Fall 1-21-2014

Predicting the Severity of Highway User Crashes on Public At-Grade Highway-Rail Crossings

Elias Wondwossen Haile

PREDICTING THE SEVERITY OF HIGHWAY

USER CRASHES ON PUBLIC

AT-GRADE HIGHWAY-RAIL CROSSINGS

by

ELIAS WONDWOSSEN HAILE

A thesis submitted in partial fulfillment of the
requirements for the degree of
Master of Science in Civil Engineering
Department of Civil Engineering

Wei Fan, Ph.D., P.E, Committee Co-chair

J. Torey Nalbone, Ph.D., CIH, Committee Co-chair

College of Engineering and Computer Science

The University of Texas at Tyler
December 2013

The University of Texas at Tyler
Tyler, Texas

This is to certify that the Master's Thesis of

ELIAS WONDWOSSEN HAILE

has been approved for the thesis requirements on
November 15th, 2013
for the degree of Master of Science in Civil Engineering

Wei Fan

Thesis Co-chair: Wei Fan, Ph.D., P.E.
Department of Civil Engineering

Thesis Co-chair: J. Torey Nalbone, Ph.D., CIH
Department of Civil Engineering

Deborah Koslover

Member: Deborah Koslover, Ph.D.
Department of Mathematics

Member: Michael Gangone, Ph.D.
Department of Civil Engineering

Chair, Department of Civil Engineering: J. Torey Nalbone, Ph.D., CIH.

Dean, College of Engineering and Computer Science: James Nelson, Ph.D., P.E.

Acknowledgements

I would like to express my gratitude to several people that helped me along the way and made this thesis possible. First of all, I want to praise God for giving me the strength to finish my thesis. I want to thank the Civil Engineering department, especially Dr. Torey Nalbone, for his support of my graduate studies. I want also to thank TxDOT for financially supporting me through the first year of my graduate studies. Also, I would like to thank my advisor, Dr. Wei Fan, for his patience and wisdom. He has had a profound influence on the research presented within this thesis and has been an excellent advisor, teacher, and friend. In addition, a thank you to Dr.Michael Gangone and Dr. Diborah Koslover who had contributed to my thesis. Lastly, I want to thank my wife, Meron, for all her love and understanding over the past two years.

Contents

List of Tables

List of Figures

Abstract

PREDICTING THE SEVERITY OF HIGHWAY
USER CRASHES ON PUBLIC AT GRADE
HIGHWAY-RAIL CROSSINGS

Elias Wondwossen Haile

Thesis Co-chair: Wei Fan, Ph.D., PE.
Thesis Co-chair: J. Torey Nalbone, Ph.D., CIH

The University of Texas at Tyler
December 2013

Highway user-rail crashes have a significant effect on highway user safety rating. However, very little attention is garnished on the subject. An understanding of the factors contributing to the levels of injury severity is an important step toward making the transportation system safer and more reliable. Numerous studies have applied statistical models for crash injury severity. The main goal of this thesis is to explore the impact of various factors involved in highway user crashes on Highway-Rail at Grade Crossings (HRGCs) and provide appropriate mitigation measures. The logistic regression modeling approach (specifically ordered and unordered logit models) was applied to predict the three levels of highway user crash severity on HRGC as a function of various factors involved. A comparison was also performed between the two logit models. The explanatory variables were obtained from the USDOT crossing inventory

and HRGCs crash data. The study revealed that some variables such as type of crash circumstance type, pedestrian gender, adverse weather condition, train speed, vehicle speed, HRGC surface type, traffic volume and number of traffic lanes were found to be statistically significant factors contributing to highway user crashes on HRGC. In addition, ordered logit model were identified to be better in estimating the highway user crash severity level on HRGCs.

Chapter One

Introduction

1.1 Background and problem definition

Fatality resulting from motor vehicle crashes is the leading cause of death in the United States. Data from the national Highway Traffic Safety Administration indicates that since 1949 more than 30,000 (40,000 average) fatalities result from motor vehicle crashes every year. However, the current trend shows this number is declining. For example, a 1.9-percent decrease in fatality related crashes was observed in 2011 as compared to 2010. Nonetheless, injury related crashes are still large in number. In 2011, an estimated 2.22 million people were injured in motor vehicle traffic crashes and 2.24 million in 2010 (*1*). Fatal crashes on HRGCs contributed to 261 deaths in 2010 and 251 in 2011 (*2*).

Highway-Rail at Grade Crossings (HRGCs) are conflict points between highway users and rail equipment (e.g locomotive, freight car, caboose or service equipment car operated by a railway company) which has contributed to a considerable amount of crashes in U.S. history. There are approximately 240,000 grade crossings in the United States. Over 39 percent (94,400) are private HRGCs. HRGC conflicts include any impact between a rail and highway user (both motor vehicles and other users of the crossing) at a designated crossing site (*3*). Though the trend of highway user crashes with rail equipment is showing a decrease in numbers, much has yet to be done to improve the safety of HRGCs.

Unlike highway traffic accidents, a significantly high percentage of vehicle-rail crashes lead to fatality and injury to vehicle users. For example, data in the past eight years (2005-2012) indicates that 8.55 percent of vehicle-rail crashes were fatal and 26.68 percent resulted in injury (*2*). However, in the case of highway traffic accidents, the percentage of fatal crashes is no more than two percent. Similarly, a majority of rail-pedestrian crashes have lead to fatal and severe injury. As the Federal Railroad Administration statistics indicate, in the last eight years, a total of around 968 pedestrian crashes were reported on United States department of Transportation (USDOT) public HRGC sites, out of which 534 fatalities and 326 injuries, accounting for 55.2 percent and 33.7 percent respectively. Only 108 (11.1 percent) crashes resulted in no injury (*1*).

Despite the fact that highway user-rail crashes had a significant effect on highway user safety, the subject still receives little attention and is under-reported. An understanding of the factors contributing to the levels of injury severity is an important step toward making the transportation system safer and more attractive. Responsible jurisdictions may use the results of this research to derive road user safety measures and policies.

One of the most important task in improving road safety is to uncover influential factors and then to develop countermeasures. The relationship between the injury severity of traffic crashes and factors such as driver and passenger characteristics, pedestrian age and gender, vehicle type, environmental conditions, and traffic and geometric conditions has attracted much attention. Better understanding of this relationship is necessary and very important for improving facility design so that accidents can be reduced. It is important to note that reducing crash frequency and reducing crash-injury severity may necessitate different strategic approaches.

The development of effective countermeasures requires a thorough understanding of the factors that affect the likelihood of a crash occurring or, given that a crash has occurred, the characteristics that may mitigate or exacerbate the degree of injury sustained by crash-involved road users. To gain such an understanding, safety researchers have applied a wide variety of methodological techniques over the years.

Numerous studies have applied statistical models for crash injury severity study. Among them, the ordered probit, ordered logit and their variations are the most often used models. Savolainen et al. (*4*) briefly discussed and summarized the wide range of methodological tools applied to study the impact various factors on motor vehicle crash-injury severities. As presented in the paper, ordered logit and probit, multinomial logit, binary logit and binary probit and nested logit are some of the frequently used statistical methodologies. However, most of these researches dealt with crashes occurring among various types of road users. There are only few studies conducted that considers crashes involving rail equipment. Khattak (*5*) recently investigated the impact of different factors on crash severity levels of pedestrian crashes on HRGCs in the U.S.

Logistic regression has been widely applied to model crash severity levels. Variables such as elements of geometric design, traffic operational measures, and environmental conditions are considered as independent variables to estimate the severity. This study also applied the logistic regression modeling approach (specifically ordered and unordered logit models) to estimate the three levels of highway user crash severity on HRGC as a function of various factors involved. Comparison was performed between the two logit models. The explanatory variables were obtained from the USDOT crossing inventory and crash data.

Before considering methods for ordinal outcomes, it is important to note that simply because the values of a variable can be ordered does not imply that the variable should be analyzed as ordinal. A variable might be ordered when considered for one purpose, but be unordered or ordered differently when used for another purpose. When the proper ordering of a variable is ambiguous, the models for nominal variable should be considered in addition to the models for ordinal variables (*6*).

Modeling ordinal outcome dependent variable using nominal variable will lead to loss of efficiency as a result of ignoring information. In the reverse, modeling nominal variable using ordinal variable will give biased or sometimes irrational estimates. The loss of information in the ordinal data can be outweighed by avoiding the bias. The primary advantage of the nominal outcome multinomial logit model is its ability to avoid the parallel effect regression assumption unlike the ordered outcome regression model. Uncertainty from considering the dependent variable as ordered outcome can also be avoided by using the nominal outcome multinomial logit model (MNLM) (*6*).

As discussed, crashes occurring at HRGCs had significant effect on highway user safety and the importance of conducting research in such areas is evident. However, this subject receives less attention and little research efforts have been made in this particular area. As such, the objective of this research is to explore the impacts of various factors contributing to different levels of crash severity to vehicle users as a result of vehicle-rail crashes on HRGCs.

1.2 Objectives of the study

The main goal of this thesis is to explore the impact of various factors involved in highway user crashes on HRGCs. The tasks of this thesis can be stated as:

- To apply statistical approaches and identify major contributing factors of highway users crashes on HRGCs.
- To develop statistical models that relate crash severity levels and significant contributing factors involved in highway user crashes on HRGCs.
- To identify and provide mitigation solutions based on the results of the study.

The following specific objectives are required to achieve these aims:

1) A literature review on highway user accident statistics, HRGC accident statistics/rate, existing studies regarding crash severity modeling.
2) Identify appropriate data, select variable to be considered in the analysis and perform descriptive statistics.
3) Produce statistical models that relate crash severity levels and various factors involved.
4) Determine the marginal effect and/or elasticity of variables included in the models.
5) Statistical interpretation of the models developed.
6) Perform comparison between different models and suggest the best one.
7) Identify counter measures to mitigate the problem.

1.3 Thesis outline

This thesis is divided into six chapters. A brief outline of this thesis is given below. Following the introductory chapter, Chapter 2 contains a literature review consisting of two parts: HRGC crash statistics and existing studies on crash severity modeling. Part one briefly describe the historical and current statistics of highway user crashes on HRGCs. Part two presents the review of various past studies conducted on crash severity modeling.

Chapter 3 describes the material used and the analysis method applied in this research. In the first part of this chapter, data description and descriptive statistics of the data is presented. The second part of this chapter describes the various models applied in this research. The results obtained from the research are described in Chapter 4. The results of analysis using various models, together with model statistics are presented in this chapter.

Chapter 5 presents the discussion of the results obtained in this research. The outputs of each model are discussed, together with a discussion of model comparison. The implications of this research for HRGC safety improvement are discussed in Chapter 6. Finally, Chapter 7 presents the conclusions of this research and gives suggestions for future work.

Chapter Two

Literature Review

2.1 Introduction

This chapter is divided in to three sections: section 2.2 presents a general review of highway user crash statistics on HRGCs in the US, section 2.3 presents a review of past studies conducted on pedestrian crash severity modeling and in section 2.4 looks at past studies conducted on vehicle crash severity modeling will be discussed.

2.2 HRGC inventory and crash review

According to the Federal Railroad Administration (FRA), train related accidents are generally divided into three major groups as follows:

1. Train accidents: They are safety-related events involving on-track rail equipment (both standing and moving), causing monetary damage to the rail equipment and track above a prescribed amount.
2. Highway-rail grade crossing incidents: These include any impact between a rail and highway user (both motor vehicle and other users of the crossing) at a designated crossing site, including walkways, sidewalks, etc., associated with the crossing.
3. Other incidents: These include any death, injury or occupational illness of a railroad employee that is not the result of a "train accident" or "highway-rail incident."

Highway-Rail at Grade Crossings (HRGCs) are conflict points between highway users and rail equipments which has a considerable amount of crashes in the U.S. history.

Currently, 240,000 at-grade highway-rail crossings exist in the United States of which over 60 percent are public highway-rail grade crossings.

Public grade crossings are those under the jurisdiction of a public authority whereas private grade crossings are located on privately roadways such as farm or industrial area. Figure 1 shows the distribution of public at-grade crossings by warning device type. As can be seen from the figure, most public HRGCs are equipped with cross bucks, gates and flashing lights.

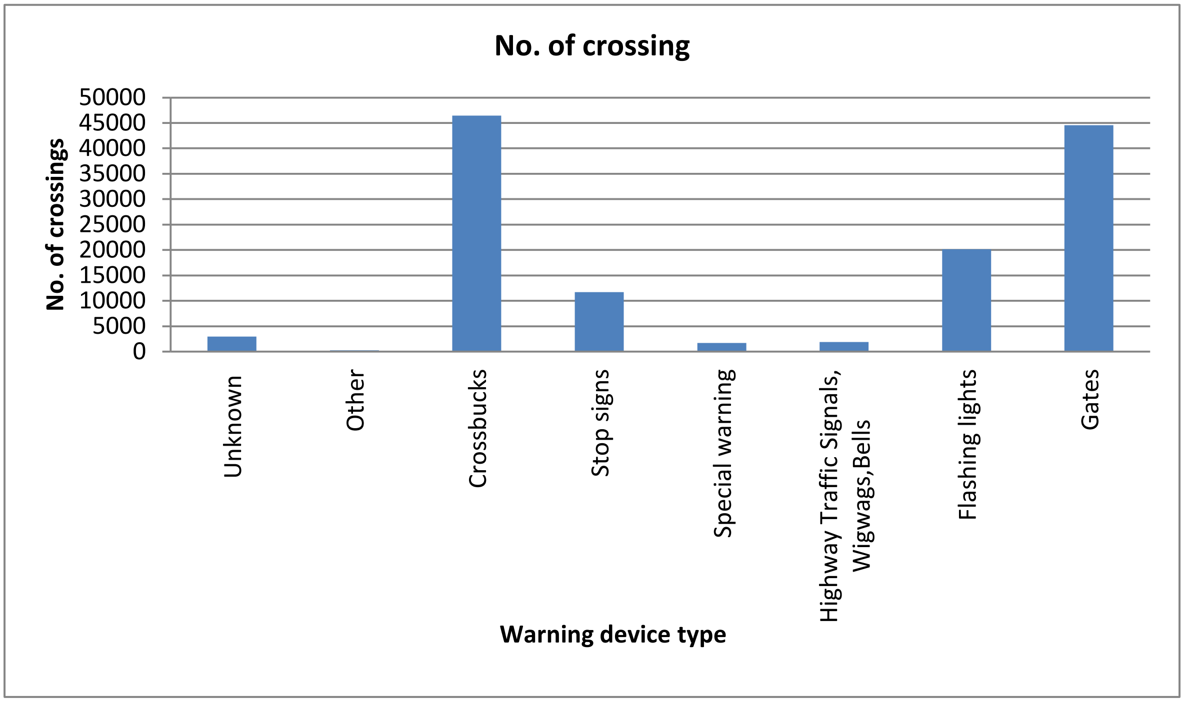

Figure 1. At-grade HRGC by Warning Device Type, 2012

Highway-rail grade crossing conflicts include any impact between a rail and highway user (both motor vehicles and other users of the crossing) at a designated crossing site (*3*). These conflicts have been decreasing in number over years as depicted in Figure 2.

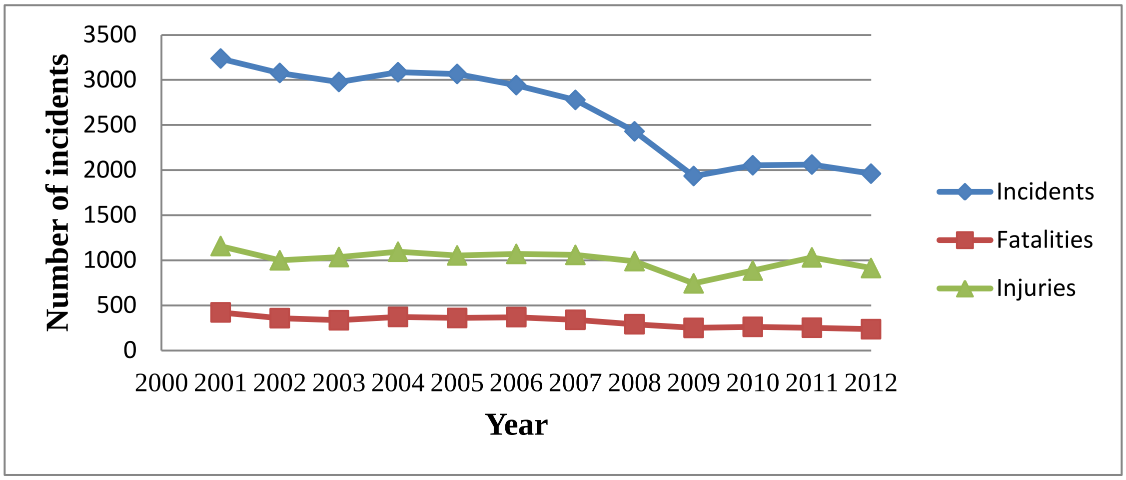

Figure 2. All Highway-Rail Incidents at Public, 2000-2012

Though the trend of highway user crashes with rail equipment is declining, there is still much to do to improve the safety of HRGCs and minimize the consequences of a crash. Unlike highway traffic accidents, a significantly high percentage of vehicle-rail crashes lead to fatality and injury to vehicle occupants. For example, data from 2005-2012 indicates that 8.55 percent of vehicle-rail crashes were fatal and 26.68 percent were resulting injury (*2*). However, in the case of highway traffic accidents, the percentage of fatal crashes is roughly two percent. Similarly, in rail- pedestrian crashes, the majority of pedestrian-rail crashes tend to lead to fatal or severe injury. The Federal Railroad Administration indicate, in this eight year period, a total of 968 pedestrian crashes were reported on USDOT public HRGC sites, of which, 534 (55.2 percent) resulted in fatality, 326 (33.7 percent) in injury and 108(11.1 percent) no injury.(*1*).

2.3 Pedestrian crash severity modeling

Several studies exist regarding pedestrian crash severity level modeling, the majority of which apply a statistical approach (*6*). Explanatory variables such as roadway condition, highway operational condition, environmental factors, pedestrian characteristics and driver characteristics are factors considered for modeling crash severity. The relation between the dependent variable and the explanatory variables can be modeled and tested using disaggregate models such as logistic regression. However, there are limited numbers of studies available which investigate the pedestrian crash severity models for pedestrian-rail crashes occurring on HRGCs. Some of the studies on pedestrian crash severity modeling are reviewed and presented in this section.

Using the ordered probit model, Khattak (*5*) investigated the impact of various factors on the three severity levels of pedestrian injuries at highway-rail grade crossings using an ordered probit modeling (OPM) technique. The three severity levels considered were no injury, injury and fatality. Crash and crossing inventory data of the USDOT at grade road-rail crossings were used in the study. The results indicate that a majority of severe injuries were related to factors such as higher train speed, rail equipment that struck pedestrians and regions which are commercially developed. Greater number of traffic lanes, HRGCs equipped with standard flash light signals and clear weather were related with lower severity levels. Lobb (*7*) conducted a review on railway pedestrian safety research regarding train-pedestrian collisions. The review indicated the need for more research to understand factors contributing to railway trespassing accidents.

Davis (*8*) developed a statistical model incorporating actual data to relate the stuck pedestrian to the speed of the striking vehicle. The model enabled the prediction of the probability distribution across categories of pedestrian injury severity for a given vehicle impact speed. The injury severity model developed for children (0-14 years) and adult groups (15-59 years) shows similar injury severity distributions while that for the elderly group (60+ years) differed. It was also reported that older pedestrians have a higher chance of suffering from severe injuries at lower speed than children and adult groups. Kwigizile et al. (*9*) conducted a study to identify factors contributing to the level of pedestrian injury severity and assess the consistency of the ordered probit and multinomial logit models. The two models were applied to the same data set and a comparison was made by calculating the marginal effects and elasticity. As indicated in the results, the two models are consistent for the lowest and highest levels of injury severity but are inconsistent in describing the impact individual factors have on intermediate injury levels. Hence, pedestrian safety measure can be derived and will be consistent by avoiding the use of intermediate outcome results. The lowest and highest injury levels are no injury or possible injury and fatal injury respectively whereas the intermediate levels of injury are non-capacitating and incapacitating injuries.

Zahabi at al. (*10*) developed an injury severity model as a function of such factors as various environment characteristics, road geometry, weather conditions, type of vehicle involved and vehicle movement to determine the impact of these variables have on pedestrian-vehicle and cyclist-vehicle crash severity. The effects of the variables were estimated using an ordered logit model. As reported in the study, motor vehicle drivers are significantly influenced by the geometry of the road while posted speed limit is not a significant factor. Collisions that occur at an intersection have a lower chance of injury or death for pedestrians as compared to collisions outside of intersections. Rosen and Sander (*11*) developed a logistic regression model to describe adult (15 years or older) pedestrian fatality risk as a function of impact speed. Effects of other variables such as pedestrian age, gender, height and weight on fatality risk were also investigated. The result of the study indicated fatality risk is strongly associated with impact speed. According to the study, the risk at 50 km/h is more than twice as high as the risk at 40 km/h and more than five times higher than the risk at 30 km/h, indicating the necessity of lowering impact speeds within city areas.

Sze and Wong (*12*) developed a binary logistic regression model and explored contributing factors to fatality and severe injury. As the study indicated, male pedestrians below age 15 in crashes that occurred in overcrowded or obstructed foot paths, in the day time and on congested road sections showed lower risk of fatality and severe injury. On the other hand, a higher risk of fatality or severe injury was noticed pedestrians above the age of 65 who were involved in a crash on the crossing and where within 15m of a crosswalk. This was particularly the case in locations where the speed limit was above 50 km/h, at signalized intersections, or where there were two or more traffic lanes. Zajac and Ivan (*13*) explored the effects of roadway and area type features on the severity of injury to pedestrian involved in crashes in a rural part of Connecticut by using an ordered probit model. The study indicated that several variables, including the clear width of the roadway, vehicle type, driver and pedestrian sobriety, and pedestrians' age of 65 and above, significantly influenced the severity of the pedestrian injury. It is also found that roads passing through villages, downtown fringe and low-density residential areas generally resulted in a more severe injury to the pedestrian.

Applying data mining techniques such as classification trees and association rules Montella et al. (*14*) conducted exploratory analysis on 56,014 pedestrian crashes that occurred in Italy from 2006-2008 to investigate interdependence and differences among crash patterns. As reported in the study, crash severity is a more sensitive response variable to crash patterns than the vehicles involved and road alignments. Road type, pedestrian age, lighting conditions, vehicle type and several interactions of these factors were reported as the most influential crash patterns.

By using a mixed logit model for pedestrian-injury severity in pedestrian-vehicle crashes, Kim et al. (*15*) tested the possibility of unobserved heterogeneity of random parameters on variables. According to the paper, variables such as darkness in locations without street lights, vehicle type, freeway versus highway, and speed with driver sobriety increased the probability of a fatal pedestrian injury by 400%, 370%, 330% and 360% respectively.

Ulfarsson et al. (*16*) used a multinomial logit model to assign a fault in pedestrian-motor vehicle crashes. As indicated in the report, pedestrians were found at fault 59% of the crashes, drivers 32% and the remaining 9% both were at fault. Driver turning or merging, vehicle speed, driver blood-alchol level, driver backing up and number of pedestrians in the area were the largest factors associated with driver being found solely at fault. On the other hand, pedestrians crossing the street, pedestrian dart/dash, pedestrians 12 years or younger, freeway, and pedestrian intoxication level were the largest factors associated in pedestrian being found solely at fault. Speed, driver backing up, driver turning/merging, both driver and pedestrian intoxication level and pedestrian walking along road were largest effects associated with both the driver and pedestrian being found jointly at fault.

By applying a multinomial logit model, Tay (*17*) studied the impact of different factors in determining the severity of pedestrian crashes in South Korea. According to the paper, male drivers are more likely to be involved in severe and fatal crashes than female drivers. Drivers with age older than 65 years are less likely to be involved in either severe or fatal crashes as compared to middle age drivers. Intoxicated drivers were reported with a higher chance of being involved in a fatal crash.

The objective of this study is to develop a pedestrian-rail crash severity model and explore the impacts of various factors involved in the crash. A nominal response multinomial logit model with three levels of severity (fatal injury, serious injury and no injury) were used to model the impact of various factors including, but not limited to, pedestrian characteristics, environmental factors, highway-rail crossing characteristics, highway characteristics and land use type. The SAS PROC LOGISTICS procedure was used to develop the model.

2.4 Vehicle crash severity modeling

Several studies have been conducted to model crash severity and investigate the impacts of various factors involved in the crashes. As briefly discussed and summarized by Savolainen et al. (*4*), a wide range of methodological tools have been applied to study the impact of various factors on motor vehicle crash-injury severities. As presented in the paper, ordered logit and probit, multinomial logit, binary logit and binary probit and nested logit are some of most frequently used statistical methodologies.

Mercier et al. (*18*) conducted a study and tested the hypothesis that older drivers and passengers would suffer more severe injuries than younger adults in presence of broadside and angle collisions of automobiles on rural highways. Logistic modeling, Hierarchical Regression Analysis and Principal Components Regression, were analysis tools applied. Injury severity levels, fatal, major and minor were considered as dependent categorical variable. Some of the independent variables considered were occupant age, occupant position relative to point of impact and protection. According to the study, age is reported as a significant predictor of injury severity and is slightly higher for females than males. It was also identified that use of lap and shoulder restrains reduces injury severity and is less certain for females. For females only, air bags deployed were reported as significant injury severity predictors.

By using sequential binary logistic regression, Dissanayake and Lu (*19*) modeled crash severity for single-vehicle fixed object crashes involving young drivers. The five crash severity categories considered were no injury, possible injury, non capacitating injury, incapacitating injury and fatal. As reported in the study, factors such as alcohol or drug influence, ejection in the crash, point of impact, rural crash locations, existence of curve or grade at the crash location and speed of vehicle significantly increased the probability of more severe crashes. On the other hand, restraint device usage and drivers being of male gender were reported to reduce the chance of high severity crashes. It was also indicated that factors such as weather condition, residence location and physical condition have no significant relation in the model.

Duncan et al. (*20*) conducted a study to investigate car occupant injury severity in two-vehicle passenger car-truck rear-end collisions by using an ordered probit model. As reported in the study, factors such as darkness, high speed differentials, high speed limits, grades, being in a car struck to the rear, driving while drunk and being female increased the passenger vehicle occupant injury severity. On the other hand, factors such as snowy or icy roads, being in a child restraint, congested roads decreased the severity level. It was also indicated that interaction effects of cars being struck to the rear with high speed differentials and car rollovers were significant.

Donnell and Mason (*21*) conducted a study and developed median-related crash severity models. Three crash severity classes, fatal, injury and property damage only (PDO) were considered as independent variable outcome. Both ordinal and nominal response logistic regression models were developed in the study. As indicated in the report, ordinal response model gave more attractive results for cross-median crashes. On the other hand, nominal response model gave better result for median-barrier crashes. Furthermore, variables such as highway surface conditions, use of drugs or alcohol, presence of an interchange entrance ramp, horizontal alignment, crash type and average daily traffic volume were reported to have effect on crash severity.

By using paired comparison analysis and ordered probit model, Renski at al. (*22*) conducted a study to test the hypothesis that speed a limit increase will result in an increase in driving speed and produce higher crash severity. The study was focused on single-vehicle crashes on interstate roadways in North Carolina. As reported in the study, increasing speed limits from 55 to 60 mph and from 55 to 65 mph increased the probability of sustaining minor and non-capacitating injuries. However, the study indicated that increasing speed limits from 65 to 70 mph did not showed significant effect on crash severity.

Huang et al. (*23*) investigated effects of road diets in which four-lane undivided roads were converted into three lanes. A road diet, is also called a lane reduction or road rechannelization, is a technique in transportation planning whereby a road is reduced in number of travel lanes and/or effective width in order to achieve systemic improvement. Twelve road diets and 25 comparison sites were considered in the study. A “yoked comparison” study was applied in a “before” and “after” analysis and it was reported that road-diet crashes occurred during the “after” period was observed to be lower by 6 percent than that of the comparison sites. Khattak (*24*) conducted a study that investigated the effect of vehicle technologies on crash injury severity. Three separate ordered probit models were developed for the three drivers, Driver 1 (leading), Driver 2 (striking) and Driver 3 (striking in a three-vehicle crash). As indicated in the study, in a two-vehicle rear-end collision the leading driver is more likely to be injured whereas in a three-vehicle collision the driver in the middle is more likely to be injured. It was also stated that being in a newer vehicle protects the driver in rear-end collisions. Moreover, the study showed the benefit of technological improvements on driver’s safety.

Mercier at al. (*25*) performed a study and tested the hypothesis that older drivers and passengers would suffer more severe injuries than younger adults in presence of head-on collisions of automobiles on rural highways. Logistic modeling, Hierarchical Regression Analysis and Principal Components Regression were applied. Injury severity levels fatal, major and minor were considered as dependent categorical variable. The independent variables considered included, among others, occupant age, occupant position relative to point of impact and level of protection. As stated in the study, age was an important factor in predicting injury severity for both men and women. The study concluded that older drivers and passengers experienced more severity injury than any of other age groups. Use of lap and shoulder devices was reported to be more important for men than women while the reverse is true for deployed air bags.

Chira-Chavala et al. (*26*) investigated the characteristics and probable causes of light rail transit system accidents and developed a crash severity model for the Santa Clara County Transit Agency. A binary logit model was applied to predict the probability of injury accident as a function of explanatory variables such as speeds before collision of light rail vehicles and motor vehicles, movement of the motor vehicle before collision, etc. As reported in the study, left-turn vehicle movements, higher speeds of the motor vehicle or the LRV and accident occurring during peak hours increased the probability of injury accidents.

Chen and Jovanis (*27*) developed and tested the variable-selection procedure that avoids problems occurring due to the presence of large number of potential factors, complex nature of crash etiology and outcomes and large number of categories in crash-severity modeling. The procedure consisted of the chi-squared automated interaction detection (CHAID) method to collapse categories, Person chi-square test to assess relationship between dependent and independent variables, and log-linear modeling techniques. As indicated in the study, the log-linear model showed that late-night or early-morning driving increased the risk of severe injury crashes for bus drivers. It was also stated that bus accidents involving large truck or tractor-trailers increased the risk of severe injury crashes.

By using ordered probit model, Khattak at al. (*28*), explored factors contributing to more severe older driver (age of 65 and above) crash injury severity. According to the study, older male drivers are more prone to injury as compared to older female drives. It was stated that older drivers under the influence of alcohol experienced more severe injuries. It was also indicated that older driver injuries in farm vehicles are more severe as compared to other vehicle types. Xie et al. (*29*) conducted a study that demonstrated application of Bayesian ordered probit model in drivers' injury severity analysis. In the Bayesian probit model, prior distributions such means and variances were included reflecting the analysts' prior knowledge about the data. Comparison was made between Bayesian ordered probit and conventional ordered probit models. As reported in the study, for large data size, model fitting results obtained from the Bayesian and the conventional probit model have no significant differences. It was also reported that for small sample size, Bayesian probit model produced parameter estimates with better prediction performance than the conventional ordered probit model.

This purpose of this study is to analyze severity of vehicle crashes at HRGCs and to investigate the impact of various factors involved in the crashes. A nominal response multinomial logit model with three levels of severity was used to model the impact of various factors that includes vehicle driver characteristics, environmental factors, rail-road crossing characteristics, highway characteristics, land use type and more. The three levels of responses considered were fatality, injury and no injury. The SAS PROC LOGISTICS procedure was used to develop the model.

Chapter Three

Methodology

3.1 Highway-Rail At-Grade Crossing incident statistics

The Federal Rail Road maintains the Highway-Rail Grade Crossing (HRGC) Incident Report (GXIR) Database that contains data describing impacts between railway equipment and highway users. All highway-rail grade crossing incidents are submitted by the railroads. The GXIR data are also available in the annual FRA-RRR publication, Railroad Safety Statistics Annual Report, and online at the FRA's Web Site. Below are some of the HRGC incident statistics from 2005 to 2012 as obtained from the FRA data base.

3.1.1 Highway user type

As can be seen in Figure 3, between 2005 and 2012, automobiles were the major highway user involved in public crossing incidents, sharing almost one-third of the total public crossing incidents. Truck-trailers in combination with pick-up truck were the second largest highway user groups involved in crashes representing 29 percent of the total incidents. In terms of fatalities, automobiles contribute nearly 38 percent and trucks in combination share nearly 25 percent of the total. In the case of pedestrians, despite representing less than 6 percent of the total incidents they contributed close to 25% of total fatalities (*2*).

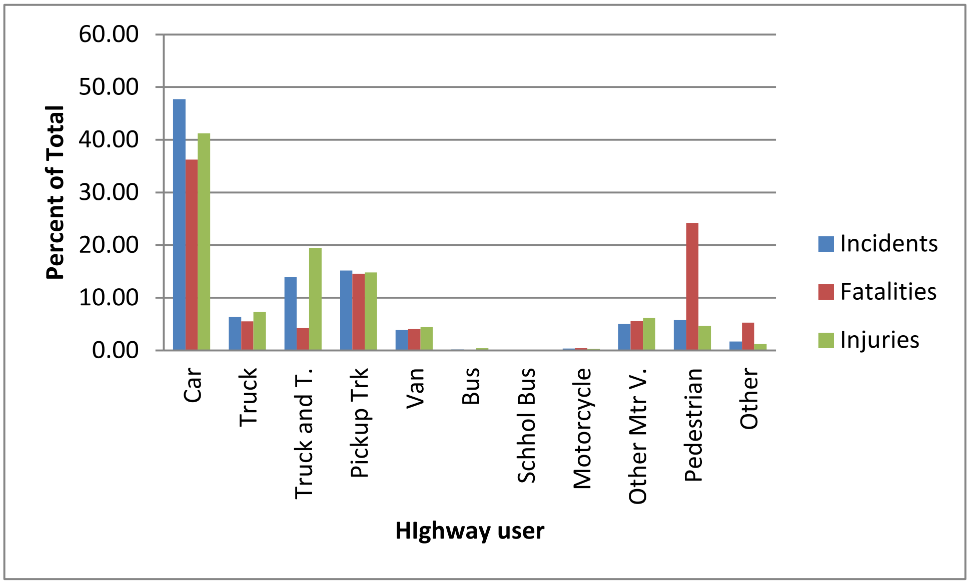

Figure 3. Number of Incidents, Fatalities and Injuries by Highway User Types, 2005-2012 (*2*)

3.1.2 Warning device type

As seen in Figure 4, highest number of incidents as well as fatalities occurred at public crossings equipped with gates and crossings with cross bucks and flashing lights were the second and third highest respectively (*2*).

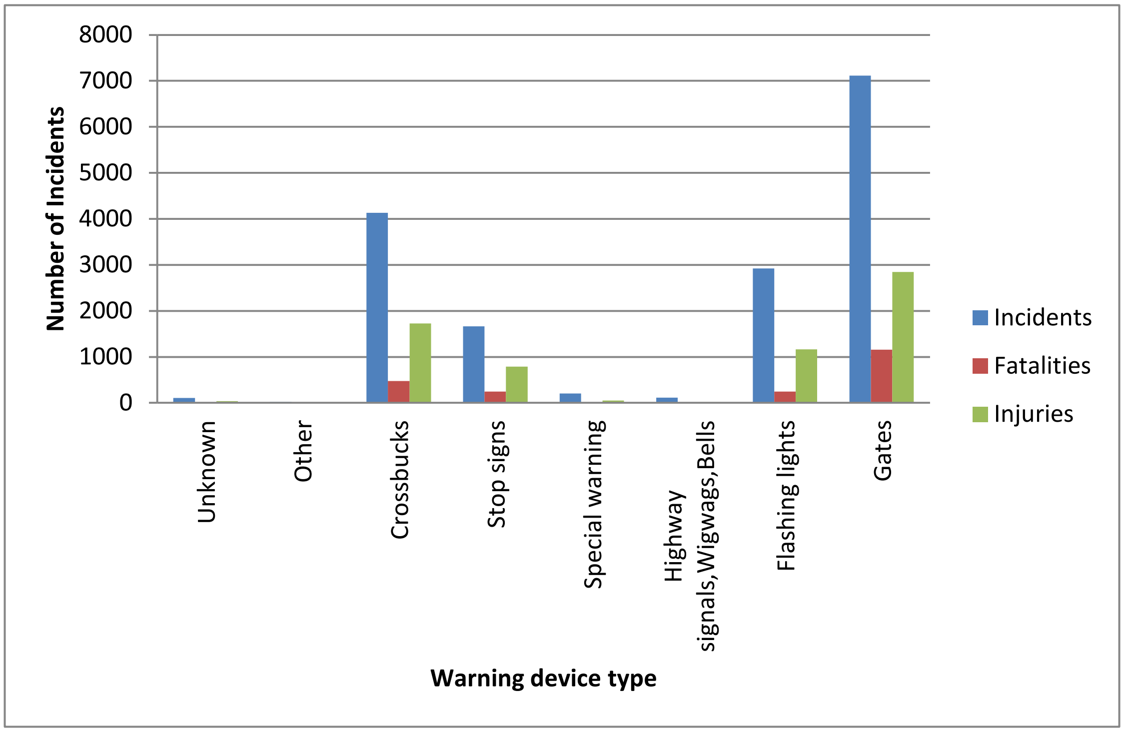

Figure 4. Number of Incidents, Fatalities and Injuries by Warning Device, 2005-2012 (*2*)

3.1.3 Weather type

HRGC incidents occur in good as well as bad weather conditions. As can be seen from Figure 5, the highest number of incidents occurred under clear weather and cloudy conditions. Cloudy weather condition is most common among the bad weather conditions that crossing incidents are dominantly occurring (*2*).

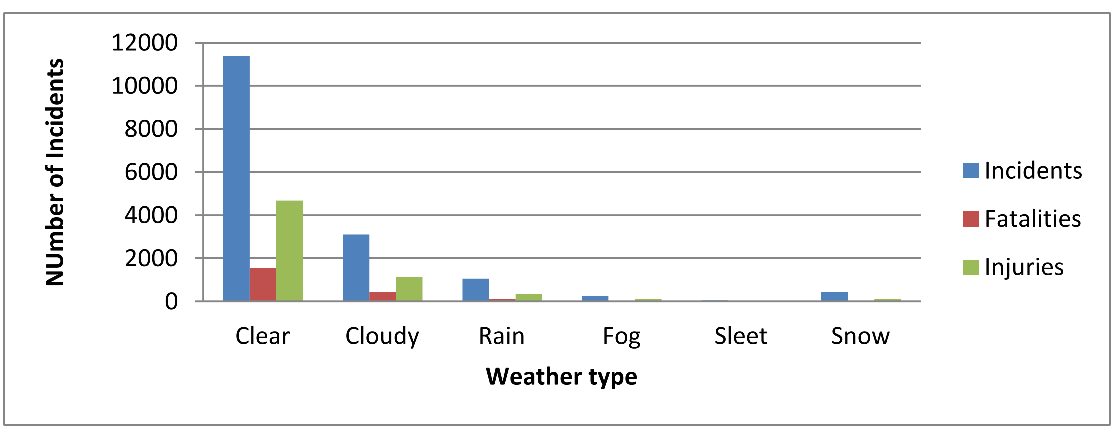

Figure 5. Number of Incidents, Fatalities and Injuries by Weather Type, 2005-2012 (*2*)

3.1.4 Train speed

Grade crossing crashes occur at different train speed levels ranging from slow to high. As shown in Figure 6, the highest number of incidents on public crossings involved trains traveling between 40 and 49 mph and those with less than 9 mph (*2*).

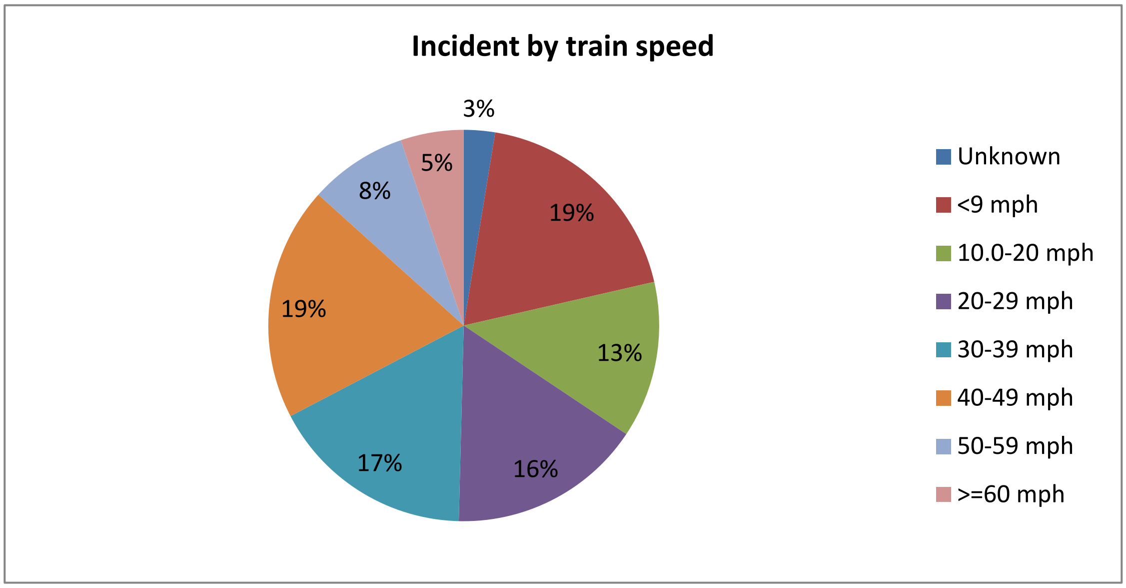

Figure 6. Number of Incidents, Fatalities and Injuries by Train Speed, 2005-2012 (*2*)

3.1.5 Vehicle speed

As shown in Figure 7, between 2005 and 2012, most crossing incidents involved slow speed (less than 9 mph) vehicles (*2*).

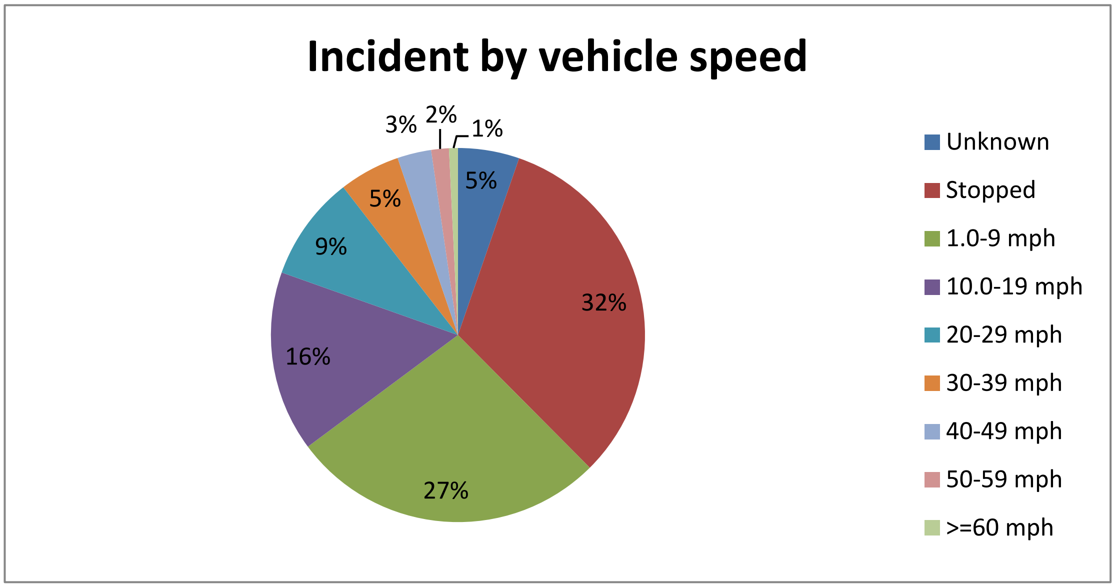

Figure 7. Number of Incidents, Fatalities and Injuries by Vehicle Speed, 2005-2012 (*2*)

3.2 Data assembly

Vehicle-rail and pedestrian-rail crash data on USDOT public crossing sites from 2005 to 2012 were used in this study. In order to acquire more explanatory variables, the USDOT highway-rail crossing inventory was also included. The crash data and the crossing inventory data were merged based on the USDOT identification number. The SAS PROC SQL was used to merge and clean the data. After the data merging and cleaning process, a total of 7,391 records were obtained and used in the modeling stage. The data used to create the data set were obtained from the Federal Railroad Administration (*2*).

Table 1 presents the descriptive statistics of some of the variables from such HRGC crash and inventory data. As shown, the distribution of vehicle-rail crash severity is 8.55%, 26.86% and 64.77% for fatal, injury and no injury respectively. This distribution of crash severity indicates around 35.41% of vehicle crashes at HRGC sites lead to fatality or injury, in which the figures are much higher as compared to those of multi-vehicle crashes in highway traffic. The majority (78.61%) of vehicle-rail crashes at HRGC sites occurred when the train equipment struck the vehicle while the remaining (21.39%) were when vehicle struck the rail equipment. It is shown in the table that a majority (53.02%) of vehicles involved in the vehicle-rail crashes are cars. It is also shown that the majority (71.02%) of vehicle crashes had occurred in clear weather conditions.

Table 1. Descriptive Statistics of Variables from HRGC Crash and Inventory Data (Pedestrian-rail Crash)

Variable	Category	Frequency	Percent
Crash Characteristics			
INJURY (crash severity level)	3=Fatal crashes	631	8.55
	2=Injury crashes	1969	26.68
	1=No Injury crashes	4780	64.77
TYPACC (Type of accident)	1=Train struck vehicle	5810	78.61
	2=Vehicle struck train	1581	21.39
Vehicle characteristics			
TYPVEH (Type of vehicle)	1=Auto	3919	53.02
	2=Truck	540	7.31
	3=Truck trailer	1296	17.53
	4=Pickup truck	1315	17.79
	5=Van	306	4.14
	6=Bus	10	0.14
	7=School Bus	5	0.07
VEHSPD (Vehicle speed)	1=<25mph	6292	85.13
	2=25-45mph	827	11.19
	3=>45mph	272	3.68
(AADT) (Average annual daily traffic)	1=<10,000	6505	88.0
	2=10-20,000	599	8.1
	3=20,000-30,000	177	2.39
	4=>30,000	110	1.49
Train Characteristics			
TRNSPD (Train speed)	1=<25mph	2984	40.37
	2=25-45mph	2541	34.38
	3=>45mph	1866	25.25

Table 1. (Cont'd) Descriptive Statistics of Variables from HRGC Crash and Inventory Data (Pedestrian-rail Crash)

Variable	Category	Frequency	Percent
Vehicle Driver Characteristics			
DRVAGE	1=<25 Years	1127	18.41
	2=25-60 Years	3965	64.77
	3=>60 Years	1030	16.82
DRIVGEN (Vehicle driver gender)	1=Male	5605	75.84
	2=Female	1786	24.16
Highway Characteristics			
HWYPVED (Highway surface type)	1=Paved	6021	81.46
	2=Unpaved	1370	18.54
HWYSGNL (Highway signal)	1=Not present	7192	97.31
	2=Present	199	2.69
TRAFICLN (No. of traffic lane)	1=1 Lane	644	8.71
	2=2 Lanes	5540	74.96
	3=3 Lanes	87	1.18
	4=4 Lanes	869	11.76
	5=≥5 Lanes	251	3.4
Environmental Characteristics			
DEVELTYP (Development area type)	1=Open space	2396	32.42
	2=Residential	1590	21.51
	3=Commercial	2074	28.06
	4=Industrial	1221	16.52
	5=Institutional	110	1.49
WEATHER (Weather condition)	1=Clear	5249	71.02
	2=Cloudy	1401	18.96
	3=Rain	445	6.02
	4=Fog	106	1.43
	5=Sleet	15	0.02
	6=Snow	165	2.37
TEMP (Temperature)	1=<50^{o}F	2010	27.61
	2=50-80^{o}F	3611	49.12
	3=>80^{o}F	1711	23.27
NEAREST (Intersecting IN or Near city)	1=In city	4226	57.18
	2=Near city	3165	42.82

Table 1. (Cont'd) Descriptive Statistics of Variables from HRGC Crash and Inventory Data (Pedestrian-rail Crash)

Variable	Category	Frequency	Percent
Crossing characteristics			
XSURFACE (Crossing surface type)	1=Timber	2046	27.68
	2=Asphalt	2999	40.58
	3=Asphalt & Flange	441	5.97
	4=Concrete	920	12.45
	5=Concrete & Rubber	266	3.6
	6=Rubber	413	5.59
	7=Metal	3	0.04
	8=Unconsolidated	256	3.46
	9=Other	47	0.64
XBUCK (Cross bucks)	1=Not Present	2342	31.69
	2=Present	5049	68.31
FLASH (Flashlight)	1=Not present	3543	48.01
	2=Present	3836	51.99
GATES (Gates)	1=Not Present	4661	63.06
	2=Present	2730	36.94

The HRGC sites where crashes occur are located in different development areas. As one can see from Table 1, 32.42% of the crossings are located in open space areas, 21.51% in residential areas, and 28.06% in commercial areas. The rest are found in industrial and institutional development areas. The majority (74.96%) of the HRGCs where accidents occur cross two lane highways. Descriptive statistics of other variables are also shown in the table. As many variables as possible are considered in this study. Some of the continuous variables are converted in to categorical variable and the MNLM is applied to estimate the model parameters.

As shown in Table 2, the distribution of pedestrian crash severity is 53.37%, 30.91% and 15.72% for fatal, severe injury and no injury respectively. This distribution of severity indicates around 85% of crashes at HRGC sites lead to fatality or injury. The majority of pedestrian crashes (95%) at HRGC sites occur when the train equipment struck the pedestrian rather than when the pedestrian struck the rail equipment (5%). It is shown in the table that there are more male pedestrians (64.34%) than female (19.55%) in the crash data. Nearly 75% of the pedestrian crashes occurred in a clear weather condition as compared to others.

Table 2. Descriptive Statistics of Variables from HRGC Crash and Inventory Data (Vehicle -Rail Crash)

Variable	Category	Frequency	Percentage
Crash Characteristics			
SEVERITY (crash severity level)	3=Fatal crashes	404	53.37
	2=Injury crashes	234	30.91
	1=No Injury crashes	119	15.72
TYPACC (Type of accident)	1=Rail equipment struck pedestrian	719	94.98
	2=Pedestrian struck rail equipment	38	5.02
Train Characteristics			
TRNSPD (Train speed)	1=Less than 25 mph	184	24.31
	2=Between 25 and 45 mph	297	39.23
	3=Between 45 and 60 mph	164	21.66
	4=Greater than 60 mph	112	14.80
Pedestrian Characteristics			
PEDESTRNGEN (Pedestrian gender)	1=Male	487	64.34
	2=Female	148	19.55
	1=Missing	122	16.11
Crossing Characteristics			
XANGLE (Smallest crossing angle)	1=0-29 degree	25	3.30
	2=30-59 degree	86	11.36
	3=60-90 degree	646	85.34
STDFLASH (Standard flash light)	1=Not present	157	20.74
	2=Present	600	79.26
GATES (Gates)	1=Not Present	209	27.61
	2=Present	548	72.38
XBUCK (Crossbucks)	1=Not present	518	68.43
	2=Present	239	31.58

Table 2. (Cont'd) Descriptive Statistics of Variables from HRGC Crash and Inventory Data (Vehicle-Rail Crash)

Variable	Category	Frequency	Percentage
XSURFACE (crossing surface type)	1=Timber	267	35.27
	2=Asphalt	323	42.67
	3=Asphalt and flange	25	3.30
	4=Concrete	42	5.55
	5=Concrete and rubber	1	0.13
	6=Rubber	89	11.76
	7= Metal	0	0
	8=Unconsolidated	5	0.66
	9=Others	5	0.66
ADVWARN (RR advanced warning	1=Not present	526	69.48
	2=Present	231	30.52
Highway Characteristics			
TRUCKLN (Truck pullout lanes)	1=Not Present	38	5.02
	2=Present	719	94.98
HIWYSGNL (Traffic signal)	1=Not present	735	97.09
	2=Present	22	2.91
TRAFICLAN (Number of traffic lane)	1=1 lane	23	3.04
	2=2 lanes	477	63.01
	3=3 lanes	28	3.70
	4=4 lanes	183	24.17
	5=≥5 lanes	46	6.08
HWYPVED (Highway pavement type)	1=Unpaved	730	96.43
	2=Paved	27	3.57
Environmental Characteristics			
DEVELTYP (Type of development)	1=Open space	61	8.06
	2=Residential	162	21.40
	3=Commercial	405	53.50
	4=Industrial	114	15.06
	5=Institutional	15	1.98
NEAREST (Intersecting In or Near City)	0=In city	656	86.66
	1=Near city	101	13.34
TEMP (Temperature)	1=<50 degree Fahrenheit	191	25.30
	2=50-80 degree Fahrenheit	466	61.72
	3=>80 degree Fahrenheit	98	12.98
WEATHER (Weather condition)	1=Clear	566	74.77
	2=Cloudy	143	18.89
	3=Rain	32	4.23
	4=Fog	8	1.06
	5=Snow	8	1.06

The HRGC sites where crashes occurred are located in different development areas. As the data indicates 53.5% of the crossings are located in commercial areas and 21.4% in residential areas. The rest are found in open space, industrial and institutional development areas. The majority of the HRGCs where accidents occurred cross two lane highways (about 63%). Descriptive statistics of the other variables, mainly of HRGC characteristics, are also shown in the table. These include the number of highway traffic lanes, highway signal, crossing surface types, presence of flash lights, advanced warning signs, gates, crossbucks, etc. As many variables as possible are considered in this study.

3.3 Mathematical model formulation

In general, crash severity level is ordinal data and as a result of this most researchers applied ordered logistic regression models in their study. Before considering methods for ordinal outcomes, it is important to note that simply because the values of a variable can be ordered does not imply that the variable should be analyzed as ordinal. A variable might be ordered when considered for one purpose, but be unordered or ordered differently when used for another purpose. When the proper ordering of a variable is ambiguous, the models for nominal variable should be considered in addition to the models for ordinal variables (*6*).

Modeling ordinal outcome dependent variable using nominal variable will lead to loss of efficiency as a result of ignoring information. In the reverse, modeling nominal variable using ordinal variable will give biased or sometimes irrational estimates. The loss of information in the ordinal data can be outweighed by avoiding the bias. The primary advantage of nominal outcome multinomial logit model is its ability to avoid the parallel effect regression assumption unlike the ordered outcome regression model.

Uncertainty to as to what to make the dependent variable to consider as ordered outcome can also be avoided by using the nominal outcome multinomial logit model (*6*). The mathematical formulation of the multinomial ordered and unordered logit models are briefly described in the following section.

3.3.1 Multinomial Logit Model (MNLM)

The MNLM formulation is well discussed by Long (*6)*. MNLM can be modeled using various approaches such as probability model, odds ratio model and discrete choice model. Regardless of which approach is used to derive the model, the equation for the probability of an outcome is the same. The probability model is described in the following paragraphs.

If *y* is the response variable with *J* nominal outcomes, then the assumption of the multinomial logit model is that category 1 through *J* are not ordered. Also, let Pr(*y*=*m*|*x*) be the probability of observing outcome *m* given the independent variable *x*. The model for *y* is constructed as follows:

- Assume that Pr(*y*=*m*|*x*) is a linear combination $x\beta_m$. The vector $\beta_m = (\beta_{0m} \ldots \beta_{km} \ldots \beta_{Km})$ contains the intercept β_{0m} and coefficients of β_{km} for the effects of x_k on outcome *m*.
- To ensure non negativity for the probabilities, the exponential of $x\beta_m$ is taken.
- For the probabilities to sum to 1, divide exp $(x\beta_m)$ by $\sum_{j=1}^{J} \exp(x_i\beta_j)$.

$$\Pr(y_i = m|x_i) = \frac{\exp(x_i\beta_m)}{\sum_{j=1}^{J} \exp(x_i\beta_j)} \qquad (1)$$

Though the probability sum gives 1, the set of parameters that generates the probabilities is not identified since more than one set of parameters can generate the same probabilities. In order to identify the set of parameters that generate the probabilities, a constant must be imposed. By imposing one of the parameter estimates equals 0 (assume β_1=0), the model can be written as follows:

$$\Pr(y_i = 1|x_i) = \frac{1}{1+\sum_{j=2}^{J} \exp(x_i\beta_j)} \quad (2)$$

$$\Pr(y_i = m|x_i) = \frac{\exp(x_i\beta_m)}{1+\sum_{j=2}^{J} \exp(x_i\beta_j)} \qquad \text{for } m > 1 \quad (3)$$

The parameter estimates are determined using maximum likelihood estimation. If the observations are independent, the likelihood equation is given by:

$$L(\beta_2, \ldots.\beta_J|y, x) = \prod_{i=1}^{N} P_i \quad (4)$$

where P_i is the probability of observing whether values of y was actually observed for the i^{th} observation. Combining the equation 1 with this equation in place of P_i the likelihood equation can be written as:

$$L(\beta_2, \ldots\beta_J|y, x) = \prod_{m=1}^{J} \prod_{y_{i=m}} \frac{\exp(x_i\beta_m)}{\sum_{j=1}^{J} \exp(x_i\beta_j)} \quad (5)$$

where $\prod_{y_{i=m}}$ is the product over all cases for which y_i is equal to m. Taking logs, we may obtain the log likelihood equation which can be maximized with numerical methods to estimate the β's.

The overall model fitness can be compared by using the model's log-likelihood at convergence with the log-likelihood of a naive model (model with all coefficients set to zero which is equivalent to assigning equal probability for all outcomes). It is also possible to compare a model with only alternative constants (assigning probability to outcomes equal to the observed share of the outcomes in the dataset):

$$\rho^2 = 1 - \frac{LL(\beta)}{LL(c)} \quad (6)$$

where LL (β) represents the log-likelihood at model convergence, LL (0) represents the log-likelihood of a naïve model (without information). The ρ^2 goes from 0 (for no improvement in the log-likelihood) to 1 for a perfect fit. A value for ρ^2 larger than 0.1 indicates meaningful improvement (*6*).

The MNLM can be interpreted by applying various kinds of approaches. One method is to determine the predicted probability. Probabilities can be computed at a variety of values and can be presented in different ways such as mean, minimum, and maximum.

The other method is to compute the marginal effect or partial change which can be determined by taking derivative of Equation 1 with respect to x_k as described in the following equation. Marginal effect is the slope of the curve relating x_k to $\Pr(y=m|x)$, holding other variables constant. Variables are held at their means, possibly with dummy variables at 0 or 1 (*6*). The value of the marginal effect depends on the value of all independent variables and on the coefficients for each outcome.

$$\frac{Pr\ (y=m|x)}{x_k} = \Pr(y = m|x)\ [\beta_{km} - \sum_{j=1}^{J} \beta_{kj} \Pr\ (y = j|x)\ ] \qquad (7)$$

Discrete change in probabilities is also an effective method of interpretation that can be applied for continuous and dummy independent variables. The change in the predicted probability when a variable x_k changes from the starting value (x_s) to the ending value (x_ϵ) can be computed as follows:

$$\frac{\Delta \Pr\ (y=m|x)}{\Delta x_k} = \Pr(y = m|x, x_k = x_\epsilon) - \Pr(y = m|x, x_k = x_s) \qquad (8)$$

Odds ratio can also be used in the interpretation of the developed model. The odds ratio is defined as the ratio of the odds of those with the risk factor to the odds for those without the risk factor. Generally, the odds ratio can be computed by exponentiating the difference of the logits between any two population profiles (*30*).

The following three equations can be used to predict the probabilities of the three severity crashes (Fatality, Injury and No injury).

$$P_{Fatal} = \frac{e^{equation(32)}}{[1+e^{equation(32)}+e^{equation(32)}]} \tag{9}$$

$$P_{Injury} = \frac{e^{equation(33)}}{[1+e^{equation(32)}+e^{equation(33)}]} \tag{10}$$

$$P_{No\ Injury} = 1 - (P_{Fatal} + P_{Injury}) \tag{11}$$

3.3.2 Ordered Logit Model (OLM)

When the absolute distance between categories of a variable is unknown, yet there is a clear ordering of the categories, the variable is considered ordinal. The ordered response logistic regression formulation is presented as discussed by Long (*6*). An ordinal logistic regression model is derived from a measurement model in which a latent variable y* is mapped to an observed variable *y*. These variables are related according to the following equation:

$$y_i = m \quad if\ \tau_{m-1} \leq y_i^* < \tau_m \quad for\ m = 1\ to\ J \tag{12}$$

The τ's are cutpoints on the measurement scale that are used to distinguish the ordinal categories. In the case of crash severity models, the ordinal response categories are fatality, injury, and no injury crashes. As shown in Figure 8, Category 1 (fatal) is defined by the open-ended interval on the lower end of the measurement scale; Category 3 (no injury) is defined as the portion of the scale above cut point τ_2 and Category 2 (injury) is the portion between the two cutpoints.

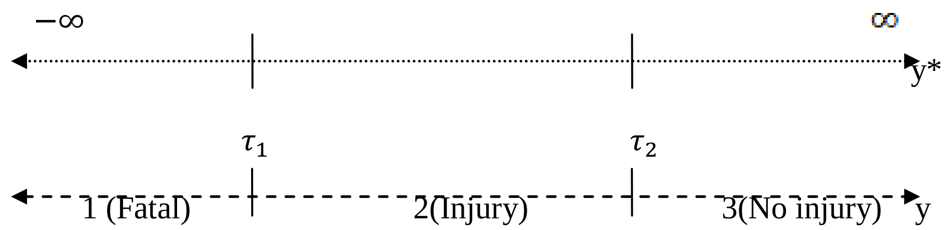

Figure 8. Ordinal Response Categories

The observed y is related to y* according to the measurement model:

$$y_i = \begin{cases} 1 = Fatal \quad & if\ \tau_o\ = -\infty \leq y_i^* < \tau_1 \\ 2 = Injury \quad & if\ \tau_1 \leq y_i^* < \tau_2 \\ 3 = No\ injury & if\ \tau_2 \leq y_i^* < \tau_3 = \infty \end{cases} \tag{13}$$

The regression equation used for an ordinal response is y_i^* = $x_i\beta+\epsilon_i$. Where x_i is a row vector (with 1 in the first column for the intercept), β is a column vector of structural coefficients (with the first element being the intercept βo), and ϵ_i is an error term.

Maximum likelihood (ML) estimation can be used to estimate the regression of *y** on *x*. In order to use ML, assumption of a specific type of error (ϵ) distribution is required. For the ordered logit model, the error term has a logistic distribution with mean zero and a variance of $\pi^2/3$. The probability density function (pdf) of the logistic distribution is given as shown in Equation (14).

$$\lambda(\epsilon) = \frac{\exp(\epsilon)}{[1+\exp(\epsilon)]^2} \tag{14}$$

And the cumulative distribution function (cdf) of the logistic distribution is given as shown in equation (15):

$$\Lambda(\epsilon) = \frac{\exp(\epsilon)}{1+\exp(\epsilon)} \tag{15}$$

After specification of the error term, the probabilities of observing values of *y* given *x* can be computed. The probability of any observed outcome y = *m* given *x* is the difference between the cdf evaluated at these values:

$$\Pr(y_i = m|X_i) = F(\tau_m - x_i\beta) - F(\tau_{m-1} - x_i\beta) \tag{16}$$

To estimate the model, let β be the vector with parameters from the structural model, with the intercept βo in the first row and let τ be the vector containing the threshold parameters. Either βo or τ_1 is constrained to 0 to identify the model. Program such as SAS's LOGISTIC procedure assumes βo and estimates τ_1. From Equation 5, the following can be obtained:

$$\Pr(y_i = m|X_i, \beta, \tau) = F(\tau_m - x_i\beta) - F(\tau_{m-1} - x_i\beta) \tag{17}$$

The probability of observing whatever value of y was actually observed for the i^{th} observation is:

$$P_i = \begin{cases} \Pr(y_i = 1|x_i, \beta, \tau) & if \quad y = 1 \\ \quad\quad\quad\vdots & \\ \Pr(y_i = m|x_i, \beta, \tau) & if \quad y = m \\ \quad\quad\quad\vdots & \\ \Pr(y_i = J|x_i, \beta, \tau) & if \quad y = J \end{cases} \tag{18}$$

If the observations are independent, the likelihood equation over the population of N observations is:

$$L(\beta, \tau|y, X) = \prod_{i=1}^{N} P_i \tag{19}$$

Combining equations 16, 17, and 18:

$$L(\beta, \tau|y, X) = \prod_{j=1}^{J} \prod_{y_i=j} \Pr\left(y_i = j \middle| x_{i,} \beta, \tau\right) = \prod_{j=1}^{J} \prod_{y_i=j} [\mathrm{F}(\tau_j - x_i\beta) - F(\tau_{j-1} - x_i\beta)] \tag{20}$$

Here, $\prod_{y_{i=j}}$ indicates multiplying over all cases where y is observed to equal j. Taking logs, the loglikelihood can be written as follows:

$$lnL(\beta, \tau | y, X) = \sum_{j=1}^{J} \sum_{y_{i=j}} \ln\left[F(\tau_j - x_i\beta) - F(\tau_{j-1} - x_i\beta)\right] \quad (21)$$

Model estimation involves maximizing Equation 10 using numerical methods to estimate the τ's and the β's.

A measure of the model goodness of fit (ρ^2) can be calculated as:

$$\rho^2 = 1 - \left[\frac{lnL_b}{lnL_o}\right] \quad (22)$$

where lnL_b is the log likelihood at convergence and lnL_o is the restricted log likelihood. The ρ^2 measure is bound by zero and one. Values of ρ^2 closer to one indicate better fit of the model.

Similarly, the MNLM can be interpreted by applying various kinds of approaches. One method is to determine the predicted probability. Probabilities can be computed at a variety of values and can be presented in different ways such as mean, minimum, and maximum.

Interpretation of ordinal response variables can be performed according to odds ratios. In this paper, the proportional odds model is used to interpret odds ratios for cumulative probabilities. The cumulative probability that the outcome is less than or equal to *m* is:

$$\Pr(y \leq m | x) = \sum_{j=1}^{m} \Pr(y = j | x) \qquad for\ m = 1, \dots, J-1 \quad (23)$$

The odds that an outcome is *m* or less versus greater than *m* given a set of explanatory variables *x* are:

$$\Omega_m(x) = \frac{\Pr(y \leq m | x)}{1 - Pr(y \leq m | x)} = \frac{\Pr(y \leq m | x)}{\Pr(y > m | x)} = \exp(\tau_m - x\beta) \quad (24)$$

Taking the log result in the logit equation:

$$ln\Omega_m(x) = \tau_m - x\beta \quad (25)$$

The marginal effects of variables x on the underlying crash severity propensity can be evaluated by taking the partial derivative of equation 25 with respect to x_k, resulting in:

$$\frac{\Pr(y=m|x)}{x_k} = \frac{F(\tau_m - x\beta)}{x_k} - \frac{F(\tau_{m-1} - x\beta)}{x_k} \tag{26}$$

or

$$\frac{\Pr(y=m|x)}{x_k} = \beta_k[f(\tau_{m-1} - x\beta) - f(\tau_m - x\beta)] \tag{27}$$

The marginal effect is the slope of the curve relating x_k to $\Pr(y = m|\mathbf{x})$, holding all other variables constant and is usually computed at the mean values of all variables. For a dummy independent variable, the derivative while treating it as a continuous variable provides an approximation.

Similar to the MNLM, the change in the predicted probability when a variable x_k changes from the starting value (x_s) to the ending value (x_ϵ) can be computed as follows:

$$\frac{\Delta\Pr(y=m|x)}{\Delta x_k} = \Pr(y = m|x, x_k = x_\epsilon) - \Pr(y = m|x, x_k = x_s) \tag{28}$$

The following three equations can be used to predict the probabilities of the three severity crashes:

$$P_{Fatal} = \frac{e^{equation(36)}}{[1+e^{equation(36)}]} \tag{29}$$

$$P_{Fatal} + P_{Injury} = \frac{e^{equation(37)}}{[1+e^{equation(37)}]} \tag{30}$$

$$P_{No\ Injury} = 1 - (P_{Fatal} + P_{Injury}) \tag{31}$$

Chapter Four

Results and Discussion

4.1 Introduction

One of the most important tasks in improving road safety is to uncover influential factors and then to develop countermeasures to improve road safety. The relationship between the injury severity of traffic crashes and factors such as driver and passenger characteristics, pedestrian age and gender, vehicle type, environmental conditions, and traffic and geometric conditions has attracted much attention. Better understanding of this relationship is necessary and very important for improving facility design so that accidents can be reduced. It is important to note that reducing crash frequency and reducing crash-injury severity may necessitate different strategic approaches. The Federal Railroad Administration has comprehensive data available for HRGC incidents, injuries and fatalities. However, the impact of various contributing factors on HRGC crash severity levels are not explored adequately.

The development of effective countermeasures requires a thorough understanding of the factors that affect the likelihood of a crash occurring or, given that a crash has occurred, the characteristics that may mitigate or exacerbate the degree of injury sustained by crash-involved road users. To gain such an understanding, safety researchers have applied a wide variety of methodological techniques over the years.

Logistic regression has been widely applied to model crash severity levels. This study also applied the logistic regression modeling approach (specifically ordered and multinomial logit models) to estimate the three levels of highway user crash severity on HRGC as a function of various factors involved. The modeling was performed separately for pedestrians and vehicle user crashes. Variables such as elements of geometric design, traffic operational measures, and environmental conditions were considered as explanatory variables in predicting crash severity levels. The explanatory variables are obtained by merging the USDOT crossing inventory and the HRGC crash data. In both of the pedestrian and vehicle user crash models, data from 2005 to 2012 was considered. In addition, comparison was performed between the ordered and multinomial logistic regression modeling approaches.

Many variables obtained from the crossing inventory and crash data were considered in developing the logistic regression models. During the final preferred model development process, some of the variables were found to be statistically insignificant and hence removed in a stepwise manner. PROC LOGISTIC procedure were applied with significance level being 0.1 to retain some of the variables. The results obtained from this study are presented in the following section of this paper.

4.2 Multinomial Logit Modeling results for pedestrian-rail crash on HRGC

Table 3 shows the coefficient estimates of the multinomial logit model for pedestrian crash severity levels on HRGCs. Among the three crash severity levels, no injury crashes were considered as the base case. A positive coefficient indicates that, given that an accident has occurred, the probability that a specific level of injury severity will occur is higher than the probability that the base level of injury severity (no injuries) will occur. A negative coefficient indicates that the probability that a specific level of injury severity will occur is less than the probability that the base level of injury severity (no injuries) will occur. For example, given that an accident has occurred, the higher that the train speed is, the chance that fatality will result is higher than the probability that no injury will result.

As illustrated in Table 3, some of the variables are not statistically significant. However, for the sake of facilitating interpretation of the results, such variables were still retained in the model if at least one of the variables in that category were significant in at least one of the models (injury and/or fatality). This actually induces reduction in efficiency of the model. In order to compensate for the reduction in efficiency, a 90 percent confidence level was considered instead of 95 percent.

Based on the parameter estimates obtained in Table 3, the MNL model can be written as follows:

$$\log\left[\frac{P(Y=Fatal)}{P(Y=No\ Injury)}\right] = -0.7175 + 1.015X_1 + 0.7851X_2 + 0.8188X_3 - 0.6152X_4 + 0.8684X_5 + 1.1033X_6 + 0.1012X_7 - 0.2294X_8 - 1.4631X_9 \quad (32)$$

$$\log\left[\frac{P(Y=Injury)}{P(Y=No\ Injury)}\right] = 1.1972 - 0.9424X_1 + 0.4101X_2 - 0.8978X_3 - 0.1331X_4 + 0.5414X_5 + 1.1033X_6 - 0.9009X_7 - 0.8188X_8 - 0.4892X_9 \quad (33)$$

where:

X_1= Train speed indicator (1 if speed train speed is 45-60mph, 0 otherwise)

X_2= Weather indicator (1 if cloud, 0 otherwise)

X_3= Type of accident indicator (1 if rail equipment struck pedestrian, 0 otherwise)

X_4= Pedestrian gender indicator (1 if male/missing, 0 otherwise)

X_5= Location of HRGC indicator (1 if located in city, 0 otherwise)

X_6= Highway pavement type indicator (1 if pavement is unpaved, 0 otherwise)

X_7= HRGC surface type (1 if surface is rubber, 0 otherwise)

X_8= Temperature indicator (1 if temperature is greater than 80^{o}F, 0 otherwise)

X_9= Highway traffic lane (1 if traffic lane is 3, 0 otherwise)

Based on the above MNL model equations, the marginal effect/value is also determined as presented in Table 4 for the explanation convenience. As can be seen in the Table 4, the sum of marginal effect gives zero which satisfies the requirement that the sum of probability is 1. The marginal effect for the remaining variables provides a great deal of valuable information for results interpretation.

Table 3. Multinomial Logistic Model Regression Results (Pedestrian-rail Crash)

	Injury		Fatality	
Parameter	**Estimate**	**P-value**	**Estimate**	**P-value**
Intercept	1.1972	0.2943	-0.7175	0.5012
TRNSPD (Ref: <25mph)				
25-45	-1.0734	0.0004	0.4302	0.1595
45-60	-0.9424	0.381	1.015	0.0054
>60	-1.8403	<.0001	0.3561	0.323
WEATHER (Ref: Clear)				
Cloudy	0.4101	0.2583	0.7851	0.0187
Rain	-0.6893	0.1862	-0.6553	0.1614
Fog	12.6421	0.9837	13.7341	0.9823
Snow	-1.8393	0.1796	-0.5642	0.5389
TYPACC (Ref: Pedestrian struck rail equipment)				
Rail equipment struck pedestrian	-0.8978	0.1222	0.8188	0.211
PEDESTRNGEN (Ref: Female)				
Male+Missing	-0.1331	0.6894	-0.6152	0.037
NEAREST (Ref: Near City)				
In City	0.5414	0.3489	0.8684	0.0041
HWYPVED (Ref: Paved)				
Unpaved	1.1033	0.5564	1.1033	0.0474

Table 3. (Cont'd) Multinomial Logistic Model Regression Results (Pedestrian-rail Crash)

	Injury		Fatality	
Parameter	**Estimate**	**P-value**	**Estimate**	**P-value**
XSURFACE (Ref :Timber)				
Asphalt	-0.0821	0.7657	0.1486	0.5535
Asphalt and flange	0.3216	0.6167	-0.401	0.5351
Concrete	1.0306	0.0912	0.388	0.5172
Concrete and rubber	0.7194	0.9998	13.4682	0.9949
Rubber	-0.9009	0.0338	0.1012	0.7747
Unconsolidated	0.5329	0.6795	-0.8596	0.5351
Others	-0.8173	0.4547	-2.1333	0.0987
TEMP (Ref: <50F)				
50-80F	-0.0666	0.8226	0.1906	0.4885
>80F	-0.8188	0.0488	-0.2294	0.5298
TRAFICLAN (Ref: 1 lane)				
2 Lanes	-0.0669	0.9309	-0.7934	0.2596
3 Lanes	-0.4892	0.6105	-1.4631	0.0937
4 Lanes	-0.0186	0.9869	-0.9677	0.1924
≥5 Lanes	0.215	0.8137	-0.9998	0.234
Number of observation= 757, ρ^2 =0.121, χ^2 for likelihood ratio =180.589, P-value for chi square = 0.000				

As one can see from Table 3, train speed was grouped into four categories. As compared to low speed train (less than 25mph), crashes with higher train speed (45-60 mph) had higher probability of resulting in fatal injury. On the other hand, train speed categories 25-45 mph and greater than 60 mph were less likely to result serious injury. Also, as shown in Table 4, the marginal effect of this variable is 0.428 on probability of fatal crash, -0.413 on probability of injury crashes and -0.015 on injury crashes. Therefore, the probability of fatal crash is 0.428 higher when the speed category is 45-60mph. Similarly, the probabilities of injury crashes and no injury crashes are 0.413 and 0.015 less when train speed category is 45-60 mph. The marginal effect for the rest of the variables can be interpreted in similar fashion.

Table 4. Marginal Effects Results for MNLM (Pedestrian-rail Crash)

Variable	P(Fatal)	P(Injury)	P(No injury)
Train speed category 2 (25-45mph)	0.200	-0.231	0.031
Train speed category 3 45-60 mph)	0.312	-0.253	-0.059
Train speed category 4 (>60 mph)	0.233	-0.287	0.054
Cloudy weather	0.119	-0.021	-0.098
Rainy weather	-0.071	-0.057	0.128
Foggy weather	0.325	-0.091	-0.233
Snow weather	0.018	-0.209	0.191
Crush circumstance(rail equipment struck vehicle)	0.258	-0.281	0.023
Pedestrian gender male	-0.116	0.048	0.068
HRGC In city	0.118	-0.011	-0.108
Highway Paved	0.130	0.092	-0.221
HRGC asphalt surface	0.041	-0.032	-0.009
HRGC asphalt and flange surface	-0.119	0.109	0.010
HRGC concrete surface	-0.054	0.151	-0.097
HRGC concrete and rubber surface	0.553	-0.321	-0.232
HRGC rubber surface	0.111	-0.154	0.043
HRGC unconsolidated surface	-0.222	0.202	0.019
HRGC other surface	-0.070	0.019	0.051
Traffic Lane (1 lane)	-0.094	0.023	0.071
Traffic Lane (2 lane)	-0.101	-0.052	0.153
Traffic Lane (3 lane)	-0.098	0.033	0.065
Traffic Lane (>=4 lane)	-0.095	0.077	0.018
Temperature category 2 (50-80F)	0.025	-0.021	-0.004
Temperature category 2 (>80Fmph)	0.003	-0.135	0.132

Among the five weather categories, cloudy weather was found to be statically significant at 90 percent confidence level and had increased the probability of fatal injury as compared to clear weather. As shown in Table 4, the marginal effect of the variable is positive on probability of fatal crashes and negative on probability of injury and no injury crashes.

Two crash circumstances (rail equipment struck pedestrian and pedestrian struck train equipment) were considered. The result showed that crash severity was more likely to be fatal when rail equipment struck a pedestrian as compared to the other crash circumstance under which pedestrian struck the rail equipment although the variable is not statistically significant. It is shown in Table 4 that the marginal effect of this variable on probability of fatal crashes is positive and it is negative on probability of injury and no injury crashes. With respect to pedestrian gender, the result indicated that male pedestrians were less likely to be involved in fatal injury crashes as compared to female pedestrians. The result study revealed that HRGC located in the city increased probability of fatal crashes as compared to those located near the city and it is statistically significant. In addition, HRGC located in the city were more likely to result serious injury though the variable is not statistically significant. The marginal effect of the variable, as shown in Table 4, is negative on the probability of fatal crashes and it is positive on probability of injury and no injury crashes.

Compared to paved highways crossing the rail line, crashes occurring on unpaved highways had higher probability of being fatal. In addition, crashes occurring on unpaved highways crossing rail line were more likely to result in serious injury though this variable is not statistically significant. As shown in Table 4, the marginal effect of this variable is positive on probability of fatal and injury crashes and it is negative on probability of no injury crashes.

Different types of crossing were considered in this study. As showed in Table 3, comparing with timber crossing surfaces, concrete surface type crossings are more likely to be associated with injury crashes and the variable is statistically significant (at 90% confidence level) whereas rubber crossing surfaces are less likely to result in injury crashes. As presented in Table 4, the marginal effect of this variable is negative on probability of injury crashes and it is positive on probability of fatal and no injury crashes.

With regards to temperature, when compared to low temperatures (less than 50^{o}F), high temperatures (greater than 80^{o}F) were less likely to result severe injury and the variable is statistically significant. The marginal effect of this variable is negative on the probability of injury crashes and it is positive on probability of fatal and no injury crashes. Finally, compared to one lane road crossing rail line, three lane highway crossing rail lines were less likely to result in fatal injury and the variable is statistically significant. The marginal effect of this variable on the probability of a fatal crash is negative and it is positive on injury and no injury crashes.

While the model gives results of intercepts and slope coefficients for serious injury and fatality, it can also be interpreted by using the odds ratio which is exponential of parameter estimate obtained from the analysis. For example, the estimated coefficient for train speed category three (45-60 mph) is 1.015 and hence the relative effect of this speed category versus the base train speed category one (<25 mph) is exp (1.015) =2.76. This indicates that the odds of pedestrian crash severity being fatal is 2.76 times higher if the speed of the train is category three compared to train speed category one. Similarly, the parameter estimate of cloudy weather, considering clear weather as a reference, is found to be 0.7851. So, the relative effect of weather inclement cloud compared to clear weather for crash severity fatal is determined as exp(0.7851) =2.193. This indicates that the odds of fatal crash severity versus no injury crashes are 2.193 times higher on cloudy weather compared to the clear weather. The other odds ratio results can also be interpreted in a similar fashion.

The probability of the three different severity levels of pedestrian crashes are determined based on the parameter estimated for the indicator variables and the probability computation equations (9), (10) and (11). Accordingly, the predicted average probability of fatal, injury and no injury severity levels are 0.476, 0.390 and 0.134 respectively. And the observed crash severity from the original data was 0.534, 0.310 and 0.157 for fatality, injury and no injury respectively. Also as shown in Table 3, the ρ^2 determined for the model is 0.121 which indicates the model has improvement over the naive model (model without covariates).

4.3 Multinomial Logit Modeling results for vehicle-rail crash on HRGC

The modeling procedure and the interpretation the MNLM vehicle-rail crash severity modeling is the same as the pedestrian crash severity model. However, some of the variables used in the pedestrian-rail crash severity model are not relevant in the vehicle-rail model and the reverse is true. Table 5 presents the result obtained from this study. In this modeling also, the three vehicle-rail crash severity levels (Fatal crashes, Injury crashes and No Injury crashes) were considered as the dependent variable. Among the three crash severity levels, No injury crashes were considered the base case. Therefore, coefficients estimated for the explanatory variables are values representing the relative effect of contributing factors on fatal or injury crashes compared to no injury crashes. Positive estimates in the model indicate that the chance of injury or fatal crash increase as the value of the independent variables increase.

As shown in Table 5, like the pedestrian-rail crash severity model, some of the variables are not statistically significant. However, for the sake of facilitating interpretation of the results, those variables were retained in the model if at least one of variables/factors in the same parameter category were significant in at least one of the models (injury and/or fatality). This actually induces reduction in efficiency of the model. In order to compensate the reduction in efficiency, a 90 percent confidence level was considered instead of 95 percent.

Table 5. Multinomial Logistic Model Regression Results (Vehicle-rail Crash)

Parameter	Injury		Fatal	
	Estimate	**P- value**	**Estimate**	**P-value**
Intercept	-1.1431	<.0001	-3.9746	<.0001
VEHSPD (Ref:<25mph)				
25-45	0.6998	<.0001	0.8189	<.0001
>45	1.0047	<.0001	1.8483	<.0001
TYPVEH (Ref: Auto)				
Truck	0.0699	0.5684	0.0738	0.6792
Truck-trailer	-0.2492	0.0074	-1.9045	<.0001
Pick-up truck	0.1508	0.0775	0.0288	0.8217
Van	0.0966	0.5273	-0.0279	0.9075
Bus	0.7142	0.4572	-10.9806	0.9838
School bus	1.0413	0.3016	-10.8854	0.9869
TYPACC (Ref: vehicle struck Rail equipment)				
Rail equipment struck vehicle	-0.0869	0.2649	0.7327	<.0001
TEMP(Ref: <50^{o}F)				
50^o-80^{o}F	0.0934	0.2174	-0.00505	0.9669
>80^{o}F	0.278	0.0016	0.2295	0.0987

Table 5. (Cont'd) Multinomial Logistic Model Regression Results (Vehicle-rail Crash)

Parameter	Injury		Fatal	
	Estimate	P-value	Estimate	P-value
WEATHER (Ref: Clear)				
Cloudy	-0.0586	0.4557	-0.1198	0.3455
Rain	-0.1843	0.1691	-0.4356	0.0773
Fog	-0.00811	0.9736	-1.0758	0.0798
Sleet	0.4497	0.4496	-11.843	0.9681
Snow	-0.647	0.0066	-0.7963	0.0601
TRNSPD (Ref: <25mph)				
25-45	0.6762	<.0001	1.7992	<.0001
>45	0.7771	<.0001	2.9039	<.0001
DRIVGEN (Ref: Female)				
Male+Missing	0.3931	<.0001	0.2286	0.0474
DEVELTYP(Ref: Open space area)				
Residential	-0.1968	0.0214	-0.2374	0.0756
Commercial	-0.3298	<.0001	-0.3529	0.0092
Industrial	-0.3918	<.0001	-0.1066	0.5259
Institutional	-0.4723	0.0648	-0.4806	0.2659
XSURFACE(Ref: Timber)				
Asphalt	-0.2359	0.0016	-0.4962	<.0001
Asphalt & Flange	-0.1417	0.3007	-0.4396	0.0588
Concrete	0.0785	0.455	-0.00019	0.9991
Concrete & Rubber	0.0908	0.6262	0.5839	0.0259
Rubber	0.0718	0.6305	-0.0533	0.8331
Metal	-0.5291	0.6725	-11.2314	0.9873
Unconsolidated	-0.3764	0.0241	-0.4246	0.0917
Other	-0.4434	0.2714	-0.6581	0.3933
AADT(Ref:<10,000)				
10,000-20,000	-0.0934	0.4353	-0.6274	0.0061
20,000-30,000	-0.596	0.0114	-0.4924	0.1693
>30,0000	-0.298	0.2415	-0.8557	0.0739

Table 5. (Cont'd) Multinomial Logistic Model Regression Results (Vehicle-rail Crash)

Parameter	Injury		Fatal	
	Estimate	P-value	Estimate	P-value
DRIVAGE(Ref:<25 Years)				
25-60 Years	0.0389	0.6326	-0.00502	0.9694
>60 Years	0.2715	0.0083	0.9462	<.0001
Number of observation= 7,391, ρ^2 =0.106, $\chi 2$ for likelihood ratio =1143.663, P-value for chi square= 0.000				

Based on the parameter estimates obtained in Table 5, the MNL model can be written as follows:

$$\log\left[\frac{P(Y=Fatal)}{P(Y=No\ Injury)}\right] = -3.9746 + 0.8189X_1 + 1.8483X_2 - 1.9045X_3 + 0.7327X_4 + 0.2295X_5 - 0.7963X_6 + 1.7992X_7 + 2.9039X_8 + 0.2286X_9 - 0.2374X_{10} - 0.3529X_{11} - 0.4962X_{12} - 0.4246X_{13} - 0.4924X_{14} + 0.9462X_{15} \quad (34)$$

$$\log\left[\frac{P(Y=Injury)}{P(Y=No\ Injury)}\right] = -1.1431 + 0.6998X_1 + 1.0047X_2 - 0.2492X_3 - 0.0869X_4 + 0.2780X_5 - 0.6470X_6 + 0.6762X_7 + 0.7771X_8 + 0.3931X_9 - 0.1968X_{10} - 0.3298X_{11} - 0.2359X_{12} - 0.3764X_{13} - 0.5960X_{14} + 0.2715X_{15} \quad (35)$$

where:

X_1 = Vehicle speed category (1 if speed train speed is 25-45mph, 0 otherwise)

X_2 = Vehicle speed category (1 if speed train speed is >45mph, 0 otherwise)

X_3 = Vehicle type indicator (1 if vehicle is truck-trailer, 0 otherwise)

X_4 = Type of accident indicator (1 if rail equipment struck vehicle, 0 otherwise)

X_5 = Temperature indicator (1 if temperature is greater than 80^{o}F, 0 otherwise)

X_6 = Weather indicator (1 if snow weather, 0 otherwise)

X_7 = Train speed category (1 if vehicle speed is 25-45mph, 0 otherwise)

X_8 = Train speed category (1 if vehicle speed is >45mph, 0 otherwise)

X_9 = Vehicle driver gender indicator (1 if male or missing, 0 otherwise)

X_{10} =Development area type indicator (1 if residential, 0 otherwise)

X_{11}= Development area type indicator (1 if commercial, 0 otherwise)

X_{12}= HRGC surface type (1 if surface is asphalt, 0 otherwise)

X_{13}= HRGC surface type (1 if surface is unconsolidated, 0 otherwise)

X_{14}= Traffic volume indicator (1 if AADT is 20,000-30,000, 0 otherwise)

X_{15}= Vehicle driver age indicator (1 if age is >60 years, 0 otherwise)

Based on the above MNL model equations, the marginal effect/value is also determined as presented in Table 6 for the explanation convenience. As can be seen in the Table 6, the sum of marginal effect gives zero which satisfies the requirement that the sum of probability is 1. The marginal effect for the remaining variables provides a great deal of valuable information for interpreting results.

As shown in Table 5, vehicle speed was one among several explanatory variables that are considered and used to estimate the vehicle-rail crash severity model. Vehicle speed was categorized in to three levels (<25mph, 25-45mph, and >45 mph). According to the result, two speed categories (25-45 mph and >45 mph) are statistically significant and they had higher probability of resulting in injury and fatal crashes. It was also shown that the parameter estimate for vehicle speed category three (>45 mph) is higher than vehicle speed category two (25-45 mph). This indicates that higher vehicle speed has a detrimental effect of increasing the chance of fatal and injury crashes.

Likewise, train speed was categorized into three levels and also found to be statistically significant. Compared to train speed category one (<25mph), both higher train speed categories (25-45mph and >45 mph) had increased probabilities of injury and fatal crashes. Like vehicle speed, higher train speed has also detrimental effects in increasing the chance of fatal and injury crashes. As shown in Table 6, the marginal effect result indicates that both probabilities of injury and fatal crashes increase as speed of vehicle and train increases. On the other hand, the probability of no injury crashes decreases as speed increases.

Table 6. Marginal Effects Results for MNLM (Vehicle-rail Crash)

Variable	**P(Fatal)**	**P(Injury)**	**P(No injury)**
Vehicle speed category 2 (25-45mph)	0.200	-0.231	0.031
Vehicle speed category 3 (45-60 mph)	0.312	-0.253	-0.059
Vehicle speed category 4 (>60 mph)	0.233	-0.287	0.054
Rainy weather	0.119	-0.021	-0.098
Foggy weather	-0.071	-0.057	0.128
Sleet weather	0.325	-0.091	-0.233
Snow weather	0.018	-0.209	0.191
Crush circumstance (rail equipment struck vehicle)	0.258	-0.281	0.023
Vehicle driver gender male	-0.116	0.048	0.068
Nearest	0.118	-0.011	-0.108
Highway Paved	0.130	0.092	-0.221
HRGC asphalt surface	0.041	-0.032	-0.009
HRGC asphalt & flange surface	-0.119	0.109	0.010
HRGC concrete surface	-0.054	0.151	-0.097
HRGC concrete & rubber surface	0.553	-0.321	-0.232
HRGC rubber surface	0.111	-0.154	0.043
HRGC metal surface	-0.222	0.202	0.019
HRGC unconsolidated surface	-0.070	0.019	0.051
HRGC others	-0.094	0.023	0.071
Traffic volume (AADT of 10,000-20,000)	-0.101	-0.052	0.153
Traffic volume (AADT of 20,000-30,000)	-0.098	0.033	0.065
Traffic volume (AADT of >30,000)	-0.095	0.077	0.018
Temperature category 2 (50-80F)	0.025	-0.021	-0.004
Temperature category 2 (>80Fmph)	0.003	-0.135	0.132

Seven vehicle categories (ranging from automobile to truck-trailer to school bus types) were considered in this study. Among these seven categories, truck-trailer was found to be statistically significant. As shown in Table 5, truck-trailer vehicles were less likely to result in injury and fatal crashes as compared to automobiles. The marginal effect result as shown in Table 6 indicates that truck-trailer vehicles increase injury and no injury crashes while they decrease fatal crashes.

Two crash circumstances (rail equipment struck vehicle and vehicle struck rail equipment) were considered. The crash circumstance under which vehicle struck rail equipment was considered a reference (i.e., base) for comparison. As shown in Table 5, when rail equipment struck vehicle, crash severity were more likely to be fatal. On the other hand, this crash circumstance is less likely to result in injury crashes. As shown in Table 6, the marginal effect results of crash circumstance (i.e., when rail equipment struck vehicle) indicate an increase in the probability of fatal crashes and a decrease in the probability of injury and no injury crashes.

Compared to low temperature (less than 50^oF), vehicle-rail crashes occurring at higher temperature (greater than 80^oF) had increased the probability of injury and fatal crashes. As presented in Table 6, the marginal effect results clearly indicate that higher temperature increases injury and fatal crashes while decreasing no injury crashes. Regarding weather condition, snow weather was found to be statistically significant. As presented in Table 5, snowy weather conditions were less likely to result in injury and fatal crashes as compared to clear weather condition. The marginal effect results of snow weather also show decreases in the probability of injury and fatal crashes and an increase in no injury crashes. In addition, rainy and foggy weather conditions, as compared to a clear weather, were less likely to result in fatal crashes and they were statistically significant.

Five different types of development area types were considered in this study. As compared to open space development areas, HRGCs located in commercial areas and residential areas were less likely to result in injury and fatal crashes and they were found to be statistically significant. The marginal effect results show that HRGCs located in both industrial and commercial areas decrease the probability of injury and fatal crashes while the probability of no injury crashes increases.

Various types of HRGC surfaces were investigated in this study. A timber crossing surface was considered a reference to which other crossing surface types were compared. As shown in Table 5, vehicle-rail crashes occurring on asphalt crossing surface were found to be less likely to result in injury and fatal crashes and the variable was found statistically significant. In addition, unconsolidated crossing surface types were also found to be statistically significant and crashes occurring on such surfaces are less likely to result in injury or fatality. Asphalt and flange crossing surface types were also found to be statistically significant and less likely to result in fatal crashes. As shown in Table 6, the marginal effects of both asphalt and unconsolidated crossing surface types show decrease in the probability of injury and fatal crashes whereas the probability of no injury crashes increases.

The Average Annual Daily Traffic (AADT) was also considered in order to investigate the effect of traffic volume on the crash severity. The AADT was categorized in to four categories. Among the four categories, category three (AADT of 20,000-30,000) was found to be statistically significant and it is less likely to result in injury and fatal crashes compared to category one (AADT less than 10,000). The marginal effect result shows this AADT category decreases the probability of injury and fatal crashes and increases that of no injury crashes.

Vehicle driver characteristics such as age and gender were considered in the study as explanatory variables. With respect to driver gender, as the result revealed, male driver are more likely to be involved in injury and fatal crashes as compared to female drivers and the variable is found to be statistically significant. The age of vehicle drivers was grouped in to three categories. Vehicle driver age below 25 was considered a reference for comparison purpose. As shown in Table 5, drivers with the age of above 60 years had higher probability of being involved in injury and fatal crashes. As shown in Table 6, the marginal effects of male vehicle drivers and age above 60 years clearly increase the probability of injury and fatal crashes while decreasing that of no injury crashes.

In addition to the model results of intercepts and slope coefficients for serious injury and fatality, the model can be interpreted by using the odds ratio which is exponential of parameter estimates obtained from the analysis. For example, the estimated coefficient for train speed category three (>45 mph) is 1.0047 and hence the relative effect of this speed category versus train speed category one (< 25mph) is exp (1.015)=2.76. This indicates that the odds of pedestrian crash severity being injury is 2.76 times higher if the speed of the train is category three compared to train speed category one. Similarly, the parameter estimate of vehicle driver age above 60 years, considering diver age below 25 years as a reference, is found to be 0.2715. So, the relative effect of drivers age of above 60 years to age of below 25 years on injury crashes is determined as exp(0.2715) =1.31. This indicates that the odds of injury crash severity versus no injury crashes are 1.31 times higher for drivers with the age of above 60 years compared to driver age below 25 years. The odds ratio results of the rest of variable can also be interpreted in a similar fashion.

The probability of the three different severity levels of vehicle-rail crashes are determined based on the parameter estimated for the indicator variables and the probability equations shown above. Accordingly, the predicted average probability of fatal, injury and no injury severity levels are 0.131, 0.190 and 0.679 respectively. And the observed crash severity from the original data was 0.085, 0.267 and 0.648 for fatality, injury and no injury respectively. Also as shown in Table 5, the ρ^2 determined for the model is 0.106 which indicates the model has improvement over the naïve model (model without covariates).

4.4 Ordered Logit Modeling results for pedestrian-rail crash on HRGC

Table 7 and 8 present the result obtained from OLM for pedestrian-rail crash severity on HRGC. The three vehicle-rail crash severity levels (Fatal crashes, Injury crashes and No Injury crashes) were considered as the dependent variable. The interpretation of the coefficient is different from the MNL model. A positive coefficient indicates that increase in the value of a variable will increase the probability of highest severity level (i.e. fatal) and decrease the lowest severity level (no injury). On the other hand, a negative coefficient indicates that a decrease in the value of a variable will increase the probability of the highest severity level and decrease probability of lowest severity level. For the intermediate severity level (i.e. injury), an increase in the value of a variable may decrease or increase the probability of occurring. As shown in Table 7, some of the variables are not statistically significant. However, for the sake facilitating interpretation of the results, those variables were retained in the model if at least one of the categories in the same factor was significant.

Based on the parameter estimates obtained in Table 7, the ordered logit model can be written as follows:

$$\log\left[\frac{P(Y=Fatal)}{1-P(Y=No\ Injury)}\right] = -1.5465 + 0.6905X_1 + 1.2315X_2 + 0.8999X_3 + 0.5204X_4 + 1.8357X_5 + 0.1593X_6 - 0.4919X_7 + 0.6462X_8 + 0.7651 - 1.5517X_{10} - 1.0547X_{11} - 0.7940X_{12} \quad (36)$$

$$\log\left[\frac{P(Y=Fatal)+P(Y=Injury)}{P(Y=No\ Injury)}\right] = 0.1298 + 0.6905X_1 + 1.2315X_2 + 0.8999X_3 + 0.5204X_4 + 1.8357X_5 + 0.1593X_6 - 0.4919X_7 + 0.6462X_8 + 0.7651 - 1.5517X_{10} - 1.0547X_{11} - 0.7940X_{12} \quad (37)$$

where:

X_1 = Train speed category (1 if speed train speed is 25-45mph, 0 otherwise)

X_2 = Train speed category (1 if speed train speed is >45mph, 0 otherwise)

X_3 = Train speed category (1 if speed train speed is >60mph, 0 otherwise)

X_4 = Weather indicator (1 if cloudy weather, 0 otherwise)

X_5 = Weather indicator (1 if foggy weather, 0 otherwise)

X_6 = Type of accident indicator (1 if rail equipment struck pedestrian, 0 otherwise)

X_7 = Pedestrian gender indicator (1 if male, 0 otherwise)

X_8 = HRGC location indicator (1 if located in city, 0 otherwise)

X_9 = Highway surface type indicator (1 if unpaved, 0 otherwise)

X_{10}= HRGC surface type (1 if surface is other, 0 otherwise)

X_{11} =No. of Highway traffic lane indicator (1 if 3 lane highway, 0 otherwise)

X_{12} = Highway surface type indicator (1 if 4 lane highway, 0 otherwise)

Based on the above ordered logit model equations, the marginal effects/values are also determined as presented in Table 8. As seen, the sum of the marginal effect gives zero which satisfies the requirement that the sum of probability is 1. The marginal effect for the remaining variables provides a great deal of valuable information for results interpretation.

As depicted in Table 7, Train speed was among several explanatory variables that are considered and used to estimate the vehicle-rail ordered logit crash severity model. Train speed was categorized in to four levels, <25 mph, 25-45 mph, 45 mph-60 mph and >60 mph. According to the result, the three higher train speed categories are statistically significant as they are positive coefficients indicating that increase in the train speed would result in an increase in the probability of higher level severity crashes and decrease in the lower severity level crashes. In addition, as can be seen in Table 8, the marginal effect of these higher train speed categories showed that they had increased the probability fatal crashes and had decreasing effect on the probability of injury and no injury severity levels. This indicates that higher vehicle speed has a detrimental effect of increasing the chance of fatal and injury crashes.

Table 7. Ordered Responses Logistic Model Regression Results (Pedestrian-rail Crash)

Parameter	Estimate	P-value
Intercept (1)	-1.5465	0.0192
Intercept (2)	0.1298	0.8435
TRNSPD (Ref: <25mph)		
25-45	0.6905	0.0002
45-60	1.2315	<.0001
>60	0.8999	0.0002
WEATHER (Ref: Clear)		
Cloudy	0.5204	0.008
Rain	-0.403	0.2496
Fog	1.8357	0.0953
Snow	0.2683	0.711
TYPACC (Ref: Pedestrian struck rail equipment)		
Rail equipment struck pedestrian	0.1593	0.0277
PEDESTRNGEN (Ref: Female)		
Male+Missing gender	-0.4919	0.0108
NEAREST (Ref: Near City)		
In City	0.6462	0.0036
HWYPVED (Ref: Paved)		
Unpaved	0.7651	0.0799
XSURFACE (Ref: Timber)		
Asphalt	0.1691	0.3107
Asphalt & Flange	-0.4619	0.2487
Concrete	-0.1336	0.6801
Concrete & Rubber	11.2459	0.9786
Rubber	0.3306	0.1906
Unconsolidated	-0.757	0.422
Other	-1.5517	0.0708
TRAFICLAN (Ref: 1 lane)		
2 Lanes	-0.6243	0.1797
3 Lanes	-1.0547	0.0748
4 Lanes	-0.794	0.1054
≥5 Lanes	-0.8469	0.1235
Likelihood Ratio Test : $\chi^2 = 83.145(22\ d.f.)$; p-value is <0.0001 Score Test for Proportional Odds Assumption $\chi^2 = 81.6159(37\ d.f.)$; p-value is <0.0001 Akaike Information Criterion (AIC) =1462.04		

Two crash circumstances (when the rail equipment struck a vehicle and when a vehicle struck the train equipment) were considered. The crash circumstance under which the vehicle struck the rail equipment was considered as a reference for comparison. As shown in Table 7, the variable is statistically significant and the coefficient is positive indicating that the crash circumstance had increased the probability fatal severity crashes. As shown in Table 8, the marginal effect results of the crash circumstance (i.e., when the rail equipment struck vehicle) indicates an increase in the probability of fatal crashes and a decrease in the probability of injury and no injury crashes.

Regarding weather condition, cloudy and foggy weather conditions were found to be statistically significant. As presented in Table 7, cloudy and foggy weather was observed to have an increasing effect on the fatal crashes as compared to clear weather condition. In addition, as shown in table 8, the marginal effect results of both cloudy and foggy weather showed an increase in the probability of fatal crashes and an decrease in injury and no injury crashes.

Various types of HRGC surfaces were investigated in this study. A timber crossing surface was considered as a reference to which other surface types were compared. As shown in Table 7, pedestrian-rail crashes occurring on other crossing surface types, such as metallic, were found to be statistically significant. As shown in Table 8, the marginal effects of other crossing surface types show a decrease in the probability of fatal and injury crashes whereas the probability of no injury crashes increased. This indicates that crossing surface type is associated with lower level severity pedestrian crash on HRGCs.

Table 8. Marginal Effects Results for OLM (Pedestrian-rail Crash)

Variable	**P(Fatal)**	**P(Injury)**	**P(No injury)**
Train speed category 2 (25-45mph)	0.148	-0.041	-0.107
Train speed category 3 (>45mph)	0.275	-0.112	-0.163
Train speed category 4 (>45mph)	0.198	-0.075	-0.123
Cloudy weather	0.117	-0.041	-0.076
Rainy weather	-0.086	0.016	0.070
Foggy weather	0.371	-0.195	-0.176
Snow weather	0.060	-0.020	-0.040
Crush circumstance(rail equipment struck vehicle)	0.145	-0.018	-0.127
Pedestrian gender male	-0.111	0.038	0.073
HRGC In city	0.147	-0.056	-0.092
Highway Paved	0.158	-0.015	-0.142
HRGC asphalt surface	0.038	-0.011	-0.027
HRGC asphalt and flange surface	-0.099	0.017	0.082
HRGC concrete surface	-0.029	0.007	0.022
HRGC concrete and rubber surface	0.570	-0.346	-0.223
HRGC rubber surface	0.074	-0.024	-0.050
HRGC unconsolidated surface	-0.156	0.014	0.142
HRGC other surface	-0.278	-0.041	0.319
2 Lanes	-0.136	0.040	0.096
3 Lanes	-0.209	0.005	0.203
4 Lanes	-0.167	0.029	0.138
≥5 Lanes	-0.173	0.015	0.158

The effect of number of highway traffic lanes crossing the railway on the various crash severity levels was also considered. The numbers of highway traffic lanes were itemized in to five categories. Among the five categories, a one lane road is considered as a reference for comparison. As shown in Table 7, three lane and four lane highways were found to be statistically significant and the coefficient is negative indicating that they increase the probability of no injury crashes and decrease the probability of fatal crashes.

The marginal effect result in Table 8 indicate the two categories decreases the probability of fatal crashes and increases that of injury and no injury crashes.

The variable corresponding to pedestrian gender was found to be statistically significant and the coefficient for the male pedestrian is negative implying that it increases the probability of no injury crashes and decreases the probability of fatal crashes as compared to female pedestrians. The marginal effect for this variable, as shown in Table 8, being a male pedestrian increases the probability of injury and no injury crashes and decreases the probability of a fatal crash.

The result of the study revealed that HRGC located in the city increased the probability of fatal crashes as compared to those located near the city and it is statistically significant. The marginal effect of the variable, as shown in Table 8, is positive on the probability of fatal crashes and it is negative on the probability of injury and no injury crashes. Compared to paved highways crossing the rail line, crashes occurring on unpaved highways had higher probability of being fatal. As shown in Table 8, the marginal effect of this variable is positive on probability of fatal and injury crashes and it is negative on probability of no injury crashes.

In addition to the model results of intercepts and slope coefficients for injury and fatality, the model can be interpreted by using the odds ratio which is the exponential of the parameter estimate obtained from the analysis. For example, the estimated coefficient for train speed category three (>45 mph) is 1.4234 and hence the relative effect of this speed category versus train speed category one (<25 mph) is exp (1.4234) = 4.151. This indicates that the odds of pedestrian crash severity being injury is 4.151times higher if the speed of the train is category three compared to train speed category one. Similarly, the parameter estimate of a vehicle driver above age 60 years, considering diver age below 25 years as a reference, is found to be 0.2715. So the relative effect of drivers over 60 years to age of below 25 years on injury crashes is determined as exp(0.5519) =1.737. This indicates that the odds of no injury crash severity versus injury and fatal crashes are 1.737 times higher on drivers with the age not above 60 years compared to driver below 25 years of age. The other odds ratio results can also be interpreted in similar fashion.

The probability of the three different severity levels of rail-vehicle crashes are determined based on the parameter estimated for the indicator variables and the probability equation shown above. Accordingly, the predicted average probability of fatal, injury and no injury severity levels are 0.431, 0.346 and 0.223 respectively. And the observed crash severity from the original data was 0.534, 0.310 and 0.157 for fatality, injury and no injury respectively.

Also as shown in Table 7, the ρ^2 determined for the model is 0.078 which indicates the model has some improvement over the naïve model (model without covariates). Moreover, the test score for the proportional odds assumption has a *p*-value of 0.0001 (22 degrees of freedom), which indicates that the proportional odds model adequately fits the data because the hypothesis that the regression lines for cumulative logits are parallel is not rejected. The likelihood ratio test *p*-value of <0.0001 (22 degrees of freedom) indicates that the null hypothesis is rejected, and the conclusion is that the predictor variables given in the model affect the severity of vehicle-rail crashes, or the model with independent variables is statistically better than the model with only the intercept.

4.5 Ordered Logit Modeling results for vehicle-rail crash on HRGC

Table 9 and 10 present the result obtained from this study. The three vehicle-rail crash severity levels (Fatal crashes, Injury crashes and No Injury crashes) were considered as the dependent variable. The interpretation of the coefficient is different from the MNL model. A positive coefficient indicates that an increase in the value of a variable will increase the probability of highest severity level (fatal) and decrease the lowest level severity level (no injury). On the other hand, a negative coefficient indicates that a decrease in the variable will increase the probability of the highest severity level and decrease the probability of the lowest severity level. For the intermediate severity level (injury), an increase in the value of a variable may decrease or increase the probability of it occurring. As shown in Table 9, some of the variables are not statistically significant. However, for the sake of facilitating the interpretation of the results, those variables were retained in the model if at least one of the categories in the same factor was significant.

Based on the parameter estimates obtained in Table 9, the ordered logit model can be written as follows:

$$\log\left[\frac{P(Y=Fatal)}{1-P(Y=No\ Injury)}\right] = -3.1208 + 0.6797X_1 + 1.1451X_2 - 0.6734X_3 + 0.1593X_4 + 0.2375X_5 - 0.2602X_6 - 0.7047X_7 + 0.846X_8 + 1.4234X_9 + 0.5519X_{10} + 0.3033X_{11} - 0.2013X_{12} - 0.3161X_{13} - 0.271X_{14} - 0.4916X_{15} - 0.3054X_{16} - 0.359X_{17} - 0.2381X_{18} - 0.52981X_{19} - 0.4453X_{20} \quad (38)$$

$$\log\left[\frac{\mathrm{P(Y=Fatal)+P(Y=Injury)}}{\mathrm{P(Y=No\ Injury)}}\right] = -1.1575 + 0.6797\mathrm{X}_1 + 1.1451\mathrm{X}_2 - 0.6734\mathrm{X}_3 +$$

$$0.1593X_4 + \quad 0.2375X_5 - 0.2602X_6 - 0.7047X_7 + 0.846X_8 + 1.4234X_9 + \ 0.5519X_{10} + 0.3033X_{11} - 0.2013X_{12} - 0.3161X_{13} - 0.271X_{14} - 0.4916X_{15} - 0.3054X_{16} - 0.359X_{17} - 0.2381X_{18} - 0.52981X_{19} - 0.4453X_{20} \qquad (39)$$

where:

X_1 = Vehicle speed category (1 if speed train speed is 25-45mph, 0 otherwise)

X_2 = Vehicle speed category (1 if speed train speed is >45mph, 0 otherwise)

X_3 = Vehicle type indicator (1 if vehicle is truck-trailer, 0 otherwise)

X_4 = Type of accident indicator (1 if rail equipment struck vehicle, 0 otherwise)

X_5 = Temperature indicator (1 if temperature is greater than 80^oF, 0 otherwise)

X_6 = Weather indicator (1 if rainy weather, 0 otherwise)

X_7 = Weather indicator (1 if snow weather, 0 otherwise)

X_8 = Train speed category (1 if vehicle speed is 25-45mph, 0 otherwise)

X_9 = Train speed category (1 if vehicle speed is >45mph, 0 otherwise)

$X_{1\ 0}$=Vehicle driver age indicator (1 if age is >60 years, 0 otherwise)

X_{11} =Vehicle driver gender indicator (1 if male, 0 otherwise)

X_{12} =Development area type indicator (1 if residential, 0 otherwise)

X_{13}= Development area type indicator (1 if commercial, 0 otherwise)

X_{14}= Development area type indicator (1 if industrial, 0 otherwise)

X_{15}= Development area type indicator (1 if institutional, 0 otherwise)

X_{16}= HRGC surface type (1 if surface is concrete and rubber, 0 otherwise)

X_{17}= HRGC surface type (1 if surface is unconsolidated, 0 otherwise)

X_{18}= Traffic volume indicator (1 if AADT is 10,000-20,000, 0 otherwise)

X_{19}= Traffic volume indicator (1 if AADT is 20,000-30,000, 0 otherwise)

X_{20}= Traffic volume indicator (1 if AADT is >30,000, 0 otherwise)

Based on the above ordered logit model equations, the marginal effect/value is also determined as presented in Table 10 for the explanation convenience. As can be seen in the Table 9, the sum of marginal effect gives zero which satisfies the requirement that the sum of probability is 1. The marginal effect for the remaining variables provides a great deal of valuable information for results interpretation.

Vehicle speed and train speed, shown in Table 9, were among several explanatory variables that are considered and used to estimate the vehicle-rail crash severity model. Vehicle speed was categorized into three levels (<25mph, 25-45 mph, and >45 mph). According to the results, two speed categories (25-45mph and >45mph) are statistically significant and they had higher probability of resulting in injury and fatal crashes. It was also shown that the parameter estimate for vehicle speed category three (>45mph) is higher than vehicle speed category two (25-45 mph). This indicates that higher vehicle speed has detrimental effects of increasing the chance of fatal and injury crashes. Likewise, train speed was categorized into three categories and found to be statistically significant. Compared to train speed category one (<25 mph), both higher train speed categories (25-45 mph and >45 mph) had increased probabilities of injury and fatal crashes. Like vehicle speed, higher train speed has also detrimental effects in increasing the chance of fatal and injury crashes. As shown in Table 10, the marginal effect result indicates that both probabilities of injury and fatal crashes increase as speed of vehicle and train increases. On the other hand, the probability of no injury crashes decreases as speed increases.

Seven vehicle categories (ranging from automobile to truck-trailer) were considered in this study. Among these seven, truck-trailer was found to be statistically significant. As shown in Table 9, truck-trailer vehicles were less likely to result injury and fatal crashes as compared to automobiles. The marginal effect result as shown in Table 10 indicates that truck-trailer vehicles increase no injury crashes while they decrease injury related and fatal crashes.

Table 9. Ordered Responses Logistic Model Regression Results (Vehicle-rail Crash)

Parameter	Estimate	P-value
Intercept (1)	-3.121	<.0001
Intercept (2)	-1.158	<.0001
VEHSPD (Ref:<25mph)		
25-45	0.6797	<.0001
>45	1.1451	<.0001
TYPVEH (Ref: Auto)		
Truck	0.0666	0.5338
Truck-trailer	-0.6734	<.0001
Pick-up truck	0.0984	0.1911
Van	0.0522	0.7023
Bus	0.423	0.659
School bus	0.5357	0.586
TYPACC (Ref: vehicle struck Rail equipment)		
Rail equipment struck vehicle	0.1593	0.0277
TEMP(Ref: <50oF)		
50-80	0.0493	0.468
>80	0.2375	0.0025
WEATHER (Ref: Clear)		
Cloudy	-0.0898	0.2047
Rain	-0.2602	0.0358
Fog	-0.25	0.2892
Sleet	-0.0975	0.8719
Snow	-0.7047	0.0012

Table 9. (Cont'd) Ordered Responses Logistic Model Regression Results (Vehicle-rail Crash)

Parameter	Estimate	P-value
TRNSPD(Ref: <25mph)		
25-45	0.846	<.0001
>45	1.4234	<.0001
DRIVAGE(Ref:<25ZYears)		
25-60Yeras	0.0229	0.754
>60Years	0.5519	<.0001
DRIVGEN (Ref: Female)		
Male+Missing	0.3033	<.0001
DEVELTYP(Ref: Open space area)		
Residential	-0.2013	0.0085
Commercial	-0.3161	<.0001
Industrial	-0.271	0.0025
Institutional (schools, hospital etc.)	-0.4916	0.0351
XSURFACE(Ref: Timber)		
Asphalt	-0.3054	<.0001
Asphalt & Flange	-0.2412	0.0532
Concrete	0.0554	0.5502
Concrete & Rubber	0.2976	0.0619
Rubber	0.0247	0.8555
Metal	-0.6736	0.5949
Unconsolidated	-0.359	0.0149
Other	-0.4474	0.2265
AADT(Ref:<10,000)		
10,000-20,000	-0.2381	0.0325
20,000-30,000	-0.5291	0.0103
>30,0000	-0.4453	0.0575
Likelihood Ratio Test : $\chi^2 = 893.5693(37\ d.f.)$; p-value is <0.0001 Score Test for Proportional Odds Assumption $\chi^2 = 217.3357(37\ d.f.)$; p-value is <0.0001 Akaike Information Criterion (AIC) =9988.045 -2LogL=9912.045		

Two crash circumstances (rail equipment struck by vehicle and vehicle struck by train equipment) were considered. The crash circumstance under which vehicle struck the rail equipment was considered a reference (base) for comparison. As shown in Table 9, when rail equipment struck the vehicle, the severity of the crash was more likely to result in injury and fatality. On the other hand, this crash circumstance is less likely to result in injury crashes. As shown in Table 10, the marginal effect results of the crash circumstance (i.e., when the rail equipment struck the vehicle) indicate an increase in the probability of injury and fatal crashes and a decrease in the probability of no injury crashes.

Five different types of development area types were considered in this study. As compared to open space development areas, HRGCs located in commercial areas, residential areas, industrial areas and institutional areas were less likely to result in injury and fatal crashes and they were found to be statistically significant. The marginal effect results show that HRGCs located in those developed areas decrease the probability of injury and fatal crashes while the probability of no injury crashes increases.

As compared to low temperatures (less than 50^{o}F), vehicle-rail crashes occurring at higher temperatures (greater than 80^{o}F) had increased probability of injury and fatal crashes. As presented in Table 10, the marginal effect results clearly indicate that higher temperature increases injury and fatal crashes while decreasing no injury crashes.

Regarding weather condition, rainy weather and snow weather were found to be statistically significant. As presented in Table 9, both rain and snow were less likely to result in injury and fatal crashes as compared to clear weather condition. The marginal effect results of snow also show decreases in the probability of injury and fatal crashes and an increase in no injury crashes.

Various types of HRGC surfaces were investigated in this study. Timber crossing surface was considered as a reference to which other surface types were compared. As shown in Table 9, vehicle-rail crashes occurring on asphalt crossing surface were found to be less likely to result in injury and fatality and the variable was found to be statistically significant. In addition, unconsolidated crossing surface types were also found to statistically significant and crashes occurring on such surfaces are less likely to result in injury or fatal. As shown in Table 10, the marginal effects of both asphalt and unconsolidated crossing surface types show a decrease in the probability of injury and fatal crashes whereas the probability of no injury crashes increased.

Table 10. Marginal Effects Results for OLM (Vehicle-rail Crash)

Variable	**P(Fatal)**	**P(Injury)**	**P(No injury)**
Indicator for vehicle speed is category 2 (25-45mph)	0.003	0.011	-0.013
Indicator for vehicle speed is category 3 (>45mph)	0.004	0.018	-0.023
Indicator for vehicle type truck-trailer	-0.003	-0.011	0.013
Indicator for rail equipment struck vehicle	0.001	0.003	-0.003
Indicator for higher temperature (>80^{o}F)	0.001	0.004	-0.005
Indicator for rainy weather	-0.001	-0.004	0.005
Indicator for snow weather	-0.003	-0.011	0.014
Indicator for train speed is category 2 (25-45mph)	0.003	0.014	-0.017
Indicator for train speed is category 3 (>45mph)	0.005	0.023	-0.028
Indicator for vehicle driver age >60 years	0.002	0.009	-0.011
Indicator for vehicle driver gender male	0.001	0.005	-0.006
Indicator for residential development area type	-0.001	-0.003	0.004
Indicator for commercial development area type	-0.001	-0.005	0.006
Indicator for industrial development area type	-0.001	-0.004	0.005
Indicator for institutional development area type	-0.002	-0.008	0.010
Indicator for HRGC concrete and rubber surface type	-0.001	-0.005	0.006
Indicator for HRGC unconsolidated surface type	-0.001	-0.006	0.007
Indicator for traffic volume (AADT of 10,000-20,000)	-0.001	-0.004	0.005
Indicator for traffic volume (AADT of 20,000-30,000)	-0.002	-0.008	0.010
Indicator for traffic volume (AADT of >30,000)	-0.002	-0.007	0.009

The Average Annual Daily Traffic (AADT) was also considered in order to investigate the effect of traffic volume on the crash severity. The AADT was categorized into four categories. Among the four categories, the three categories (AADT of 10,000-20,000, 20,000-30,000 and greater than 30,000) were found to be statistically significant and less likely to result injury and fatal crashes compared to category one (AADT less than 10,000). The marginal effect result shows this AADT category decreases the probability of injury and fatal crashes and increases that of no injury crashes.

Vehicle driver characteristics such as age and gender were considered in the study as explanatory variables. With respect to driver gender, as the result revealed, male drivers are more likely to be involved in injury and fatal crashes as compared to female drivers and the variable is found to be statistically significant. The age of vehicle drivers was grouped into three categories and age below 25 was considered as a reference for comparison purpose. As shown in Table 9, drivers with an age above 60 years had higher probability to being involved in injury and fatal crashes. As shown in Table 10, the marginal effects of male vehicle drivers and people above age 60 clearly increase the probability of injury and fatal crashes while decreasing that of no injury crashes.

In addition to the model results of intercepts and slope coefficients for injury and fatality, the model can be interpreted by using the odds ratio which is the exponential of the parameter estimate obtained from the analysis. For example, the estimated coefficient for train speed category three (>45mph) is 1.4234 and hence the relative effect of this speed category versus train speed category one (<25mph) is exp (1.4234) = 4.151. This indicates that the odds of pedestrian crash severity being injury is 4.151 times higher if the speed of the train is category three compared to train speed category one. Similarly, the parameter estimate of vehicle driver age above 60 years, considering diver age below 25 years as a reference, is found to be 0.2715. So, the relative effect of drivers above 60 years to that of drivers below 25 for injury related crashes is determined as exp(0.5519) =1.737. This indicates that the odds of no injury crash severity versus injury and fatal crashes are 1.737 times higher on drivers with the age of not above 60 years compared to driver age below 25 years. The other odds ratio results can also be interpreted in a similar fashion.

The probability of the three different severity levels of rail-vehicle crashes are determined based on the parameter estimated for the indicator variables and the probability equation shown above. Accordingly, the predicted average probability of fatal, injury and no injury severity levels are 0.090, 0.375 and 0.625 respectively. And the observed crash severity from the original data was 0.085, 0.267 and 0.648 for fatality, injury and no injury respectively. Also as shown in Table 9, the ρ^2 determined for the model is 0.083 which indicates the model has some improvement over the naïve model (model without covariates). Moreover, the test score for the proportional odds assumption has a *p*-value of 0.0001 (37 degrees of freedom), indicating that the proportional odds model adequately fits the data because the hypothesis that the regression lines for cumulative logits are parallel is not rejected. The likelihood ratio test *p*-value of <0.0001 (37 degrees of freedom) indicates that the null hypothesis is rejected, and the conclusion is that the predictor variables given in the model affect the severity of vehicle-rail crashes, or the model with independent variables is statistically better than the model with only the intercept.

4.6 Pedestrian-rail crash severity model comparison

For the MNLM, the lowest injury severity level (no injury) was considered as a comparison group. Therefore, the estimated coefficient of injury and fatal severity models is compared to the case of no injury severity level. A positive estimated coefficient indicates that the probability of injury or a fatal crash increased as compared to the no injury severity level. Furthermore, a negative estimated coefficient indicated that the probability of injury and fatal crash decreased as compared to the base case.

The coefficient estimated for the ordered logistic model is as presented Table 7. The interpretation of the coefficient is different from the MNL model. A positive coefficient indicates that an increase in the value of a variable will increase the probability of the highest level severity level (fatal) and decrease the lowest level severity level (no injury). On the other hand, a negative coefficient indicates that a decrease in the variable will increase the probability of the highest severity level and decrease probability of lowest severity level. For the intermediate severity level (injury), an increase in the value of a variable may decrease or increase the probability of occurring.

Assessment and comparison of the two models cannot be performed simply based on the estimated coefficients of the models. Marginal effect of the variables on the probability of severity levels is computed for the two models in Table 4 and Table 6 and used for comparison purpose. The positive sign in estimated marginal effect indicates that the probability of a given crash severity level increases when the variable changes and the converse is true for negative sign. And the value of the number indicates the magnitude of shift in the probability.

The shifting direction of the probability in the two models was used for comparison of the impacts of each variable on the probability of injury severity outcomes as shown in Table 11. As the results indicate, all the variables are consistent except the variable crash circumstance for the case of intermediate severity level (injury).

Table 11. Comparison of Marginal Effect on Variables for OLM and MNLM (Pedestrian-rail Crash)

Variable	**Fatal**		**Injury**		**No injury**	
	OLM	**MNLM**	**OLM**	**MNLM**	**OLM**	**MNLM**
Train speed category 2 (25-45mph)	+	+	-	-	-	+
Train speed category 3 (>45mph)	+	+	-	-	-	-
Train speed category 4 (>45mph)	+	+	-	-	-	+
Cloudy weather	+	+	-	-	-	-
Rainy weather	-	-	-	-	+	+
Foggy weather	+	+	-	-	-	-
Snow weather	+	+	-	-	-	-
Crush circumstance(rail equipment struck vehicle)	+	+	-	-	-	+
Pedestrian gender male	-	-	+	+	+	+
HRGC In city	+	+	-	-	-	-
Highway Paved	+	+	-	+	-	-
HRGC asphalt surface	+	+	-	-	-	-
HRGC asphalt and flange surface	-	-	+	+	+	+
HRGC concrete surface	-	-	+	+	+	+
HRGC concrete and rubber surface	+	+	-	-	-	-
HRGC rubber surface	+	+	-	-	-	+
HRGC unconsolidated surface	-	-	+	+	+	+
HRGC other surface	-	-	-	+	+	+
Traffic Lane (1 lane)	-	-	+	+	+	+
Traffic Lane (2 lane)	-	-	+	-	+	+
Traffic Lane (3 lane)	-	-	+	+	+	+
Traffic Lane (>=4 lane)	-	-	+	+	+	+
Temperature category 2 (50-80^{o}F)		+		-		-
Temperature category 3 (>80^{o}F)		+		-		+

Empty cells indicate that the variable is not significant even at the 90 percent confidence level.

Both temperature category 2 and 3 found to be not statistically significant in the OLM case where as they are statistically significant in the MNLM case. The other variables have the same effect on the probability of severity levels except variable representing crash circumstance. This indicates that the variables in two models have almost similar effect on the probability of crash severity levels. Another method to compare the two models is application of the Akaike Information Criteria (AIC). The AIC value of the two models is 9,810 and 9,988 for the MNLM and OLM respectively. The larger the AIC value the stronger the model is in estimating the coefficients. Thus, our result indicates that the OLM is better than the MNLM.

4.7 Vehicle crash severity model comparison

Variables obtained from the crossing inventory and crash data were used in developing the nominal response MNLM and ordered logistic regression model. During the final preferred model development process, some of the variables were found to be statistically insignificant and hence removed in a stepwise manner. PROC LOGISTIC procedures were applied with significance level being 0.1 to retain some of the variables. Table 7 and 9 present the result obtained from this study. The three vehicle-rail crash severity levels (Fatal crashes, Injury crashes and No Injury crashes) were considered as the dependent variable. Among the three crash severity levels, no injury crashes were considered the base case. Therefore, coefficients estimated for the explanatory variables are values representing the relative effect of contributing factors on fatal or injury crashes compared to no injury crashes. Positive estimates in the model indicate that the chance of injury or fatal crash increase as the value of the independent variables increases. As shown in Table 7, some of the variables are not statistically significant. However, for the sake of facilitating interpretation of the results, those variables were retained in the model if at least one of variables/factors in the same parameter category were significant in at least one of the models (injury and/or fatality). This actually induces reduction in efficiency of the model.

For the MNLM, the lowest injury severity level (no injury) was considered as a comparison group. Therefore, the estimated coefficient of injury and fatal severity models is as compared to no injury severity level. A positive estimated coefficient indicates that the probability of injury or fatal crash increased as compared to the no injury severity level. Furthermore, a negative estimated coefficient indicated that the probability of a injury or fatal crash decreased as compared to the base case.

The coefficient estimated for the ordered logistic model is as presented Table 9. The interpretation of the coefficient is different from the MNL model. A positive coefficient indicates that increase in the value of a variable will increase the probability of highest level severity level (fatal) and decrease the lowest level severity level (no injury). On the other hand, a negative coefficient indicates that a decrease in the value a variable will increase the probability of the highest severity level and decrease probability of lowest severity level. For the intermediate severity level (injury), an increase in the value of a variable may decrease or increase the probability of occurrence.

Assessment and comparison of the two models cannot be performed simply based on the estimated coefficients of the models. Instead, marginal effect of the variables on the probability of severity levels is computed for the two models as shown in Table 8 and

Table 10. The positive sign in the estimated marginal effect indicates that the probability of a given crash severity level increases when the variable changes and the converse is true for negative sign. And the value of the number indicates the magnitude of shift in the probability.

Table 12. Comparison of Marginal Effect on Variables for OLM and MNLM (Vehicle-rail Crash)

Variable	Fatal		Injury		No injury	
	OLM	MNLM	OLM	MNLM	OLM	MNLM
Vehicle speed category 2 (25-45mph)	+	+	+	+	-	-
Vehicle speed is category 3 (>45mph)	+	+	+	+	-	-
Vehicle type truck-trailer	-	-	-	+	+	+
Circumstance rail equipment struck	+	+	+	-	-	-
Temperature (>80oF)	+	+	+	+	-	-
Rainy weather	-		-		+	
snow weather	-	-	-	-	+	+
Train speed category 2 (25-45mph)	+	+	+	+	-	-
Train speed category 3 (>45mph)	+	+	+	+	-	-
Vehicle driver age >60 years	+	+	+	+	-	-
Vehicle driver gender male	+	+	+	+	-	-
Residential development area type	-		-		+	
Commercial development area type	-	-	-	-	+	+
Industrial development area type	-	-	-	-	+	+
Institutional development area type	-		-		+	
HRGC asphalt surface type	-	-	-	-	+	+
HRGC unconsolidated surface type	-	-	-	-	+	+
Traffic volume (AADT of 10,000-20,000)	-		-		+	
Traffic volume (AADT of 20,000-30,000)	-	-	-	-	+	+
Traffic volume (AADT of >30,000)	-		-		+	

The shifting direction of the probability in the two models was used for comparison of the impacts of each variable on the probability of injury severity outcomes as shown in Table 12. The result indicates, that all the variables are consistent except crash circumstance for the case of intermediate severity level (injury). Empty cell indicate that the variable is not significant for 90 percent confidence level.

Some of the variables in the MNLM namely rainy weather, residential development area type, institutional development type, AADT of 10,000-20,000 and AADT of >30,000 are found not to be statically significant where as they are statistically significant in the case OLM. The other variables have the same effect on the probability of severity levels except variable representing a crash. This indicates that the variables in two models have almost similar effect on the probability of crash severity levels. However, the AIC of the two models is 9,810 and 9,988 for the MNLM and OLM respectively. This indicates that the OLM is better than the MNLM.

Chapter Five

Conclusion

5.1 Research summary

Highway user crash severity levels of at-grade highway-rail crossing were modeled using logistic regression techniques. In addition, comparison was conducted between the MNLM and OLM crash severity level models that are developed using the same data set. As described in the methodology, only vehicle and pedestrian crashes on HRGC were considered in this research. The three crash severity levels, fatality, injury and no injury were considered as dependent variables. Pedestrian characteristics, vehicle and vehicle user characteristics, environmental factors, type of development area, highway-rail crossing characteristics, highway traffic characteristics, vehicle speed and train speed were the explanatory variables used in predicting the crash severity levels. The analysis was conducted using SAS PROC LOGISTICS procedure. In order to retain some of the variables, those within 90 percent confidence level were considered statistically significant. Some of the variables were found to be statistically significant even at 95 percent confidence level.

The main goals of this research were to model the crash severity levels and to identify the various factors contributing to different highway user crashes on HRGCs. The result obtained from the pedestrian crash severity modeling indicated that higher train speed as compared to lower train speeds are associated with fatal pedestrian crashes.

Reducing the train speed near the HRGC sites would minimize the chance of a pedestrian crash to be fatal. As the study showed, HRGCs located in the city have increased the chance of fatal crashes as compared to those located near the city. Hence considerable train speed reduction in the cities would help in reducing fatal crashes. It was also observed that female pedestrians are more likely to involve in a crash as compared to male pedestrians. Educating pedestrians through various communication means would help in reducing the number of female victims in pedestrian-rail crashes. The majority of crashes occurred are when the rail equipment struck a pedestrian which indicates that train speed reduction would give pedestrians sufficient time to leave the crossing and possibly avoid the crash. Crossing surface types and outdoor temperature result in more severe injury type crashes. Improving the crossing surface types and educating pedestrians would help to minimize the impact of such factors.

The result obtained from the vehicle crash severity model indicated that both higher vehicle and train speed increased the chance of injury and fatal crashes. The majority of crashes occurred are when the rail equipment struck a vehicle. In particular, these types of crash circumstances increased the chance of fatal crashes. Hence, reducing train and vehicle speed on HRGC would minimize the chance of more severe crashes. Among the various vehicle types, a pick-up truck is observed to increase the chance of injury related crash. It was also observed that male drivers with the age of above 60 years are more likely to be involved in an injury related or fatal crash compared to female vehicle drivers. Moreover, crashes that occurred at higher outside temperature increased the probability of fatal and injury crashes as compared to crashes occurred at low temperature. Therefore, educating vehicle drivers to increase their awareness towards the problem would help in reducing chances of severe vehicle-rail crashes. Concrete and rubber crossing surface type is related with fatal crashes. Improving the crossing surface type would minimize the number of fatal crashes.

The other goal of this thesis was to perform model comparison between multinomial and ordered logit models. As discussed, in the ordered logit pedestrian-rail crash severity level model, some variables which are statistically significant in the MNLM were found to be statistically not significant. In addition, there are some inconsistencies observed in some other variables. The multinomial logit model has an increasing effect on the probability of lower severity level where as the ordered logit model has a decreasing effect. On the other hand, the ordered logit model has an increasing effect on the probability of intermediate severity level and the reverse is true in the case of multinomial logit model. Furthermore, based on the AIC, it was found that the OLM is better in estimating the vehicle crash severity levels on HRGCs. Therefore, the researcher recommends the OLM to be applied rather than the MNLM in modeling pedestrian crash severity levels on HRGCs.

In the ordered logit vehicle-crash severity level model also, some variables that are not statistically significant in the MNLM were found to be statistically significant. Apart from this, almost all variables were found to have the same effect on the probability of crash severity levels except one variable (crash circumstance). Based on the AIC, it was found that the OLM is better in predicting vehicle crash severity levels on HRGCs. Similarly to the pedestrian crash severity model, the researcher recommends the OLM be applied rather than the MNLM in predicting severity levels of vehicle-rail crashes on highway-rail at-grade crossings.

5.2 Future research

There are various alternative modeling techniques in addition to the models used in this research. Savolainen et al. (*4*) briefly discussed and summarized the wide range of methodological tools applied to study the impact of various factors on motor vehicle crash-injury severities. As presented in the thesis, ordered logit and prohibit, multinomial logit, binary logit and binary probit and nested logit are some of frequently used statistical methodologies. In addition to models adopted in this research, those various models should be applied in modeling crash severity levels as a function of different factors involved in the highway-rail crashes on at-grade crossings. In general, there is little research conducted on highway-rail at –grade crossing crash severity level modeling. Therefore, the researcher recommends intensive studies to be conducted in modeling highway user crashes on HRGCs.

In this research, the crash severity modeling was conducted only for crashes on public HRGCs. The modeling should also be extended for crashes on private crossings. In addition, this research focused on vehicle users and pedestrian crash with rail on highway-rail at-grade crossings. The study should also be extended for other highway user categories such as bicycle and motorcycle users.

References

1. National Highway Traffic Safety Administration, US Department of transportation Traffic Safety Facts Research Note, 2011 Motor Vehicle Crashes: Overview, 2012. Available at: http://www-nrd.nhtsa.dot.gov/Pubs/811701.pdf. Accessed on May 20, 2013.

2. Federal Railroad Administration, US Department of Transportation. Railroad Safety Statistics, 2012 Preliminary Annual Report, 2012. Available at: http://safetydata.fra.dot.gov/OfficeofSafety/publicsite/Prelim.aspx. Accessed on May 20, 2013.

3. Federal Railroad Administration, US Department of Transportation. Private Highway-Rail Grade Crossing Safety Research and Inquiry, Final Report, 2010.

4. Savolainen, P., Mannering, F., Lord, D. and Quddus, M. The Statistical Analysis of Highway Crash-injury Severities: A Review and Assessment of Methodological Alternatives. *Accident Analysis and Prevention*, Vol. 43, No. 5, 2011, pp. 1666-1676.

5. Khattak, A. Severity of Pedestrian Crashes at Highway-Rail Grade Crossings. Paper No. 13-4587, Proceedings of the 92nd Annual Meeting of the *Transportation Research Board*, Washington D.C., January 13-17, 2012.

6. Long, J. S. Regression Models for Categorical and Limited Dependent Variables. Thousand Oaks, CA: Sage Publications, 1997.

7. Lobb, B. Trespassing on the Tracks: A Review of Railway Pedestrian Safety Research. *Journal of Safety Research*, Vol. 37, No. 4, 2006, pp. 359-365.

8. Davis, G.A. Relating Severity of Pedestrian Injury to Impact Speed in Vehicle-Pedestrian Crashes: Simple Threshold Model. *Transportation Research Record*, Vol. 1773, 2001, pp. 108-113.

9. Kwigizile, V., Sando, T. and Chimba, D. Inconsistencies of Ordered and Unordered Probability Models for Pedestrian Injury Severity. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2264, 2011, pp. 110-118.

10. Zahabi, A.A.H., Strauss, J., Manaugh, K. and Miranda-Moreno, L.F. Estimating Potential Effect of Speed Limits, Built Environment, and Other Factors on Severity of Pedestrian and Cyclist Injuries in Crashes. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2247, 2011, pp. 81-90.

11. Rosén, E., Sander, U. Pedestrian Fatality Risk as a Function of Car Impact Speed. *Accident Analysis and Prevention*, Vol. 41, No. 3, 2009, pp. 536-542.

12. Sze, N.N. and Wong, S.C. Diagnostic Analysis of the Logistic Model for Pedestrian Injury Severity in Traffic Crashes. *Accident Analysis and Prevention*, Vol. 39, No. 6, 2007, pp. 1267-1278.

13. Zajac, S.S. and Ivan, J.N. Factors Influencing Injury Severity of Motor Vehicle-Crossing Pedestrian Crashes in Rural Connecticut. *Accident Analysis and Prevention*, Vol. 35, No. 3, 2003, pp. 369-379.

14. Montella, A., Aria, M., D'Ambrosio, A. and Mauriello, F. Data-Mining Techniques for Exploratory Analysis of Pedestrian Crashes. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2237, 2011, pp. 107-116.

15. Kim, J-K, Ulfarsson, G.F., Shankar, V.N. and Mannering, F.L. A Note on Modeling Pedestrian-Injury Severity in Motor-Vehicle Crashes with the Mixed Logit Model. *Accident Analysis and Prevention*, Vol. 42, No. 6, 2010, pp. 1751-1758.

16. Ulfarsson, G.F., Kim, S., Booth, K.M. Analyzing Fault in Pedestrian–Motor Vehicle Crashes in North Carolina. *Accident Analysis and Prevention*, Vol. 42, No. 6, 2010, pp. 1805-1813.

17. Tay, R., Choi, J., Kattan, L. and Khan, A. A Multinomial Logit Model of Pedestrian–Vehicle Crash Severity. *International Journal of Sustainable Transportation*, Vol. 15, No. 4, 2011, pp. 233-249.

18. Mercier, C., Shelley, M., Adkins, G. and Mercier, J. Age and Gender as Predictors of Injury Severity in Broadside and Angle Vehicular Collisions. *Transportation Research Record*, Vol. 1693, 1999, pp. 50-61.

19. Dissanayake, S. and Lu, J. Analysis of Severity of Young Driver Crashes: Sequential Binary Logistic Regression Modeling. *Transportation Research Record*, Vol. 1784, 2002, pp. 108-114.

20. Duncan, C., Khattak, A. and Council, F. Applying the Ordered Probit Model to Injury Severity in Truck-Passenger Car Rear-End Collisions. *Transportation Research Record*, Vol. 1635, 1998, pp. 63-71.

21. Donnell, E.T., and Mason, J.M. Predicting the Severity of Median-Related Crashes in Pennsylvania by Using Logistic Regression. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 1897, 2004, pp. 55-63.

22. Renski, H., Khattak, A.J. Council, F.M. Effect of Speed Limit Increases on Crash Injury Severity Analysis of Single-Vehicle Crashes on North Carolina Interstate Highways. *Transportation Research Record*, Vol. 1665, 1999, pp. 100-108.

23. Huang, H.F, Stewart, J.R. and Zegeer, C.V. Evaluation of Lane Reduction "Road Diet" Measures on Crashes and Injuries. *Transportation Research Record*, Vol. 1784, 2002, pp. 80-90.

24. Khattak, A.J. Injury Severity in Multivehicle Rear-End Crashes. *Transportation Research Record*, Vol. 1746, 2001, pp. 59-68.

25. Mercier, C., Shelley, M., Rimkus, J. and Mercier, J. Age and Gender as Predictors of Injury Severity in Head-on Highway Vehicular Collisions. *Transportation Research Record*, Vol. 1581, 1997, pp. 37-46.

26. Chira-Chavala, T., Coifman, B., Porter, C. and M. Hansen, M. Light Rail Accident Involvement and Severity. *Transportation Research Record*, Vol. 1521, 1996, pp. 147-155.

27. Chen, W. and Jovanis, P. Method for Identifying Factors Contributing to Driver-Injury Severity in Traffic Crashes. *Transportation Research Record*, Vol. 1717, 2000, pp. 1-9.

28. Khattak, A., Pawlovich, M., Souleyrette, R., and Hallmark, S. Factors Related to More Severe Older Driver Traffic Crash Injuries. *Journal of Transportation Engineering*, Vol. 128, No. 3, 2002, pp. 243-249.

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

30. SAS/STAT User's *Guide*, *Version 9.2*. SAS Publishing, Cary, N.C., 2008.