6-2017

Injury Severity of Truck Drivers in Crashes at Highway-Rail Grade Crossings

Waleed Ali Khan
University of Nebraska-Lincoln, waleedalikhan@hotmail.com

INJURY SEVERITY OF TRUCK DRIVERS IN CRASHES AT HIGHWAY-RAIL GRADE CROSSINGS

By

Waleed Ali Khan

A THESIS

Presented to the faculty of

The Graduate College at the University of Nebraska

In Partial Fulfillment of Requirements

For the Degree of Master of Science

Major: Civil Engineering

Under the Supervision of Professor Aemal J. Khattak

Lincoln, Nebraska

June, 2017

INJURY SEVERITY OF TRUCK DRIVERS IN CRASHES AT HIGHWAY-RAIL GRADE CROSSING

Waleed Khan, M.S.

University of Nebraska, 2017

Advisor: Aemal J.Khattak

There is a noticeable difference between different road users, specifically between passenger vehicles and heavy vehicles such as its length and weight. The majority of previous research were focused on general highway traffic that included passenger cars, trucks, buses, motorcycles, etc. Moreover, HRGC safety studies of specific types of vehicles are relatively few and heavy vehicle safety at grade crossing is even more under-explored.

This research thus focuses on the following objectives: Identify factors related to different injury severity levels of heavy-vehicle drivers (truck/truck-trailer) drivers in crashes reported at HRGCs; to identify a more suitable statistical model for injury severity modeling of truck involved crashes. This study considered variables that have not been explored in previous injury severity studies of truck-involved crashes at HRGCs. Three unordered response models: Multinomial Logit model (MNL), Nested Logit model (NL) and Mixed Logit model (RPL) were evaluated to investigate injury severity of drivers of heavy-vehicles involved in crashes at HRGCs.

Based on criteria used for judging the models and the dataset used in this study, it was concluded that the RPL was most suitable for modeling truck drivers' injuries in crashes reported at HRGCs amongst the models considered. Truck drivers' injuries in crashes reported at HRGCs are positively associated with speed of train and road user

(truck/trailer), truck-train crash, when train strike road user (truck/trailer), hazardous materials by either one or both users, driver behavior “went around the gates”, age of driver, crashes reported in rural areas and crashes at minimum crossing angle of 60-90 degrees. Whereas truck drivers’ injuries are negatively associated with train detection system, gates, if the track is signaled, when the track is obstructed, HRGCs within 500 feet of a highway and position of vehicle “heavy vehicle stopped on the crossing”.

ACKNOWLEDGEMENTS

I would like to express my appreciation to my academic advisor, Dr. Aemal Khattak, for his continuous professional guidance, encouragement and support through out my two years of education at the University of Nebraska-Lincoln. I also thank Dr. Khattak for expanding my horizon and preparing me for my career.

I would also like to thank my committee members, Dr. Laurence R. Rilett and Dr. John Sangster for their support and advice.

Thanks to the Civil Engineering Department and the Nebraska Transportation Center for providing me the financial support, resources and facilities.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1 INTRODUCTION

1.1 Background

In the United States (US), the FHWA report for the year 2013 states that 122.5 million households, 7.5 million business establishments, and over 90,000 governmental units are part of the economy (FHWA: Freight Facts and Figures report FFF-2015), for which the efficient movement of freight is critical. Major freight transportation modes include highway, rail, water and pipelines. The 2012 Commodity Flow Survey (CFS-2012), jointly conducted by the US Census Bureau and the Bureau of Transportation Statistics (BTS), estimated shipping of about 11.3 billion tons of freight valued at more than $13 trillion over the nation's freight transportation system and generating 3.3 trillion ton-miles of travel in 2013 (US DOT 2015). Freights transportation by roads(Table 1), continued to dominate the nation's movement of freight for value and tonnage, accounting for 73.1% of the value ($10.1 trillion) and about 71% of weight (8.1 billion tons). Truck and rail each accounted for 1.2 trillion ton-miles. Single mode truck was the dominant mode of freight transportation, accounting for at least 60% of the total value of shipments for 43 states in the US. According to Freight Facts and Figures 2015 report (US DOT 2015), total shipments are expected to increase to 28.5 billion tons, with domestic shipments of about 23 billion tons by 2040 (Table-1).

Freight transportation has made important contributions to the growth of the national economy but these have come at the price of traffic crashes, injuries and fatalities. Truck and train traffic is expected to increase due to the expected growth in the demand for freight. This will likely increase the risk of conflicts between these two modes of transport thereby exacerbating a multimodal safety issue.

Table 1 The weight of shipments by transportation mode (millions of tons). Source: US DOT: Freight Facts and Figures, 2015

	2013				2040			
	Total	**Domestic**	**Exports**	**Imports**	**Total**	**Domestic**	**Exports**	**Imports**
Truck	13955	13732	120	103	18786	18083	368	335
Rail	1858	1681	82	94	2770	2182	388	201
Water	808	410	89	309	1070	559	164	347
Air, air & truck	15	3	5	7	53	6	20	27
Multiple modes & Mail	1554	459	559	536	3575	645	1546	1383
Pipeline	1539	1391	11	137	1740	1257	17	467
Other & unknown	333	274	47	13	526	362	130	34
Total	20062	17950	913	1199	28520	23094	2633	2794

Collisions at highway-rail grade crossings (HRGCs), although relatively rare events are nonetheless a safety concern as crashes at these locations tend to be more severe in terms of fatalities, injuries and property damage, compared to crashes reported elsewhere on the transportation network. Federal Railroad Administration (FRA) crash data shows that the total number of reported HRGC crashes decreased by 25.7% from 2007-2015 (Figure 1). However, it can be observed (Figure 1) that there has been an increase (15.4%) in the number of crashes between 2012 and 2014 (1,987 crashes in 2012 to 2,293 crashes in 2014). According to the FRA crash data, there have been relatively small changes in the number of injuries and fatalities from the year 2007 to 2015. In fact, the number of injuries and fatalities have slightly increased from the year 2012 with 231 fatalities and 971 injuries, to 2015 with 237 fatalities and 1,003 injuries.

In 2015, of the 2,063 crashes at grade crossings, 317 (15.4%) involved heavy vehicles (truck, trailer) on public crossings with 10 truck driver fatalities constituting 4.2% of the total fatalities reported at HRGCs. Figure 2 presents details of heavy-vehicle involved crashes over the nine-year period (2007-2015) while Figure 3 shows its comparison by different severity levels with the total number of HRGC crashes. These two figures show no appreciable decrease in truck-involved crashes at HRGCs over the years.

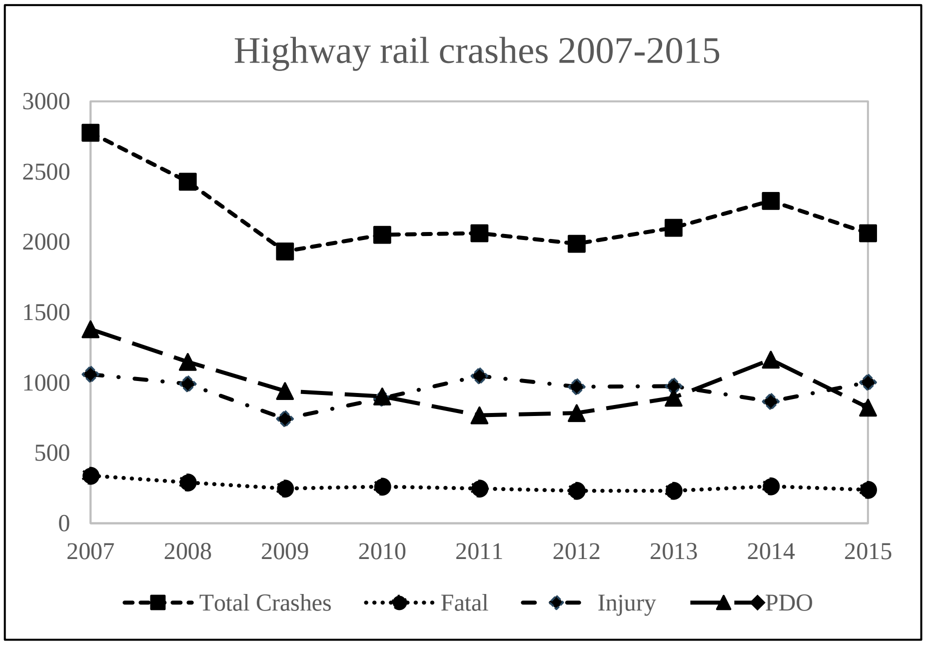

Figure 1 Highway rail crashes in U.S 2007-2015

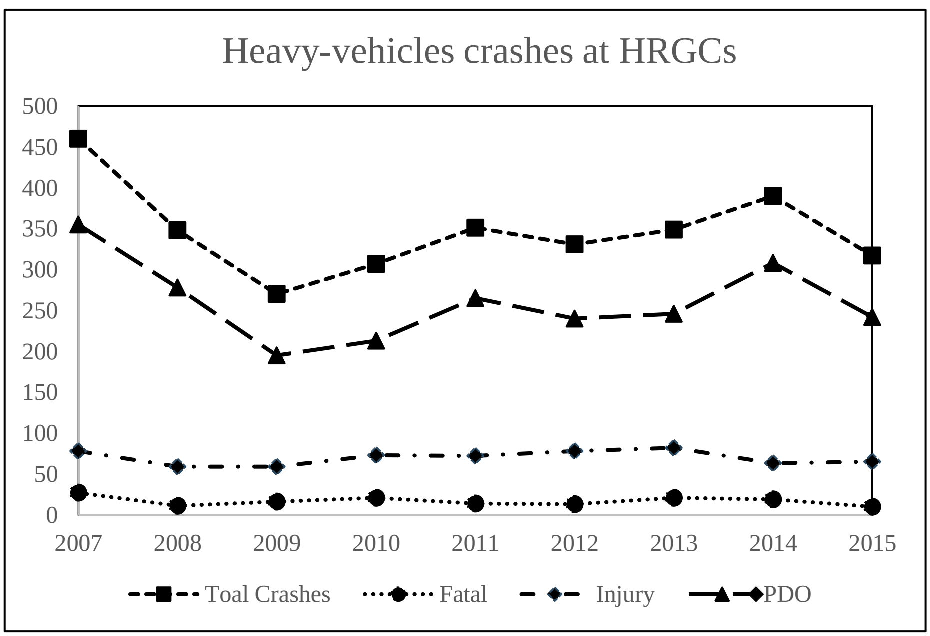

Figure 2 Heavy-vehicle crashes at HRGC in U.S 2007-2015

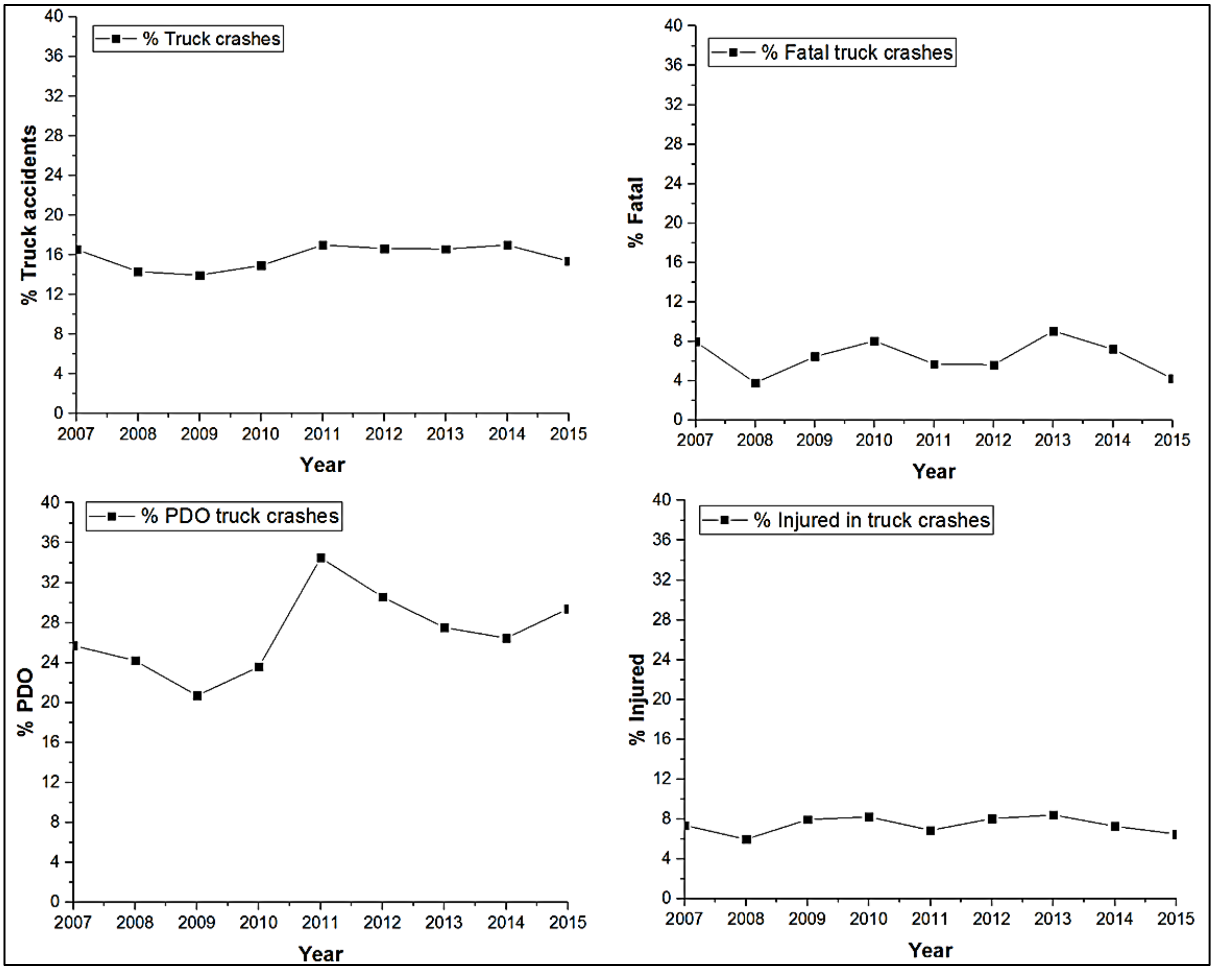

Figure 3 Percent of heavy-vehicle crashes w.r.t each category of the total HRGC crashes in

1.2 Problem Statement

Truck-involved crashes at HRGCs are important to investigate because they are not only vulnerable to more severe injury, but also can potentially disrupt both the highway and the rail network. Current safety research is mostly focused on crashes reported at non-HRGC locations while research on crashes reported at HRGCs is not specifically focused on trucks- it is mostly focused on mixed traffic or pedestrians. Trucks have unique characteristics compared to other motor vehicles in terms of size, weight, and acceleration characteristics. However, limited literature was found on truck-involved crashes at HRGCs and therefore they require attention. In light of the above, the problem statement for this research is as follows:

Due to the unique characteristics of trucks compared to other motor vehicles and the greater potential for injury and disruption of multimodal networks due to truck-involved crashes at HRGCs, there is a need to study reported truck crashes at HRGCs. Specifically, truck driver injury severity and appropriate models for studying factors related to truck driver injury severity need investigation.

1.3 Research Objectives

a) Identify factors related to different injury severity levels (fatal, injury, PDO) of truck drivers in crashes reported at HRGCs.

This study will consider variables that have not been explored in previous injury severity studies of truck-involved crashes at HRGCs. They include variables such as the railroad class, distance to nearby intersecting highway, percentage of school buses and train traffic at HRGCs, primary obstruction of track view, active and passive warning devices, and different behavioral characteristics of the highway user (truck driver) prior to the crash.

b) To identify a more suitable statistical model for injury severity modeling of truck involved crashes

This study will evaluate three unordered response models: Multinomial Logit model (MNL), Nested Logit model (NL) and Mixed Logit model (RPL) to investigate injury severity of drivers of heavy-vehicles involved in crashes at HRGCs.

1.4 Research Outline

This thesis consists of six chapters. Chapter 1 introduces background of this study, problem statement, and outlines the structure of this thesis. Chapter 2 presents a comprehensive review of published literature related to this topic. Reviewed topics include HRGC safety studies, injury severity studies of road users at HRGCs, safety studies related to truck drivers, and potential modeling approaches used for crash injury severity.

Chapter 3 introduces the statistical models used in this study and the general framework for model estimation. Chapter 4 describes the source of data, its formulation and provides descriptive statistics of the dataset used for model estimation. Chapter 5 presents the three estimated models (MNL, NL, and RPL), comparison of the three models including model classification accuracy and discussion of the different independent variables that were found associated with driver injury severity of heavy-vehicles at HRGCs. Chapter 6 summarizes this study, presents conclusions from the analysis, provides recommendations for safety improvement at HRGCs and proposes future research.

CHAPTER 2 LITERATURE REVIEW

This literature review covers HRGC safety analysis, injury severity of different road users at HRGCs and different factors associated with it. It also covers different statistical models and methods used to identify key factors related to injury severity at HRGCs.

2.1 Highway-Rail Grade Crossing Safety

According to the latest FRA national HRGC inventory there are 133,825 public rail grade crossings whereas 82,921 crossings are situated on private property in the US. Special highway traffic control devices, such as advance warnings, flashing lights, gates, stop signs, pavements markings, bells, cross bucks and their combinations are regulated for installation by local, state and federal authorities, to ensure safe and efficient operation of both highway and railroad traffic system at HRGCs. Crossings with a history of crashes can be examined and upgraded to more restrictive warning devices. Railroads and transportation agencies work together to close unsafe crossings or grade-separate them with the goal to balance cost with risk reduction. Nelson (2010) encapsulates many strategies currently in use for reducing the risk of crashes at grade crossings. These include upgraded lights and gates, alternate technologies such as in-pavement flashers, and closure and consolidation.

The North Carolina DOT (NCDOT) and Illinois DOT (IDOT) implement the sealed corridor concept on 216 and 311 HRGCs respectively (Bien-Aime, 2009, Hellman and Ngamdung, 2009). This concept was developed as a way to upgrade conventional rail lines to accommodate higher-speed passenger trains. FRA requires crossings to have approved barrier systems that can prevent infiltration of motor vehicles. Obstacle detection systems are also recommended to alert oncoming trains if a motor vehicle is stuck on the tracks. The use of appropriate technologies and its requirements are summarized in the document "Highway-Rail Grade Crossing Guidelines for High-Speed Passenger Rail" (2009). NCDOT projected that the implementation of the sealed corridor concept saved 19 lives between 2004 and 2009. As mentioned earlier, the goal of obstacle detection systems is to identify motor vehicles or persons on the crossing and warn approaching trains in time to allow train stoppage (Glover 2009). Glover discussed that obstacle detection should provide a feasible way to attenuate grade crossing risk. However, due to short amount of time for the system to react and the train to stop, there may be limited benefits. Hall (2007) on the other hand suggested that benefits of an obstacle detection system may still exist although it may not necessarily prevent crashes at HRGCs as trains may possibly slow down reducing crash severity.

Low-cost warning devices provide similar level of safety as conventional devices; in this respect Hellman and Ngamdung (2010) demonstrated several low-cost warning devices for HRGCs that satisfied FRA's requirements. Several studies have been conducted to identify the reactions of different people to warning signs at HRGCs (Lenne et al., 2011, Tey et al., 2011a, Tey et al., 2011b). Drivers exhibit lower compliance at passive crossings in response to warning signs than at active crossings. The addition of warning signs, especially active warning signs has reduced crashes at HRGCs. Chadwick et al., 2014 performed in-depth analysis of relevant research through an extensive literature review and addressed safety enhancing strategies at HRGCs as well as limitations of those strategies.

2.2 Injury Severity

Safety at HRGCS is a significant concern because the severity of crashes at these locations is usually higher than those reported at non-HRGCs. Although many studies have been conducted on crash injury severity analysis, the majority of the research published is on injury severities on road segments or intersections.

Most of the research focused on HRGCs used MNL, OP, Ordered Logit (OL) and mixed logit model (mixed generalized logit model) to identify different aspects of crash injury severity at level crossings. Hu et al. (2010) conducted a study in Taiwan using 592 highway-rail crossings. A generalized logit model was estimated using different characteristics of crossings, highways, railway traffic controls and land use. Results indicated that the likelihood of more severe crash injuries increased with an increase in the number of trucks and daily trains. Highway obstacle and separation detection devices were also found to be associated with more severe crash. A latent segmentation based ordered logit model was developed by Eluru et al. (2012) using the FRA crash data (1997-2006). In this model, HRGCs were assigned probabilistically to different segments based on their characteristics with a separate injury severity component for each segment. The results indicated that time of the crash, the presence of snow and/or rain, driver age, driver behavior before the crash and vehicle role in the crash were the key factors influencing injury severity.

Hao and Daniel (2013) determined different factors influencing driver's injury severity at HRGCs, using OP model by utilizing FRA 2002-2011 data. Factors related to higher injury severity of vehicle driver at HRGCs included adverse weather conditions, low visibility, train and vehicle speeds greater than 50mph, highways with AADT over

10,000, crashes reported in open areas, and crashes involving trucks and semi-trailers. Russo and Savolainen (2013) used an ordered logit model using the FRA data to identify different factors of rail, highway, traffic and driver characteristics associated with the frequency and injury severity of HRGCs crashes. Factors that were found to be positively associated with more severe injury included females, drivers aged over 60 years, motorists behavior: did not stop at crossings and trains with speed greater than 60 mph. A MNL was used by Fan and Haile (2014), to identify various factors that increased injury severity of crashes at HRGCs by using the FRA 2005-2012 crash data. Drivers aged 25 years and older, pickup trucks and crossing surfaces with concrete or rubber were found related to more severe crashes. Foggy and snowy weather conditions, truck-trailers, certain land development types and higher AADT were found associated with less severe crashes.

A study was conducted in Australia to identify the effect of active and passive controls, in which participants drove the Monash University Accident Research Center (MUARC) advanced driving simulator for 30min. The study found that traffic signals at HRGCs did not appear to offer safety benefits beyond those provided by the use of flashing lights, the reduction in vehicle speeds at crossings with flashing lights was greater than crossings with signals. It was concluded that vehicle speed was significantly lower when approaching a stop sign, compared to both red flashing lights and traffic signals (Lennéet al. 2011). Hao et al. 2016 identified different factors affecting driver injury severity of vehicle driver at highway-rail grade crossing under different weather conditions using mixed logit model. The result showed that injury severity was more prevalent in crashes involving vehicles or trains with high speeds. Light condition and unpaved surfaces also increased injury severity.

Freight transport by rail and road (trucks) has increased, and will likely keep on increasing in the future. As a result, more and more trucks will surmount HRGCs, thereby increasing the chances of conflict betweet trains and trucks. Several studies conducted on safety at HRGCs have identified heavy vehicles as one of the factors contributing to HRGC crashes (Hu et al. 2010, Hao and Daniel 2013, Fan and Haile 2014). Due to the disparity in mass between train and motor vehicles, the impact is usually extensive leading to traumatic scenes. A recent trend of heavy vehicle involvement in these crashes, in Australia at least, has led to risk the train and its passengers, in addition to the road vehicle, with the potential for catastrophic outcomes (Australian Transport Safety Bureau, 2008). With growing numbers of longer and heavier freight vehicles using the road network, coupled with increased train services and speeds, this catastrophic risk may be increasing. A study found that the passing time for heavy-vehicles at rail crossing is about four time greater than the passing time of an automobile at the same location. Due to its physically large size and weight, the behavior of large vehicles at HRGCs is different than other motor vehicles, hence the topic requires an investigation that can identify the potential factors associated with truck driver injury severity. Limited research was found on safety analysis of trucks at HRGCs, few studies have been conducted on driver behavior at grade crossing and the type of violations. The majority of reviewed research found was focused on general highway traffic that includes passenger cars, buses, trucks, motorcycles etc. Highway-rail safety studies are relatively few and heavy-vehicle safety at grade crossing is even more under-explored.

Human errors are primarily considered as a cause of railway crossing crashes. A study conducted in Australia focused on understanding the design issues and behavior issues that affect at-grade crossings safety and may cause heavy vehicle-train collisions by conducting a series of group discussions. A selected group of train and truck drivers were selected for the discussion, it was concluded that the vehicle driver visibility (line of sight & angles of approach) and effective vehicle clearance (impeded acceleration, length of carriage maneuverability) was affected by the configuration of level crossings. However, the driver compliance towards violation of crossing protocols was often due to saving time or due to high familiarity with the crossing (Davey, Jeremy, et al 2008).

Ishak et al. 2011 introduced Petri nets- a graphical and mathematical modeling tool in assessing risk at HRGCs when heavy vehicles were passing through intersecting areas. Results indicated that factors associated with heavy vehicle collisions at level crossings included traffic level of service (LOS), the percentage of heavy vehicles and the distance of grade crossing to or from the nearest intersection (Ishak et al. 2011). Driver behavior was identified as one of the potential factors in crashes, especially truck driver behavior, which was not only different than passenger drivers but more critical due to long hours of driving, sleep factor, consciousness, frustration level etc. The behavior of truck driver led to violations of traffic laws, hence increasing the risk of a crash.

A study was conducted on the frequency, type of crossing gate-related violations by truck drivers and the contributing factors at gated HRGCs in Nebraska (Khattak and Miao 2012). The analysis indicated that violations increased at crossings with longer time between onset of flashing lights and train arrivals and with greater truck traffic at the HRGCs. The results also indicated that most of the violations occurred during night time. Jun Liu et al (2016), conducted a detail safety analysis of truck involved crashes, to identify the factors associated with driver's behavior before the collision. The study also explored several key factors on different crash outcomes. The results indicated that truck-involved crashes occurring at HRGCs equipped with gates were generally less severe, compared to those occurring at crossings without gates. The correlates of pre-crash behaviors revealed that the truck drivers at crossings without gates, are more likely failed to make an appropriate stop, or proceeded after a short stop, or even stopped on crossing before crash occurance.

2.3 Potential Modeling Approaches

Injury severity data may be considered as nominal or ordinal and relevant modeling techniques may be used. Frequently used nominal models include MNL, NL and mixed logit models (RPL model), while GOL model, OL, and OP models are commonly used ordinal models. The modeling approach for injury severity depends on the quality and quantity of data available for the analysis. A number of data characteristics and its limitations have been identified in past that may be critical in development and application of a statistical model. Hence it is important to identify the most suitable model to overcome data limitations to the extent possible.

Some of the commonly used models for modeling injury severities in the past decade are OL/OP model (O'Donnell and Connor, 1996; Kockelman and Kweon, 2002; Kweon and Kockelman, 2003), MNL (Carson and Mannering, 2001; Lee and Mannering, 2002; Khorashadi et al., 2005) and NL (Lee and Mannering, 2002). Abdel-Aty (2003) compared OP, MNL and NL model, in addition to identifying different factors associated with injury severity at intersections and roadway sections. The OP model was comparatively simple and produced better results in terms of model's goodness of fit and number of significant variables entered in the model specification. Abdel-Aty and Abdel Wahab (2004) compared results from an Artificial Neural Network (ANN) and an OP model. The test of difference in proportion revealed that ANN showed more accurate prediction capabilities and performed better than the OP model.

A bivariate response model was used by Yamamoto and Shankar (2004) to capture different levels of crash severity and most severely injured passengers. Yau (2004) used a logistic regression model with stepwise variable selection to identify different factors affecting the severity of single-vehicle traffic crashes. To count for unobserved effects associated with driver and highway characteristics, Milton et al. (2008) used mixed logit model. Mixed logit model overcomes the limitation induced by MNL model i.e. allowing heterogeneous effect and correlation in unobserved factors. Hu et al. (2010) developed a generalized logit model by using HRGC data in Taiwan.

A study conducted by Haleem and Abdel-Aty (2010) on traffic crash injury severity at un-signalized intersection concluded that binary probit model showed better goodness of fit compared to the disaggregated OP and NL models. Yasmin and Eluru (2013) compared different ordered and unordered response models for driver injury severity of crashes involved in traffic. The models used for nominal response were MNL, NL and order generalized extreme value logit (OGEV) where as OL and GOL model were used for the ordinal response framework. The criteria used to compare performances of the estimated models included in the study are; Akaike information criterion corrected (AICc), Bayesian Information Criterion (BIC) and Ben-Akiva and Lerman's adjusted likelihood ratio (BL) test. It was found that OGEV and NL models reduced to simple MNL model. However, GOL model comparatively performed better in terms of data fit than OL and

MNL models. Eluru (2013) also examined the performance of the GOL and MNL models by examining the issues by data generation perspective. In conclusion, the author discussed that it was not possible to conclude which of the two models was better without considering the dataset structure. The results indicated the emergence of the GOL model as a true equivalent ordered response model to the MNL model for ordinal discrete variables.

Yasmin et al. (2014) attempted to identify a better model framework for injury severity of pedestrian by comparing three order response models: OL model, latent segmentation based ordered logit model (LSOL) and GOL. The results indicated that LSOL performed better than the GOLand LSOL model for identifying factors associated with different injury severity levels of pedestrians. The effect of sample size on model development was investigated (Ye and Lord 2014) by using a Monte-Carlo analysis based on simulated and observed data. The three models estimated in the study are OP, MNL and RPL models and the criteria used for comparison of these three models are: total root-mean-square-error (RMSE) and maximum APB and absolute-percentage-bias (APB).The results indicated that RPL model required largest sample size than the other two models whereas OP model required the smallest. In terms of model interpretations, RPL model performed better than the MNL model, whereas MNL model had superior interpretation power compare to the order probit model. However, the OP model had better goodness-of-fit than the other two models (RPL & MNL), and the RPL had better goodness-of-fit than the MNL model.

Zhao and Khattak (2015) recenly used the FRA crash data to identify different variables associated with driver injury severity of train-motor vehicle crashes at grade crossings. The study compared OP, MNL and RPL models, in an attempt to identify a suitable model to explore factors related to different severity levels of driver in train-vehicle crash. The following criteria was used for model superiority: number of statistically significant parameters, model goodness-of-fit, model's interpretation power and classification accuracy. It was concluded that the RPL model and the MNL model performed better for injury severity analysis of motor vehicle drivers involved in crashes at highway-rail grade crossings.

2.4 Gaps in Literature

The majority of previous research were focused on general highway traffic that included passenger cars, trucks, buses, motorcycles, etc. HRGC safety studies of specific types of vehicles are relatively few and heavy vehicle safety at grade crossing is even more under-explored. There is a noticeable difference between different road users, specifically between passenger vehicles and heavy vehicles such as length and weight. This may affect the time, a heavy vehicle takes to cross the crossing and its impact on the level of severity, if a collision occurs between train and heavy vehicles specifically in the presence of any hazardous materials, the result of collision can be catastrophic. There is a research gap for investigation of injury severity of heavy vehicles at HRGCs, some of the limited literature previously found did not consider all the characteristics in the investigation. Previous studies majorly included driver and operational characteristics.

Because there is limited research available on heavy vehicle injury severity at HRGCs, it provides an opportunity to investigate different statistical models utilizing the FRA HRGC crash dataset to identify the modeling framework suitable for the injury severity of heavy-vehicle crashes at HRGCs. For dependent variable (i-e injury severity) with multiple response outcomes, injury severity is divided into three levels (PDO, injury, fatal) from low to high. This study considered unordered (i.e. treat injury severity as discrete outcomes and neglect ordering in the severity) response models that were found to be vital in the literature by overcoming some of the limitations of the available dataset. This study will use MNL, NL and mixed logit model (RPL) for unordered response modeling of injury severity of heavy-vehicle crashes at level crossings.

CHAPTER 3 METHODOLOGY

To achieve the objectives of this study, it is important to identify a suitable model for truck driver injury severity. This chapter presents model selection criteria and introduction of each model considered in this research. A model selection discussion is presented in Section 3.1. A brief introduction of crash injury severity models used in this study is presented in Section 3.2. Section 3.3 provides the estimation procedure for each model. Details of model estimating and results are provided in Chapter 4.

3.1 Model Selection

A variety of methodological techniques have been employed to analyze crash injury severity data. The statistical methods applied by researchers have primarily relied on methodological issues associated with the data. Because driver injury severity is discrete, discrete outcome models were selected for this study. The three models selected for this study are: MNL (Multinomial Logit) model, NL (Nested Logit) model and a mixed logit model, also known as RPL (Random Parameter Logit) model. The MNL model was selected because it is by far the most widely used due to its simplicity and ease of estimation. A prominent limitation of this model is a property known as "Independence of Irrelevant Alternatives (IIA)" and identically distribution (IID) assumption.

The IIA property states that the ratio of the choice probabilities of any pair of alternatives is independent of the presence or absence of any other alternative in a choice set. A particularly important behavioral implication of IIA is that all pairs of alternatives are equally similar or dissimilar. This amounts to assuming that all the information in the random components is identical in quantity for the set of attributes that are not observed and the relationship between pairs of alternatives and hence across all alternatives (IID condition). In addition to not accounting for the ordinal nature of injury severity, the MNL is particularly vulnerable to correlations of unobserved effects from one injury severity level to the next. This causes a violation of the model's IIA property (Washington et al., 2011). The IIA property neglects unobserved heterogeneity which leads to an inferior model specification and a spurious interpretation of the model.

The NL models offer a partial relaxation of the IID and IIA assumptions of the MNL model, this relaxation occurs in the variance components of the model together with some correlations within subsets of alternatives, but the IID problem still exists within the groups, however the NL model is relatively straightforward to estimate and offers the added benefit of being a closed form solution. RPL model is more complex model and it offers relaxation of the IIA property. The three models mentioned in this sections are used to achieve the best results possible.

3.2 Multinomial Logit Model

The MNL model is a traditional discrete outcome model that does not explicitly consider the ordering nature that may be present in the outcomes. It is a special case of a general model of utility maximization. The general framework used to model the degree of injury severity of a crash begins by a linear function U_{ij}. According to NLOGIT version 5 (Greene 2002) reference guide, consider driver i in a crash experiencing an injury severity level j, the severity function for the outcome is:

$$U_{ij} = \partial_j + \beta_j X_{ij} + \varepsilon_{ij} \qquad (1)$$

Where,

U_{ij} = function of covariates that determines the severity level j for driver i

∂_j = constant parameter for injury severity level of j

β_j = vector of coefficients to be determined for severity level j

X_{ij} = vector of independent variable values for driver i for the severity level of j

ε_{ij} = represents a random error term

The error terms are assumed to be independent and identically distributed with identical type 1 extreme value distribution. Based on the above specification, let $P_{i(j)}$ represents the probability of driver i experiencing injury severity level j in a crash. The probability of MNL model is expressed in eq-2, where EXP represents the base of natural logarithm.

$$P_i(j) = \frac{\exp\left(\partial_j + \beta_l X_{ij}\right)}{\sum_{j=1}^{J} \exp\left(\partial_j + \beta_j X_{ij}\right)} \tag{2}$$

3.3 Nested Logit Model

A class of models known as generalized extreme value models (GEV) were developed by McFadden (1981) to address the IIA limitation. The NL model is one of the commonly used model in this class. It is the generalization of the MNL model that is based on the idea that some alternatives may be joined in several groups called nests. The error term may represent some correlations within the nest, but different nests are still uncorrelated. It overcomes the IIA limitation of the MNL model and potentially improves upon the sequential logit model by allowing for correlations among error terms across different severity levels (Savolainen et. al 2011). Assuming the disturbances are generalized extreme value distributed, the NL model can be written as (McFadden, 1981):

$$P_n(i)=\frac{\exp\left[\beta_i X_{in}+\phi_i LS_{in}\right]}{\sum_{\forall l}\exp\left[\beta_l X_{\ln}+\phi_l LS_{\ln}\right]} \tag{3}$$

$$P_n(j\mid i)=\frac{\exp\left\lfloor\beta_{j|i} X_n\right\rfloor}{\sum_{\forall l}\beta_{J|i} X_{Jn}} \tag{4}$$

$$LS_{in}=LN\left[\sum_{\forall J}\exp\left(\beta_{J|i} X_{Jn}\right)\right] \tag{5}$$

Where,

P_n = unconditional probability of crash n resulting in injury outcome i

β = vectors of estimable parameters

X = it represents the vectors of measurable characteristics that determine the probability of injury severities.

$P_{n(j/i)}$ = the probability of observation n have injury severity level j conditioned on the outcome being in the outcome category i

For example in the nested structure shown in Fig 4, the outcome category i will be "injury" and $P_{n(j/i)}$ would be the binary logit model of injury severity outcomes; Non-fatal (injury) and fatal, whereas j is the conditional set of outcomes i-e conditioned on i and i is the unconditional set of outcome categories (the upper two branches of fig 4 i-e no injury & injury).

LN_{in} is the exclusive value (logsum), and ø is an estimable parameter. This equation system implies that the probability (unconditional) of having outcome j is,

$$P_n(j)=P_n(i)*P_n(j\mid \mathrm{i}) \tag{6}$$

The marginal distribution for term ε_s are still univariate extreme value, but there is some correlation within the nests. 1-λ is a measure of the correlation i.e. λ_m = 1 indicates no correlation.

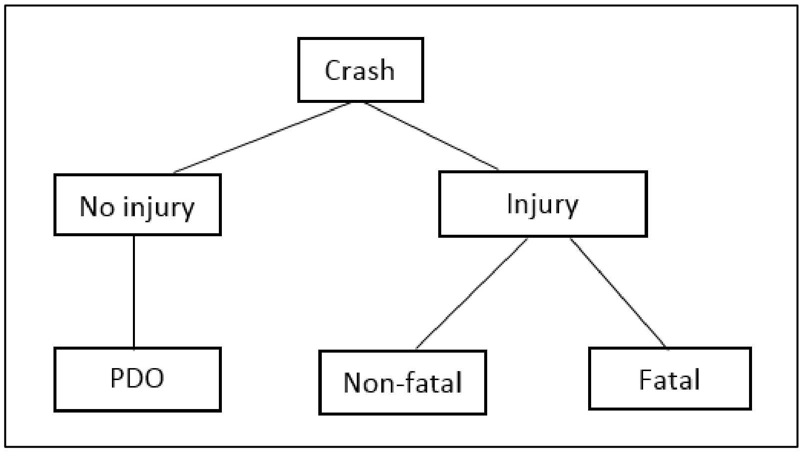

Figure 4 NL model structure

3.4 Mixed Logit Model

Mixed logit model also called random parameter logit model (RPL) or hybrid model is a relatively recent development for the analysis of discrete data (McFadden and Train, 2000). The random parameter model addresses a weakness of standard MNL model by allowing parameter values to vary across observations. For the derivation and application of the standard MNL model, it is assumed that parameters are fixed across all observations. When this assumption is incorrect, the parameter estimates and outcome probabilities are inconsistent (Washington et. al 2010).

Random parameter logit model is appropriate to account for the possibility of variation of different parameters across individual observations. Following the work presented by McFadden and Train (2000) to develop the RPL modeling approach, consider a function determining discrete outcome probabilities as;

$$T_{in} = \beta_i X_{in} + \varepsilon_{in} \qquad (7)$$

Where,

β_i = vector of estimable parameter for discrete outcome i

X_{in} = a vector of the observable characteristics (covariates) that calculate discrete outcomes for observation n,

$\mathcal{E}_{in}$ = disturbance term.

As mentioned in the previous section (eq-2), the standard MNL form can be written as

$$P_i(j) = \frac{\exp\left(\partial_j + \beta_l X_{ij}\right)}{\sum_{j=1}^{J} \exp\left(\partial_j + \beta_j X_{ij}\right)} \tag{8}$$

Where,

P_i (j) = the probability of observation i having discrete outcome j (j denoting all possible outcomes for observation n). By defining a mixed model (with a mixing distribution) whose outcome probabilities are defined as Pi (j) with

$$P_i\ (j) = \int P_i(j) f(\beta \mid \varphi) d\beta \tag{9}$$

Where $f(\beta \mid \varphi)$ represents the density function of β and φ, refers to the mean and variance of the density function, all other terms are previously defined. By putting the values of eq-7 in eq-8 we get

$$P_i\ (j) = \int \frac{\exp\left(\partial_j + \beta_l X_{ij}\right)}{\sum_{j=1}^{J} \exp\left(\partial_j + \beta_j X_{ij}\right)} f(\beta \mid \varphi) d\beta \tag{10}$$

Equation 8 indicates that the mixed logit probabilities P_i (j) are the weighted average of the standard MNL probabilities with the weights determined by the density function. In case of $f(\beta \mid \varphi)$ =1, the model reduces to simple MNL. The term β of eq-8, can now account for observation-specific variations of the effect of X on outcome probabilities, with the density function $f(\beta|\varphi)$ used to determine β. Different types distribution (normal, uniform, triangular distribution) can be used as a density function for β. RPL probabilities are thus a weighted average of different values of β across different observation where some elements of parameter vector β are random parameters and some are fixed.

3.5 Modeling Procedure

This section provides a general procedural approach to analyze and estimate the three models used in this study. The three models were estimated by using the NLOGIT-5 software package (Econometric Software, Inc). NLOGIT is widely used for data analysis in different fields such as transportation, economics, marketing, statistics and other social sciences. The details of estimating each model will be discussed in Chapter 5, however, the reader can refer to Applied Choice Analysis by Hensher et. al (2005).

The estimating procedure using NLOGIT of all the three models used in this study are discussed in detail. An initial model with independent variables was calibrated, each model was then revised by removing the non-significant variables (P-value > 0.1) and adding new variables. Fig 5 represents a general idea of the approach used to estimate each model.

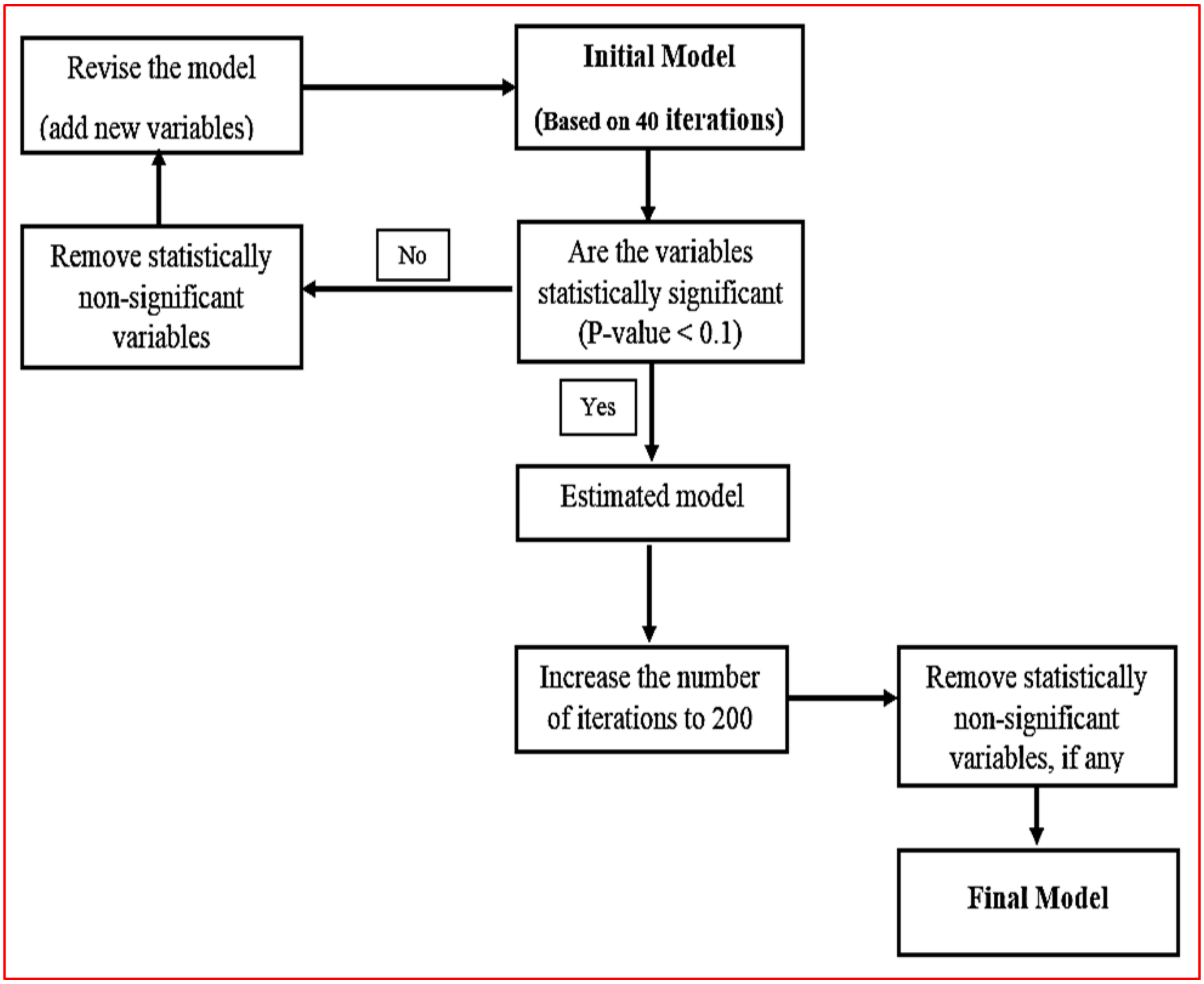

Figure 5 General procedure adopted for model estimation

CHAPTER 4 DATA PROCESSING

The dataset utilized in this study was extracted from the FRA highway-rail grade crossing inventory and crash databases. This chapter focuses on the data used for the analysis, the manipulation of the data extracted from the FRA database into the final dataset used for model estimation. The first section (4.1) of this chapter introduces the FRA database. This section introduces the different database files used to extract the information related to crash data. Section 4.2 details the merging procedures of the different files and data clean up. It further details the description and frequencies of the dependent variable and all independent variables utilized in this study.

4.1 Data Source

The FRA started an original national highway-rail crossing inventory database on January 1, 1975. This database includes both current and historical records with 80,000 to 100,000 crossings updated per year (Woll, 2007). The database contains three major data files; highway-rail crossing accident file, highway-rail crossing history file and highway-rail crossing inventory file. These three files are linked to each other by a unique crossing ID number that is common amongst the three files.

The highway-rail crossing accident database provides a history file of all the crashes reported at highway-rail crossings and the surrounding conditions at that time. This sub-database consists of records of all yearly crashes starting from 1975 to-date. This file has details such as speed of train and vehicle involved in the crash, type of train, type of materials carried (by freight vehicles), type of vehicle, crash circumstances, time of day, environmental conditions, and driver attributes etc.

The highway-rail crossing inventory file provides current crossing inventory information, which reflects the current state of each crossing with reference attributes. The highway-rail crossing history file reflect the changes made to the crossing including a reason for the update and an effective date of the update. The history file contains previous records of every crossing before any changes were made to the crossing, this is helpful to understand or to get inventory information of crossing before the changes were made at a particular crossing. The inventory file contains information such as average annual daily traffic (AADT), active and passive warnings, warning type, area type, geometric characteristics and coordinates of the crossings.

In order to get inventory information for the year a crash occurred, both highway-rail crossing inventory and highway-rail crossing history files were utilized. The data was substantially checked and cleaned for consistency, some IDs were missing in the highway-rail crossing inventory but were found in the accident files. In such case, the crossings were removed from the final data set.

4.2 Data Formulation

Initially, crashes at highway-rail crossings were extracted from highway-rail crossing accidents database for the year 2007-2015. The unique ID number between the three data sets were then used to extract inventory information for each accident/incident. The total number of accidents/incident were 19,689 and this number includes all kinds of crashes reported at crossings such as auto truck, passenger vehicles, pedestrians, school bus, motorcycle, at-grade and grade separated crashes etc. Heavy-vehicle (truck & truck-trailer) involved crashes at grade crossings were then extracted from the dataset, which contributed to about 15.2% (2,980) of the total accidents/incidents occurred at crossings for the year 2007-2015.

The heavy-vehicles dataset was then divided into two subsets; subset-I consisting of crashes from the year 2007-2014 with a total of 2664 observations (each observation representing a single crash) for model estimation. Subset-II consisted of 315 crashes (10.6% of total heavy-vehicle crashes from 2007-2015) for model validation . Fig 6 shows the steps towards the final dataset used in this study.

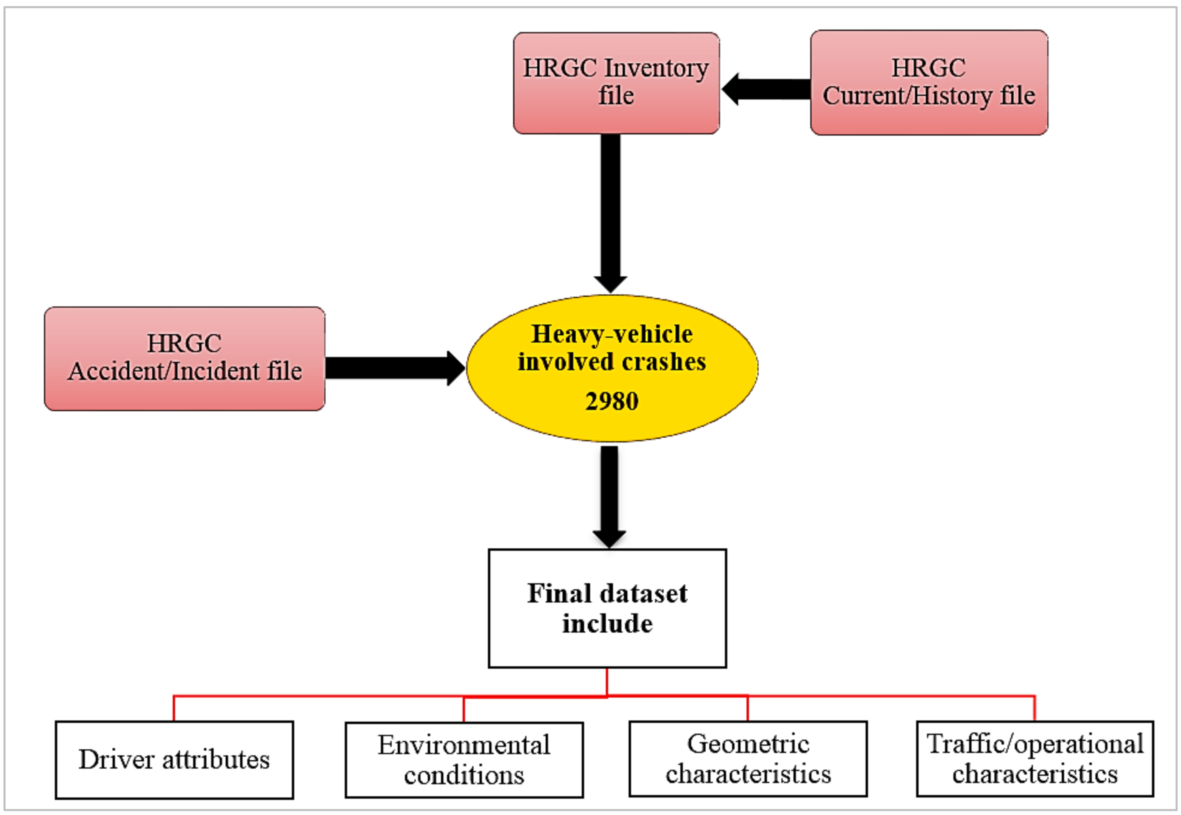

Figure 6 Data processing of HRGC crash data (2007-2015)

4.3 Data Description

The dependent variable i.e., injury severity consisted of three severity levels, property damage, injury and fatal. The three levels were coded as 0-property damage

(PDO), 1-injury (INJ) and 2-fatal. The estimating data set (subset-I) consisted of 2,005 PDOs (75.26%), 525 injury (19.7%) and 134(5%) fatal crashes. Table 2 and 3 presents details of some of the variables used in model estimation.

The parameters used in modeling were mostly related to crossing geometric characteristics, traffic-related variables such as different types of passive and active controls, truck driver attributes, environmental aspects and some crash specific details such driver behavior, circumstances of the crash, hazardous materials released if carried by either train or truck involved in the crash etc. Details of some important variables based on the analysis are presented in table 2 and 3. The more detailed form of these tables can be found in the appendix which includes all the parameters used in the process of model estimation.

Table 2 Descriptive statistics for the variables incorporated in the injury severity models

Variable Type	Description and Coding	Frequency Yes=1 No=0	Mean	Standard Deviation
Dependent Variable				
Truck driver injury severity	PDO	2005	-	-
	Injured	525	-	-
	Fatal	134	-	-
Independent Variables				
Motor Characteristics				
Vehicle type	Truck	832	0.31	0.46
	Truck-trailer	1832	0.69	0.46
Vehicle speed (mph)		NA	7.79	11.54
Hazardous materials carried	Yes	711	0.27	0.44
	No	1950		
Railway Characteristics				
Train speed (mph)		NA	30.67	18.36
Primary obstruction of track view	Yes=1	97	0.04	0.19
	No=0	2566		
Driver Attributes				
Driver age (years)		NA	45.70	13.61
Driver gender	Male	2472	0.96	0.04
	Female	96	0.19	0.19
Truck driver behavior/action of highway user	Driver went around the gates	199	0.07	0.26
	Standing RR equipment/ did not stop	1364	0.51	0.50
	Stopped on crossing	784	0.29	0.46
	Went around/ through temporary barricade	317	0.12	0.32
Traffic Characteristics				
Active controls				
Gates available (indicator)	Yes=1	1091	0.92	1.19
	No=0	1573		
Is track signaled	Yes=1	1430	0.55	0.50
	No=0	1183		
Highway traffic signal controling crossing	Yes=1	74	0.03	0.16
	No=0	2590		
Nearby hwy intersection have traffic signals	Yes=1	271	0.19	0.39
	No=0	1163		
Train detection system indicator	Yes=1	1369	0.52	0.50
	No=0	1247		
Indicator for availability of bells	Yes=1	1375	0.52	0.50
	No=0	1289		
Passive controls				
Stop sign available	Yes=1	383	0.26	0.66
	No=0	2281		
Pavement marking indicator (stop line/RR xing symbols)	Yes=1	1275	0.49	0.50
	No=0	1349		
Crossbuck assemblies available indicator	Yes=1	1894	0.71	0.45
	No=0	770		

Table 3 Descriptive statistics for the variables incorporated in the injury severity models

Variable Type	Description and Coding	Fraquency Yes=1 No=0	Mean	Standard Deviation
Environmetal Characteristics				
Weather	Clear	2436	0.91	0.28
	Rain	157	0.06	0.24
	Snow/sleet	71	0.03	0.16
Geometric Characteristics				
No of Lanes Crossing Railroad indicator	2-Lanes	2248	0.86	0.35
	4-Lanes	280	0.11	0.31
	More than 4 Lanes	88	0.03	0.18
Highway paved	Yes=1	2098	0.80	0.40
	No=0	517		
Intersecting roadway within 500ft	Yes=1	1603	0.61	0.49
	No=0	1007		
Smallest crossing angle	0-29	74	0.03	0.17
	30-59	338	0.13	0.34
	60-90	2205	0.84	0.36
Functional classification of road at crossing:	Rural = 1	1504	0.58	0.49
	Urban = 1	1090	0.42	0.49
Crash Specifications				
Hazardous material released by both (Highway user/ Rail equipment)	Yes=1	35	0.01	0.10
	No=0	2628		
Circumstances of accidents	Indicator for rail equipment struck highway user	2338	0.88	0.33
	Indicator for rail equipment struck by highway user	326	0.12	0.33
Indicator for position of the vehicle	Stalled or stuck on crossing/blocked on crossing by gates	302	0.11	0.32
	Stopped on crossing	660	0.25	0.43
	Moving over crossing	1647	0.62	0.49
	Trapped on crossing by traffic	54	0.02	0.14

CHAPTER 5 DATA ANALYSIS AND RESULTS

Section 5.1 presents the model estimation procedure of each model and its results. Section 5.2 shows comparisons between the three models based on the number of significant parameters, Akaike Information Criteria (AIC), log-likelihood function and model prediction accuracy. Section 5.3 presents discussion pertaining to the results obtained from the modeling and comparison of the three models.

5.1 Model Estimation

NLOGIT 5 was used for estimating the models by using data subset-I, consisting of 2,664 observations from 2007-2015. The dependent variable representing injury severity levels of truck driver was named “injury”. For MNL model, NLOGIT utilized single line data i.e., each observation representing a single crash. However, the data was converted to multi line format for NL and RPL model i.e., three rows represented each crash with each row representing an injury severity level. Therefor for NL and RPL models, the number of rows were 7992. The independent variables included in the model estimating process were based on previous research and their statistical significance in the modeling process.

5.1.1 Multinomial Logit Model

The category of PDO (coded as 0) was set as the baseline category for the MNL model. Different independent parameters were tried and those statistically not significant were removed from the final model. Model estimation removed observations with missing data and the final output is based on 2,156 observations.

Table 4 presents the results of final MNL model estimated for the injury severity of truck drivers at HRGCs. This table contains the estimated coefficients of the significant parameters and the standard error of these coefficients. A positive coefficient indicates increased likelihood toward a particular crash injury severity category compared to the no injury (PDO).

The results indicate that driver's injury severity increased with higher train speed and vehicle speed (truck, truck trailer); both findings being rational as higher speeds are known to result in severe injuries. After examining both rural and urban area, it was found that higher injury severity was more likely in rural areas. Since this study is focused on truck and truck-trailers crashes, the model revealed that trucks were more vulnerable to higher injuries compared to truck-trailers. Freight transport (either train or heavy vehicle) carrying hazardous materials was positively associated with injury severity of truck drivers. Thus carrying hazardous materials increased the likelihood of more severe crashes. After examining different driver characteristics, driver age and driver behavior while crossing were found statistically significant. Driver age was strongly associated with fatal crashes at 95% confidence level, indicating that older truck drivers are more vulnerable to fatal crashes.

Driver behavior that significantly increased the likelihood of severe crashes were crossing violation at HRGCs; the motorist attempts to drive around the gates when gates are closed. However, the presence of gates at the crossing was found to statistically significantly reduce the likelihood of a severe crash at a significance level of 95%. HRGCs with a minimum crossing angle of 60^0-90^0 were found positively associated with crash severity outcome injury but negatively associated with fatal crashes.

Table 4 MNL model results

Multinomial Logit Model				
Log likelihood function	-1345.42			
Chi squared	463.98			
McFadden Pseudo R-squared	0.147			
Akaike Information Criterion (AIC)	2734.8			
No. of observations	2156			
Injury Severity	Coefficient	Standard Error	z	Prob. Z>Z
Injury Severity Level : Injury				
Constant	-3.21317	0.30515	10.53	0.000
Hazardous Material	0.32804	0.12427	2.64	0.0083
Speed of Train	0.03183	0.0035	9.08	0
Rural Area	0.33552	0.12854	2.61	0.009
Indicator for Gates availability	-1.03754	0.15556	-6.67	0
Motorist Behavior: the motorist went around the gates	1.15276	0.22236	5.18	0
Speed of Vehicle (truck/ truck trailer)	0.01934	0.00484	4	0.0001
Age of Driver	N/S	0.00409	0.38	0.7003
Smallest crossing; $60^0 – 90^0$	0.3439	0.164	2.1	0.036
Truck indicator in crash	0.88658	0.12072	7.34	0
Train Detection System indicator	-0.2393	0.12902	1.86	0.0636
Injury Severity Level : Fatal				
Constant	-7.25993	0.60271	-12.05	0
Hazardous Material	0.40507	0.21565	1.88	0.0603
Speed of Train	0.06468	0.00637	10.15	0
Rural Area	0.81145	0.26517	3.06	0.0022
Indicator for Gates availability	-1.07314	0.30452	-3.52	0.0004
Motorist Behavior: the motorist went around the gates	1.50018	0.37524	4	0.0001
Speed of Vehicle (truck/ truck trailer)	0.03737	0.00756	4.94	0
Age of Driver	0.0225	0.00695	3.24	0.0012
Smallest crossing; $60^0 – 90^0$	N/S	0.25432	-0.75	0.4518
Truck indicator in crash	1.48149	0.208	7.12	0
Train Detection System indicator	-0.0738	0.23991	0.31	0.7582

Note: dependent variable = injury severity of truck drivers is coded as; PDO = 0, injury=1 and fatal = 2

5.1.2 Nested Logit Model

The NL model permits partial relaxation in the IID assumption of the MNL model by permitting for differential variation in the unobserved effects across partitions (nests) of alternatives but not with in same partitions. That is with only a minor complexity of model estimation (Hensher et. al 2005).The NL model is estimated in the form of a tree (i.e., alternatives are separated in different nests). NLOGIT has the ability to estimate NL models with up to four nest levels. However, the majority of NL models estimated as part of choice studies have only two levels or in some cases three levels. The three highest levels of NL tree structure are named, from the highest level to the lowest level, as trunk, limbs, and branches. This general concept of NL model can be found in the Applied Choice Analysis (Hensher et. al 2005) and Statistical and Econometric Methods for Transportation Data Analysis (Washington et. al 2010).

Different tree structure can be formulated in NL models, some branches can even have one alternative called degenerate branches. There exists a unique Inclusive Value (IV) parameter for each trunk, limb and branch specified as part of the tree structure in the NL model. For model estimation, one can constrain or normalize several of the IV parameters.

Different tree structures were tested to develop the best possible structure for NL model estimation. The tree structures tested in this study are shown in Fig 7 and the final NL model tree structure with a degenerate branch selected is Fig 7(d). It is common in many applications to have partition or nests with only one alternative within the nest referring to it as a degenerate branch and we had a similar situation. The tree structure performing better has a degenerate branch (No injury) with only one alternative i-e PDO. Whereas the nest of branch “Injury” has two alternatives; non-fatal and fatal injury.

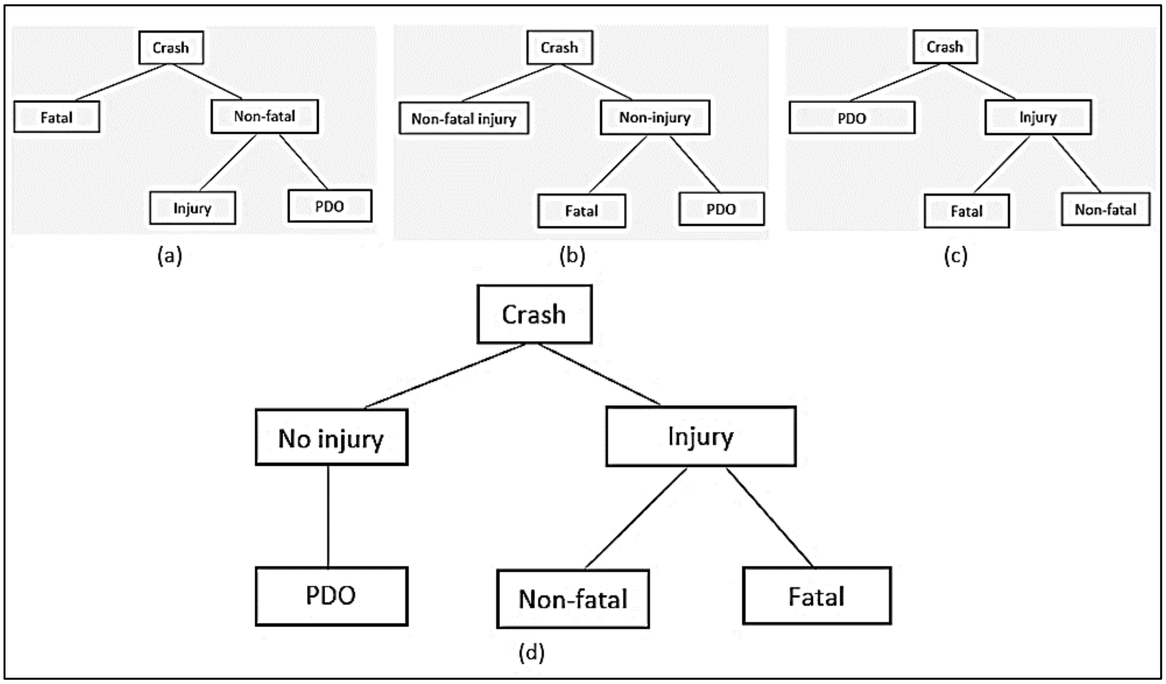

Figure 7 Different tree structures examined for NL model estimation

Given that PDO is the only alternative with in the nest, it follows that the conditional choice probability at level one for PDO must be equal to 1. Table 5 presents the details of final NL model estimated. The NL model was estimated using an estimation technique known as Full Information Maximum Likelihood (FIML). For NL models with two to four levels, it is common to use simultaneous estimation techniques which provide statistically efficient parameters estimates. The simultaneous estimation of braches, trunks and limbs of NL model is achieved using FIML (Hensher et. al 2005). Initially, for testing different independent variables, the maximum number of iterations were set to 40. However, the maximum number of iterations were then increased to 200 for the final model estimation. As mentioned earlier, NLOGIT feeds on a number of observations based on the number of outcomes. Since there were three outcomes for the dependent variable, the number of observations were three times, thus across all 2644 choice sets (observations), there was a total of 7992 alternatives.

As per requirement for the degenerate branch, "No Injury" (table 5) was set to 1. To test if the IV value is statistically different than 0 and 1, two test are required. The tests are undertaken to see if there exists an evidence for a partition of the tree structure at this section of the model. This procedure was repeated by using the different tree structures mentioned earlier. To identify if the IV is statistically different than zero at 95% confidence level (alpha=0.05), the IV estimated is divided by its associated standard error and is compared with the critical value of ± 1.96. If the parameter is found not to be significant (zero), the parameter remains in the 0-1 bound. By doing so, it was found that the parameter is significantly different than zero (7.757/1.6668=4.65 > 1.96). This indicates that the two scale parameters taken from different levels to form the IV parameter are not statistically different.

A second test is required to see if the parameter estimate is different than 1 (Greene 2005). This is done by using the Wald-test, which is undertaken with a simple modification to the test conducted to determine whether the parameter is statistically equal to zero.

$$\text{Wald-test} = \frac{IV_{parameter} - 1}{Std.error} \tag{11}$$

The IV parameter for "Injury" branch was found to be statistically different than zero. To determine if it is different than 1, eq-9 is used. By comparing the test-statistics of 4.05 to the critical value of ± 1.96 (i.e., at alpha equal to 0.05), we reject the null hypothesis that branch (injury) is statistically equal to one. This indicates that the two branches should not collapse into a single branch.

$$\text{Wald-test} = \frac{7.75720-1}{1.6668} = 4.05$$

The results obtained from the NL model had some similarity with the MNL model results. In addition, some new parameters were also found to be statistically significant. The total number of independent parameters found to be significant at 90% confidence interval are 14.The results indicate that crossing angle of 60^0-90^0 and motorist behavior (went around the gates) were positively associated with severity level; injury. Whereas train and vehicle speed, hazardous materials carried, the age of truck driver and crashes reported to occur in rural area are positively associated with fatal crashes. Crashes occurring in the rural area and older drives increased the likelihood of more severe crashes. Two circumstances of a crash (rail equipment struck highway user and rail equipment struck by highway user) were examined and it was found that crash circumstance in which highway user (truck/truck trailer) was hit by rail equipment, increase the likelihood of a fatal crash. This finding is reasonable as driver's injuries would be more severe when the train (being larger in size) strikes truck or trailer.Other factors that were found positively associated with injury severity were; hazardous materials carried, the position of truck/trailer i.e., when it was moving over the crossing. Trucks involved in crashes at HRGCs were also found more severe. However, the presence of gates and location of crossing near the highway (i.e., with in 500ft) decreased the likelihood of severe crashes.

Table 5 NL model results

Nested Logit Model				
Log likelihood function	-1330.35			
Chi squared	1147.43			
McFadden Pseudo R-squared	0.30131			
Akaike Information Criterion (AIC)	2696.7			
No. of observations	2153			
Injury Severity	Coefficient	Standard Error	z	Prob. Z>Z
Injury Severity level : Injury				
Constant	-0.39718	0.06554	6.06	0.000
Speed of Train	0.00213	0.00104	2.04	0.0416
Indicator for Gates availability	-0.1086	0.03889	-2.79	0.0052
Motorist Behavior: the motorist went around the gates	0.12342	0.04899	2.52	0.0118
Smallest crossing; $60^0 – 90^0$	0.04345	0.02628	1.65	0.0982
Indicator for primary obstruction of track view	-0.1050	0.06089	-1.73	0.0845
Highway near intersection (500ft)	-0.0392	0.0203	-1.93	0.0536
Train Detection System indicator	-0.0444	0.0212	2.09	0.037
Injury Severity level : Fatal				
Constant	-3.99	0.4829	-8.26	0
Hazardous Material	0.19764	0.0755	2.62	0.0089
Speed of Train	0.0159	0.00348	4.59	0
Rural Area	0.29178	0.09836	2.79	0.003
Speed of Vehicle (truck/ truck trailer)	0.00678	0.00312	2.17	0.0297
Age of Driver	0.00556	0.00254	2.19	0.0287
Truck indicator in crash	0.61336	0.12303	4.99	0
Circumstances of Crash: rail equipment struck highway user	0.23114	0.11827	1.95	0.0506
Position of vehicle: moving over crossing	0.75791	0.18783	4.04	0.0001
IV Parameter				
NoINJ	1	Fixed parameter		
Injury	7.757	1.6668	4.65	0

Note: dependent variable = injury severity of truck drivers is coded as; PDO = 0, injury=1 and fatal = 2

5.1.3 Random Parameter Logit Model

The RPL model also known as the mixed logit model offers the ability to overcome the limitation imposed by the MNL and NL model, as discussed in Chapter 3. The RPL model comparatively represents the latest development in the econometric toolkit available to the choice modeler. It provides flexibility to estimate different parameters as random. The analyst can test different parameters in the data set for random effects by using the function (;fcn) command. Different distribution can be assigned to the random parameters, to improve the overall performance of the model. In the RPL model estimation, all the independent parameters were first assumed random and both the uniform and normal distribution were tested for randomness. The random parameters that were not found statistically significant at 90% confidence interval for both normal and uniform distribution were then kept as fixed parameters in the model specifications and examined.

The historic approach used in the estimation of RPL models has been, to use R random draws from some derived empirical distributions. However, to get satisfactory results a large number of random draws is computationally time-consuming. Another limitation cited by using random draws in estimating RPL model is that random draws may over-sample (in assigning parameters over the sampled population) from the areas of distributions while leaving the other areas of the distribution under-sampled (Hensher et. al 2005). To overcome this, a number of intelligent draws methods have been introduced which have been shown to provide no discernible degradation in model results.

Unlike random draws, intelligent draw methods are designed to sample the entire parameter space in accordance with the empirical distribution imposed. NLOGIT provides two types of intelligent draws; Standard Halton Sequence (SHS) and Shuffled Uniform

Vectors (Hess et. al 2003). Bhat (2001) compared the results of models estimated by using SHS intelligent draws and random draws. It was reported that by using Halton draws to estimate the model, the results can be obtained with only one-tenth of the total number of random draws. Thus, SHS intelligent draws were selected for RPL model estimation. Initially, the number of Halton draws and iterations were set to 40, to identify significant random and fixed parameters at a confidence level of 90% (p-value=0.10). The final model was then revised by increasing the number of draws (SHS) and maximum iterations to 200. Table 6 presents results of the final RPL model estimated. The original output of NLOGIT for the final estimated model can be found in the Appendix-II.

For injury crash level in the RPL model, vehicle position (i-e stopped on the crossing) was found to follow normal random distribution implying that the parameters can vary from crash to crash. All other independent variables were restricted to fixed parameters. A total of 16 parameters (including random parameter) were found statistically significant at overall 90% confidence level. The parameters that were found to increase the likelihood of crash severity at 90% confidence level (at alpha=0.05) are; vehicle and train speed, crashes occurring in rural area, crossing angle of 60^0-90^0, driver age, crash involving trucks at HRGCs, hazardous materials carried by either train or road user, motorist behavior; went around the gates (violation) and crashes circumstances in which train strikes roadway user (truck/truck trailer). However, primary obstruction of track view, crossings within 500ft of the highway were found negatively associated with injury severity of heavy-vehicle drivers in crashes at HRGCs. These parameters were also found to have similar behavior with crash injury severity in MNL and NL model results. Two additional variables that were found statistically significant in the RPL model were the position of the vehicle; stopped on crossing (random parameter) and the presence of signal equipment. The position of vehicle i.e., the vehicle stopped on crossings seemed to reduce the likelihood of crash severity level "injury". This result appears reasonable in light of the common practice of abandoning the vehicle when stalled on a crossing.

Table 6 RPL model results

Random Parameter Logit Model				
Log likelihood function	-1329.9			
Chi-squared	2070.7			
McFadden Pseudo R-squared	0.4377			
AIC	2707.8			
No. of observations	2153			
Injury Severity	**Coefficient**	**Standard Error**	**z**	**Prob. Z>Z**
Random parameter in utility functions				
Position of vehicle: Stopped on crossing	-1.39064	0.84437	-1.65	0.0996
Injury Severity level : Injury				
Constant	-2.899	0.2165	2.1	0.000
Hazardous Material	0.36489	0.13203	2.76	0.0057
Speed of Train	0.03656	0.00424	8.61	0
Rural Area	0.30245	0.1365	2.22	0.0267
Indicator for Gates availability	-0.83641	0.17072	-4.9	0
Motorist Behavior: the motorist went around the gates	0.83721	0.22786	3.67	0.0002
Truck indicator in crash	0.93508	0.1291	7.24	0
Speed of Vehicle (truck/ truck trailer)	0.0116	0.00516	2.25	0.0246
Smallest crossing; $60^0 – 90^0$	0.3872	0.16674	2.32	0.0202
Indicator for primary obstruction of track view	-0.72544	0.37753	-1.92	0.0547
Highway near intersection (500ft)	-0.27709	0.1208	-2.29	0.0218
Indicator if track is signaled	-0.2904	0.13186	-2.2	0.0276
Train Detection System indicator	-0.3059	0.1349	2.27	0.0233
Injury Severity level : Fatal				
Constant	-9.117	0.66607	-13.7	0
Hazardous Material	0.45507	0.21652	2.1	0.0356
Speed of Train	0.06253	0.00642	9.74	0
Rural Area	0.79671	0.255	3.12	0.0018
Speed of Vehicle (truck/ truck-trailer)	0.02959	0.00843	3.51	0.0005
Age of Driver	0.02202	0.00668	3.3	0.001
Truck indicator in crash	1.46426	0.20902	7.01	0
Circumstances of Crash: rail equipment struck highway user	0.58321	0.33823	1.72	0.0847
Position of vehicle: vehicle moving over crossing	1.56492	0.3222	4.86	0
Distns. Of Standard deviation or limits of triangular				
Position of vehicle: Stopped on crossing (Normal distribution)	1.79577	1.00448	1.79	0.0738

Note: dependent variable = injury severity of truck drivers is coded as; PDO = 0, injury=1 and fatal = 2

5.2 Model Comparison

The approach for model comparison was adopted from previous research (Abdel-Aty and Abdel Wahab, 2004; Yasmin and Eluru, 2013; Zhao and Khattak, 2015). The following criteria were used in model comparison: number of significant parameters, models classification accuracy, model's interpretation power and model's goodness-of-fits.

Table 7 represents the results of all three models. The RPL model had the highest number of statistically significant parameters (16), compared to NL model (14) and MNL model (10). The greater number of significant parameters in the model comparatively leads to a better model in terms of higher adjusted R-square; MNL (0.142), NL (0.298), RPL (0.4346). It also helps identify additional explanatory variables impacting or associated with the dependent variable. The RPL model overcomes individual variation issues compare to MNL model and does not exhibit the IIA (Independence of irrelevant alternatives) property. However, NL model represents a partial relaxation of the IIA property. In terms of interpretation, RPL model had more flexibility in estimation and thus, performed better compared to the NL and the MNL models. The parameter found to vary across individual crash was the position of the vehicle (i-e stopped on the crossing), it was found to be normally randomly distributed.

5.2.1 Likelihood-Ratio Test

To examine the model fit, the likelihood ratio test and AIC (Akaike Information Criteria) were compared. The likelihood ratio test is conducted at 95% confidence level (alpha=0.05) with a degree of freedom equal to the difference between the significant parameters between the two models. The null hypothesis is that there is not statistical difference between the two models. The general form of likelihood ratio for comparing two models can be shown as; LL ratio test = -2($LL_{largest}$ - $LL_{smallest}$)

˜ X(difference in the number of parameters estimated between the two models)

The LL-ratio test indicated that the NL model was statistically better than the MNL model in this case. That is the LL-ratio value (i.e., 30) was larger than the critical value (9.487) at 95% significance level. Similar results were found between the RPL and the MNL model. Which is obvious, because the LL-ratio test between RPL and NL model indicated that the RPL model was not significantly better than the NL model. That is, the LL-ratio statistics for RPL and NL with 2 degree of freedom was 2.0, which was smaller than the Chi-square critical value of 5.99 at the 95% significance level. The AIC values for MNL, NL and RPL models were 2734.8, 2696.7 and 2707.8 respectively. Models with lower AIC values are preferable, therefore RPL model and NL model were superior to the MNL model in this case. The NL model had slightly better model fit than the RPL model based on the AIC criteria.

- Likelihood ratio test between MNL & NL model (df = 14-10= 4)

  LL ratio test = -2[-1345-(-1330)] = 30

  Chi-square critical value at 95% confidence level (df=4) = 9.487)

- Likelihood ratio test between NL & RPL model (df = 16-14= 2)

  LL ratio test = -2[-1330-(-1329)] = 2

  Chi-square critical value at 95% confidence level (df=2) = 5.99

- Likelihood ratio test between MNL & RPL model (df = 16-10= 6)

  LL ratio test = -2[-1345-(-1329)] = 32

  Chi-square critical value at 95% confidence level (df=6) = 12.59)

Table 7 Driver injury severity: MNL, NL and RPL models

Variables	**MNL**		**NL**		**RPL**	
	Injury	**Fatal**	**Injury**	**Fatal**	**Injury**	**Fatal**
Constant	-3.2131	-7.2599	-0.3972	-3.99	-2.899	-9.117
Vehicle Characteristics						
Speed of Train	0.03183	0.06468	0.00213	0.0159	0.03656	0.06253
Hazardous Material Carried	0.32804	0.40507	N/S	0.19764	0.36489	0.45507
Truck indicator in crash	0.88658	1.48149	N/S	0.61336	0.3059	1.46426
Speed of Vehicle (truck/ truck trailer)	0.01934	0.03737	N/S	0.00678	0.0116	0.02959
Driver Attributes						
Motorist Behavior: the motorist went around the gates	1.15276	1.50018	0.12342	N/S	0.83721	N/S
Age of Driver	0.00157	0.0225	N/S	0.00556	N/S	0.0222
Crash Specific Characteristics						
Circumstances of Crash: rail equipment struck highway user	N/S	N/S	N/S	0.23114	N/S	0.58321
Position of vehicle: vehicle moving over crossing	N/S	N/S	N/S	0.75791	N/S	1.56492
Position of vehicle: Stopped on crossing (Normal distribution)	-	-	-	-	-1.39	N/S
Standard deviation of distribution	-	-	-	-	1.79577 (1.0045)	N/S
Traffic Characteristics						
Indicator if track is signaled	N/S	N/S	N/S	N/S	-0.2904	N/S
Indicator for Gates availability	-1.0375	-1.0731	-0.1086	N/S	-0.83641	N/S
Train Detection System indicator	-0.2394	-0.0738	-0.0444	N/S	-0.3059	N/S
Geometric Characteristics						
Rural Area	0.33552	0.81145	NS	0.29178	0.30245	0.79671
Smallest crossing; $60^0 – 90^0$	0.3439	-0.1914	0.04345	N/S	0.3872	N/S
Indicator for primary obstruction of track view	N/S	N/S	-0.1050	N/S	-0.72544	N/S
Highway near intersection (500ft)	N/S	N/S	-0.0392	N/S	-0.27709	N/S
Inclusive Value (NL model)						
NoINJ	-	-	1		-	-
Atleast injury	-	-	7.757		-	-
Model Characteristics						
Number of Significant parameters	10		14		16	
Log likelihood function	-1345.42		-1330.35		-1329.9	
Chi squared	463.98 (df=20)		1147.4 (df=18)		2070.78 (df=24)	
McFadden Pseudo R-squared	0.14707		0.3013		0.4377	
Adjusted R-square	0.142		0.2984		0.4346	
AIC	2734.8		2696.7		2707.8	
Inf. Cr. AIC	1.268		1.253		1.258	
Note: N/A is not applicable, whereas N/S implies not significant at 10% level. All other values are statistically significant at 10% level.						

5.2.2 Model Prediction

The prediction accuracy of the three models was compared using subset-II which consisted of heavy-vehicle crashes at HRGC reported in 2015. As mentioned before, the testing data (subset-II) had 315 HRGC crashes which constituted about 10.6% of the total reported crashes between 2007 and 2015.The severity outcomes of the 2015 crashes were consistent with the 2007-2014 crashes, there was 75.26% PDOs, 19.7% injury crashes and 5% fatal crashes, while the corresponding percentages in the 2015 crash dataset were 76.8%, 19.7% and 5% respectively. The prediction success and failures for the three models are shown in Table 8. The row value represents the actual injury outcome while the column value is the model predicted value.

Comparison of the model prediction indicated that the MNL model correctly classified 74.8% of the 2015 observations while the NL and the RPL models correctly classified 75.95 and 75.2% of the observations, respectively. Hence, there is not much difference in the overall prediction accuracy of the three models. However, for fatal crashes, the MNL and RPL model performed better in terms of classification compared to the NL model. The prediction accuracy of an individual crash severity level for each model is presented in Fig 8. It was observed that for prediction of fatal crashes, the NL model underperformed (did not classify fatal crashes). However, the MNL model and the RPL model had similar results. Thus, it was concluded that MNL and RPL model had better prediction accuracy in this case.

Table 8 Prediction success table for MNL, NL & RPL model using 2015 crash

MNL Model				
Category	**Predicted**			
Actual	**PDO**	**Injury**	**Fatal**	**Total/actual observed**
PDO	183	7	0	190
INJ	46	9	2	57
FATAL	6	4	1	11
TOTAL	235	20	3	258
Percentage correctly classified =		193(100)/258		74.81
NL Model				
Category	**Predicted**			
Actual	**PDO**	**Injury**	**Fatal**	**Total/actual observed**
PDO	183	7	0	190
INJ	44	13	0	57
FATAL	4	7	0	11
TOTAL	231	27	0	258
Percentage correctly classified =		196(100)/258		75.97
RPL Model				
Category	**Predicted**			
Actual	**PDO**	**Injury**	**Fatal**	**Total/actual observed**
PDO	184	5	1	190
INJ	46	9	2	57
FATAL	7	3	1	11
TOTAL	237	17	4	258
Percentage correctly classified =		194(100)/258		75.20

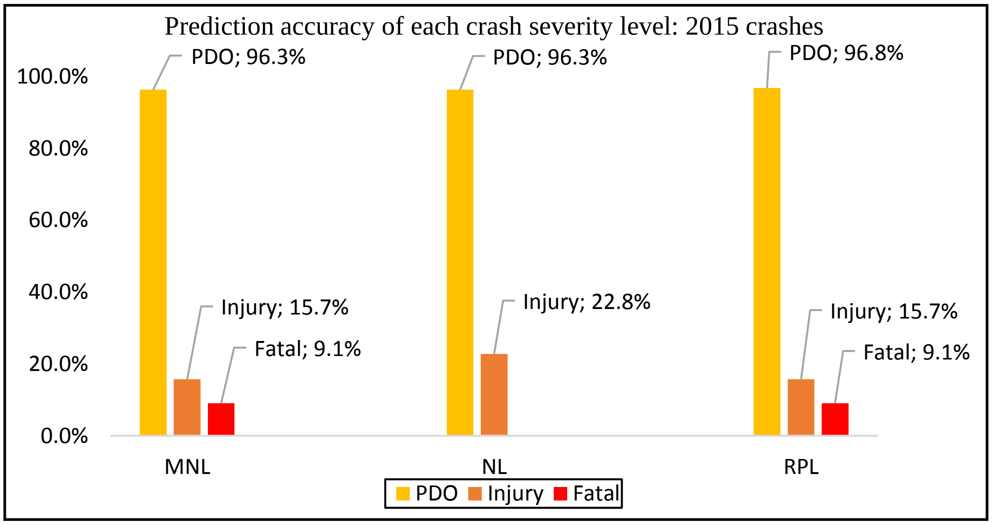

Figure 8 Prediction comparison of MNL, NL and RPL model in percentage

5.3 Results and Discussion

Comparison of all three models revealed that the RPL model had the most significant parameters included in its specification and had the best interpretation power compare to the other two models due to more flexible parameter estimates (randomly assigned with different distributions). In terms of goodness-of-fit, the RPL and NL model were significantly better than the MNL model. However, there was no significant difference found between the RPL and the NL model. Although the overall prediction accuracy of all three models were found to be similar, but it can be said that the MNL and the RPL model performed better in terms classifying fatal crashes. Overall, the RPL model performed slightly better than the MNL and the NL model for driver's injury severity analysis of heavy vehicle involved crash at highway-rail grade crossings. Thus, the factors associated with driver's injury severity at HRGCs identified by the RPL model are discussed below.

Sixteen independent variables were identified as being statistically significant with different driver's injury severity levels of train-heavy vehicle crashes at HRGCs based on the RPL model at the 90% significance level. The results indicated that train speed and the vehicle speed were positively associated with injury severity at the 99% significance level. Both findings were found to have similar association with injury severity in literature (Ishak et. Al 2011, Hao and Daniel 2015, Zhao and Khattak, 2015, Jun Liu et., al 2015) and were reasonable as higher speeds are commonly associated with more severe injuries (Ishak et. Al 2011, Hao and Daniel 2015, Zhao and Khattak, 2015, Jun Liu et., al 2015).

Truck involved crashes at HRGCs significantly increased the likelihood of a more severe crash. The total number of truck-train crashes consisted of 9.5% fatal and 27.8% of injury crashes, whereas as truck-trailer although had a higher number of total crashes (68.7%) consist of 3% fatal and 16% of injury crashes. The dummy variable indicating hazardous materials carried by either road user or train significantly increased the likelihood of a more severe crash at 99% significance level.

Different geometric characteristics that were statistically significant to driver injury severity included; crossings at the rural road, crossing angle 60^0-90^0, the intersecting roadway within 500ft of the crossing and primary obstruction of the track view. Crashes that occoured in the rural areas were more severe. About 58% of the total crashes (2594) were reported in rural areas in which the total number of injury and fatal crashes were 24% and 7.2% respectively. Primary obstruction of track view and roadway located within 500ft of the crossing were found to be negatively associated with severity level; injury. According to the data, no fatal or injury crashes were reported when truck view was obstructed. Similarly, 17.3% and 4.6% of total crashes reported at crossings within 500ft of intersecting roadway were injury and fatal respectively, whereas 23.7% and 6% of total crashes occurring at crossing not within 500ft of roadway were injury and fatal respectively. This explains the negative sign associated with the two variables for ‘injury’ severity level.

Different active and passive traffic controls were examined in model estimation and three types of passive control devices were found to reduce the likelihood of injury severity; presence of rain detection system, gates installed and if the track was signaled. The model results indicates that the availability of gates decreases the likelihood of a severe crash of heavy-vehicle drivers at HRGCs; this finding is consistent with previous studies

(Jun Liu et al., 2015). Three types of train detection systems were specified in the data set; constant warning, motion detection and direct current track circuit.

Driver attributes that significantly increased injury severity of truck drivers were the age of driver and motorist (truck/truck-trailer) action “went around the gates while crossing”. Other crash specific characteristics that increased the likelihood of crash injury severity were; when trains struck road user and vehicle moving over the crossing. Both of the findings are reasonable and consistent with each other. Drivers will be more vulnerable to severe injury when train strikes the road user. About 62% of total crashes (2007-2014 crashes) were reported when road user was moving over the a crossing, in which 26% where injury and 7.3% were fatal crashes. The parameter representing the position of the vehicle (i.e., stopped on the crossing), was found to follow normal random distribution implying that it varied from crash to crash. This parameter was negatively associated with “injury” category of severity level.

CHAPTER 6 CONCLUSIONS AND FUTURE RESEARCH

This chapter presents a summary of the research, including a brief discussion of the results. Based on the research findings, this chapter presents the conclusions. It also provides infomraton on limitations of this research study and recommendations for future research on truck driver safety at HRGCs.

6.1 Research Summary

Heavy-vehicle crashes account for 14% to 17% of the yearly crashes reported at HRGCs in the US; the estimated cost of the damages from these crashes is about $49 million. No substantial decrease was observed in truck-involved crashes at HRGCs during 2007-2015. Heavy-vehicle crashes at HRGC reported between 2007-2015 were utilized in this study. A total of 2664 observations (2007-2014) were used for model estimation. The models estimated in this study were MNL, NL and RPL. Criteria used for comparison of the estimated models were AIC, model interpretation power, goodness-of-fit, the number of significant parameters and models prediction accuracy (using 2015 crash data). For dependent variables with three injury severity levels, sixteen independent variables were statistically significant at 90% confidence level (alpha=0.10).

6.2 Results and Discussion

Comparison of the three models revealed that the RPL model performed better than the MNL and NL models. Statistically significant parameters that were positively associated with injury severity included speed of train and road user (truck/trailer), truck-train crash, hazardous materials carried by either one or both users, driver behavior " went around the gates", age of driver, crashes reported in rural areas and crashes at minimum crossing angle of 60-90 degrees. Crash specific characteristics increasing the likelihood of fatalities included when train struck heavy-vehicle and when the vehicle was moving over the crossing.

Higher speeds were commonly associated with more severe injury. This finding is reasonable and consistent with previous injury severity research. The total number of truck-train crashes reported were comparatively lower (Table 2) than trailer-train crashes. However, truck-train crashes constituted 9.5% fatal and 27.8% injury crashes. Whereas truck-trailer crashes at HRGC consisted of 3% fatal and 16% injury crashes (2007-2014).

The dataset included 200 crashes resulting from the heavy-vehicle driver going around crossing gates. Thus resulting in about 28% injury and 9.6% fatal crashes. About 58% of the total crashes were reported in rural areas, which consisted about 7.2% fatal and 23.8% injury crashes. Heavy-vehicles moving over HRGC i.e., it failed to make a stop for the oncoming train, turned to be more severe. Examples of such instances include truck drivers unaware of oncoming trains due to poor visibility, the absence of appropriate traffic warnings and driver inattention. Heavy-vehicles hit by a train while moving over the crossing consisted of 26% injury and 7.3% fatal crashes. Age of driver and when train strikes the road user (truck/trailer) increased the likelihood of a severe crash. This finding is reasonable and consistent with injury severity of motor vehicle at HRGC (Zhao and Khattak 2015).

Variables that significantly decreased the likelihood of a severe crash were; crossing with gates, if the track is signaled, train detection system, if the track was obstructed and crashes in which heavy vehicles stopped on the crossing. The variables representing the position of the vehicle (i.e., stopped on the crossing), was found to follow normal random distribution implying that it varied from crash to crash. This parameter was negatively associated with "injury" category of severity level.

6.3 Conclusions

This research was undertaken with the objectives to: 1) identify factors associated with injury severity of heavy-vehicle drivers in crashes reported at HRGCs and 2) identify a more suitable model for modeling heavy-vehicle drivers' injury severities in crashes reported at HRGCs. Based on the results both objectives were successfully achieved. The following conclusions are drawn:

- Truck drivers' injuries in crashes reported at HRGCs are positively associated with the following factors: speed of train and road user (truck/trailer), truck-train crash, when train strike road user (truck/trailer), hazardous materials carried by either one or both users, driver behavior "went around the gates", age of driver, crashes reported in rural areas and crashes at minimum crossing angle of 60-90 degrees.
- Truck drivers' injuries in crashes reported at HRGCs are negatively associated with the following factors: train detection system, gates, if track is signaled, when the track is obstructed, HRGCs within 500 feet of a highway and position of vehicle "heavy vehicle stopped on the crossing".

The RPL was most suitable for modeling truck drivers' injuries in crashes reported at HRGCs amongst the models considered, based on criteria used for judging the models, and the dataset used in this study.

6.4 Limitation and Future Research

This research investigated different factors associated with driver injury severity of heavy-vehicles but did not consider the injury severity of the most severe person in the crash. Furthermore, results indicated that driver behavior had a strong relationship with injury severity. However, this study did not consider truck drivers' physical and personality characteristics such as health/illness, financial and educational levels, driving experience, past traffic citations, etc. These chracteristics were not available for this study but future research should attempt to include such data in evaluating truck drivers' safety at HRGCs.

Truck drivers going around crossing gates and moving over crossing were positively associated with injury severity. Future research can build on this finding by identifying factors that are associated with such unsafe driving behavior, e.g., driver age, gender, driving speed range, visibility and environmental conditions. Such research will allow for more targeted information campaigns and educational activities aimed at improving HRGC safety.

This study includes three models but future studies can consider other types of models and techniques. This research considered the unordered response of the dependent variable, ordered response models such as OP and GOL etc. may be considered. Other methods such as Artificial Neural Network (ANN) and different data mining techniques were used in the past (Abdelwahab and Abdel-Aty. 2001, Chang and Wang. 2006, Chimba and Sando. 2009). Such methods may be used to investigate truck drivers' injury severity in crashes reported at HRGCs.

REFERENCES

1. Abdel-Aty, M. A., and H. T. Abdel Wahab. Predicting Injury Severity Levels in Traffic Crashes: A Modeling Comparison. Journal of Transportation Engineering, Vol. 130, No. 2, 2004, pp. 204–210.

2. Abdelwahab, Hassan, and Mohamed Abdel-Aty. "Development of artificial neural network models to predict driver injury severity in traffic accidents at signalized intersections." *Transportation Research Record: Journal of the Transportation Research Board* 1746 (2001): 6-13.

3. Abdel-Aty, M. Analysis of driver injury severity levels at multiple locations using ordered probit models. Journal of Safety Research, Vol. 34, No. 5, 2003, pp. 597–603.

4. Bien-Aime, P. (2009). North Carolina" sealed Corridor" Phase I, II, and III Assessment (No. DOT-VNTSC-FRA-09-08). US Department of Transportation, Federal Railroad Administration, Office of Research and Development.

5. Chadwick, S. G., N. Zhou, and M. R. Saat. Highway-rail grade crossing safety challenges for shared operations of high-speed passenger and heavy freight rail in the US Safety Science, Vol. 68, 2014, pp. 128–137.

6. Chambers, M., Goworowska, J., Rick, C., & Sedor, J. (2015). Freight Facts and Figures 2015.

7. Chang, Li-Yen, and Hsiu-Wen Wang. "Analysis of traffic injury severity: An application of non-parametric classification tree techniques." *Accident Analysis & Prevention* 38, no. 5 (2006): 1019-1027.

8. Carson, J., and F. Mannering. The effect of ice warning signs on ice-accident frequencies and severities. Accident Analysis & Prevention, Vol. 33, No. 1, 2001, pp. 99–109.

9. Chimba, D., and T. Sando. "Neuromorphic prediction of highway injury severity." *Advances in Transportation Studies* 19, no. 1 (2009): 17-26.

10. Davey, J., A. Wallace, N. Stenson, and J. Freeman. The experiences and perceptions of heavy vehicle drivers and train drivers of dangers at railway level crossings. Accident Analysis & Prevention, Vol. 40, No. 3, 2008, pp. 1217–1222.

11. Eluru, N. Evaluating alternate discrete choice frameworks for modeling ordinal discrete variables. Accident Analysis & Prevention, Vol. 55, 2013, pp. 1–11.

12. Eluru, N., M. Bagheri, L. F. Miranda-Moreno, and L. Fu. A latent class modeling approach for identifying vehicle driver injury severity factors at highway-railway crossings. Accident Analysis & Prevention, Vol. 47, 2012, pp. 119–127.

13. Fan, W. D., M. R. Kane, and E. Haile. Analyzing Severity of Vehicle Crashes at Highway-Rail Grade Crossings: Multinomial Logit Modeling. Journal of the Transportation Research Forum, Vol. 54, No. 2, 2015, pp. 39–56.

14. FHWA, Manual on uniform Traffic Control Devices, in: U. S. D. O. T, Federal Highway Administration (ed.), Washington, D.C, 2007.

15. Freight analysis framework - FHWA freight management and operations,". [Online]. Available: http://www.ops.fhwa.dot.gov/freight/freight_analysis/faf/. Accessed: Nov. 2, 2016.

16. Glover, J. The Level Crossing Conundrum. Rail Professional, 2009 (152).

17. Hall, S. Reducing risk at automatically operated level crossings on public roads. IET Seminar on Reducing Risk at the Road Rail Interface, 2007, pp. 1–5.

18. Hensher, David A., John M. Rose, and William H. Greene. *Applied choice analysis: a primer*. Cambridge University Press, 2005.

19. Hellman, A., and T. Ngamdung. Illinois High-Speed Rail Four-Quadrant Gate Reliability Assessment. 2010 Joint Rail Conference, Volume 1, 2010, pp. 445–454.

20. Hao, W., and J. Daniel. Driver injury severity related to inclement weather at highway–rail grade crossings in the United States. Traffic Injury Prevention, Vol. 17, No. 1, 2015, pp. 31–38.

21. Hao, W., and J. Daniel. Motor vehicle driver injury severity study under various traffic control at highway-rail grade crossings in the United States. Journal of Safety Research, Vol. 51, 2014, pp. 41–48.

22. Hu, S.-R., C.-S. Li, and C.-K. Lee. Investigation of key factors for accident severity at railroad grade crossings by using a logit model. Safety Science, Vol. 48, No. 2, 2010, pp. 186–194.

23. Ishak, S. Z., S. Somenahalli, and W. L. Yue. An Assessment of Heavy Vehicle Safety at Level Crossing Using Petri Nets: South Australia Case Studies. Proceedings of the Eastern Asia Society for Transportation Studies, Vol. 8, 2011, p. 359.

24. Kweon, Y.-J., and K. M. Kockelman. Overall injury risk to different drivers: combining exposure, frequency, and severity models. Accident Analysis & Prevention, Vol. 35, No. 4, 2003, pp. 441–450.

25. Kallberg, V.-P., M. Anila, K. Pajunen, M. Kallio, and J. Hytönen. Assessment and Improvement of Safety at Finnish Railway-Road Grade Crossings. Transportation Research Record: Journal of the Transportation Research Board, Vol. 1801, 2002, pp. 54–60.

26. Kockelman, K. M., and Y.-J. Kweon. Driver injury severity: an application of ordered probit models. Accident Analysis & Prevention, Vol. 34, No. 3, 2002, pp. 313–321.

27. Khattak, A., and M. Gao. Truck safety at highway-rail grade crossings. Final report to the Mid-America Transportation Center, 2012

28. Khorashadi, A., D. Niemeier, V. Shankar, and F. Mannering. Differences in rural and urban driver-injury severities in accidents involving large-trucks: An exploratory analysis. Accident Analysis & Prevention, Vol. 37, No. 5, 2005, pp. 910–921.

29. Lee, J., and F. Mannering. Impact of roadside features on the frequency and severity of run-off-roadway accidents: an empirical analysis. Accident Analysis & Prevention, Vol. 34, No. 2, 2002, pp. 149–161.

30. Liu, Jun, Asad J. Khattak, Stephen H. Richards, and Shashi Nambisan. "What are the differences in driver injury outcomes at highway-rail grade crossings? Untangling the role of pre-crash behaviors." Accident Analysis & Prevention 85 (2015): 157-169.

31. Liu, J., X. Wang, A. J. Khattak, J. Hu, J. Cui, and J. Ma. How big data serves for freight safety management at highway-rail grade crossings? A spatial approach fused with path analysis. Neurocomputing, Vol. 181, 2016, pp. 38–52.

32. Lenné, M. G., C. M. Rudin-Brown, J. Navarro, J. Edquist, M. Trotter, and N. Tomasevic. Driver behaviour at rail level crossings: Responses to flashing lights, traffic signals and stop signs in simulated rural driving. Applied Ergonomics, Vol. 42, No. 4, 2011, pp. 548–554.

33. Lenné, M. G., C. M. Rudin-Brown, J. Navarro, J. Edquist, M. Trotter, and N. Tomasevic. Driver behaviour at rail level crossings: Responses to flashing lights, traffic signals and stop signs in simulated rural driving. Applied Ergonomics, Vol. 42, No. 4, 2011, pp. 548–554.

34. McFadden, Daniel, and Kenneth Train. "Mixed MNL models for discrete response." *Journal of applied Econometrics* (2000): 447-470.

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

36. Nelson, A. (2010). Breaking down the barriers to safer crossings. IRJ-International Railway Journal, 50(5).

37. O'donnell, C., and D. Connor. Predicting the severity of motor vehicle accident injuries using models of ordered multiple choice. Accident Analysis & Prevention, Vol. 28, No. 6, 1996, pp. 739–753.

38. Raub, R. Examination of Highway-Rail Grade Crossing Collisions Nationally from 1998 to 2007. Transportation Research Record: Journal of the Transportation Research Board, Vol. 2122, 2009, pp. 63–71.

39. Russo, B., and P. T. Savolainen. Examination of Factors Affecting Frequency and Severity of Crashes at Rail-Grade Crossings. Transportation Research Board 92nd Annual Meeting, Vol. 13, 2013.

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

41. Tey, L.-S., L. Ferreira, and A. Wallace. Measuring driver responses at railway level crossings. Accident Analysis & Prevention, Vol. 43, No. 6, 2011a, pp. 2134–2141.

42. Tey, L.-S., G. Wallis, L. Ferreira, and A. T. Hojati. Australasian Transport Research Forum 2011 Proceedings. Australasian Transport Research Forum 2011 Proceedings, 2011b.

43. US Department of Transportation: Freight facts and figures 2015 http://www.rita.dot.gov/bts/sites/rita.dot.gov.bts/files/FFF_complete.pdf (accessed 11/08/2016).

44. Washington, Simon P., Matthew G. Karlaftis, and Fred Mannering. *Statistical and econometric methods for transportation data analysis*. CRC press, 2010.

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

46. Yau, K. K. Risk factors affecting the severity of single vehicle traffic accidents in Hong Kong. Accident Analysis & Prevention, Vol. 36, No. 3, 2004, pp. 333–340.

47. Yasmin, S., and N. Eluru. Evaluating alternate discrete outcome frameworks for modeling crash injury severity. Accident Analysis & Prevention, Vol. 59, 2013, pp. 506–521.

48. Yasmin, S., N. Eluru, and S. V. Ukkusuri. Alternative Ordered Response Frameworks for Examining Pedestrian Injury Severity in New York City. Journal of Transportation Safety & Security, Vol. 6, No. 4, 2014, pp. 275–300.

49. Yamamoto, T., and V. N. Shankar. Bivariate ordered-response probit model of driver's and passenger's injury severities in collisions with fixed objects. Accident Analysis & Prevention, Vol. 36, No. 5, 2004, pp. 869–876.

50. Yau, K. K. Risk factors affecting the severity of single vehicle traffic accidents in Hong Kong. Accident Analysis & Prevention, Vol. 36, No. 3, 2004, pp. 333–340

51. Zhao, S., and A. Khattak. Motor vehicle drivers' injuries in train–motor vehicle crashes. Accident Analysis & Prevention, Vol. 74, 2015, pp. 162–168.

APPENDIX A: DATA CHARACTERISTICS

Variable Type	Description and Coding	Fraquency yes=1 No=0	Mean	Standard deviation
Dependent Variables				
Injury severity	PDO	2005		
	Injured	525		
	Fatal	134		
Independent Variables				
Motor Charateristics				
Vehicle type (mph)	Truck	832	0.31	0.46
	Truck-Trailer	1832	0.69	0.46
Vehicle Speed		NA	7.79	11.54
Hazardous materials carried	Yes	711	0.27	0.44
	No	1950		
Railway Charateristics				
Train Speed (mph)		NA	30.67	18.36
Type of Train	Freight train	820	0.78	0.42
	Passenger Train, passenger train (pushing)	43	0.05	0.22
	commuter train/work train/commuter train (pushing)	25	0.02	0.16
	yard/switching, light loco, main/inspec car, special MoW equipment, EMU, DMU	135	0.15	0.35
Primary Obstruction of Track view	Yes=1	97	0.04	0.19
	No=0	2566		
Crossing surface type	Timber	744	0.28	0.45
	ASPHALT AND TIMBER	628	0.24	0.43
	Asphalt and timber	466	0.18	0.38
	Concrete	511	0.20	0.40
	Concrete & rubber	52	0.02	0.14
	Rubber	138	0.05	0.22
	MetalL	4	0.00	0.04
	Timber	41	0.02	0.13
Driver Attributes				
Driver Age (years)		NA	45.70	13.61
Driver Gender	Male	2472	0.96	0.04
	Female	96	0.19	0.19
Motorist behavior/Action of highway user	Driver went around the gates	199	0.07	0.26
	Standing RR equipment/ did not stop	1364	0.51	0.50
	Stopped on crossing	784	0.29	0.46
	Went around/ through temporary barricade	317	0.12	0.32

Variable Type	Description and Coding	Fraquency yes=1 No=0	Mean	Standard deviation
Environmental Characteristics				
Temperature	degree Fahrenheit	NA	61.74	22.87
Visibility	Day	2150	0.81	0.39
	Dark	514	0.19	0.39
Weather	Clear	2436	0.91	0.28
	Rain	157	0.06	0.24
	Snow/Sleet	71	0.03	0.16
Roadway conditions indicators	Dry	1070	0.85	0.35
	Wet/Water (standing, moving)	90	0.09	0.29
	Snow/Slush/Ice	92	0.05	0.23
Traffic Characteristics				
Location of warning	Both Sides	2542	0.96	0.19
	Single Side	99	0.04	0.19
Are there Signs or Signals	Yes=1	2601	0.98	0.15
	No=0	61		
AADT			4326	18355.509
Active controls				
Count of roadway gate arms	0-8	NA	0.41	0.49
Gates availeble (indicator)	Yes=1	1091	0.92	1.19
	No=0	1573		
Crossing warning Interconnected with Highway Signal	Connected (1)	286	0.13	0.34
	Not Connected (0)	1797		
Crossing illuminated by street Lights or Special Lights	Yes=1	569	0.23	0.42
	No=0	1878		
Whistle ban in effect	Yes=1	53	0.04	0.19
	No=0	1325		
Is track signaled	Yes=1	1430	0.55	0.50
	No=0	1183		
Highway traffic signal controling crossing	Yes=1	74	0.03	0.16
	No=0	2590		
Nearby hwy intersection have traffic signals	Yes=1	271	0.19	0.39
	No=0	1163		
Highway traffic signal interconnection	Connected	245	0.40	0.49
	Not-Connected	366		
Train detection system indicator	Yes=1	1369	0.52	0.50
	No=0	1247		
Emergency Notification system (ENS) sign displayed	Yes=1	1205	0.82	0.39
	No=0	272		
Is crossing illuminated	Yes=1	475	0.31	0.46
	No=0	1057		
No of Bells			0.82	0.91
Indicator for availability of bells	Yes=1	1375	0.52	0.50
	No=0	1289		
Mast mounted flash light indicator	Yes=1	1108	0.42	0.49
	No=0			
Passive controls				
Stop sign available	Yes=1	383	0.26	0.66
	No=0	2281		
Number of crossbuck assemblies available (number 0-9)			1.54	1.16
Pavement Marking indicator (stop line/RR xing symbols)	Yes=1	1275	0.49	0.50
	No=0	1349		
Crossbuck assemblies available indicator	Yes=1	1894	0.71	0.45
	No=0	770		

APPENDIX B: NLOGIT ESTIMATED MODELS OUTPUTS

Multinomial Logit Model

```
NLogit command:
skip$
|-> LOGIT; LHS=INJ_SEV;
RHS=ONE, HAZARD, TRNSPD, RURAL, GATESD, MOTR_A, VEHSPD, DRIVAGE, ANGLE_C,
TRUCK, TRNDTC; MARGINAL;
CROSSTAB$
```

Dependent Variables

INJ_SEV: Injury severity of driver
0 = PDO
1 = Injured
2 = Fatal

Independent Variables:

1. HAZARD: Indicator for Hazardous materials carried by one or both i-e train and truck.
2. TRNSPD: Speed of Train
3. RURAL: Functional classification of road at crossing (Rural Area)
4. GATESD: Indicator of gates availability at the crossings
5. TRUCK: Indicator of Truck involved in the crash
6. DRIVAGE: Age of driver
7. ANGLE_C: Smallest crossing angle (Angle = $60^0 – 90^0$)
8. MOTR_A: Motorist behavior (MOTR_A = Went around the gates)
9. TRNDTC: Train detection system indicator
10. VEHSPD: Speed of vehicle

```
----------------------------------------------------------------
Deleted    508 observations with missing data. N is now   2156
----------------------------------------------------------------
Normal exit:   6 iterations. Status=0, F=    1345.422

-------------------------------------------------------------------------
-----
Multinomial Logit Model
Dependent variable              INJ_SEV
Log likelihood function     -1345.42190
Restricted log likelihood   -1577.41561
Chi squared [  20 d.f.]       463.98742
Significance level               .00000
McFadden Pseudo R-squared      .1470720
```

```
Estimation based on N =   2156, K =  22
Inf.Cr.AIC  =   2734.8 AIC/N =    1.268
--------+-------------------------------------------------------------
-----
        |                  Standard            Prob.      95% Confidence
 INJ_SEV|  Coefficient       Error       z    |z|>Z*         Interval
--------+-------------------------------------------------------------
-----
        |Characteristics in numerator of Prob[INJ_SE=1]
Constant|   -3.21317***      c    -10.53  .0000    -3.81124  -2.61509
  HAZARD|     .32804***      .12427     2.64  .0083      .08448    .57159
  TRNSPD|     .03183***      .00350     9.08  .0000      .02496    .03869
   RURAL|     .33552***      .12854     2.61  .0090      .08359    .58745
  GATESD|   -1.03754***      .15556    -6.67  .0000    -1.34243   -.73265
  MOTR_A|    1.15276***      .22236     5.18  .0000      .71693   1.58858
  VEHSPD|     .01934***      .00484     4.00  .0001      .00986    .02882
 DRIVAGE|     .00157         .00409      .38  .7003     -.00644    .00959
 ANGLE_C|     .34390**       .16400     2.10  .0360      .02246    .66534
   TRUCK|     .88658***      .12072     7.34  .0000      .64997   1.12319
  TRNDTC|    -.23935*         .12902     1.86  .0636     -.01353    .49222
        |Characteristics in numerator of Prob[INJ_SE=2]
Constant|   -7.25993***      .60271   -12.05  .0000    -8.44122  -6.07864
  HAZARD|     .40507*        .21565     1.88  .0603     -.01758    .82773
  TRNSPD|     .06468***      .00637    10.15  .0000      .05219    .07717
   RURAL|     .81145***      .26517     3.06  .0022      .29172   1.33117
  GATESD|   -1.07314***      .30452    -3.52  .0004    -1.66999   -.47629
  MOTR_A|    1.50018***      .37524     4.00  .0001      .76471   2.23564
  VEHSPD|     .03737***      .00756     4.94  .0000      .02255    .05220
 DRIVAGE|     .02250***      .00695     3.24  .0012      .00888    .03612
 ANGLE_C|    -.19136         .25432     -.75  .4518     -.68983    .30711
   TRUCK|    1.48149***      .20800     7.12  .0000     1.07382   1.88916
  TRNDTC|    -.07384         .23991      .31  .7582     -.39637    .54406
--------+-------------------------------------------------------------
-----
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
----------------------------------------------------------------------
----------------------------------------------------------------------
-----
Partial derivatives of probabilities with
respect to the vector of characteristics.
They are computed at the means of the Xs.
Observations used for means are  All Obs.
A full set is given for the entire set of
outcomes, INJ_SEV   =   0 to INJ_SEV   =   2
Probabilities at the mean values of X are
  0= .772 1= .200 2= .028
--------+-------------------------------------------------------------
-----
        |     Partial                          Prob.      95% Confidence
 INJ_SEV|      Effect    Elasticity      z    |z|>Z*         Interval
--------+-------------------------------------------------------------
-----
        |Marginal effects on Prob[INJ_SE=0]
  HAZARD|    -.05932***     -.02101    -2.85  .0043     -.10006   -.01857
  TRNSPD|    -.00630***     -.25331   -10.76  .0000     -.00745   -.00515
   RURAL|    -.06924***     -.05301    -3.22  .0013     -.11134   -.02715
  GATESD|     .18311***      .09400     7.20  .0000      .13328    .23294
  MOTR_A|    -.21010***     -.01981    -5.69  .0000     -.28251   -.13769
  VEHSPD|    -.00379***     -.03953    -4.65  .0000     -.00538   -.00219
 DRIVAGE|    -.00073        -.04305    -1.06  .2871     -.00207    .00061
 ANGLE_C|    -.04888*       -.05336    -1.81  .0708     -.10191    .00415
   TRUCK|    -.16866***     -.06684    -8.35  .0000     -.20827   -.12905
  TRNDTC|    -.03849*       -.02535    -1.78  .0750     -.08086    .00388
```

```
        |Marginal effects on Prob[INJ_SE=1]
  HAZARD|     .05014**       .06876     2.56  .0104       .01179    .08848
  TRNSPD|     .00472***      .73476     8.62  .0000       .00365    .00580
   RURAL|     .04907**       .14541     2.42  .0154       .00939    .08874
  GATESD|    -.15974***     -.31745    -6.64  .0000      -.20692   -.11257
  MOTR_A|     .17577***      .06414     5.07  .0000       .10781    .24372
  VEHSPD|     .00288***      .11636     3.78  .0002       .00139    .00437
 DRIVAGE|     .00013         .02879      .20  .8452      -.00114    .00139
 ANGLE_C|     .05600**       .23663     2.17  .0303       .00534    .10666
   TRUCK|     .13335***      .20456     7.02  .0000       .09614    .17057
  TRNDTC|     .03782*        .09643     1.86  .0628      -.00202    .07766
        |Marginal effects on Prob[INJ_SE=2]
  HAZARD|     .00918         .08984     1.59  .1110      -.00211    .02046
  TRNSPD|     .00158***     1.75478     7.83  .0000       .00118    .00198
   RURAL|     .02018***      .42686     2.94  .0033       .00673    .03362
  GATESD|    -.02337***     -.33157    -2.86  .0043      -.03941   -.00733
  MOTR_A|     .03433***      .08944     3.42  .0006       .01463    .05403
  VEHSPD|     .00091***      .26176     4.17  .0000       .00048    .00133
 DRIVAGE|     .00060***      .98423     3.16  .0016       .00023    .00098
 ANGLE_C|    -.00712        -.21472    -1.05  .2943      -.02042    .00618
   TRUCK|     .03531***      .38668     5.61  .0000       .02298    .04763
  TRNDTC|    -.00067         .01222      .10  .9164      -.01187    .01321
--------+-------------------------------------------------------------------
-----
z, prob values and confidence intervals are given for the partial effect
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
----------------------------------------------------------------------------
-----
Marginal Effects Averaged Over Individuals
--------+---------+---------+---------+
Variable|INJ_SE=0 |INJ_SE=1 |INJ_SE=2 |
--------+---------+---------+---------+
ONE     |   .6367 |  -.3488 |  -.2879 |
HAZARD  |  -.0554 |   .0423 |   .0131 |
TRNSPD  |  -.0061 |   .0036 |   .0025 |
RURAL   |  -.0680 |   .0354 |   .0327 |
GATESD  |   .1694 |  -.1381 |  -.0313 |
MOTR_A  |  -.1970 |   .1472 |   .0497 |
VEHSPD  |  -.0037 |   .0022 |   .0014 |
DRIVAGE |  -.0009 |  -.0002 |   .0011 |
ANGLE_C |  -.0405 |   .0567 |  -.0162 |
TRUCK   |  -.1608 |   .1067 |   .0541 |
TRNDTC  |  -.0341 |   .0353 |  -.0012 |
--------+---------+---------+---------+
Averages of Individual Elasticities of Probabilities
--------+---------+---------+---------+
Variable|INJ_SE=0 |INJ_SE=1 |INJ_SE=2 |
--------+---------+---------+---------+
ONE     |  1.1236 | -2.0895 | -6.1363 |
HAZARD  |  -.0314 |   .0584 |   .0795 |
TRNSPD  |  -.4229 |   .5652 |  1.5852 |
RURAL   |  -.0899 |   .1085 |   .3900 |
GATESD  |   .0790 |  -.3325 |  -.3466 |
MOTR_A  |  -.0380 |   .0459 |   .0712 |
VEHSPD  |  -.0736 |   .0823 |   .2277 |
DRIVAGE |  -.0799 |  -.0080 |   .9474 |
ANGLE_C |  -.0560 |   .2339 |  -.2174 |
TRUCK   |  -.1349 |   .1365 |   .3186 |
TRNDTC  |  -.0258 |   .0960 |   .0118 |
```

Nested Logit Model

```
NLogit command:
|-> SKIP$
|-> NLOGIT; LHS=INJSEV;
    CHOICES= PDO, INJ, FATAL;
    TREE= CRASH[NOINJ(PDO), ATLEAST(INJ,FATAL)];
    IVSET:(NOINJ)=[1];
    MODEL:
    U(INJ)= C_I+TRNSPD1*TRNSPD+GATESD1*GATESD+ANGLE_C1*ANGLE_C
    +HWYNEAR1*HWYNEAR+MOTR_A1*MOTR_A+TRNDTC1*TRNDTC+VIEW1*VIEW/
    U(FATAL)= C_F+TRNSPD2*TRNSPD+VEHSPD2*VEHSPD
    +TRUCK2*TRUCK+POSI_C2*POSI_C+RURAL2*RURAL+
    SRKUSR2*SRKUSR +DRIVAGE2*DRIVAGE+ HAZARD2*HAZARD;
    PTS=200;
    MAXIT=200;
    HALTON;
    CROSSTAB$
```

Dependent Variables

INJ_SEV: Injury severity of driver

0 = PDO

1 = Injured

2 = Fatal

Independent Variables:

1. HAZARD: Indicator for Hazardous materials carried by one or both i-e train and truck.
2. TRNSPD: Speed of Train
3. RURAL: Functional classification of road at crossing (Rural Area)
4. GATESD: Indicator of gates availability at the crossings
5. TRUCK: Indicator of Truck involved in the crash
6. DRIVAGE: Age of driver
7. ANGLE_C: Smallest crossing angle (Angle = $60^0 – 90^0$)
8. MOTR_A: Motorist behavior (MOTR_A = Went around the gates)
9. TRNDTC: Train detection system indicator
10. VEHSPD: Speed of vehicle
11. HWYNEAR: Indicator for Intersecting Roadway within 500ft
12. VIEW: Indicator for Primary Obstruction of Track view
13. POSI_C: Vehicle moving over crossing
14. SRKUSR: Rail equipment struck highway user

```
+-----------------------------------------------------+
|WARNING:   Bad observations were found in the sample. |
|Found 511 bad observations among    2664 individuals. |
|You can use ;CheckData to get a list of these points. |
+-----------------------------------------------------+

Normal exit:   6 iterations. Status=0, F=    1378.580

-----------------------------------------------------------------------
-----
Discrete choice (multinomial logit) model
Dependent variable                Choice
Log likelihood function      -1378.58020
Estimation based on N =    2153, K =  17
Inf.Cr.AIC  =    2791.2 AIC/N =    1.296
Model estimated: Apr 19, 2017, 22:18:44
R2=1-LogL/LogL* Log-L fncn R-sqrd R2Adj
Constants only  -1576.4480  .1255 .1218
Chi-squared[15]          =    395.73551
Prob [ chi squared > value ] =   .00000
Response data are given as ind. choices
Number of obs.=  2664, skipped  511 obs
--------+--------------------------------------------------------------
-----
        |                  Standard            Prob.      95% Confidence
  INJSEV|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------
-----
     C_I|   -2.04120***      .19504   -10.47  .0000    -2.42346  -1.65893
 TRNSPD1|     .03137***      .00324     9.69  .0000      .02502    .03771
 GATESD1|   -1.26539***      .14754    -8.58  .0000    -1.55456   -.97621
ANGLE_C1|     .35373**       .15830     2.23  .0254      .04347    .66400
HWYNEAR1|    -.35839***      .11156    -3.21  .0013     -.57705   -.13973
 MOTR_A1|    1.29070***      .21065     6.13  .0000      .87784   1.70356
 TRNDTC1|     .19055         .12311     1.55  .1217     -.05074    .43183
   VIEW1|    -.65174*        .36720    -1.77  .0759    -1.37144    .06795
     C_F|   -8.74870***      .65873   -13.28  .0000   -10.03978  -7.45762
 TRNSPD2|     .05962***      .00632     9.44  .0000      .04724    .07200
 VEHSPD2|     .02438***      .00811     3.00  .0027      .00847    .04028
  TRUCK2|    1.11544***      .20193     5.52  .0000      .71966   1.51121
 POSI_C2|    1.47775***      .31752     4.65  .0000      .85541   2.10008
  RURAL2|     .71229***      .25122     2.84  .0046      .21992   1.20467
 SRKUSR2|     .60469*        .34125     1.77  .0764     -.06414   1.27353
DRIVAGE2|     .02269***      .00671     3.38  .0007      .00954    .03584
 HAZARD2|     .30432         .20944     1.45  .1462     -.10617    .71481
--------+--------------------------------------------------------------
-----
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.

FIML Nested Multinomial Logit Model
Dependent variable                INJSEV
Log likelihood function      -1330.35725
Restricted log likelihood    -1904.07530
Chi squared [  18 d.f.]       1147.43611
Significance level                .00000
McFadden Pseudo R-squared       .3013106
Estimation based on N =    2153, K =  18
Inf.Cr.AIC  =    2696.7 AIC/N =    1.253
Model estimated: Apr 19, 2017, 22:19:00
R2=1-LogL/LogL* Log-L fncn R-sqrd R2Adj
No coefficients -1904.0753  .3013 .2984
Constants only  -1576.4480  .1561 .1526
```

```
At start values -1378.5802  .0350 .0309
Response data are given as ind. choices
The model has 2 levels.
Nested Logit form:IVparms=Taub|l,r,Sl|r
& Fr.No normalizations imposed a priori
Number of obs.=  2664, skipped  511 obs
--------+----------------------------------------------------------
-----
        |                  Standard            Prob.      95% Confidence
  INJSEV|  Coefficient       Error       z    |z|>Z*         Interval
--------+----------------------------------------------------------
-----
        |Attributes in the Utility Functions (beta)
     C_I|    -.39718***      .06554    -6.06  .0000     -.52564   -.26871
 TRNSPD1|     .00213**       .00104     2.04  .0416      .00008    .00417
 GATESD1|    -.10860***      .03889    -2.79  .0052     -.18483   -.03237
ANGLE_C1|     .04345*        .02628     1.65  .0982     -.00805    .09495
HWYNEAR1|    -.03919*        .02030    -1.93  .0536     -.07898    .00060
 MOTR_A1|     .12342**       .04899     2.52  .0118      .02740    .21944
 TRNDTC1|     .04441**       .02130     2.09  .0370      .00267    .08615
   VIEW1|    -.10504*        .06089    -1.73  .0845     -.22439    .01430
     C_F|   -3.99089***      .48291    -8.26  .0000    -4.93737  -3.04442
 TRNSPD2|     .01599***      .00348     4.59  .0000      .00916    .02281
 VEHSPD2|     .00678**       .00312     2.17  .0297      .00067    .01290
  TRUCK2|     .61336***      .12303     4.99  .0000      .37223    .85449
 POSI_C2|     .75791***      .18783     4.04  .0001      .38978   1.12605
  RURAL2|     .29178***      .09836     2.97  .0030      .09900    .48456
 SRKUSR2|     .23114*        .11827     1.95  .0506     -.00065    .46294
DRIVAGE2|     .00556**       .00254     2.19  .0287      .00058    .01053
 HAZARD2|     .19764***      .07555     2.62  .0089      .04956    .34571
        |IV parameters, tau(b|l,r),sigma(l|r),phi(r)
  NOINJ|         1.0      .....(Fixed Parameter).....
 ATLEAST|    7.75720***     1.66681     4.65  .0000     4.49032  11.02409
--------+----------------------------------------------------------
-----
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
Fixed parameter ... is constrained to equal the value or
had a nonpositive st.error because of an earlier problem.
--------+------------------------------------------------------
NLOGIT Cross Tabulation for 3 outcome Multinomial Choice Model
XTab_Prb|           PDO           INJ          FATAL          Total
--------+------------------------------------------------------
     PDO|       1220.00       276.000        63.0000        1559.00
     INJ|       281.000       140.000        46.0000        467.000
   FATAL|       58.0000       48.0000        21.0000        127.000
   Total|       1559.00       464.000        130.000        2153.0
+--------------------------------------------------------+
| Cross tabulation of actual y(ij) vs. predicted y(ij)   |
| Row indicator is actual, column is predicted.          |
| Predicted total is N(k,j,i)=Sum(i=1,...,N) Y(k,j,i).   |
| Predicted y(ij)=1 is the j with largest probability.   |
+--------------------------------------------------------+

--------+------------------------------------------------------
NLOGIT Cross Tabulation for 3 outcome Multinomial Choice Model
XTab_Frq|           PDO           INJ          FATAL          Total
--------+------------------------------------------------------
     PDO|       1479.00       80.0000        .000000        1559.00
     INJ|       362.000       105.000        .000000        467.000
   FATAL|       65.0000       62.0000        .000000        127.000
   Total|       1906.00       247.000        .000000        2153.00
```

Random Parameter Logit Model

```
NLogit command:
CALC;RAN(12345)$
|-> SKIP$
|-> NLOGIT;
    LHS=INJSEV;
    CHOICES= PDO, INJ, FATAL;
    MODEL:
    U(FATAL)= C_F+VEHSPD*VEHSPD+TRNSPD*TRNSPD
    +POSI_C*POSI_C+RURAL*RURAL+TRUCK*TRUCK+
    SRKUSR*SRKUSR +DRIVAGE*DRIVAGE+HAZARD*HAZARD/
    U(INJ)= C_I+TRNSPD1*TRNSPD+VEHSPD1*VEHSPD
    +GATESD1*GATESD+MOTR_A1*MOTR_A+TRUCK1*TRUCK+
    TRNDTC1*TRNDTC+ANGLE_C1*ANGLE_C+HAZARD1*HAZARD+
    RURAL1*RURAL+VIEW1*VIEW+POSI_B1*POSI_B
    +HWYNEAR1*HWYNEAR+SGNLEQP1*SGNLEQP;
    RPL;
    PARAMETER;
    PTS=200;
    MAXIT=200;
    HALTON;
    FCN= POSI_B1(N);
    CROSSTAB$
```

Dependent Variables

INJ_SEV: Injury severity of driver

0 = PDO

1 = Injured

2 = Fatal

Independent Variables:

1. HAZARD: Indicator for Hazardous materials carried by one or both i-e train and truck.
2. TRNSPD: Speed of Train
3. RURAL: Functional classification of road at crossing (Rural Area)
4. GATESD: Indicator of gates availability at the crossings
5. TRUCK: Indicator of Truck involved in the crash
6. DRIVAGE: Age of driver
7. ANGLE_C: Smallest crossing angle (Angle = $60^0 – 90^0$)
8. MOTR_A: Motorist behavior (MOTR_A = Went around the gates)
9. TRNDTC: Train detection system indicator
10. VEHSPD: Speed of vehicle
11. HWYNEAR: Indicator for Intersecting Roadway within 500ft
12. VIEW: Indicator for Primary Obstruction of Track view
13. POSI_C: Vehicle moving over crossing
14. SRKUSR: Rail equipment struck highway user
15. SGNLEQP: Indicator if track is signaled

16. Posi_B: Stopped on the crossing

```
Normal exit:   6 iterations. Status=0, F=    1330.634

------------------------------------------------------------------------
-----
Start values obtained using MNL model
Dependent variable                Choice
Log likelihood function     -1330.63377
Estimation based on N =   2153, K =  23
Inf.Cr.AIC  =   2707.3 AIC/N =    1.257
Model estimated: Apr 19, 2017, 22:18:52
R2=1-LogL/LogL* Log-L fncn R-sqrd R2Adj
Constants only  -1576.4480  .1559 .1512
Chi-squared[21]          =    491.62837
Prob [ chi squared > value ] =   .00000
Response data are given as ind. choices
Number of obs.=  2664, skipped  511 obs
--------+---------------------------------------------------------------
-----
```

INJSEV	Coefficient	Standard Error	z	Prob. |z|>Z*	95% Confidence Interval	
POSI_B1	-.53407***	.16675	-3.20	.0014	-.86090	-.20724
C_F	-9.06479***	.66292	-13.67	.0000	-10.36410	-7.76549
VEHSPD	.02942***	.00843	3.49	.0005	.01290	.04594
TRNSPD	.06206***	.00640	9.70	.0000	.04952	.07461
POSI_C	1.51925***	.31919	4.76	.0000	.89365	2.14486
RURAL	.79961***	.25471	3.14	.0017	.30038	1.29883
TRUCK	1.45335***	.20845	6.97	.0000	1.04479	1.86191
SRKUSR	.58519*	.33861	1.73	.0839	-.07847	1.24885
DRIVAGE	.02212***	.00668	3.31	.0009	.00902	.03521
HAZARD	.44507**	.21580	2.06	.0392	.02211	.86803
C_I	-2.82576***	.23757	-11.89	.0000	-3.29139	-2.36012
TRNSPD1	.03419***	.00374	9.13	.0000	.02685	.04153
VEHSPD1	.01158**	.00511	2.26	.0236	.00156	.02160
GATESD1	-.81318***	.15934	-5.10	.0000	-1.12548	-.50088
MOTR_A1	.83278***	.22253	3.74	.0002	.39663	1.26892
TRUCK1	.90041***	.12139	7.42	.0000	.66249	1.13833
TRNDTC1	-.27423**	.12718	2.16	.0311	.02497	.52349
ANGLE_C1	.38787**	.16037	2.42	.0156	.07354	.70219
HAZARD1	.35542***	.12510	2.84	.0045	.11023	.60061
RURAL1	.30521**	.12984	2.35	.0187	.05073	.55969
VIEW1	-.73257**	.37352	-1.96	.0499	-1.46466	-.00048
HWYNEAR1	-.25280**	.11521	-2.19	.0282	-.47861	-.02699
SGNLEQP1	-.26146**	.12454	-2.10	.0358	-.50555	-.01737

```
--------+---------------------------------------------------------------
-----
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
------------------------------------------------------------------------
-----

Normal exit:  34 iterations. Status=0, F=    1329.921

------------------------------------------------------------------------
-----
Random Parameters Logit Model
Dependent variable                INJSEV
Log likelihood function     -1329.92063
Restricted log likelihood   -2365.31226
```

```
Chi squared [  24 d.f.]      2070.78326
Significance level               .00000
McFadden Pseudo R-squared      .4377399
Estimation based on N =   2153, K =  24
Inf.Cr.AIC  =   2707.8 AIC/N =    1.258
Model estimated: Apr 19, 2017, 22:22:58
R2=1-LogL/LogL* Log-L fncn R-sqrd R2Adj
No coefficients -2365.3123  .4377 .4346
Constants only  -1576.4480  .1564 .1517
At start values -1330.6338  .0005-.0051
Response data are given as ind. choices
Replications for simulated probs. = 200
Used Halton sequences in simulations.
Number of obs.=  2664, skipped  511 obs

--------+------------------------------------------------------------
-----
        |                  Standard            Prob.      95% Confidence
  INJSEV|  Coefficient       Error       z    |z|>Z*         Interval
--------+------------------------------------------------------------
-----
        |Random parameters in utility functions
 POSI_B1|   -1.39064*        .84437    -1.65  .0996    -3.04557    .26429
        |Nonrandom parameters in utility functions
     C_F|   -9.11763***      .66607   -13.69  .0000   -10.42310  -7.81216
  VEHSPD|     .02959***      .00843     3.51  .0005      .01306    .04612
  TRNSPD|     .06253***      .00642     9.74  .0000      .04994    .07511
  POSI_C|    1.56492***      .32220     4.86  .0000      .93341   2.19643
   RURAL|     .79671***      .25500     3.12  .0018      .29692   1.29650
   TRUCK|    1.46426***      .20902     7.01  .0000     1.05458   1.87394
  SRKUSR|     .58321*        .33823     1.72  .0847     -.07972   1.24613
 DRIVAGE|     .02202***      .00668     3.30  .0010      .00892    .03511
  HAZARD|     .45507**       .21652     2.10  .0356      .03069    .87945
     C_I|   -2.89959***      .25239   -11.49  .0000    -3.39427  -2.40491
 TRNSPD1|     .03656***      .00424     8.61  .0000      .02824    .04488
 VEHSPD1|     .01160**       .00516     2.25  .0246      .00148    .02172
 GATESD1|    -.83641***      .17072    -4.90  .0000    -1.17102   -.50180
 MOTR_A1|     .83721***      .22786     3.67  .0002      .39060   1.28381
  TRUCK1|     .93508***      .12910     7.24  .0000      .68205   1.18811
 TRNDTC1|    -.30590**       .13490     2.27  .0233      .04151    .57029
ANGLE_C1|     .38720**       .16674     2.32  .0202      .06040    .71399
 HAZARD1|     .36489***      .13206     2.76  .0057      .10605    .62373
  RURAL1|     .30245**       .13649     2.22  .0267      .03494    .56997
   VIEW1|    -.72544*        .37753    -1.92  .0547    -1.46539    .01450
HWYNEAR1|    -.27709**       .12080    -2.29  .0218     -.51385   -.04032
SGNLEQP1|    -.29039**       .13186    -2.20  .0276     -.54883   -.03195
        |Distns. of RPs. Std.Devs or limits of triangular
NsPOSI_B|    1.79577*       1.00448     1.79  .0738     -.17297   3.76451
--------+------------------------------------------------------------
-----
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
---------------------------------------------------------------------
-----

+-----------------------------------------------------+
| Cross tabulation of actual choice vs. predicted P(j)  |
| Row indicator is actual, column is predicted.         |
| Predicted total is F(k,j,i)=Sum(i=1,...,N) P(k,j,i).  |
| Column totals may be subject to rounding error.       |
+-----------------------------------------------------+

--------+---------------------------------------------------
NLOGIT Cross Tabulation for 3 outcome Multinomial Choice Model
```

```
XTab_Prb|           PDO           INJ         FATAL         Total
--------+----------------------------------------------------------
     PDO|       1214.00       284.000       61.0000       1559.00
     INJ|       284.000       140.000       43.0000       467.000
   FATAL|       61.0000       43.0000       23.0000       127.000
   Total|       1559.00       467.000       127.000       2153.00


--------+----------------------------------------------------------
NLOGIT Cross Tabulation for 3 outcome Multinomial Choice Model
XTab_Frq|           PDO           INJ         FATAL         Total
--------+----------------------------------------------------------
     PDO|       1488.00       65.0000       6.00000       1559.00
     INJ|       361.000       97.0000       9.00000       467.000
   FATAL|       78.0000       37.0000       12.0000       127.000
   Total|        1927.00          199.000          27.0000          2153.00
27.0000       2153.00
```

END OF DOCUMENT