EXPLORING SINGLE VEHICLE CRASH SEVERITY ON RURAL, TWO-LANE HIGHWAYS WITH CRASH-LEVEL AND OCCUPANT-LEVEL MULTINOMIAL LOGIT MODELS

by

Yunqi Zhang

A thesis submitted to the faculty of
The University of Utah
in partial fulfillment of the requirements for the degree of

Master of Science

Department of Civil and Environmental Engineering

The University of Utah

August 2011

The University of Utah Graduate School

STATEMENT OF THESIS APPROVAL

The thesis of **Yunqi Zhang**

has been approved by the following supervisory committee members:

Richard Porter	, Chair	**06/02/2011** Date Approved
Peter Martin	, Member	**06/02/2011** Date Approved
Xuesong Zhou	, Member	**06/02/2011** Date Approved

and by **Paul Tikalsky** , Chair of

the Department of **Civil and Environmental Engineering**

and by Charles A. Wight, Dean of The Graduate School.

ABSTRACT

This thesis is conducted to compare a crash-level severity model with an occupant-level severity model for single-vehicle crashes on rural, two-lane roads. A multinomial logit model is used to identify and quantify the main contributing factors to the severity of rural, two-lane highway, single-vehicle crashes including human, roadway, and environmental factors. A comprehensive analysis of 5 years of crashes on rural, two-lane highways in Illinois with roadway characteristics, vehicle information, and human factors will be provided. The modeling results show that lower crash severities are associated with wider lane widths, shoulder widths, and edge line widths, and larger traffic volumes, alcohol-impaired driving, no restraint use will increase crash severity significantly. It is also shown that the impacts of light condition and weather condition are counterintuitive but the results are consistent with some previous research. Goodness of fit test and IIA (independence of irrelevant alternatives) test are applied to examine the appropriateness of the multinomial logit model and to compare the fit of the crash-level model with the occupant-level model. It is found that there are consistent modeling results between the two models and the prediction of each severity level by crash-level model is more accurate than that of the occupant-level model.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Background

Motor vehicle travel is the primary means of transportation in the United States, providing an unprecedented degree of mobility. The National Highway Traffic Safety Administration have reported that in the year of 2008, vehicle travel resulted in 5,811,000 police reported crashes and 37, 261 fatalities on highways, and those who manage to survive crashes are faced with such potential consequences as mental trauma, pain, and expensive medical costs. The society as a whole is also at a loss, both economically and emotionally, because of these incidents. Fortunately, statistical analyses of the likelihood of motor vehicle accidents have the ability to predict motor vehicle safety, thus reducing injuries or fatalities and mitigating the loss by crashes. Such analyses could help identify factors that one can control to keep motor vehicle risks at an acceptable level, thereby saving lives, preventing injuries, and making motor vehicle travel a more competitive mode of travel.

The Highway Safety Manual (HSM) provides analytical tools and techniques for quantifying the potential effects on crashes as a result of decisions made in planning, design, operations, and maintenance. There is no such thing as absolute safety. There is risk in all highway transportation. A universal objective is to reduce the number and severity of crashes within the limits of available resources, science, and technology, while meeting legislatively mandated priorities. The information in the HSM is provided to assist agencies in their effort to integrate safety into their decision-making processes. Most of the research used to develop the HSM relied on accident frequency methods for analyzing road safety.

The following equations show how the HSM predicts crash frequency:

$$N_{predicted} = N_{SPF} * C_r * (AMF_{1r} * AMF_{2r} * \ldots AMF_{12r}) \quad (1.1)$$

$$N_{SPF} = AADT * L * 365 * 10^{-6} * e^{-0.312} \quad (1.2)$$

where

$N_{predicted}$ =predicted average crash frequency for an individual roadway segment for a specific year

N_{SPF} = predicted average crash frequency for an individual roadway segment that has base condition

C_r =calibration factor for roadway segments of a specific type developed for a particular jurisdiction or geographic area

$AMF_{1r} \ldots AMF_{12r}$ = Accident Modification Factor for rural two-way two-lane roadway segments that modify the safety prediction based on how the segment deviates from base conditions

AADT= average annual daily traffic (vehicles per day)

L=segment length (miles)

The base condition for an individual roadway segment represents geometric design features and traffic control conditions that are quantified for some specific values to be set as basic. For example, it is base condition when lane width equals 12 feet or the roadside hazard rating is 3.

Accident Modification Factors (AMFs) represent the relative change in crash frequency due to a change in one specific condition (when all other conditions and site characteristics remain constant). AMFs are the ratio of the crash frequency of a site under two different conditions. Therefore, an AMF may serve as an estimate of the effect of a particular geometric design or traffic control feature or the effectiveness of a particular treatment or condition.

Therefore, in the Highway Safety Manual, the crash frequency is predicted based on roadway characteristics; any change in geometric attributes of a roadway can be converted to a change of crash frequency.

Table 1 shows that the HSM method for estimating accident severity on rural, two-lane roads is limited to using default distribution for crash severity level based on California data. Crash severity likely varies from place to place, and the distribution of severity in California may not represent severity distributions in other places. The HSM method focuses mainly on roadway characteristic factors for predicting crash frequency or crash severity. Driver and occupant factors are briefly mentioned in the HSM, but they do not play key roles in the safety prediction algorithms.

Frequency-based modeling methods (i.e., negative binomial, Poisson regression model) are sometimes used to estimate the expected numbers of crashes on road segments for each severity level (e.g., one model for fatal plus injury crashes, one model for total crashes) and then the proportion of each severity level can be estimated.

The Poisson approach for predicting crash frequency is:

$$P(n_i)=\frac{\lambda_i^{ni}\exp(-\lambda_i)}{n_i!} \tag{1.3}$$

$$\lambda_i = \exp(\beta X_i) \tag{1.4}$$

where

$P(n_i)$= the probability of n accidents occurring on a highway section i over one year time period

λ_i = the expected accident frequency for highway section i

X_i = a vector of explanatory variable

β = a vector of estimable coefficients

The potential limitation of this approach is that this method could introduce estimation bias because it assumes the accident frequencies of each severity level are independent from each other (Milton et al. 2008). It is reasonable to think that the change in the frequency of one accident severity may have some effects on the frequencies of other severity levels.

There are other limitations with accident frequency-based approaches (the number of accidents), or the use of frequency-dominated approaches (the number of accidents with consideration to resulting injury severity usually only at the fatality level). Accident frequency approaches tend to favor locations where accidents are more likely to occur at the expense of some locations that may have fewer but more severe accidents. The frequency-based models used to address crash severity often include aggregating different severity levels (e.g., one outcome is PDO, the other outcome is fatal plus injury). An injury crash has potential outcomes ranging from possible injury to near death.

1.2 Classification of Crash Severity

Identifying the level of crash severity is an important step to model crash severity. There are several ways to classify crash severity. The advantages and disadvantages of different classifications are listed below.

1.2.1 Fatal Plus Injury And No Injury

This classification divides crashes into those resulting in any level of injury for fatality, and those resulting in property damage only. This division of severity level is too vague because injury has so many forms from possible injury to near death.

1.2.2 AIS (Abbreviated Injury Scale)

AIS classifies injuries by body part, specific lesion, and severity. AIS is on a 6-point scale (0-6) and the level is determined by comparing injury diagnosed by a medical expert to a defined scale. Usually the AIS level determined soon after an accident is not a final outcome because the injuries from other levels may turn out to be death. AIS level is based on medical criteria and does not reflect how the same injury will affect different individuals.

1.2.3 MAIS (Maximum Abbreviated Injury Scale)

Sometimes people suffer from injuries to more than one body part, so a maximum AIS (MAIS) is used to cover different injured regions of the body and reflect the most serious injured regions. The advantage of AIS and MAIS is that they are determined by physicians so they tend to be more accurate. However, they need experts to examine the injuries and then put the findings into files so it may take a longer time period.

1.2.4 KABCO

The KABCO scale is the most commonly used by the crash reporting officer, and is therefore the most readily available scale in databases received from state agencies. However, measurement biases are made by police officers. The most serious problem is overrating of the severity level by the police officer. The bias in police reported crash severities has been explored in Farmer (2003).

1.3 Research Objectives

The objective of this thesis is to explore the relationship between crash severity and driver, vehicle, and roadway variables on two-lane rural highways. The data used will be from commonly maintained crash and roadway databases from state transportation agencies. However, it is a first step towards a more detailed explanation of crash severities that will be conducted during a doctoral thesis. The objective of this thesis will be achieved through the following tasks:

1. Provide a literature review of discrete choice methods used to estimate the relationship between crash severity and contributing factors;
2. Identify and quantify the main contributing factors to the severity of rural, two-lane highway, single-vehicle crashes including human, roadway, and environmental factors. A comprehensive analysis of 5 years of crashes on rural, two-lane highways in Illinois with roadway characteristics, vehicle information, and human factors will be provided. The multinomial logit model will be used to identify factors that are

associated with the severity outcomes of single-vehicle crashes. Key assumptions of the multinomial logit model will be tested.

3. Compare a crash-level severity model with an occupant-level severity model for single-vehicle crashes on rural, two-lane roads.

Table 1 Distribution for Crash Severity on Rural Two-Lane Segments

Crash severity level	Percentage of total roadway segment crashes
Fatal	1.3
Incapacitating	5.4
Nonincapacitating	10.9
Possible injury	14.5
Total plus injury	32.1
Property damage only	67.9
Total	100

CHAPTER 2

LITERATURE REVIEW

2.1 Multinomial Logit Model

The multinomial logit model is widely used to estimate accident severity.

Shankar and Mannering (1996) attempted to address the potential bias that univariate analyses creates by presenting a multinomial logit model of motorcycle-rider accident severity in single vehicle collisions. They concluded that the multinomial model is a promising approach to evaluate the determinants of motorcycle accident severity. Savolainen and Mannering (2007) researched a similar topic (motorcyclists' injury severities in single- and multivehicle crashes) using a multinomial logit model for multi-vehicle crashes. It is concluded that collision type, roadway characteristics, alcohol consumption, helmet use, and unsafe speeds play significant roles in crash-injury outcomes.

The injury severity of male and female drivers in single and two-vehicle accidents for different types of vehicles was explored by Ulfarsson and Mannering (2004) using a multinomial logit model. The results suggest that there are important behavioral and physiological differences between male and female drivers that must be explored further and addressed in vehicle and roadway design.

Multinomial logit models have been used to explore the differences between urban and rural driver injuries in accidents that involve large trucks by Niemeier et al. (2005). The results showed that many variables were found to be significant in either the rural or the urban model, but not both because of the different perceptual, cognitive, and response demands placed on drivers in rural versus urban areas.

2.2 Nested Logit Model

Generalized extreme value (GEV) models constitute a large class of models that exhibit a variety of substitution patterns. The unifying attribute of these models is that the unobserved portions of utility for all alternatives are jointly distributed as a generalized extreme value. This distribution allows for correlations over alternatives and, as its name implies, is a generalization of the univariate extreme value distribution that is used for standard multinomial logit models described above. When all correlations are zero, the GEV distribution becomes the product of independent extreme value distributions and the GEV model becomes standard multinomial logit. The class therefore includes logit but also includes a variety of other models.

Hypothesis tests on the correlations within a GEV model can be used to examine whether the correlations are zero, which is equivalent to testing whether standard logit provides an accurate representation of the substitution patterns.

The most widely used member of the GEV family is called nested logit. This model has been applied by many researchers in a variety of situations, including energy, transportation, housing, and telecommunications. Its functional form is simple compared to other types of GEV models. Nested logit models allow partial relaxation of the IIA property. Sometimes different alternatives may share the same unobserved terms. The nested logit model can overcome the restriction of the MNL model that requires the error term for different alternatives, ε_{in}, to be independent from each other.

Shankar et al. (1996) presented a nested logit formulation as a means for determining accident severity on rural highways given that an accident has occurred. They concluded that a nested logit model, which accounted for shared unobservables between property damage and possible injury accidents, provided the best structural fit for the observed distribution of accident severities.

Chang and Mannering (1999) studied occupancy/injury severity relationship in truck- and non-truck-involved accidents using the nested logit model. The findings of this study demonstrated that the nested logit model, which was able to take into account vehicle occupancy effects and identify a broad range of factors that influence occupant injury, is a promising methodological approach.

Holdridge et al. (2005) analyzed the in-service performance of roadside hardware on the entire urban State Route system in Washington State by developing multivariate nested logit models of injury severity in fixed-object crashes. The models showed the contribution of guardrail leading ends toward fatal injuries and also indicated the importance of protecting vehicles from crashes with rigid poles and tree stumps.

2.3 Ordered Logit Model

Wang and Abdel-Aty (2008) examine left-turn crash injury severity using an ordered logit model. This study found that neither the total approach volume, nor the entire intersection volume, but rather the specific vehicle movements affected crashed injury significantly.

2.4 Ordered Probit Model

Duncan et al. (1998) examined the impact of various factors on injuries to passenger car occupants involved in truck-passenger car rear-end collisions and demonstrated the use of the ordered probit model in the complex highway safety problem. They concluded that the ordered probit model is flexible because it allows the injury severity probabilities to vary differently across categories.

Klop and Khattak (1999) explored the effect of a set of roadway, environmental, and crash variables on bicycle injury severity using the ordered probit model. The model results showed that variables that significantly increase injury severity include straight grades, curved grades, darkness, fog, and speed limit.

Quddus et al. (2002) used an ordered probit model to examine factors that affect the injury severity of motorcycle accidents and the severity of damage to the motorcycles and vehicles involved in those crashes. They concluded that factors leading to increased probability of vehicle and motorcycle damage included some similar factors and different factors.

Kockelman and Kweon (2002) described the use of ordered probit models to examine the risk of different injury levels sustained under all crash types, two-vehicle crashes, and single-vehicle crashes. This work suggested that the manner of collision, number of vehicles involved, driver gender, vehicle type, and driver alcohol use played major roles in terms of the crash severity.

Adbel-Aty (2003) analyzed driver injury severity at locations of roadway sections, signalized intersections, and toll plazas using the ordered probit model. This study illustrated the similarities and differences in the factors that affect injury severities at different locations.

Donnell and Cornor (1996) use both an ordered logit model and ordered probit model to predict the severity of motor vehicle accident injuries. They concluded that occupant age, vehicle speed, seat position, blood alcohol level, and type of collision had affected the probabilities of serious injury and death.

2.5 Mixed Logit Model

Gkritza and Mannering (2008) demonstrated a mixed logit approach that can be used to better understand the use of safety belts in single- and multi-occupant vehicles. They concluded that the mixed logit model can provide a much fuller understanding of the interaction of the numerous variables which correlate with safety-belt use.

Milton et al. (2008) analyzed the injury-severity distributions of accidents on highway segments, and the effect that traffic, highway, and weather characteristics have on these distributions using a mixed logit model. Their results showed that the mixed logit model has considerable promise as a methodological tool in highway safety programming.

Pai et al. (2009) estimated mixed logit models to investigate the contributory factors to motorists' right of way violations in different crash types. It was found that motorcycles' right of way was more likely to be violated on non-built-up roads, and in diminished light conditions.

Kim et al. (2010) applied a mixed logit model to analyze pedestrian-injury severity in pedestrian-vehicle crashes to address possible unobserved heterogeneity. It was found that several factors increased the fatal injury level significantly, including darkness, drunk driving, and speeding. They found that the effect of pedestrian age was normally distributed across observations, and that as pedestrians became older, the probability of fatal injury increased substantially.

Eluru et al. (2008) has developed an ordered mixed logit to examine pedestrian and bicyclist injury severity in traffic crashes. They concluded that the ordered mixed model does not produce inconsistent estimates of the effects of some variables as does the ordered probit model. The analysis also suggested that the general pattern and relative magnitude of elasticity effects of injury severity determinants are similar for pedestrians and bicyclists.

2.6 Crash-level Model Versus Occupant-level Model Analysis

Lenguerrand et al. (2006) use the multilevel logistic model, generalized estimating equation and logistic model to estimate the hierarchical structure of road crash data—it is believed that correlations of injury severity can be found for drivers and occupants in the same car or in the same accident. They concluded that the MLM is the most efficient model while both GEE and LM underestimate parameters and confidence intervals. MLM methods divide the crash data into 3 categories—crash level, car level and then occupant level so it provides a more precise estimation for the crash data. It is also concluded that the Lenguerrand study is in agreement with others studies not taking the hierarchical road crash structure into account because in practice, the departures from more appropriate and more complex models are minor and the results from the LM model is acceptable. One of the main objectives of this paper is to compare a crash-level severity model with an occupant-level severity model for single-vehicle crashes on rural, two-lane roads.

2.7 Summary of Literature Review

There are several commonly used discrete choice models for predicting crash severity such as the multinomial logit model, nested logit model, ordered probit model, and mixed logit model. These approaches have been applied to crash severity analysis by researchers on the relationship between crash severity and its contributing factors. Table 2 shows a summary of commonly used discrete choice models. Advantages and limitations as well as important assumptions of these models are presented.

Table 2 Summary of Crash Severity Predicting Models

Model Type	Previous Research	Advantage	Limitation	Assumptions
Multinomial Logit	Shankar and Mannering (1996); Ulfarsson and Mannering (2004); Niemeier et al. (2005); Savolainen and Mannering (2007)	Readily interpretable; Allows coefficients of variables to vary between different categories	Susceptible to correlation of unobserved effects from one injury severity level to the next (IIA property); Does not recognize the ordering of injury severity outcomes	The error terms should be independently and identically distributed
Nested Logit	Shankar et al. (1996); Chang and Mannering (1999); Holdridge et al. (2005)	Relaxes IIA assumption	Does not recognize the ordering of injury severity outcomes	The error terms should be generalized extreme value distributed
Ordered Logit	Donnell and Cornor (1996); Wang and Abdel-Aty (2008)	Recognizes the ordering of injury severity outcomes	The shift in thresholds are restricted to move in the same direction	Parallel slope assumption
Ordered Probit	Duncan et al. (1998); Klop and Khattak (1999); Quddus et al. (2002); Kockelman and Kweon (2002); Adbel-Aty (2003)	Recognizes the ordering of injury severity outcomes	The shift in thresholds are restricted to move in the same direction	Parallel slope assumption; The error terms should be normally distributed
Mixed Logit	Eluru et al. (2008); Gkritza and Mannering (2008); Milton et al. (2008); Pai et al. (2009); Kim et al. (2010)	It is highly flexible that it obviates the limitations of standard logit	Does not recognize the ordering of injury severity outcomes	

CHAPTER 3

METHODOLOGY

3.1 Model Selection

The multinomial logit model was selected to estimate the relationship between crash severity and contributing factors in this masters thesis. The multinomial logit results in choice probabilities that take a closed form and is readily interpretable. Also, the multinomial logit model allows the coefficients of variables to vary for different categories so that the different impact of variables for each severity category is clearly shown. A full understanding of the multinomial logit is an important transition to my doctoral studies.

3.2 Modeling Approach

The logit model was first derived by Luce (1959), and it is the most widely used model because of the fact that the choice probabilities take a closed form and is readily interpretable.

In the multinomial logit model, the probability that accident n will have severity i is given by:

$$p_n(i) = \exp(\beta_i X_n) / \sum_I \exp(\beta_I X_n) \tag{3.1}$$

where

p_n (i) =the probability that crash n will be in severity level i

X_n = a set of variables that will determine the crash severity

β_i = a vector of parameters to be estimated

Utility functions defining the severity likelihoods are defined as:

$$S_{in} = \beta_i X_n + \varepsilon_{in} \tag{3.2}$$

where

ε_{in} = error terms that account for unobserved variable.

The error terms for each choice should follow independent extreme value distributions (also called Gumbel or type I extreme value). The key assumption is that the errors are independent of each other. This independence means that the unobserved portion of utility for one alternative is unrelated to the unobserved portion of utility for another alternative.

If the researcher thinks that the unobserved portion of utility is correlated over alternatives, then there are three options: (1) use a different model that allows for correlated errors, such as the nested logit or mixed logit model, (2) respecify the representative utility so that the source of the correlation is captured explicitly and thus the remaining errors are independent, or (3) use the logit model under the current specification of representative utility, considering the model to be an approximation.

3.3 Goodness-of-Fit Test

In logit model, the method used to test how well the model fits the data is Goodness-of-Fit, which is known as the likelihood ratio index. Stated more precisely, the statistic measures how well the model, with its estimated parameters, performs compared with a model in which all the parameters are zero (which is usually equivalent to having no model at all).

The likelihood ratio index is defined as:

$$\rho = 1 - \frac{LL(\beta)}{LL(0)} \tag{3.3}$$

where

ρ = likelihood ratio index

LL (β) = the value of the log-likelihood function at the estimated parameters

LL(0) = the value when all the parameters are set equal to zero.

Values for this goodness of fit value have generally varied from 0.2 to 0.5 for crash severity analysis using multinomial logit models estimated by other researchers such as Shankar and Mannering (1996), and Ulfarsson and Mannering (2004).

3.4 IIA Test

Whether IIA (independence of irrelevant alternatives) holds is an important issue for the application of the multinomial logit model. If IIA holds, the ratio of probabilities for any two alternatives is entirely unaffected by the systematic utilities of any other alternatives. Tests of IIA were first developed by McFadden et al. (1978). Under IIA, the ratio of probabilities for any two alternatives is the same whether or not other alternatives are available. As a result, if IIA holds in reality, then the parameter estimates obtained on the subset of alternatives will not be significantly different from those obtained on the full set of alternatives. A test of the hypothesis that the parameters on the subset are the same as the parameters on the full set constitutes a test of IIA (Mcfadden and Hausman 1984).

Denote the following:

a= full set alternatives

b= specified subset of alternatives

β_b = estimates from b, an n*r matrix in which n represents number of categories and r represents number of parameters in each category

Ω_b = estimated covariance matrix of b

β_a = estimates from a, an n*r matrix that n represents number of categories (consistent with b) and r represents number of parameters in each category

Ω_a = estimated covariance matrix of a

Null hypothesis: the coefficients of variables are equal for full set alternatives and subset alternatives.

The null hypothesis can be tested by:

$$(\beta_a - \beta_b)'(\Omega_b - \Omega_a)^{-1}(\beta_a - \beta_b) \tag{3.4}$$

The quadratic has a chi-square distribution with the degrees of freedom equal to the number of coefficients estimated in the constrained model. If the null hypothesis is not rejected, then the IIA assumptions hold and the multinomial logit model is appropriate. Table 3 shows the important variables used in this chapter and their notations.

Table 3 Notations of Variables Used in Chapter 3

variables	notations
p_n (i)	probability that crash n will be in severity level i
X_n	a set of variables that will determine the crash severity
β_i	a vector of parameters to be estimated
ρ	likelihood ratio index
LL (β)	the value of the log-likelihood function at the estimated parameters
LL(0)	the value when all the parameters are set equal to zero
a	full set alternatives
b	specified subset of alternatives
β_a	estimates from a
Ω_a	estimated covariance matrix of a
β_b	estimates from b
Ω_b	estimated covariance matrix of b

CHAPTER 4

CRASH-LEVEL MODEL DATA

4.1 Description of Data

The data used for this study come from the Highway Safety Information System (HSIS) for crashes that occurred on rural, two-lane highways in Illinois from 2001 through 2005. There are four datasets that can be merged together using linkage variables. The accident file has information about crashes including where and when the accident occurred as well as the characteristics of the crash. Details like roadway condition, accident type, traffic control condition, and weather are included in the accident file. The road file includes basic characteristics of the roadway segment where the accident occurred. Information such as lane width, shoulder width, average annual daily traffic (AADT), speed limit, and horizontal curvature are included in road file. The vehicle file includes vehicle type and the driver characteristics and conditions such as driver age, driver sex, physical condition of the drivers, and restraint use of drivers. The occupant file includes information about the vehicle occupants involved in the crash other than drivers, including occupant age, restraint use, and their seat position when the accident occurred.

4.1.1 Merging Data

The accident file and road file were linked by using the “milepost” and “cntyrte” variables. These variables describe the county, route number, and location along the route where the accident occurred. Only those accidents and segments labeled as "rodwycls=8" were selected representing rural, two-lane roads. After road and accident files were merged, the occupant and vehicle files were matched to the combined accident and road file. The accident file includes information about the most severe injured occupant while the vehicle and occupant files includes the severity level experienced by the driver and all the occupants. Therefore the dataset was expanded after the occupant and vehicle file were merged to the accident and roadlog file. This merging was done using the accident case number.

For the accident-level database used for this study, one row represented one accident with all the variables associated with that accident. If there were three persons involved in the accident, the accident severity was coded as the most severe injury sustained by all of the occupants. For the occupant-level database, one row represents one person involved in the crash with all the variables associated with that occupant and crash. If there were three persons involved in the accident, there were three rows containing information concerning the condition of each specific person and the level of injury severity that each particular person sustained.

4.1.2 Variable Definitions and Descriptive Statistics

Table 4 shows all the contributing factors used in the severity analyzing model. There are 21 independent variables altogether. Roadway characteristics variables include lane width, shoulder width, edgeline width, speed limit (whether the speed limit is 55mph or less than that), horizontal curvature, and average annual daily traffic. There are three types of collisions for single vehicle crashes: animal collisions, fixed-object collisions, and rollover collisions. The animal collisions were set to be the base crash type, with indicator variables included for fixed-object collisions and rollover collisions. Light conditions and weather conditions were also important factors potentially influencing crash severity. Driver sex, driver age, restraint use of occupants, an indicator for alcohol-impaired driving, and the numbers of occupants by sex were also included in the models. Table 5 shows the summary statistics for all the variables. A total of 9194 observations were included in the model. The observations for which “unknown” was recorded were excluded from further analysis.

Table 4 Description of Variables Used in the Accident-Level Model

variables	description
male	Number of male occupants other than driver
Female	Number of female occupants other than driver
No_Back_rest	Restraint use for back seat occupants (0=yes 1=no)
No_Front_rest	Restraint use for front seat occupants (0=yes 1=no)
drvage	Driver age
alcohol	Alcohol impaired driving (0=no 1=yes)
drvsex	Driver sex (0=female 1=male)
drvrest	Restraint use for driver (0=yes 1=no)
fixed	Fixed object collision (0=no 1=yes)
rollover	Rollover collision (0=no 1=yes)
daylight	Light condition (0=good 1=dark)
weather	Weather condition (0=normal 1=rainy 2=snowy 3=foggy)
lanewid	Lane width (ft)
R_shdr_wid	Right shoulder width (ft)
Spd_limt	Speed limit indicator (0=speed limit less than 55mph 1=speed limit equal to 55mph)
Edg_line_wid	Edgeline width (inches)
horizontal	Horizontal curve (0=no 1=yes)
aadt	average annual daily traffic (1000 vehicles per day)

Table 5 Summary Statistics for Variables in the Accident- Level Model

Variable	Mean	Standard deviation	Min	Max
Male	0.144	0.441	0	5
Female	0.188	0.513	0	10
No_Back_rest	0.006	0.075	0	1
No_Front_rest	0.01	0.098	0	1
drvage	38.468	16.074	12	96
alcohol	0.064	0.245	0	1
drvsex	0.616	0.486	0	1
drvrest	0.042	0.201	0	1
animal	0.629	0.483	0	1
fixed	0.269	0.443	0	1
rollover	0.102	0.302	0	1
daylight	0.622	0.485	0	1
rainy	0.087	0.281	0	1
snowy	0.06	0.238	0	1
foggy	0.019	0.137	0	1
lanewid	11.758	0.733	9	16
R_shdr_wid	6.446	2.503	0	14
Spd_limt	0.912	0.283	0	1
Edg_line_wid	0.129	0.335	0	1
horizontal	0.078	0.269	0	1
aadt	3.715	2.351	0.1	25.9

CHAPTER 5

CRASH-LEVEL MODEL ESTIMATION RESULTS

5.1 Modeling Results

Table 6 shows the magnitude of relative impact of variables on each severity level. The coefficients of the estimated model can be interpreted as follows. A positive significant coefficient on a variable indicates that the variable is associated with a higher probability of being in that group choice relative to the reference group. The implication is that the probability of a crash at that level of severity is greater than the probability of placing it in the reference group. The negative sign means that the probability of a crash at that level of severity is smaller than the probability of placing it in the reference group. For example, "3.881" means that compared to animal collision, fixed object collision increase the log odds of fatality by 3.881. Of all the variables, number of female occupants, front seat occupants restraint use, snow weather fixed object collision, and rollover crash are significant. Driver sex and alcohol impaired driving are significant for most severity categories.

5.2 Marginal Effects

The marginal effects, defined as the derivative of the probability with respect to an independent variable, have substantive behavioral meaning, and are provided below to explain the role of each parameter. For continuous variables, a marginal effect is the influence a one unit change in an explanatory variable has on the probability of selecting a particular outcome. For dummy variables, the marginal effects are the derivative of the probability given a change in the dummy variable and thus represent the influence of a change in the variable upon the probability of choosing a given outcome. Table 7 shows the marginal effects outcomes in the accident-level model.

5.3 Fit of the Model

The Goodness of Fit test result is 0.254.

The expected probability for each severity level is calculated based on the mean level of all variables. We can see from Table 8 that although the result of goodness of fit test is not very high compared to least squares modeling in more controlled environments, the predicted probability for each severity level is very close to the actual condition. The predicted result for fatal crashes is not as accurate as the others probably due to the relatively small sample size of fatal crashes.

5.4 IIA Test Results

An important assumption of the multinomial logit model is that outcome categories for the model have the property of independence of irrelevant alternatives (IIA) (described above). Stated simply, this assumption requires that the inclusion or exclusion of categories does not affect the relative risks associated with the regressors in the remaining categories. Hausman test is used to test IIA property.

If the p-value is greater than 0.05 or chi2 statistic is actually negative, we might interpret this result as some evidence that we cannot reject the null hypothesis, that is, the IIA assumption holds. Table 9 indicates that for each of the categories being eliminated, the IIA property holds, so the multinomial logit model is good for estimating the accident-level model without violating its important assumptions.

Table 6 Modeling Results for Accident-Level Model

	Possible Injury		Non-incapacitating		Incapacitating		Fatal	
Variables	Coefficient	P-value	Coefficient	P-value	Coefficient	P-value	Coefficient	P-value
constant	-4.225	0.000	-2.447	0.001	-4.898	0.000	-6.771	0.004
male	-0.280	0.140	0.128	0.173	0.200	0.126	0.400	0.141
Female	0.399	0.000	0.302	0.000	0.260	0.032	0.500	0.050
No_Back_rest	0.307	0.709	0.051	0.922	1.048	0.047	0.394	0.659
No_Front_rest	1.668	0.022	1.878	0.000	2.137	0.000	3.091	0.000
drvage	-0.003	0.570	-0.002	0.451	0.010	0.008	0.030	0.001
alcohol	-0.224	0.509	0.932	0.000	1.034	0.000	1.871	0.000
drvsex	-0.458	0.002	-0.413	0.000	-0.242	0.060	0.019	0.953
drvrest	0.549	0.227	1.755	0.000	2.911	0.000	3.216	0.000
fixed	1.613	0.000	2.326	0.000	2.539	0.000	3.881	0.000
rollover	2.646	0.000	3.343	0.000	3.535	0.000	3.766	0.000
daylight	-0.061	0.689	-0.231	0.016	-0.301	0.024	-0.272	0.379
rainy	0.182	0.409	-0.072	0.616	-0.272	0.202	-0.830	0.179
snowy	-0.441	0.081	-0.764	0.000	-1.207	0.000	-2.327	0.024
foggy	-1.517	0.135	-0.212	0.512	0.280	0.453	-0.070	0.947
lanewid	0.061	0.511	-0.054	0.344	0.020	0.805	-0.172	0.334
R_shdr_wid	0.013	0.661	-0.043	0.021	-0.073	0.005	-0.030	0.622
Spd_limt	-0.022	0.931	0.131	0.410	0.266	0.240	-0.224	0.608
Edg_line_wid	-0.169	0.486	0.141	0.320	-0.046	0.826	-0.984	0.124
horizontal	0.114	0.641	0.219	0.127	0.024	0.903	0.493	0.198
aadt	-0.132	0.000	-0.054	0.006	-0.041	0.126	-0.024	0.673

Table 7 Marginal Effects of Independent Variables in the Accident- Level Model

variable	No injury	Possible Injury	Non-incapacitating	Incapacitating	Fatal
male	-0.005	-0.007	0.007	0.004	0.0004
Female	-0.030	0.009	0.016	0.005	0.0005
No_Back_rest	-0.044	0.007	0.000	0.037	0.0005
No_Front_rest	-0.337	0.054	0.177	0.092	0.014
drvage	-0.00006	-0.00006	-0.0001	0.0002	0.00003
alcohol	-0.099	-0.007	0.070	0.031	0.005
drvsex	0.038	-0.010	-0.023	-0.004	0.00007
drvrest	-0.351	0.001	0.138	0.198	0.014
fixed	-0.320	0.037	0.186	0.084	0.013
rollover	-0.609	0.070	0.364	0.164	0.011
daylight	0.020	-0.0009	-0.012	-0.006	-0.0003
rainy	0.005	0.005	-0.004	-0.005	-0.0007
snowy	0.056	-0.008	-0.031	-0.016	-0.001
foggy	0.021	-0.019	-0.010	0.008	-0.00005
lanewid	0.001	0.001	-0.003	0.0005	-0.0002
R_shdr_wid	0.003	0.0004	-0.002	-0.001	-0.00003
Spd_limt	-0.010	-0.0008	0.007	0.005	-0.0003
Edg_line_wid	-0.003	-0.004	0.008	-0.001	-0.0008
horizontal	-0.016	0.002	0.013	0.0001	0.0007
aadt	0.006	-0.003	-0.003	-0.0007	-0.00002

Table 8 Percentage Correct for Accident- Level Model

	Real	Predicted	%correct
No injury	81.84%	81.4%	99.46%
possible	2.65%	2.82%	93.97%
Nonincapacitating	9.22%	9.83%	93.79%
Incapacitating	5.21%	5.1%	97.89%
fatal	1.08%	0.85%	78.7%

Table 9 IIA Test Result for Accident- Level Model

Omitted severity level	Chi2	p-value	IIA property
possible	-5.8		holds
nonincapacitating	-3.78		holds
incapacitating	9.9	1	holds
fatal	-2.59		holds

CHAPTER 6

CRASH-LEVEL MODEL INTERPRETATION

6.1 Interpretation of Impacts of Variables

There were 21 variables identified by the crash-level model as contributing to the severity. In this part, the analysis conducted to investigate the impact of these factors on crash severity and the possible reason why there is such impact is reported.

We can see from modeling results that, the more occupants there are in the vehicle, the more likely at least one or more of them experience a more severe injury in single vehicle crashes. It is probable that when the driver or one occupant is severely injured, the others have higher probability to sustain similar level of severity. The results show that male drivers suffer less severe injury than female in a single vehicle accident. However, the crash is more likely to be fatal if the driver is male. The older the driver is, the higher the possibility that the crash is more severe. Although they are experienced and may be more careful while driving, once involved in an accident, they tend to more severely injured because of more fragile bodies. Undoubtedly, using restraint will decrease the injury severity. Figure 1 shows the percentage of drivers, front seat occupants, and back seat occupants not using restraint by different levels of accident severity. We can see from Figure 1 that with the increase of crash severity, the percentage of occupants not using restraint increases significantly. The percentage of drivers that do not use restraint in fatal crashes is about 35% while for PDO it is only 1%.

The crash severity decreases when it is dark. This is probably because people pay more attention to the road condition in the night and they tend to lower the speed when they are driving at night. The result was consistent with Eluru and Bhat (2007) though it was a little counterintuitive. It is surprising to see that crash severities are lower in rainy, snowy, foggy weather than in normal weather because when it is bad weather, there is shorter sight distance and less reaction time. However, drivers may drive more slowly when the weather is bad. In addition, 60% of crashes in bad weather are animal collisions, which tend to be less severe. Both fixed object collision and rollover will increase crash severity compared to an animal collision.

The wider the lane width, shoulder width, and edgeline width, the less severe crash it is. The shoulder and edgeline will help mitigate the impact when the car runs off the road and hits some fixed object. When there is a horizontal curve, the crash become more severe because it will affect steering control and reduce sight distance. With the increase of AADT, the crash severity decreases because speeds are slower as volumes increase. Figure 2 shows that higher speed limit will increase the probability of nonincapacitating injury and incapacitating injury, but decrease the probability of fatal crash, likely due to more forgiving designs with higher speed limits.

Alcohol is one of the most influential factors in determining whether a crash will be fatal. Figure 3 shows that male drivers are more likely to be involved in alcohol-related crashes for each level of crash severity. The proportion difference for male and female drivers involved in alcohol-related crashes for each severity level do not change much except fatal crashes—90% of fatal crashes involving an impaired driver have male drivers. It is also shown in Figure 4 that the younger the driver is, the more likely they will drive after drinking. The numbers of alcohol-impaired drivers decrease with the increase of age. Young drivers who are below 25 have the highest rate of alcohol-impaired driving and drivers who are below 45 are dominant in alcohol-related driving. Figure 5 shows the relationship between alcohol-impaired driving and crash severity. The more severe the crash is, the higher percentage of drivers involved in that type of crash that are under the influence of alcohol. For fatal crashes, there are 60% that include drivers under the influence of alcohol.

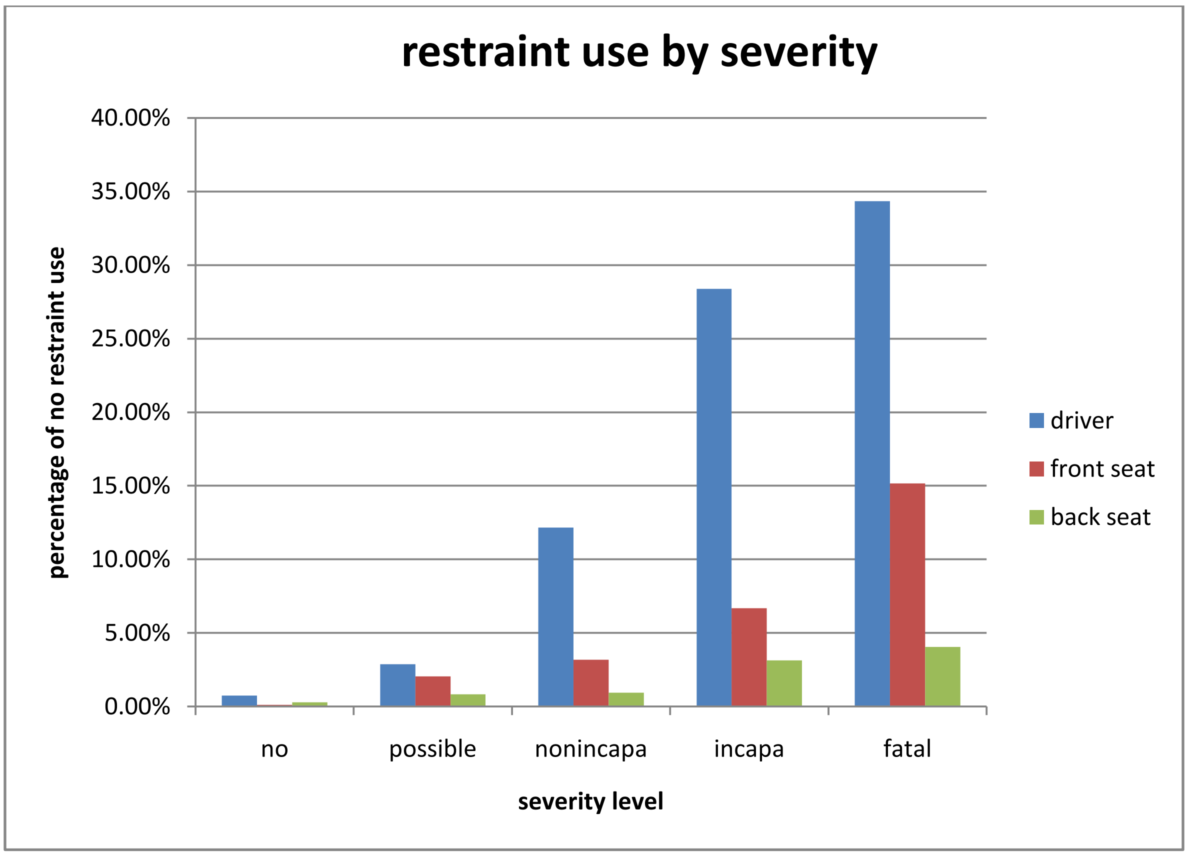

FIGURE 1 The percentage of not using restraint by different groups of people

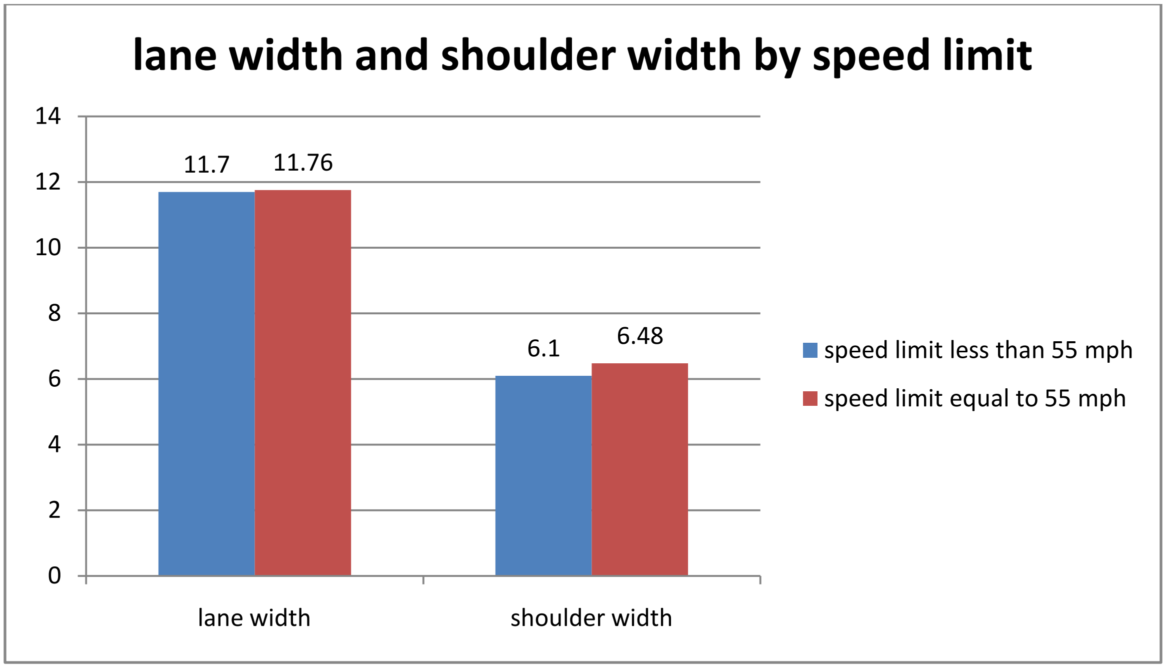

FIGURE 2 The relationship between speed limit and lane width and shoulder width

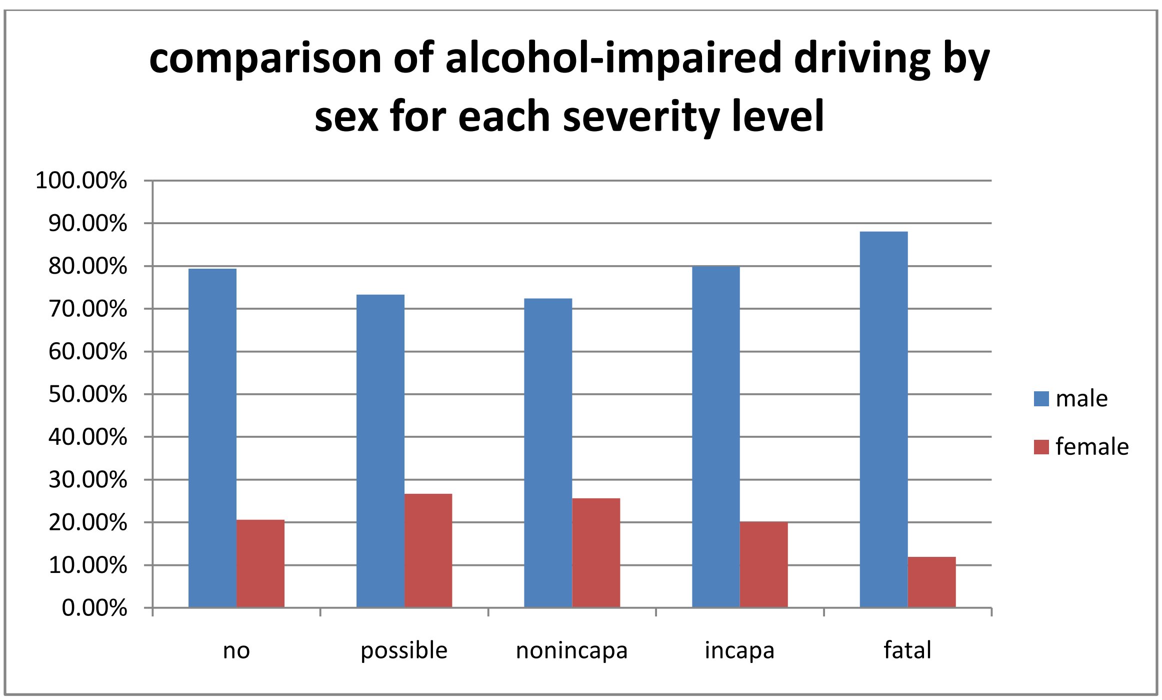

FIGURE 3 Alcohol-impaired driving by sex

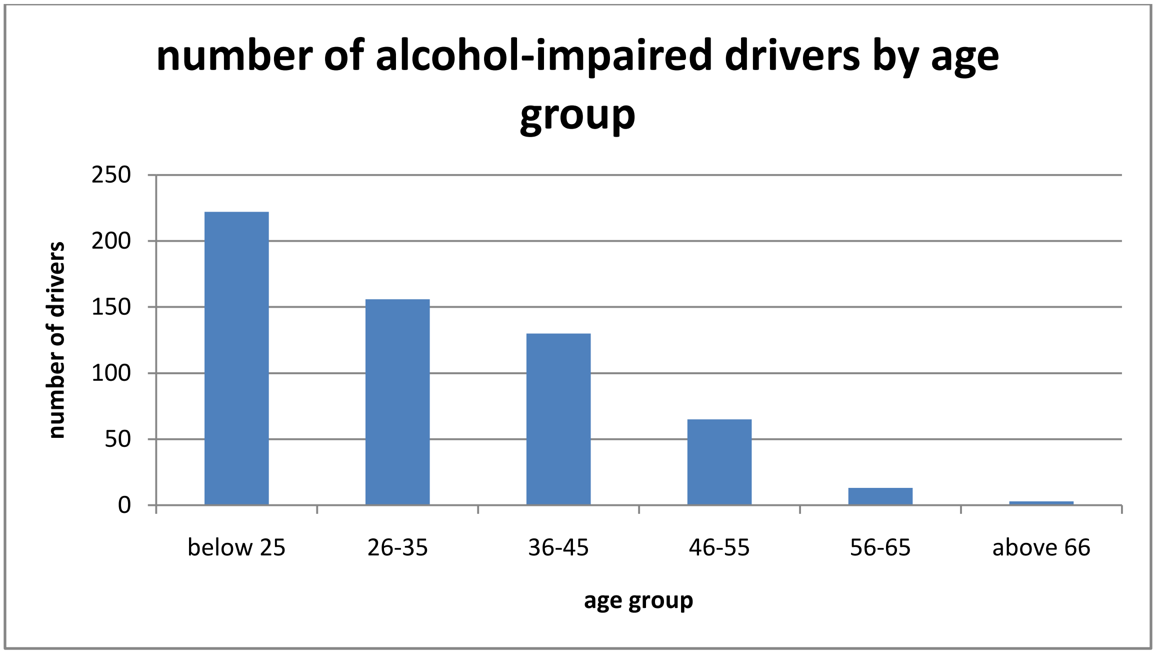

FIGURE 4 Alcohol-impaired driving by age

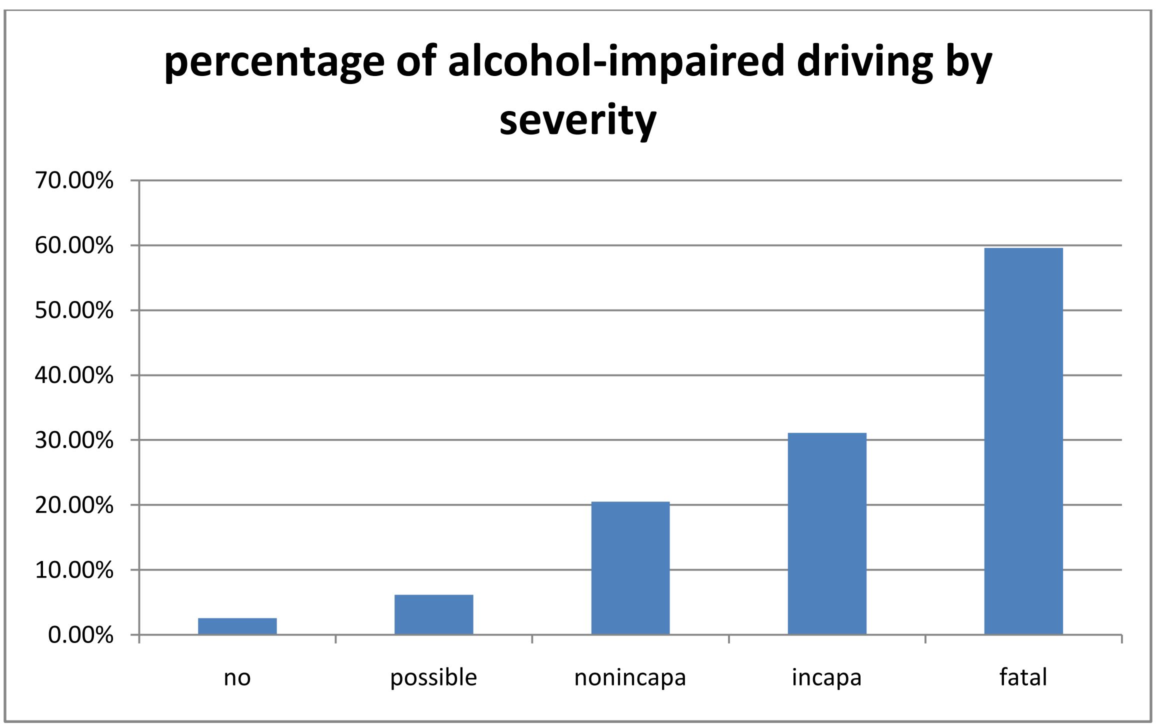

FIGURE 5 The percentage of alcohol-impaired driving for each severity leve

CHAPTER 7

OCCUPANT-LEVEL MODEL DATA

7.1 Variable Definitions and Descriptive Statistics

Table 10 shows all the contributing factors used in the occupant-level model. There are 18 independent variables altogether. Roadway variables and environmental factors are the same as for the accident-based model. For human factors, due to the reason that one row represents the condition of one person, occupant sex, occupant age, restraint use of occupants, whether it is alcohol-impaired driving, and seat position are included in the model. The variables are associated with each individual occupant, unlike the accident-level models where more general variables had to be developed (e.g., number of male occupants). Table 11 shows the summary statistics for all the variables. A total of 12243 observations were included in the model. The observations for which “unknown” was recorded were excluded from further analysis.

Table 10 Description of Variables Used in the Occupant-Level Model

variables	description
age	Age of occupant
sex	occupant sex (0=female 1=male)
norest	Restraint use for occupant (0=yes 1=no)
alcohol	Alcohol impaired driving (0=no 1=yes)
seatpos	Seat position for occupant (0=driver or front seat occupant 1=back seat occupant)
fixed	Fixed object collision (0=no 1=yes)
rollover	Rollover collision (0=no 1=yes)
daylight	Light condition (0=good 1=dark)
weather	Weather condition (0=normal 1=rainy 2=snowy 3=foggy)
lanewid	Lane width (ft)
R_shdr_wid	Right shoulder width (ft)
Spd_limt	Speed limit indicator (0=speed limit less than 55mph 1=speed limit equal to 55mph)
Edg_line_wid	Edgeline width (inches)
horizontal	Horizontal curve (0=no 1=yes)
aadt	average annual daily traffic (1000 vehicles per day)

Table 11 Summary Statistics for Variables in the Occupant-Level Model

Variable	Mean	Standard deviation	Min	Max
age	36.103	17.754	1	96
sex	0.579	0.494	0	1
norest	0.048	0.213	0	1
seatpos	0.098	0.298	0	1
alcohol	0.072	0.259	0	1
animal	0.631	0.483	0	1
fixed	0.269	0.443	0	1
rollover	0.1	0.3	0	1
daylight	0.636	0.481	0	1
rainy	0.088	0.284	0	1
snowy	0.062	0.242	0	1
foggy	0.018	0.135	0	1
lanewid	11.762	0.731	9	16
R_shdr_wid	6.443	2.509	0	14
Spd_limt	0.911	0.285	0	1
Edg_line_wid	0.128	0.334	0	1
horizontal	0.078	0.268	0	1
aadt	3.709	2.312	0.1	25.9

CHAPTER 8

OCCUPANT-LEVEL MODEL ESTIMATION RESULTS

8.1 Modeling Results

Table 12 shows the magnitude of relative impact of variables on each severity level. The coefficients of the estimated model can be interpreted as follows. A positive significant coefficient on a variable indicates that the variable is associated with a higher probability of being in that group choice relative to the reference group. The implication is that the probability of a crash at that level of severity is greater than the probability of placing it in the reference group. The negative sign means that the probability of a crash at that level of severity is smaller than the probability of placing it in the reference group. For example, the coefficient "2.404" indicates that the log odds of fatality increases 2.404 when driving impaired by alcohol.

8.2 Marginal Effects

The marginal effects, defined as the derivative of the probability with respect to an independent variable, have substantive behavioral meaning, and are provided below to explain the role of each parameter. We can see from Table 13 that older driver age, no restraint use, alcohol-impaired driving, front-seat occupant, fixed object collision and rollover collision rather than animal collision, higher speed limit, and horizontal curve are more likely to suffer from more severe accident.

8.3 Fit of the Model

The Goodness of Fit test result is 0.253 and the result is similar to the accident-level model.

The expected probability for each severity level is calculated based on the mean level of all variables. We can see from Table 14 that the predicted probability for each severity level is very close to the actual condition for "no injury" to "incapacitating injury." The predicted result for fatal crashes is not as accurate as the others, probably due to the relatively small sample size of fatal crashes. The percentage accuracy for the occupant-level model is close to that of the accident-level model, but the accident- level model has a better prediction for fatal crashes because the predicted severity proportion underestimates the fatality rate, and for the accident-level model, it only captures the most severe person in an accident, so the prediction for fatality is more appropriate for the accident-level model.

8.4 IIA Test Results

Hausman test is used to test IIA property.

If the p-value is greater than 0.05 or chi2 statistic is actually negative, we might interpret this result as some evidence that we cannot reject the null hypothesis, that is, the IIA assumption holds. Table 15 indicates that for each of the categories that are eliminated, the IIA property holds. This means that the multinomial logit model is good for estimating the data without violating its important IIA assumption.

Table 12 Single Vehicle Crashes Modeling by Occupant-Level Model

	Possible Injury		Non-incapacitating		Incapacitating		Fatal	
Variables	Coefficient	P-value	Coefficient	P-value	Coefficient	P-value	Coefficient	P-value
constant	-4.578	0.000	-3.020	0.000	-5.178	0.000	-8.173	0.001
age	-0.0002	0.960	-0.001	0.602	0.011	0.002	0.027	0.008
Sex	-0.672	0.000	-0.432	0.000	-0.373	0.002	-0.063	0.852
norest	1.144	0.000	1.782	0.000	2.813	0.000	3.658	0.000
Seatpos	-0.728	0.012	-0.401	0.01	-0.726	0.004	-0.616	0.347
alcohol	-0.217	0.466	0.951	0.000	1.085	0.000	2.404	0.000
fixed	1.749	0.000	2.282	0.000	2.673	0.000	3.139	0.000
rollover	2.707	0.000	3.354	0.000	3.647	0.000	3.037	0.000
daylight	-0.046	0.752	-0.212	0.018	-0.197	0.13	-0.519	0.130
Rainy	0.049	0.816	-0.114	0.398	-0.316	0.126	-1.025	0.168
snowy	-0.545	0.020	-0.727	0.000	-1.207	0.000	-14.387	0.978
Foggy	-1.586	0.117	-0.060	0.837	0.734	0.019	-13.184	0.987
lanewid	0.106	0.223	-0.054	0.739	0.032	0.676	-0.015	0.937
R_shdr_wid	0.009	0.736	-0.028	0.104	-0.059	0.019	0.001	0.982
Spd_limt	-0.252	0.258	0.121	0.406	0.275	0.217	0.157	0.780
Edg_line_wid	-0.159	0.492	0.282	0.029	0.095	0.625	-0.411	0.481
horizontal	0.026	0.912	0.200	0.136	0.189	0.304	0.267	0.519
aadt	-0.138	0.000	-0.047	0.009	-0.081	0.003	-0.184	0.018

Table 13 Marginal Effects for Occupant-Level Model

variable	No injury	Possible Injury	Non-incapacitating	Incapacitating	Fatal
age	-0.0001	0.000	-0.00008	0.0002	0.000
sex	0.041	-0.014	-0.021	-0.006	0.000
norest	-0.317	0.021	0.144	0.146	0.005
seatpos	0.037	-0.011	-0.017	-0.009	-0.0001
alcohol	-0.09	-0.006	0.068	0.026	0.002
fixed	-0.291	0.040	0.174	0.076	0.002
rollover	-0.591	0.071	0.366	0.152	0.001
daylight	0.015	-0.0006	-0.011	-0.003	-0.0001
rainy	0.009	0.001	-0.005	-0.004	-0.0002
snowy	0.049	-0.008	-0.028	-0.012	-0.0007
foggy	0.002	-0.017	-0.003	0.018	-0.0003
lanewid	-0.002	0.002	-0.001	0.0005	0.000
R_shdr_wid	0.002	0.0002	-0.001	-0.0009	0.000
Spd_limt	-0.004	-0.006	0.006	0.004	0.0004
Edg_line_wid	-0.014	-0.003	0.016	0.001	-0.00009
horizontal	-0.014	0.0002	0.011	0.003	0.00007
aadt	0.006	-0.002	-0.002	-0.001	-0.00005

Table 14 Percentage Correct for Occupant-Level Model

	Real	Predicted	%correct
No injury	83.38%	82.86%	99.38%
possible	2.43%	2.64%	92.04%
Non-incapacitating	8.72%	9.45%	92.27%
Incapacitating	4.61%	4.44%	96.31%
fatal	0.85%	0.61%	71.76%

Table 15 IIA Result for Occupant-Level Model

Omitted severity level	Chi2	p-value	IIA property
possible	Negative value		holds
Non-incapacitating	1.41	0.7	holds
incapacitating	2.36	0.98	holds
fatal	-0.15		holds

CHAPTER 9

OCCUPANT-LEVEL MODEL INTERPRETATION

9.1 Interpretation of Impacts of Variables

There were 17 variables identified by occupant-level model as contributing to the severity. In this part, the analysis conducted to investigate the impact of these factors on crash severity and the possible reason why there is such impact is reported.

The results show that male occupants suffer less severe injuries than female occupants. Compared to the accident-based model, occupant sex has less significant impact on fatal crashes than driver sex. Older occupants are more likely to experience more severe injuries when involved in a crash. Compared to the accident-level model, occupant age does not have as significant impact on fatal crashes as driver age. The drivers and front seat occupants (occupants who sit beside the drivers) are more likely to have more severe injuries than back seat occupants when there is an accident. We can see from the model results that if the drivers drink alcohol, both the drivers and the occupants in the vehicle have a higher chance to suffer from more severe injuries when involved in a crash.

Undoubtedly, using restraint will decrease the injury severity of any occupant. Figure 6 shows the percentage of occupants not using restraint for different levels of crash severity. We can see from Figure 6 that with an increase of crash severity, the percentage of occupants not using restraint is much higher. For fatal crashes, the percentage of occupants that do not use restraint is over 40%, while for PDO it is only around 1%.

The injury severity of occupants decreases when it is dark. This is probably because people pay more attention to the road when at night and they tend to lower the speed when they are driving at night. It is surprising to see that occupants experience less severe injuries in rainy, snowy, foggy weather than normal weather because when it is bad weather, there is shorter sight distance and less reaction time. However, drivers may drive slower in bad weather and 60% of bad weather crashes are animal collisions which tend to be less severe. Both fixed object collision and rollover will increase the injury severity of occupants compared to occupants involved in animal collisions.

There is not a consistent modeling result for the impact of lane width in the occupant-based model. The widening of the lane width increase the chance of an incapacitating injury while at the same time has insignificant negative impact on the chance of a fatality. The wider the shoulder width is, the less severe the occupant injury is. The shoulder will help mitigate the impact when the car runs off the road and hits some fixed object. The impact of edgeline is a little inconsistent for the occupant-level model. Compared to the previous model result, the occupant-based modeling shows that the widening of edgeline width will help to decrease fatal outcomes but increase the possibility of incapacitating injuries while the accident-level model shows that the widening of edgeline width will help to decrease both. The modeling results also show that higher speed limit will increase the probability of non-incapacitating injury, incapacitating injury, and fatality to the vehicle occupants. When there is horizontal curve, the injury becomes more severe for occupants because it will affect steering control and reduce sight distance. With the increase of AADT, the crash severity decreases because speeds tend to decrease when volumes increase.

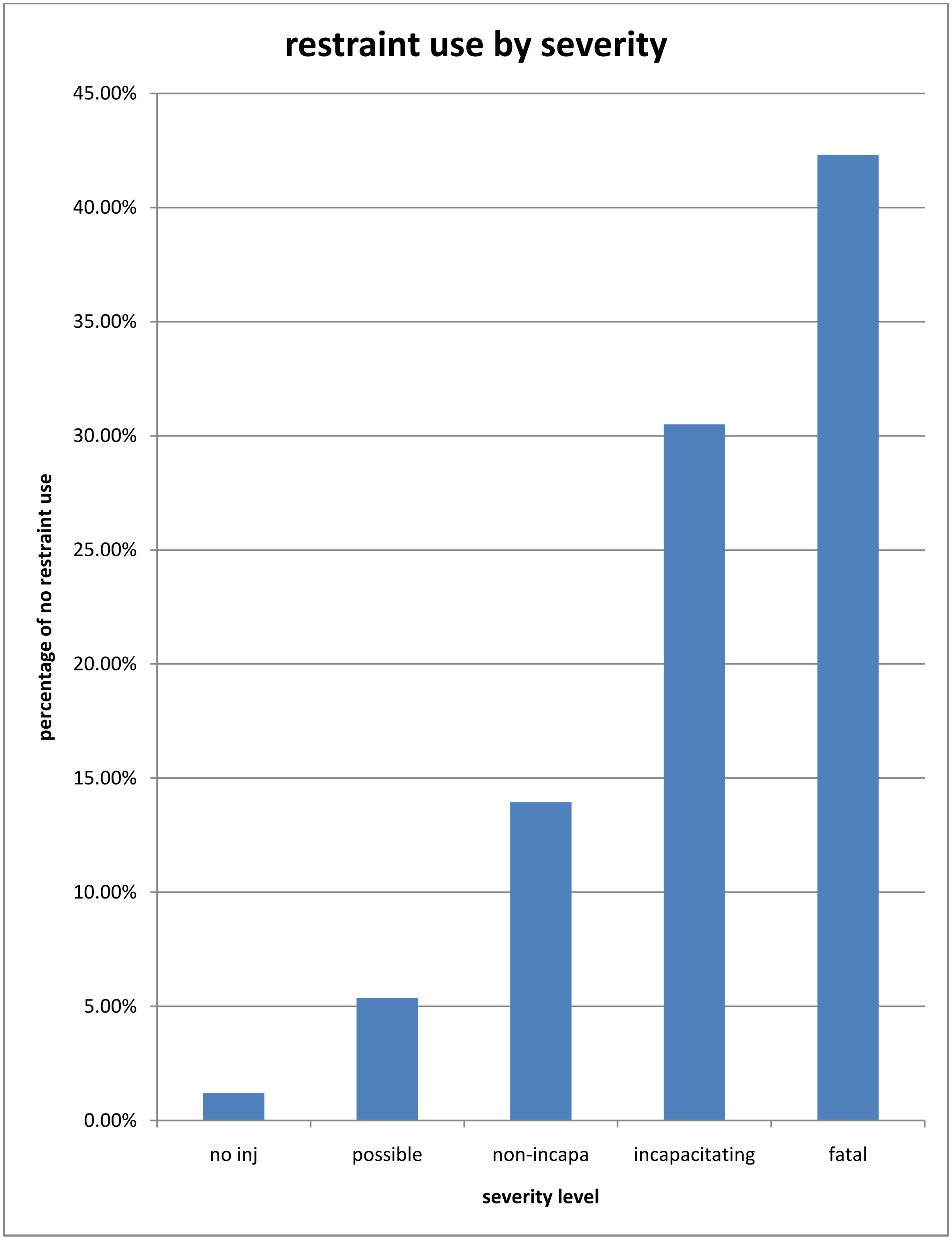

FIGURE 6 The percentage of not using restraint for each severity category

CHAPTER 10

CONCLUSIONS AND RECOMMENDATIONS

10.1 Conclusions

This study examined the effect of roadway, environmental, and human factors on injury severity in single vehicle collisions. A multinomial logit model for injury severity was estimated using the HSIS data set for rural, two-lane roadways. An accident-based model and an occupant-based model were estimated to provide different perspectives of crash analysis. The model parameters and the marginal effects of variables were used to examine the influence of roadway and crash characteristics on injury severity.

Although the magnitudes of impacts of independent variables on injury level changes slightly between the two models, there was consistent modeling results related to the impact of independent variables on severity level, probably due to the fact that large proportions of vehicles (77.45%) are single-occupant (only driver is included) so it is not a big difference whether it is an accident-level model or occupant-level model.

For roadway characteristics and environmental factors, lower crash severities are associated with wider lane widths, shoulder widths, and edge line widths, and larger traffic volumes. Crashes that occurred in darkness, or on rainy, snowy, and foggy days tended to be less severe than crashes occurred in clear, daylight conditions. Crashes occurring on horizontal curves, and on road segments with lower than a 55 mph speed limit were more severe. Fixed object collisions and rollover collisions were more severe than animal collisions. For human factors in the occupant-based model, older occupants, female occupants, an alcohol-impaired driver, and occupants not using restraint or improperly using restraint suffered from more severe injuries. The driver and front seat occupants were more likely to sustain more severe injuries than back seat occupants. The accident-based model shows similar results, but it indicates that males are more likely to be involved in fatal crashes because they are more likely to drive while impaired. It also shows that the more occupants there are in the vehicle, the more severe the crash is likely to be.

The modeling results show that alcohol-impaired driving and no restraint use have significant impacts on crash severity in the directions expected. More strict restrictions or policies that help to reduce the incidence of driving under the influence of alcohol and not using restraint should be an effective way to reduce crash severity. Also, providing or widening shoulder and edgeline would also help to mitigate the impact when the car runs off the road and hits some fixed object and further decrease the possibility of having more severe crashes.

Some of the findings were counterintuitive, such as those associated with the light condition and weather condition. It is probably because people tend to be more careful when driving in bad environmental conditions and due to the fact that over 60% of the crashes are animal collisions, which are usually less severe than fixed object or rollover collisions. So countermeasures based on these findings—such as providing warning signs in rainy or foggy days and street lights where it is too dark to have enough sight distance to enhance visibility—should be examined more deeply.

Goodness of fit test and the comparison of predicted distribution of severity with actual distribution of severity were used to assess the model fit. The goodness of fit test shows two models perform equivalent, and the accident-based model estimates the distribution of severity more accurately than the occupant-level model.

10.2 Recommendations

The methodological approach demonstrated for severity analysis needs to be applied to larger and more detailed databases, resulting in more precise safety insights to factors affecting crash severity. Also, the crash related data used were collected from police accident reports. The severity level is recorded according to the police officer judgments rather than using actual medical records so it may cause some statistical biases. And for some environmental factors, previous studies (Shinar et al. 1983) have shown the records of police officers to be unreliable (e.g., how to tell if it is a foggy day or how to distinguish rainy day from snowy day). The existence of missing data would also decrease the accuracy of the modeling results. The use of more professional records (e.g., records from medical experts for severity level) would permit an improvement of any potential biases that may result from using police officer judgments that are commonly available in national accident databases.

REFERENCES

Abdel-Aty, M., (2003). Analysis of Driver Injury Severity Levels at Multiple Locations Using Ordered Probit Models. *Journal of Safety Research 34(5), 597– 603*

Abdel-Aty, M., Keller, J., (2005). Exploring the Overall and Specific Crash Severity Levels at Signalized Intersections. *Accident Analysis and Prevention 37(3), 417–425*

Abdel-Aty, M., Abdelwahab, H., (2004). Modeling Rear-End Collisions Including the Role of Driver's Visibility and Light Truck Vehicles Using A Nested Logit Structure. *Accident Analysis and Prevention 36(3), 447–456*

Boufous, S., Finch, C., Hayen, A., Williamson, A., (2008). The Impact of Environmental, Vehicle and Driver Characteristics on Injury Severity in Older Drivers Hospitalized as a Result of A Traffic Crash. *Journal of Safety Research 39(1), 65–72*

Carson, J., Mannering, F., (2001). The Effect of Ice Warning Signs on Ice-Accident Frequencies and Severities. *Accident Analysis and Prevention 33(1), 99–109*

Chang, L., (2005). Analysis of Freeway Accident Frequencies:Negative Binomial Regression Versus Artificial Neural Network. *Safety Science 43(8), 541–557*

Chang, L., Mannering, F., (1999). Analysis of Injury Severity and Vehicle Occupancy in Truck- and Non-Truck-Involved Accidents. *Accident Analysis and Prevention 31(5), 579–592*

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

Eluru, N., Bhat, C., (2007). A Joint Econometric Analysis of Seat Belt Use and Crash-Related Injury Severity. *Accident Analysis and Prevention 39(5), 1037–1049*

Eluru, N., Bhat, C., Hensher, D., (2008). A Mixed Generalized Ordered Response Model for Examining Pedestrian and Bicyclist Injury Severity Level in Traffic Crashes. *Accident Analysis and Prevention 40(3), 1033–1054*

Farmer, C., (2003). Reliability of Police-Reported Information for Determining Crash and Injury Severity. *Traffic Injury Prevention 4(1), 38-44.*

Gkritza, K., Mannering, F., (2008). Mixed Logit Analysis of Safety-Belt Use in Single- and Multi-Occupant Vehicles. *Accident Analysis and Prevention 40(2), 443–451*

Holdridge, J., Shankar, V., Ulfarsson, G., (2005). The Crash Severity Impacts of Fixed Roadside Objects. *Journal of Safety Research 36(2), 139 – 147*

Khorashadi, A., Niemeier, D., Shankar, V., Mannering, F., (2005). Differences in Rural and Urban Driver-Injury Severities in Accidents involving Large-Trucks: An Exploratory Analysis. *Accident Analysis and Prevention 37(5), 910–921*

Kim, J., Ulfarsson, G., Shankar, V., Mannering, F., (2010). A Note on Modeling Pedestrian-Injury Severity in Motor-Vehicle Crashes With the Mixed Logit Model. *Accident Analysis and Prevention 42(6), 1751-1758*

Klop, J., Khattak, A., (1999). Factors Influencing Bicycle Crash Severity on Two-Lane, Undivided Roadways in North Carolina. *Transportation Research Record 1674(1), 78-85*

Kockelman, K., Kweon, Y., (2002). Driver Injury Severity: An Application of Ordered Probit Models. *Accident Analysis and Prevention 34(3), 313–321*

Kweon, Y., Kockelman, K., (2003). Overall Injury Risk to Different Drivers: Combining Exposure, Frequency, and Severity Models. *Accident Analysis and Prevention 35(4), 441–450*

Lee, J., Mannering, F., (2002). Impact of Roadside Features on the Frequency and Severity of Run-Off-Roadway Accidents: An Empirical Analysis. *Accident Analysis and Prevention 34(2), 149–161*

Lenguerrand, E., Laumon, J.L., (2006). Modelling the Hierarchical Structure of Road Crash Data—Application to Severity Analysis. *Accident Analysis and Prevention 38(1), 43–53*

Lord, D., Mannering, F., (2010). The Statistical Analysis of Crash-Frequency Data: A Review and Assessment of Methodological Alternatives. *Transportation Research Part A 44(5), 291-305*

Malyshkina, N., Mannering, F., (2009). Markov Switching Multinomial Logit Model: An Application to Accident-Injury Severities. *Accident Analysis and Prevention 41(4), 829–838*

Milton, J., Shankar, V., Mannering, F., (2008). Highway Accident Severities and the Mixed Logit Model: An Exploratory Empirical Analysis. *Accident Analysis and Prevention 40(1), 260–266*

O'Donnell, C.J., Connor, D.H., (1996). Predicting The Severity of Motor Vehicle Accident Injuries Using Models OF Ordered Multiple Choice. *Accident Analysis and Prevention 28(6), 739-753*

Pai, C., Hwang, K., Saleh, W., (2009). A Mixed Logit Analysis of Motorists' Right-Of-Way Violation in Motorcycle Accidents at Priority T-Junctions. *Accident Analysis and Prevention 41(3), 565–573*

Quddus, M., Noland, R., Chin, H. An Analysis of Motorcycle Injury and Vehicle Damage Severity Using Ordered Probit Models. *Journal of Safety Research 33(4), 445–462*

Savolainen, P., Mannering, F., (2007). Probabilistic Models of Motorcyclists' Injury Severities in Single- and Multi-Vehicle Crashes. *Accident Analysis and Prevention 39(5), 955–963*

Shankar, V., Mannering, F., (1996). An Exploratory Multinomial Logit Analysis of Single-Vehicle Motorcycle Accident Severity. *Joumal of Safety Research 27(3), 183-194*

Shankar, V., Mannering, F., Barfield, W., (1996). Statistical Analysis of Accident Severity on Rural Freeways. *Accident Analysis and Prevention 28(3), 391~401*

Shinar, D., Treat, J., McDonald, S., (1983). The Validity of Police Reported Accident Data. *Accident Analysis and Prevention 15(3), 175-191*

Ulfarsson, G., Mannering, F., (2004). Differences in Male and Female Injury Severities in Sport-Utility Vehicle, Minivan, Pickup and Passenger Car Accidents. *Accident Analysis and Prevention 36(2), 135–147*

Wang, X., Abdel-Aty, M., (2008). Analysis of Left-Turn Crash Injury Severity by Conflicting Pattern Using Partial Proportional Odds Models. *Accident Analysis and Prevention 40(5), 1674–1682*

Yamamoto, T., Shankar, V., (2004). Bivariate Ordered-Response Probit Model of Driver's and Passenger's Injury Severities in Collisions with Fixed Objects. *Accident Analysis and Prevention 36(5), 869–876*