Analyzing Severity of Vehicle Crashes at Highway-Rail Grade Crossings: Multinomial Logit Modeling

by Wei (David) Fan, Martin R. Kane, and Elias Haile

The purpose of this paper is to develop a nominal response multinomial logit model (MNLM) to identify factors that are important in making an injury severity difference and to explore the impact of such explanatory variables on three different severity levels of vehicle-related crashes at highway-rail grade crossings (HRGCs) in the United States. Vehicle-rail and pedestrian-rail crash data on USDOT highway-rail crossing inventory and public crossing sites from 2005 to 2012 are used in this study. A multinomial logit model is developed using SAS PROC LOGISTICS procedure and marginal effects are also calculated. The MNLM results indicate that when rail equipment with high speed struck a vehicle, the chance of a fatality resulting increased. The study also reveals that vehicle pick-up trucks, concrete, and rubber surfaces were more likely to be involved in more severe crashes. On the other hand, truck-trailer vehicles in snow and foggy weather conditions, development area types (residential, commercial, industrial, and institutional), and higher daily traffic volumes were more likely to be involved in less severe crashes. Educating and equipping drivers with good driving habits and short-term law enforcement actions, can potentially minimize the chance of severe vehicle crashes at HRGCs.

INTRODUCTION

Fatalities resulting from motor vehicle crashes is the fifth leading cause of death in the United States. Data from the National Highway Traffic Safety Administration indicate that since 1949 more than 30,000 (40,000 average) fatalities result from motor vehicle crashes every year. However, the current trend shows this number is declining. For example, a 1.9% decrease in crash-related fatalities was observed in 2011 as compared with 2010. Nonetheless, crash-related injuries are still large in number. In 2011, an estimated 2.22 million people were injured in motor vehicle traffic crashes and 2.24 million in 2010 (NHTSA 2012). Fatal crashes on highway-rail grade crossings (HRGCs) contributed to 261 deaths in 2010 and 251 in 2011 (FRA 2012).

HRGCs are conflict points between highway users and rail equipment (e.g., locomotive, freight car, caboose, or service equipment car operated by a railway company), which have contributed to a considerable amount of crashes in U.S. history. Though the trend of highway user crashes with rail equipment is showing a decrease in numbers, much has yet to be done to improve the safety of HRGCs. Unlike highway traffic crashes, a significantly high percentage of vehicle-rail crashes lead to fatality and injury to vehicle users. For example, data in the past eight years (2005-2012) indicate that 8.55% of vehicle-rail crashes were fatal and 26.68% resulted in injury (FRA 2012). However, in the case of highway traffic crashes, the percentage of fatal crashes is no more than 2% (NHTSA 2012).

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

One of the most important tasks in improving road safety is to uncover influential factors and then to develop countermeasures. The relationship between the injury severity of traffic crashes and factors such as driver and passenger characteristics, pedestrian age and gender, vehicle type, environmental conditions, traffic, and geometric conditions has attracted much attention. A better understanding of this relationship is necessary and very important for improving facility design so that crashes can be reduced. It is important to note that reducing crash frequency and reducing crash-injury severity may necessitate different strategic approaches. The development of effective countermeasures requires a thorough understanding of the factors that affect the likelihood of a crash occurring or, given that a crash has occurred, the characteristics that may mitigate or exacerbate the degree of injury sustained by crash-involved road users. To gain such an understanding, safety researchers have applied a wide variety of methodological techniques over the years.

Numerous studies have applied statistical models for crash injury severity studies. Among them, the ordered probit, ordered logit, and their variations are the most often used models. Savolainen et al. (2011) briefly discussed and summarized a wide range of methodological tools applied to study the impact of various factors on motor vehicle crash injury severities. As presented in the paper, ordered logit and probit, multinomial logit, binary logit and binary probit, and nested logit are some of the frequently used statistical methodologies. In particular, logistic regression has been widely applied to model crash severity levels. Variables such as elements of geometric design, traffic operational measures, and environmental conditions are considered as independent variables to estimate the severity. Savolainen et al. (2011) also applied the logistic regression modeling approach (specifically an unordered logit model) to estimate the three levels of highway user crash severity on HRGC as a function of various factors involved. The explanatory variables were obtained from the USDOT crossing inventory and crash data.

This purpose of this study is to analyze the severity of vehicle crashes at USDOT public HRGCs from 2005 to 2012, and to investigate the impact of various factors involved in the crashes. The remainder of this paper is organized as follows. The second section presents a literature review on existing studies regarding vehicle crash severity modeling. The third describes the MNL modeling methodology. The fourth section discusses the data assembly and analysis of the research. Section five presents numerical results and discussion. The sixth section discusses the conclusions and recommendations are also made.

LITERATURE REVIEW

Several studies have been conducted to model crash severity and investigate the impacts of various factors involved in the crashes. Mercier et al. (1999) conducted a study (using data from the Iowa Department of Transportation for 1986 to 1993) and tested the hypothesis that older drivers and passengers would suffer more severe injuries than younger adults in the presence of broadside and angle collisions of automobiles on rural highways. Logistic modeling, Hierarchical Regression Analysis, and Principal Components Regression, were analysis tools applied. Injury severity levels, fatal, major, and minor, were considered as dependent categorical variables. Some of the independent variables considered were occupant age, occupant position relative to point of impact, and protection. According to the study, age-related variables were generally more significant predictors of injury severity for females than for males. It was also identified that use of lap and shoulder restraints reduces injury severity and is less certain for females. For females only, air bags deployed were reported as significant injury severity predictors.

By using sequential binary logistic regression, Dissanayake and Lu (2002) modeled crash severity for single-vehicle fixed object crashes involving young drivers using data from the Florida Traffic Crash Database for the two-year period (1997 and 1998). The five crash severity categories considered were no injury, possible injury, non-capacitating injury, incapacitating injury, and fatal. As reported in the study, factors such as alcohol or drug influence, ejection in the crash, point of impact, rural crash locations, existence of curve or grade at the crash location, and speed of vehicle significantly increased the probability of more severe crashes. On the other hand, restraint device usage and drivers being male were reported to reduce the chance of high severity crashes. It was also indicated that factors such as weather condition, residence location, and physical condition have no significant relation in the model.

Duncan et al. (1998) conducted a study to investigate car occupant injury severity in two-vehicle passenger car-truck rear-end collisions by using an ordered probit model. The 1993-95 Highway Safety Information System (HSIS) data for collisions between heavy trucks and passenger cars in North Carolina were used for analysis. As reported in the study, factors such as darkness, high speed differentials, high speed limits, grades, being in a car struck to the rear, driving while drunk, and being female increased the passenger vehicle occupant injury severity. On the other hand, factors such as snowy or icy roads, being in a child restraint, and congested roads decreased the severity level. It was also indicated that interaction effects of cars being struck to the rear with high speed differentials and car rollovers were significant and increased the injury severity.

Donnell and Mason (2004) conducted a study and developed median-related crash severity models using data collected from Pennsylvania Department of Transportation between 1994 and 1998. Three crash severity classes, fatal, injury, and property damage only (PDO), were considered as independent variable outcomes. Both ordinal and nominal response logistic regression models were developed in the study. As indicated in the report, the ordinal response model gave better results for cross-median crashes. On the other hand, the nominal response model gave better results for median-barrier crashes. Furthermore, variables such as highway surface conditions, use of drugs or alcohol, presence of an interchange entrance ramp, horizontal alignment, crash type, and average daily traffic volume were reported to have some significant positive or negative effects on crash severity.

By using paired comparison analysis and ordered probit model, Renski at al. (1999) conducted a study to test the hypothesis that a speed limit increase will result in an increase in driving speed and produce higher crash severity. The study was focused on single-vehicle crashes on interstate roadways in North Carolina. As reported in the study, increasing speed limits from 89 to 97 km/hour (55 to 60 mph) and from 89 km/hour (55 mph) to 105 km/hour (65 mph) increased the probability of sustaining minor and non-capacitating injuries. However, the study indicated that increasing speed limits from 105 km/hour (65 mph) to 113 km/hour (70mph) did not show a significant effect on crash severity.

Huang et al. (2002) investigated the effects of road diets in which four-lane undivided roads were converted into three lanes. A road diet, also called a lane reduction or road rechannelization, is a technique in transportation planning whereby a road is reduced in number of travel lanes and/or effective width in order to achieve systemic improvement. Twelve road diets and 25 comparison sites in California and Washington cities were analyzed in the study. A before and after analysis was conducted and it was reported that road-diet crashes that occurred during the "after" period were observed to be about 6% lower than that of the comparison sites.

Khattak (2001) conducted a study that investigated the effect of vehicle technologies on crash injury severity. The North Carolina 1994-1995 HSIS crash data were used for the analysis. Three separate ordered probit models were developed for the three drivers, Driver 1 (leading), Driver 2 (striking), and Driver 3 (striking in a three-vehicle crash). As indicated in the study, in a two-vehicle rear-end collision the leading driver is more likely to be injured, whereas in a three-vehicle collision the driver in the middle is more likely to be injured. It was also stated that being in a newer vehicle protects the driver in rear-end collisions. Moreover, the study showed the benefit of technological improvements on driver safety.

Mercier at al. (1997) performed a study and tested the hypothesis that older drivers and passengers would suffer more severe injuries than younger adults in the presence of head-on collisions of automobiles on rural highways. Study data were drawn from the Iowa Department of Transportation's accident files from 1986 through part of 1993. Logistic modeling, Hierarchical Regression Analysis, and Principal Components Regression were applied. Injury severity levels, fatal, major, and minor, were considered as dependent categorical variables (which take on one of a limited number of possible values). The independent variables considered included, among others, occupant age, occupant position relative to point of impact, and level of protection. As stated in the study, age was an important factor in predicting injury severity for both men and women. The study concluded that older drivers and passengers experienced more severe injury than any of the other age groups. Use of lap and shoulder devices was reported to be more important for men than women while the reverse was true for deployed air bags.

Chira-Chavala et al. (1996) investigated the characteristics and probable causes of light rail transit system crashes and developed a crash severity model for the Santa Clara County Transit Agency. A binary logit model was applied to predict the probability of injury accident as a function of explanatory variables such as speeds before collision of light rail vehicles and motor vehicles and movement of the motor vehicle before collision. As reported in the study, left-turn vehicle movements, higher speeds of the motor vehicle or the LRV, and accidents occurring during peak hours increased the probability of injury crashes.

Chen and Jovanis (2000) developed and tested the variable-selection procedure that avoids problems occurring due to the presence of a large number of potential factors, the complex nature of crash causes and outcomes, and a large number of categories in crash-severity modeling. Bus-involved crash data for Freeway 1 in Taiwan from 1985 through 1993 were used. The procedure consisted of the chi-squared automated interaction detection (CHAID) method to collapse categories. Person chi-square test was used to assess the relationship between dependent and independent variables and log-linear modeling techniques. As indicated in the study, the log-linear model showed that late-night or early-morning driving increased the risk of severe injury crashes for bus drivers. It was also stated that bus crashes involving a large truck or tractor-trailers increased the risk of severe injury crashes.

By using an ordered probit model, Khattak at al. (2002) explored factors contributing to more severe older driver (age 65 and above) crash injury severity by analyzing 1990-1999 crash data from Iowa. According to the study, older male drivers are more prone to injury as compared with older female drives. It was stated that older drivers under the influence of alcohol experienced more severe injuries. It was also indicated that older driver injuries involving farm vehicles are more severe as compared with other vehicle types.

Xie et al. (2009) conducted a study that demonstrated application of a Bayesian ordered probit model in drivers' injury severity analysis. In the Bayesian probit model, prior distributions such as means and variances were included, reflecting the analysts' prior knowledge about the data. Comparisons were made between Bayesian ordered probit and conventional ordered probit models. As reported in the study, for large data size, model fitting results obtained from the Bayesian and the conventional probit model have no significant differences. It was also reported that for small sample size, a Bayesian probit model produced parameter estimates with better prediction performance than the conventional ordered probit model.

Some recent research efforts were also made to the joint estimation of two dependent variables that were closely related to each other in order to improve the efficiency of uncovering the influential factors. For example, Ye et al. (2009) developed a simultaneous equations model of crash frequency by collision type for rural intersections. Ye et al. (2013) developed and presented a similar multivariate Poisson regression to model the crash frequency by severity level for freeway sections in this paper. Along this same line, a generalized Poisson model was developed to assess the effects of demographic factors, driving habits, and medicinal use on elderly driver automobile crashes (Famoye et al. 2004). Likewise, several multivariate Poisson-lognormal regression models were presented for jointly modeling crash frequency by severity and applied to a case study in California

(Park and Lord 2007). The results showed promise toward the goal of obtaining more accurate estimates by accounting for correlations and over-dispersion (Park and Lord 2007).

Gitelman and Hakkert (1997) developed a method to evaluate road-rail crossing safety with limited accident statistics when the need for grade separation was discussed using available Israeli accident data. Austin and Carson (2002) developed an alternate highway-rail crossing accident prediction model using negative binomial regression, which showed great promise. Saccomanno et al. (2004) developed risk-based models for identifying high-rail grade crossing blackspots (i.e., crossings with unacceptable risks of involving high expected collision frequencies or consequences or both). Miranda-Moreno et al. (2005) compared the relative performance of three alternative models for ranking locations for safety improvement, which included the traditional negative binomial model, the heterogeneous negative binomial model, and the Poisson lognormal model. These models were calibrated using a sample of Canadian highway-railway intersections with an accident history of five years. It was concluded that the choice of model assumptions and ranking criteria can lead to considerably different lists of hazardous locations. Saccomanno et al. (2007) conducted a research study of estimating countermeasure effects for reducing collisions at highway-railway grade crossings. Park and Saccomanno (2007) developed a propensity score method to reduce treatment selection bias for estimating treatment effects. The model was also applied to Canadian highway-railway grade crossings data to estimate reductions in collisions due to upgrades in warning devices. It was shown that the propensity score method could be used to reduce treatment selection bias. Hu et al. (2010) investigated key factors and developed a generalized logit model to estimate the accident severity at railroad grade crossings in Taiwan.

As discussed, crashes occurring at HRGCs have a significant effect on highway user safety, and the importance of conducting research in such areas is evident. However, compared with the amount of work on general highway traffic crashes, this subject receives relatively less attention, although some research efforts have been made in this particular area. As such, the objective of this research is to explore the impacts of various factors contributing to different levels of crash severity to vehicle users as a result of vehicle-rail crashes on HRGCs. A nominal response multinomial logit model with three levels of severity was used to model the impact of various factors that include vehicle driver characteristics, environmental factors, railroad crossing characteristics, highway characteristics, land use type, and more. The three levels of responses considered were fatality, injury, and no injury. The SAS PROC LOGISTICS procedure was used to develop the model.

METHODOLOGY

The MNLM formulation is well discussed by Long (1997). If y is the response variable with J nominal (i.e., categorical) outcomes (which takes on one of a limited number of possible values), then the assumption of the multinomial logit model is that category 1 through J are not ordered (i.e., not arranged in an increasing or decreasing order). Also, let $\Pr(y=m|x)$ be the probability of observing outcome m given the independent variable x. The model for y is constructed as follows:

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

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

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

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

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

The parameter estimates are determined using maximum likelihood estimation. If the observations are independent, the likelihood eq. (4) is given by:

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

Where P_i is the probability of observing whether values of y was actually observed for the i^{th} observation. Combining the eq. (1) with this eq. (4) in place of P_i the likelihood eq. (5) can be written as:

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

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

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

(6) $\rho^2 = 1 - \dfrac{\mathrm{LL}(\beta)}{\mathrm{LL}(0)}$

Where LL (β) represents the log-likelihood at model convergence, LL(0) represents the log-likelihood of a naïve model (without information). The ρ^2 goes from 0 (for no improvement in the log-likelihood) to 1 for a perfect fit. A value for ρ^2 larger than 0.1 indicates meaningful improvement (Long 1997).

The marginal effect or partial change can be determined by taking derivative of Eq. (1) with respect to x_k as described in the following eq. (7).

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

Marginal effect is the slope of the curve relating x_k to Pr(y=m|x), holding other variables constant. Variables are held at their means, possibly with dummy variables at 0 or 1.Though the computation of the change in the probability is important to interpret the effects of the MNLM, it has limitations. Firstly, the discrete change indicates the change for a particular set of values of the independent variables, which means at different levels of these variables, the changes will be different. And the second limitation is that it measures the discrete change, which does not indicate the changes among the dependent outcome due to infinitely small changes in independent variables (Long 1997).

An odds ratio can also be used in the interpretation of the developed model. The odds ratio is defined as the ratio of the odds of those with the risk factor to the odds for those without the risk factor. Generally, the odds ratio associated with a one-unit increase in the risk factor can be computed by the exponential function of the regression coefficient of that risk factor (SAS 2008).

DATA ASSEMBLY AND ANALYSIS

Vehicle-rail crash data on the USDOT public crossing sites from 2005 to 2012 are used in this study. In order to acquire more explanatory variables, the USDOT highway-rail crossing inventory was also included. The crash data and the crossing inventory data were merged based on the USDOT identification number. The SAS PROC SQL was used to merge and clean the data (i.e., removing data errors). After the data merging and cleaning process, a total of 7,414 records were obtained and used in the modeling stage. The data used to create the data set were obtained from the Federal Railroad Administration (2012).

Table 1 presents the descriptive statistics of some of the variables from such HRGC crash and inventory data. As shown, the distribution of vehicle-rail crash severity is 6.80%, 26.63% and 66.58% for fatal, injury, and no injury, respectively. This distribution of crash severity indicates 33.43% of vehicle crashes at HRGC sites lead to fatality or injury. The majority (78.64%) of vehicle-rail crashes at HRGC sites occurred when the rail equipment struck the vehicle while the remaining (21.36%) were when the vehicle struck the rail equipment. It is shown in the table that a majority (53.09%) of vehicles involved in the vehicle-rail crashes are cars, and the majority (71.01%) of vehicle crashes occurred in clear weather conditions.

The HRGC sites where crashes occur are located in different development areas. As one can see from Table 1, 32.37% of the crossings are located in open space areas, 21.51% in residential areas, and 28.10% in commercial areas. The rest are found in industrial and institutional development areas. The majority (74.99%) of the HRGCs cross two-lane highways. Descriptive statistics of other variables are also shown in the table. All variables that are available in the database are considered in this study. Some of the continuous variables are converted into categorical variables and the MNLM is applied to estimate the model parameters.

RESULTS AND DISCUSSION

Many variables obtained from the crossing inventory and crash data were used in developing the nominal response MNLM. During the final preferred model development process, some of the variables were found to be statistically insignificant and hence removed in a stepwise manner. PROC LOGISTIC procedure was applied with significance level being 0.1 to retain some of the variables.

Tables 2 and 3 present the results obtained from this study. In this modeling, three vehicle-rail crash severity levels (fatal crashes, injury crashes, and no injury crashes) were considered as the dependent variable. Among the three crash severity levels, no injury crashes were considered the base case. Therefore, coefficients estimated for the explanatory variables are values representing the relative effect of contributing factors on fatal or injury crashes compared with no injury crashes. Positive estimates in the model indicate that the chance of injury or fatal crash increases as the value of the independent variables increases, while negative estimates indicate that the chance of injury or fatal crash decreases as the value of the independent variables increases.

As shown in Table 2, some of the variables are not statistically significant. However, since the main interest of this paper is to examine how the chance of injury and fatal (both) crash increases or decreases (separately or simultaneously) corresponding to a one-unit change in the value of the independent variables, for the sake of facilitating interpretation of the results, those variables were still retained in the model if at least one of variables/factors in the same parameter category was significant in at least one of the models (injury and/or fatality). This actually induces reduction in efficiency of the model. Furthermore, a 90% confidence level was considered instead of 95% (Tay et al. 2011).

Table 1: Descriptive Statistics of Variables from HRGC Crash and Inventory Data

Variable	Category	Frequency	Percent
Crash Characteristics			
INJURY (crash severity level)	3=Fatal crashes	504	6.80
	2=Injury crashes	1974	26.63
	1=No Injury crashes	4936	66.58
TYPACC (Type of accident)	1=Train struck vehicle	5830	78.64
	2=Vehicle struck train	1584	21.36
Vehicle Characteristics			
TYPVEH (Type of vehicle)	1=Auto	3936	53.09
	2=Truck	542	7.31
	3=Truck trailer	1298	17.51
	4=Pickup truck	1317	17.76
	5=Van	306	4.13
	6=Bus	10	0.13
	7=School Bus	5	0.07
VEHSPD (Vehicle speed)	1=<40km/hour (<25mph)	6312	85.14
	2=40-72km/hour (25-45mph)	830	11.20
	3=>72km/hour (>45mph)	272	3.67
(AADT) (Average annual daily traffic)	1=<10,000	6525	88.01
	2=10,000-20,000	602	8.12
	3=20,000-30,000	177	2.39
	4=>30,000	110	1.48
Train Characteristics			
TRNSPD (Train speed)	1=<40km/hour (<25mph)	2999	40.45
	2=40-72km/hour (25-45mph)	2549	34.38
	3=>72km/hour (>45mph)	1866	25.17
Vehicle Driver Characteristics			
DRVAGE	1=<25 Years	1186	16.00
	2=25-60 Years	3978	53.66
	3=>60 Years	1029	13.88
	Missing	1221	16.47
DRIVGEN (Vehicle driver gender)	1=Male	5645	76.14
	2=Female	1769	23.86
Highway Characteristics			
HWYPVED (Highway surface type)	1=Paved	6042	81.49
	2=Unpaved	1372	18.51
HWYSGNL (Highway signal)	1=Not present	7215	97.32
	2=Present	199	2.68

Table 1 (continued)

Variable	Category	Frequency	Percent
TRAFICLN (No. of traffic lane)	1=1 Lane	644	8.69
	2=2 Lanes	5560	74.99
	3=3 Lanes	87	1.17
	4=4 Lanes	872	11.76
	5=≥5 Lanes	251	3.39
Environmental Characteristics			
DEVELTYP (Development area type)	1=Open space	2400	32.37
	2=Residential	1595	21.51
	3=Commercial	2083	28.1
	4=Industrial	1226	16.54
	5=Institutional	110	1.48
WEATHER (Weather condition)	1=Clear	5265	71.01
	2=Cloudy	1406	18.96
	3=Rain	445	6
	4=Fog	107	1.44
	5=Sleet	15	0.2
	6=Snow	176	2.37
TEMP (Temperature)	1=<10°C (50°F)	2074	27.97
	2=10°-27°C (50°-80°F)	3624	48.88
	3=>27°C (80°F)	1716	23.15
NEAREST (Intersecting IN or Near city)	1=In city	4244	57.24
	2=Near city	3170	42.76
Crossing Characteristics			
XSURFACE (Crossing surface type)	1=Timber	2049	27.64
	2=Asphalt	3015	40.67
	3=Asphalt & Flange	445	6
	4=Concrete	920	12.41
	5=Concrete & Rubber	266	3.59
	6=Rubber	413	5.57
	7=Metal	3	0.04
	8=Unconsolidated	256	3.45
	9=Other	47	0.63
XBUCK (Cross bucks)	1=Not Present	2348	31.67
	2=Present	5066	68.33
FLASH (Flashlight)	1=Not present	3475	46.87
	2=Present	3939	53.13
GATES (Gates)	1=Not Present	6371	85.93
	2=Present	1043	14.07

Table 2: Multinomial Logistic Model Regression Results

Parameter	Injury		Fatal	
	Estimate	P-value	Estimate	P-value
Intercept	-1.1553*	<.0001	-4.4843*	<.0001
VEHSPD (Ref:<40km/hour (<25mph))				
40-72km/hour (25-45mph)	0.6457*	<.0001	0.7110*	<.0001
>72km/hour (>45mph)	0.9211*	<.0001	1.6351*	<.0001
TYPVEH (Ref: Auto)				
Truck	0.0581	0.6299	0.0846	0.6604
Truck-trailer	-0.1967*	0.0316	-1.5297*	<.0001
Pick-up truck	0.1480*	0.0766	0.0385	0.7808
Van	0.0756	0.6144	-0.2670	0.3401
Bus	0.7259	0.4470	-9.9575	0.9790
School bus	1.0507	0.2969	-10.0643	0.9820
TYPACC (Ref: vehicle struck rail equipment)				
Rail equipment struck vehicle	-0.1107	0.1476	0.6935*	<.0001
TEMP(Ref: <10^{o}C (50^{o}F))				
10^o-27^{o}C (50^o-80^{o}F)	0.1029	0.1654	0.0671	0.6081
>27^{o}C (80^{o}F)	0.2520*	0.0034	0.1148	0.4494
WEATHER (Ref: Clear)				
Cloudy	-0.0399	0.6056	-0.0438	0.7463
Rain	-0.1611	0.2240	-0.4313	0.1130
Fog	0.0295	0.9021	-1.2110	0.1003
Sleet	0.4891	0.4086	-10.7328	0.9568
Snow	-0.6097*	0.0087	-0.6858	0.1285
TRNSPD (Ref: <40km/hour (<25mph))				
40-72km/hour (25-45mph)	0.6274*	<.0001	1.7280*	<.0001
>72km/hour (>45mph)	0.6433*	<.0001	2.7725*	<.0001
DRIVGEN (Ref: Female)				
Male	0.3848*	<.0001	0.2965*	0.0176
DEVELTYP(Ref: Open space area)				
Residential	-0.1907*	0.0231	-0.1882	0.1913
Commercial	-0.3342*	<.0001	-0.3510*	0.0171
Industrial	-0.4128*	<.0001	-0.1197	0.5122
Institutional	-0.4649*	0.0666	-0.5219	0.2897

Table 2 (continued)

Parameter	Injury		Fatal	
	Estimate	**P-value**	**Estimate**	**P-value**
XSURFACE(Ref: Timber)				
Asphalt	-0.2094*	0.0043	-0.4813*	0.0002
Asphalt & Flange	-0.1327	0.3229	-0.6683*	0.0143
Concrete	0.0793	0.4405	0.0422	0.8002
Concrete & Rubber	0.0897	0.6240	0.5610*	0.0428
Rubber	0.0745	0.6092	-0.3451	0.2467
Metal	-0.4770	0.7027	-10.1543	0.9825
Unconsolidated	-0.3027*	0.0669	-0.1017	0.6871
Other	-0.2763	0.4752	-0.3334	0.6653
AADT(Ref:<10,000)				
10,000-20,000	-0.0882	0.4556	-0.4342*	0.0698
20,000-30,000	-0.5348*	0.0184	-0.8054*	0.0755
>30,000	-0.2838	0.2595	-0.9880*	0.0788
DRIVAGE(Ref:<25 Years)				
25-60 Years	0.0727	0.3548	0.2983*	0.0452
>60 Years	0.2706*	0.0069	1.2399*	<.0001
Number of observation= 7,414, ρ^2=0.011, $\chi2$ for likelihood ratio =943.787, P-value for chi square= 0.000				

Based on the parameter estimates obtained in Table 2, the MNL models can be written as follows. Note that the information about driver under the influence or not is unavailable in the dataset and thus not included in the analysis.

(8) $$\log\left[\frac{P(Y = Fatal)}{P(Y = No\ Injury)}\right] = -4.4843 + 0.7110X_1 + 1.6351X_2 - 1.5297X_3 + 0.0385X_4 + 0.6935X_5 + 0.114 - 0.6858X_7 - 1.2110X_8 + 1.7280X_9 + 2.7725X_{10} + 0.2965X_{11} - 0.1882X_{12} - 0.3510X_{13} - 0.1197X_{14} - 0.5219X_{15} - 0.4813X_{16} - 0.1017X_{17} - 0.6683X_{18} + 0.5610X_{19} - 0.4342X_{20} - 0.8054X_{21} - 0.9880X_{22} + 0.2983X_{23} + 1.2399X_{24}$$

(9) $$\log\left[\frac{P(Y = Injury)}{P(Y = No\ Injury)}\right] = -1.1553 + 0.6457X_1 + 0.9211X_2 - 0.1967X_3 + 0.1480X_4 - 0.1107X_5 + 0.25 - 0.6097X_7 + 0.0295X_8 + 0.6274X_9 + 0.6433X_{10} + 0.3848X_{11} - 0.1907X_{12} - 0.4128X_{14} - 0.4649X_{15} - 0.2094X_{16} - 0.3027X_{17} - 0.1327X_{18} + 0.0897X_{19} - 0.5348X_{21} - 0.2838X_{22} + 0.0727X_{23} + 0.2706X_{24}$$

Where:

X_1 = Vehicle speed category (1 if vehicle speed is 40-72 km/hour (25-45 mph), 0 otherwise)

X_2 = Vehicle speed category (1 if vehicle speed is >72 km/hour (>45 mph), 0 otherwise)

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

X_4 = Vehicle type indicator (1 if vehicle is pick-up truck, 0 otherwise)

X_5 = Accident type indicator (1 if rail equipment struck vehicle, 0 otherwise)

X_6= Temperature indicator (1 if temperature is greater than 27°C (80°F), 0 otherwise)

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

X_8= Weather indicator (1 if foggy weather, 0 otherwise)

X_9 = Train speed category (1 if train speed is 40-72 km/hour (25-45 mph), 0 otherwise)

X_{10} =Train speed category (1 if train speed is >72 km/hour (>45 mph), 0 otherwise)

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

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

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

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

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

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

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

X_{18}= HRGC surface type (1 if surface is asphalt and flange, 0 otherwise)

X_{19}= HRGC surface type (1 if surface is concrete and rubber, 0 otherwise)

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

X_{21}= Traffic volume indicator (1 if AADT is 20,000-30,000, 0 otherwise)

X_{22}= Traffic volume indicator (1 if AADT is >30,000, 0 otherwise)

X_{23}= Vehicle driver age indicator (1 if age is 25-60 years, 0 otherwise)

X_{24}= Vehicle driver age indicator (1 if age is >60 years, 0 otherwise)

In particular, it should be noted that by dropping insignificant variables (one at a time) through conducting a series of tests, the preferred model may actually be different from the above models. However, again, since the main interest of this paper is to examine how the chance of injury and fatal (both) crash increases or decreases (separately or simultaneously) corresponding to a one-unit change in the value of the independent variables, for the convenience and consistency of illustration purposes, the MNL models developed in eqs. (8-9) are used as the final preferred models. Based on the above MNL model eqs. (8-9), the marginal effect/value is also determined and presented in Table 3. As can be seen in Table 3, the sum of marginal effect is zero, which satisfies the requirement that the sum of probability is 1. Using the values in the first row of Table 3 as an example, as vehicle speed changes from category (<40 km/hour [<25 mph]) to category (40-72 km/hour [25-45 mph]), the probability of fatal and injury crashes increases by 0.028 and 0.135, respectively, while the probability of no injury crashes decreases by -0.163. The marginal effect for the remaining variables provides a great deal of valuable information for interpreting results.

As shown in Table 2, vehicle speed was one among several explanatory variables that are considered and used to estimate the vehicle-rail crash severity level. Vehicle speed was categorized into three levels (<40 km/hour [<25 mph], 40-72km/hour [25-45mph], and >72 km/hour [>45 mph]). According to the result, two speed categories (40-72 km/hour [25-45 mph] and >72 km/hour [>45 mph]) were statistically significant and had higher probability of resulting in injury and fatal crashes. It was also shown that the parameter estimate for vehicle speed category three (>72 km/hour [>45 mph]) was higher than vehicle speed category two (40-72 km/hour [25-45 mph]). This indicates that higher vehicle speed has a detrimental effect of increasing the chance of fatal and injury crashes. In this regard, reducing vehicle speed will definitely help in reducing the chance of more severe vehicle-rail crashes at HRGCs.

Table 3: Marginal Effects Results

Variable	P(Fatal)	P(Injury)	P(No injury)
Vehicle speed (40-72 km/hour [25-45 mph])	0.028	0.135	-0.163
Vehicle speed (>72 km/hour [>45 mph])	0.026	0.323	-0.349
Vehicle type (truck-trailer)	0.020	-0.316	0.296
Vehicle type (pick-up)	0.009	0.005	-0.014
Accident type (rail equipment struck vehicle)	-0.022	0.148	-0.125
Temperature (>27^{o}C (80^{o}F))	0.014	0.019	-0.033
Weather (snow)	-0.026	-0.131	0.157
Weather (foggy)	0.028	-0.254	0.226
Train speed (40-72 km/hour [25-45 mph])	0.005	0.349	-0.353
Train speed (>72 km/hour [>45 mph])	-0.017	0.567	-0.550
Vehicle driver gender (male)	0.019	0.054	-0.073
Development area type (residential)	-0.009	-0.035	0.044
Development area type (commercial)	-0.015	-0.066	0.081
Development area type (industrial)	-0.025	-0.016	0.041
Development area type (institutional)	-0.020	-0.099	0.119
HRGC surface type (asphalt)	-0.004	-0.096	0.100
HRGC surface type (unconsolidated)	-0.018	-0.015	0.033
HRGC surface type (asphalt and flange)	0.006	-0.137	0.132
HRGC surface type (concrete and rubber)	-0.006	0.116	-0.110
Traffic volume (AADT 10,000-20,000)	0.003	-0.089	0.086
Traffic volume (AADT 20,000-30,000)	-0.018	-0.157	0.176
Traffic volume (AADT >30,000)	0.002	-0.201	0.199
Vehicle driver age (25-60 years)	-0.002	0.061	-0.059
Vehicle driver age (>60 years)	-0.009	0.254	-0.245

Likewise, train speed was categorized into three levels and also found to be statistically significant. Compared with train speed category one (<40 km/hour [<25 mph]), both higher train speed categories (40-72 km/hour [25-45 mph] and >45 mph) had increased probabilities of injury and fatal crashes. Like vehicle speed, higher train speed also has a detrimental effect of increasing the chance of fatal and injury crashes. As shown in Table 3, the marginal effect result indicates that probabilities of injury and fatal crashes increase as speed of vehicle increases. On the other hand, the probability of no injury crashes decreases as vehicle speed increases.

Seven vehicle categories (ranging from automobile to truck-trailer to school bus types) were considered in this study. Among these seven categories, both truck-trailers and pick-up trucks were found to be statistically significant. As shown in Table 2, truck-trailer vehicles were less likely to result in injury and fatal crashes as compared with automobiles. On the other hand, pickup trucks were more likely to result in injury and fatal crashes. The marginal effect result as shown in Table 3 indicates that truck-trailer vehicles increase the likelihood of fatal and no injury crashes while they decrease injury crashes; and that pickup trucks increase the likelihood of fatal and injury crashes and decrease that of no injury crashes. The reasons behind these interesting results are uncertain and need to be further investigated.

Two crash circumstances (rail equipment struck vehicle and vehicle struck rail equipment) were considered. The crash circumstance under which vehicle struck rail equipment was considered a reference (i.e., base) for comparison. As shown in Table 2, when rail equipment struck vehicle, crash severity was more likely to be fatal. On the other hand, this crash circumstance is less likely to result in injury crashes. Such results are expected and come as no surprise because fatal crashes are believed to be more likely when rail equipment struck vehicle.

Compared with low temperature (less than 10^{o}C [50^{o}F]), vehicle-rail crashes occurring at higher temperatures (greater than 27^{o}C [80^{o}F]) had increased the probability of injury and fatal crashes. As presented in Table 3, the marginal effect result also clearly indicates that higher temperature increases the chance of injury and fatal crashes while decreasing no injury crashes.

Regarding weather condition, snow and foggy conditions were found to be statistically significant. As presented in Table 2, snowy weather conditions were less likely to result in injury and fatal crashes as compared with clear weather conditions. Result also shows that foggy weather conditions were more likely to result in injury crashes but less likely to result in fatal crashes. This might suggest that, as compared with clear weather conditions, people are more likely to drive vehicles with caution under severe weather conditions (such as snow and foggy weather) and therefore the chance of resulting in more severe crashes is reduced.

Five different types of development area types were considered in this study. Compared with open space development areas, HRGCs located in commercial, residential, industrial, and institutional areas were less likely to result in injury and fatal crashes and they were all found to be statistically significant. The marginal effect results in Table 3 also confirm that HRGCs located in these development areas decrease the probability of injury and fatal crashes while the probability of no injury crashes increases. This might suggest that compared with open space development areas, vehicles are more likely to be driven with caution and therefore the probability of resulting in more severe crashes is reduced.

Various types of HRGC surfaces were investigated in this study. A timber crossing surface was considered a reference to which other crossing surface types are compared. As shown in Table 2, vehicle-rail crashes occurring on asphalt, asphalt and flange, and unconsolidated crossing surfaces were found to be less likely to result in injury and fatal crashes, and all these variables were found to be statistically significant. On the other hand, concrete and rubber crossing surface types were also found to be statistically significant; however, crashes occurring on such surfaces were more likely to be injury and fatal crashes. Similar results can also be seen in Table 3. This might indicate that compared with timber crossing surfaces, people are less likely to drive vehicles with caution on concrete surfaces and therefore the chance of resulting in more severe crashes is higher.

The Average Annual Daily Traffic (AADT) was also considered in order to investigate the effect of traffic volume on crash severity. The AADT was classified into four categories. The three AADT categories (i.e., AADT of 10,000-20,000, 20,000-30,000, and >30,000) were found to be statistically significant and they were less likely to result in injury and fatal crashes compared with category one (i.e., AADT less than 10,000). This probably suggests that compared with low traffic volume conditions, people are more likely to drive vehicles with caution under high traffic volume conditions and as a result, the chance of resulting in more severe crashes is reduced.

Finally, vehicle driver characteristics such as age and gender were considered in the study as explanatory variables. With respect to driver gender, male drivers were more likely to be involved in injury and fatal crashes as compared with female drivers and the variable was found to be statistically significant. The age of vehicle drivers is grouped into three categories. Vehicle driver age below 25 was considered a reference for comparison purpose. As shown in Table 2, driver age of 25-60 and above 60 years had higher probability of being involved in injury and fatal crashes. As shown in Table 3, the marginal effects confirm that vehicle drivers age 25-60 and above 60 years increase the probability of being involved in injury crashes while decreasing the probability of being involved in fatal and no injury crashes.

In addition to the model results of intercepts and slope coefficients for serious injury and fatality, the model can be interpreted by using the odds ratio, which is the exponential of parameter estimates obtained from the analysis. For example, the estimated coefficient for vehicle speed category three (i.e., >72 km/hour [>45 mph]) is 0.9211 and hence the relative effect of this speed category versus vehicle speed category one (i.e., <40 km/hour [<25 mph]) is =2.512. This indicates that the odds of vehicle crash severity being injury is 2.512 times higher if the speed of the vehicle is category three compared with that of vehicle speed category one. Similarly, the parameter estimate of vehicle driver age above 60 years, considering driver age below 25 years as a reference, is found to be 0.2706. So, the relative effect of drivers age above 60 years to age below 25 years on injury crashes is determined as =1.311. This indicates that the odds of injury crash severity versus no injury crashes are 1.311 times higher for drivers age above 60 years compared with those below 25 years. The odds ratio results of the rest of variables can also be interpreted in a similar fashion.

The probabilities of the three severity crashes (fatality, injury and no injury) can be predicted by using the following three eqs. (10-12).

(10) $$P_{Fatal} = \frac{e^{equation(8)}}{[1 + e^{equation(8)} + e^{equation(9)}]}$$

(11) $$P_{Injury} = \frac{e^{equation(9)}}{[1 + e^{equation(8)} + e^{equation(9)}]}$$

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

The probability of the three different severity levels of vehicle-rail crashes are determined based on the parameters estimated for the indicator variables in the models as shown in eqs. (8-9) and the probability eqs. (10-12) shown above. Accordingly, the predicted average probability of fatal, injury, and no injury severity levels are 0.072, 0.299, and 0.629, respectively by using these equations. And the observed crash severity from the original data (as discussed and shown in the "Data Assembly and Analysis" section) was 0.069, 0.334, and 0.597 for fatality, injury, and no injury, respectively. Also as shown in Table 2, the ρ^2 determined for the model is 0.011, which indicates the model has some improvement over the naïve model (model without covariates).

CONCLUSION

Highway vehicle crash severity levels at HRGCs were modeled using MNLM in this paper. The three vehicle crash severity levels (fatality, injury, and no injury) were considered as dependent variables. Vehicle and vehicle user characteristics, environmental factors, type of development area, highway-rail crossing characteristics, highway traffic characteristics, vehicle speed, and train speed were the explanatory variables used in predicting crash severity levels. The analysis was conducted using SAS PROC LOGISTICS procedure. In order to retain some of the variables, those within 90% confidence level were considered statistically significant. Some of the variables were found to be statistically significant even at 95% confidence level.

As discussed in the paper, results indicate that as vehicle and/or train speeds increase, the chance of being involved in injury and fatal crashes at HRGCs also increases. Hence, reducing train and vehicle speeds at HRGCs will certainly minimize the chance of resulting in more severe crashes. It is noted that the majority of crashes occurred when rail equipment struck vehicles. In particular, this type of accident increases the chance of resulting in fatal crashes. As for vehicle types, truck-trailer vehicles are observed to decrease the probability of fatal crashes while pickup trucks increase such chances. It is also observed that male drivers above 25 years are more likely to be involved in injury and fatal crashes. Moreover, crashes occurring at higher temperatures are more likely to be injury and fatal compared with those occurring at low temperatures. Also, higher traffic volume (i.e., higher AADT) decreases the probability of resulting in injury and fatal crashes. Results seem to suggest that people are more likely to drive vehicles with caution at commercial/residential/industrial/institutional areas as opposed to open space development areas, under severe weather conditions (such as snow and foggy weather) compared with clear weather conditions, on non-concrete surfaces as opposed to concrete surfaces, and as a result, the chances of being involved in more severe crashes are reduced. In all, educating and equipping drivers with good driving habits (such as reducing speeds or stopping their vehicle completely regardless of a train being present or not) at HRGCs and short-term law enforcement actions, can potentially minimize the chance of resulting in more severe vehicle crashes at HRGCs.

Future research may be directed toward modeling the severity level at HRGCs using ordered logit, which can be modeled by the proportional odds models (McCullagh 1980) along with the conduct of a score test for proportional odds assumption (Strokes et al. 2000). Furthermore, future direction will be on developing a simultaneous equations model of crash frequency by severity level at HRGCs since this will improve the efficiency of uncovering the influential factors. Last but not least, the reasoning based on the empirical study results will need to be fully supported by field investigation evidences in the near future.

References

Austin, Ross D. and Jodi Carson. "An Alternative Accident Prediction Model for Highway-Rail Interfaces." *Accident Analysis and Prevention* 34 (1), (2002): 31-42.

Chen, Wan-Hui and Paul Jovanis. "Method for Identifying Factors Contributing To Driver-Injury Severity in Traffic Crashes." *Transportation Research Record* 1717, (2000): 1-9.

Chira-Chavala, T., B. Coifman, C. Porter, and M. Hansen. "Light Rail Accident Involvement and Severity." *Transportation Research Record* 1521, (1996): 147-155.

Dissanayake, Sunanda and John Lu. "Analysis of Severity of Young Driver Crashes: Sequential Binary Logistic Regression Modeling." *Transportation Research Record* 2302, (2002): 108-114.

Donnell, Eric T. and John M. Mason. "Predicting the Severity of Median-Related Crashes in Pennsylvania by Using Logistic Regression." *Transportation Research Record* 1897, (2004): 55-63.

Duncan, Chandler, Asad J. Khattak, and Forrest M. Council. "Applying the Ordered Probit Model to Injury Severity in Truck–Passenger Car Rear-End Collisions." *Transportation Research Record* 1237, (1998): 63-71.

Famoye, Felix, John T. Wulu, and Karan P. Singh. "On the Generalized Poisson Regression Model with an Application to Accident Data." *Journal of Data Science* 2, (2004): 287-295.

Federal Railroad Administration. U.S. Department of Transportation. Railroad Safety Statistics, *2012 Preliminary Annual Report*. http://safetydata.fra.dot.gov/OfficeofSafety/publicsite/Prelim.aspx (2012) (accessed August 2, 2014).

Gitelman, V. and A.S. Hakkert. "The Evaluation of Road–Rail Crossing Safety with Limited Accident Statistics." *Accident Analysis and Prevention* 29 (2), (1997): 171-179.

Hu, Shou-Ren, Chin-Shang Li, and Chi-Kang Lee. "Investigation of Key Factors for Accident Severity at Railroad Grade Crossings by Using a Logit Model." *Safety Science* 48, (2010): 186-194.

Huang, Herman F., Richard J. Stewart, and Charles V. Zegeer. "Evaluation of Lane Reduction "Road Diet" Measures on Crashes and Injuries." *Transportation Research Record* 2955, (2002): 80-90.

Khattak, Asad J. "Injury Severity in Multivehicle Rear-End Crashes." *Transportation Research Record* 3466, (2001): 59-68.

Khattak, Aemal, Michael D. Pawlovich, Reginald R. Souleyrette, and Shauna L. Hallmark. "Factors Related to More Severe Older Driver Traffic Crash Injuries." *Journal of Transportation Engineering* 128 (3), (2002): 243-249.

Long, Scott J. *Regression Models for Categorical and Limited Dependent Variables*. Sage Publications. Thousand Oaks, CA, 1997.

McCullagh, Peter. "Regression Models for Ordinal Data (With Discussion)." *Journal of the Royal Statistical Society Series B* 42, (1980): 109-142.

Mercier, Cletus R., Mack C. Shelley, Geneva H. Adkins, and Joyce M. Mercier. "Age and Gender as Predictors of Injury Severity in Broadside and Angle Vehicular Collisions." *Transportation Research Record* 0607, (1999): 50-61.

Mercier, Cletus R., Mack C. Shelley, Julie B. Rimkus, and Joyce M. Mercier. "Age and Gender as Predictors of Injury Severity in Head-On Highway Vehicular Collisions." *Transportation Research Record* 0535, (1997): 37-46.

Miranda-Moreno, Luis F., Liping Fu, Frank F. Saccomanno, and Aurelie Labbe. "Alternative Risk Models for Ranking Locations for Safety Improvement." *Transportation Research Record: Journal of the Transportation Research Board* 1908, (2005): 1-8.

National Highway Traffic Safety Administration. U.S. Department of Transportation Traffic Safety Facts Research Note, "2011 Motor Vehicle Crashes: Overview," 2012. http://www-nrd.nhtsa.dot.gov/Pubs/811701.pdf (accessed August 2, 2014).

Park, Eun Sug and Dominique Lord. "Multivariate Poisson-Lognormal Models for Jointly Modeling Crash Frequency by Severity." *Transportation Research Record: Journal of the Transportation Research Board* 2019, (2007): 1-6.

Park, Peter Young-Jin and Frank Fedel Saccomanno. "Reducing Treatment Selection Bias for Estimating Treatment Effects Using Propensity Score Method." *ASCE Journal of Transportation Engineering* 133 (2), (2007): 112-118.

Renski, Henry, Asad J. Khattak, and Forrest M. Council. "Effect of Speed Limit Increases on Crash Injury Severity Analysis of Single-Vehicle Crashes on North Carolina Interstate Highways." *Transportation Research Record* 0975, (1999): 100-108.

Saccomanno, Frank F., Liping Fu, and Luis F. Miranda-Moreno. "Risk-Based Model for Identifying High-Rail Grade Crossing Blackspots." *Transportation Research Record: Journal of the Transportation Research Board* 1862, (2004): 127-135.

Saccomanno, Frank F., Peter Young-Jin Park, and Liping Fu. "Estimating Countermeasure Effects for Reducing Collisions at Highway–Railway Grade Crossings." *Accident Analysis and Prevention* 39, (2007): 406-416.

SAS/STAT User's Guide, Version 9.2. SAS Publishing, Cary, NC, U.S.A., 2008.

Savolainen, Peter T., Fred L. Mannering, Dominique Lord, and Mohammed M. Quddus. "The Statistical Analysis of Highway Crash-Injury Severities: A Review and Assessment of Methodological Alternatives." *Accident Analysis and Prevention* 43, (2011): 1666-1676.

Strokes, Maura E., Charles S. Davis, and Gary G. Koch. *Categorical Data Analysis Using the SAS System*, second ed. SAS Institute, Inc., Cary, NC, U.S.A., 2000.

Tay, Richard, Jaisung Choi, Lina Kattan, and Amjad Khan. "A Multinomial Logit Model of Pedestrian–Vehicle Crash Severity." *International Journal of Sustainable Transportation* 5, (2011): 233-249.

Xie, Yuanchang, Yunlong Zhang, and Faming Liang. "Crash Injury Severity Analysis Using Bayesian Ordered Probit Models." *Journal of Transportation Engineering* 135 (1), (2009): 18-25.

Ye, Xin, Ram M. Pendyala, Simon P. Washington, Karthik Konduri, and Jutaek Oh. "A Simultaneous Equations Model of Crash Frequency by Collision Type for Rural Intersections." *Safety Science* 47, (2009): 443-452.

Ye, Xin, Ram M. Pendyala, Venky Shankar, and Karthik Konduri. "A Simultaneous Equations Model of Crash Frequency by Severity Level for Freeway Sections." *Accident Analysis and Prevention* 57, (2013): 140-149.

Dr. Wei (David) Fan *is currently an associate professor of civil engineering at the University of North Carolina at Charlotte. He worked as an associate professor at the University of Texas at Tyler and was a senior optimization developer in the R&D department in SAS Institute Inc. after he obtained a Ph.D. degree in civil engineering from the University of Texas at Austin in 2004. His primary research interests include operations research, transportation network modeling and optimization, statistical data analysis, traffic simulation, traffic engineering and operations, transportation planning, and pavement management systems. He has 20 years of experience with statistical and optimization computer software development, has authored over 60 journal and conference publications, and has given numerous conference and invited presentations.*

Dr. Martin Kane *is currently an associate professor of civil engineering at the University of North Carolina at Charlotte. His primary research interests include traffic engineering, highway & traffic safety, public transportation, driver performance, traffic control devices, and intelligent transportation systems.*

Mr. Elias Haile *received his master's degree from the University of Texas at Tyler in December 2013. His primary research interests include statistical data analysis of highway-rail grade crossings.*