(May 2018), 226–238.

Tay, R. (2015a). “A random parameters probit model of urban and rural intersection crashes.” *Accident Analysis and Prevention*, Elsevier Ltd, 84, 38–40.

Tay, R. (2015b). “A random parameters probit model of urban and rural intersection crashes.” *Accident Analysis and Prevention*, Elsevier Ltd, 84, 38–40.

Tixier, A. J. P., Hallowell, M. R., Rajagopalan, B., and Bowman, D. (2016). “Application of machine learning to construction injury prediction.” *Automation in Construction*, Elsevier B.V., 69, 102–114.

Tollazzi, T. (2015). “First developments of different roundabout types.” *Springer Tracts on Transportation and Traffic*.

Toran, A. (n.d.). “Analysing the pedestrian crash severity at mid-blocks using classification tree models.”

Ulfarsson, G. F., and Mannering, F. L. (2004). “Differences in male and female injury severities in sport-utility vehicle, minivan, pickup and passenger car accidents.” *Accident Analysis & Prevention*, Pergamon, 36(2), 135–147.

Valdez, M., Cheu, R., and Duran, C. (2011). “Operations of Modern Roundabout with Unbalanced Approach Volumes.” *Transportation Research Record: Journal of* *the Transportation Research Board*, Transportation Research Board of the National Academies , 2265, 234–243.

Vasconcelos, L., Seco, Á., and Silva, A. (2012). “A Comparison of Roundabout Capacity Models.” *IASTED - International Association of Science and Technology for Development*.

Venkataraman, N., Shankar, V., Ulfarsson, G. F., and Deptuch, D. (2014). “A heterogeneity-in-means count model for evaluating the effects of interchange type on heterogeneous influences of interstate geometrics on crash frequencies.” *Analytic Methods in Accident Research*, Elsevier, 2, 12–20.

Venkataraman, N., Ulfarsson, G. F., and Shankar, V. N. (2013). “Random parameter models of interstate crash frequencies by severity, number of vehicles involved, collision and location type.” *Accident Analysis and Prevention*, Elsevier Ltd, 59, 309–318.

Wahab, L., and Jiang, H. (2019). “A comparative study on machine learning based algorithms for prediction of motorcycle crash severity.” *PLoS ONE*, 14(4), 1–17.

Washington, S. P., Karlaftis, M. G., and Mannering, F. L. (2010). *Statistical and Econometric Methods for Transportation Data Analysis, Second Edition*. CRC Press.

Washington, S. P., Karlaftis, M. G., and Mannering, F. L. (2011). *Statistical and Econometric Methods for Transportation Data Analysis*. *Chapman & Hall/CRC*, Chapman & Hall/CRC.

WIDOT. (2020). “Motorists’ handbook.”

de Winter, J. C. F., de Groot, S., Mulder, M., Wieringa, P. A., Dankelman, J., and Mulder, J. A. (2009). “Relationships between driving simulator performance and driving test results.” *Ergonomics*, 52(2), 137–153.

Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data*. MIT Press, Cambridge, MA.

Wu, Q., Chen, F., Zhang, G., Liu, X. C., Wang, H., and Bogus, S. M. (2014). “Mixed logit model-based driver injury severity investigations in single- and multi-vehicle crashes on rural two-lane highways.” *Accident Analysis and Prevention*, Elsevier Ltd, 72, 105–115.

Wu, Q., Zhang, G., Zhu, X., Liu, X. C., and Tarefder, R. (2016). “Analysis of driver injury severity in single-vehicle crashes on rural and urban roadways.” *Accident Analysis and Prevention*, Elsevier Ltd, 94, 35–45.

Xie, Y., Zhao, K., and Huynh, N. (2012). “Analysis of driver injury severity in rural single-vehicle crashes." *Accident Analysis and Prevention*, Elsevier Ltd, 47, 36–44.

Yap, Y. H., Gibson, H. M., and Waterson, B. J. (2013). "An International Review of Roundabout Capacity Modelling." *Transport Reviews*.

Yasmin, S., Eluru, N., and Pinjari, A. R. (2015). "Analyzing the continuum of fatal crashes: A generalized ordered approach." *Analytic Methods in Accident Research*, Elsevier, 7, 1–15.

Ye, F., and Lord, D. (2014). "Comparing three commonly used crash severity models on sample size requirements: Multinomial logit, ordered probit and mixed logit models." *Analytic Methods in Accident Research*, Elsevier, 1, 72–85.

Yu, R., and Abdel-Aty, M. (2014). "Analyzing crash injury severity for a mountainous freeway incorporating real-time traffic and weather data." *Safety Science*, Elsevier Ltd, 63, 50–56.

Zeng, Q., and Huang, H. (2014). "A stable and optimized neural network model for crash injury severity prediction." *Accident Analysis and Prevention*, Elsevier Ltd, 73, 351–358.

Zhang, J., Li, Z., Pu, Z., and Xu, C. (2018a). "Comparing prediction performance for crash injury severity among various machine learning and statistical methods." *IEEE Access*, IEEE, 6, 60079–60087.

Zhang, J., Li, Z., Pu, Z., and Xu, C. (2018b). "Comparing prediction performance for crash injury severity among various machine learning and statistical methods." *IEEE Access*, IEEE, 6, 60079–60087.

Zhao, X., Zhang, X., and Rong, J. (2014). "Study of the Effects of Alcohol on Drivers and Driving Performance on Straight Road." *Mathematical Problems in Engineering*, Hindawi, 2014, 1–9.

Ziolkowski, R. (2014). "The influence of roundabouts on drivers' speed and behaviour." *Advanced Materials Research*, 1020(Table 1), 674–679.

Zohdy, I., and Rakha, H. (2013). *Enhancing roundabout operations via vehicle connectivity*. *Transportation Research Record*.

Zubaidi, H. A., Anderson, J. C., and Hernandez, S. (2020). "Understanding roundabout safety through the application of advanced econometric techniques." *International Journal of Transportation Science and Technology*.