A COMPREHENSIVE MIXED LOGIT ANALYSIS OF CRASH TYPE CONDITIONAL ON A CRASH EVENT

Alice Chu
The University of Texas at Austin
Department of Civil, Architectural and Environmental Engineering
301 E. Dean Keeton St. Stop C1761, Austin TX 78712
Phone: 512-471-4535, Fax: 512-475-8744, Email: aliceachu@utexas.edu

Chandra R. Bhat (corresponding author)
The University of Texas at Austin
Department of Civil, Architectural and Environmental Engineering
301 E. Dean Keeton St. Stop C1761, Austin TX 78712
Phone: 512-471-4535, Fax: 512-475-8744, Email: bhat@mail.utexas.edu
and
King Abdulaziz University, Jeddah 21589, Saudi Arabia

Ram M. Pendyala
Georgia Institute of Technology
School of Civil and Environmental Engineering
Mason Building, 790 Atlantic Drive, Atlanta, GA 30332-0355
Phone: 404-385-3754, Fax: 404-894-2278, Email: ram.pendyala@ce.gatech.edu

Venu M. Garikapati
Georgia Institute of Technology
School of Civil and Environmental Engineering
Mason Building, 790 Atlantic Drive, Atlanta, GA 30332-0355
Phone: 480-522-8067, Fax: 404-894-2278, Email: vgarikapati3@gatech.edu

ABSTRACT
This paper presents a comprehensive mixed logit model of crash types, where the crash type outcomes are defined by a combination of the nature of collision and the types of vehicles involved in the crash. While prior research in the highway safety field has largely studied and modeled crashes along specific dimensions and categories, this study attempts to model the influence of various explanatory factors on crash type probabilities in a comprehensive and holistic way. The model considers 20 different crash types (alternatives) simultaneously. Using the 2011-2013 General Estimates System (GES) crash database in the United States, this research effort presents a mixed logit model that characterizes the effects of weather and seasonal variables, temporal attributes, roadway characteristics, and driver factors on the probability of observing various crash types. The model reveals the competing influences of various factors on different crash outcomes and the presence of significant unobserved heterogeneity in the manner in which variables affect crash type probabilities. The model offers a framework for developing safety measures and devices that do not result in unintended consequences where a reduction in one crash type probability is met with an increase in another crash type probability.

Keywords: crash modeling, crash types, highway safety, mixed logit model, unobserved heterogeneity.

1. INTRODUCTION

Impressive improvements have been made in the United States over the past several years when it comes to transportation safety statistics. A comparison of crash statistics between 2000 and 2010 in the US shows fatalities per 100 million vehicle miles traveled reducing from 1.5 to 1.09, fatalities per 100,000 population reducing from 15.23 to 10.35, injured persons per 100 million vehicle miles traveled reducing from 116 to 77, and injured persons per 100,000 population reducing from 1,161 to 732 (National Highway Traffic Safety Administration or NHTSA, 2015). Despite these improvements, the total number of crashes continues to register an increase; there were 5.338 million crashes in 2011 and this number creeped up to 5.687 million crashes in 2013 – thus continuing to render the goal of "towards zero deaths" elusive (NHTSA, 2015).

In an effort to enhance safety, transportation agencies and auto manufacturers continuously strive to implement safety improvements and effective counter-measures that would reduce the risk of crashes or reduce the degree of severity of the crash. Passive safety measures such as roadway improvements, barriers, signage, and striping are often utilized by roadway agencies to alert drivers to safety hazards and enhance safety. Seatbelts and airbags are examples of passive safety devices that auto manufacturers have introduced in vehicles to reduce crash severity. More recently, auto manufacturers have been introducing active safety systems that utilize sensor based technologies (such as radar, video, laser, and global positioning systems) to incorporate collision-avoidance applications such as adaptive cruise control, forward collision warning, lane departure warning, blind spot detection, and parking assist. Active safety systems may be considered as the initial steps on the path to full-fledged connectivity and automation that the auto and technology industry hope to achieve over the next several decades.

The design and deployment of effective safety countermeasures (whether passive or active) requires a knowledge of, and the ability to model and quantify the effects of various roadway, environmental, vehicular, and driver factors that contribute to crashes of various types. In this context, it is desirable to understand and model how various factors influence crash occurrence while explicitly considering the type of crash and the type(s) of vehicle(s) involved. While there is a plethora of research examining the effects of variables on specific crash types (by type of collision, or by types of vehicles involved, or by type of location), to our knowledge, there is virtually no research study that takes a comprehensive approach to modeling crash occurrence by type of collision and types of vehicles involved. This paper aims to fill this gap in the literature by presenting a comprehensive model of crash types that considers these two key dimensions that characterize crashes.

In this study, crash records for 2011-2013 from the National Automotive Sampling System-General Estimates System (GES) crash database are used to estimate a mixed random parameter multinomial logit model of crash probability by collision and vehicle type. The model accounts for roadway attributes, weather and temporal attributes, and driver behavior. The mixed logit modeling approach is adopted to test for unobserved heterogeneity in the impacts of roadway characteristic variables on crash occurrence by type. The model system offers a holistic approach to identifying how various factors influence crash occurrence by collision and vehicle type, thus offering a mechanism to identify how counter-measures may simultaneously affect multiple crash types.

The remainder of this paper is organized as follows. The next section offers a brief overview of crash modeling. The third section offers a description of the data used in this research effort while the fourth section presents the modeling methodology. The fifth section offers a discussion of model estimation results. Concluding remarks are presented in the sixth and final section of the paper.

2. MODELING CRASH OCCURRENCE

Crashes are of many different types and involve a multitude of vehicle types. According to the NHTSA, the most common types of collisions in the United States are: rear-end collision with a motor vehicle in transport; angle collision with a motor vehicle in transport; collision with a fixed object (e.g., pole, tree); and collision with a non-fixed object (e.g., parked vehicle, pedestrian, bicyclist) (NHTSA, 2015). In the years 2011 through 2013, crash statistics in the United States show that about 68 percent of all crashes are collisions with motor vehicle in transport, 15 percent are collisions with a fixed object, and 14 percent are crashes with a non-fixed object. A little over two percent are non-collision events such as rollovers. As the severity of injury in a crash is often associated with the size and weight of vehicles involved, consideration of vehicle type is important in safety research. The NHTSA (NHTSA, 2015) defines six major vehicle type categories including passenger cars, light trucks, large trucks, motorcycles, buses, and other vehicles. Passenger cars and light trucks are involved in 95 percent of all crashes in the United States, which is not surprising given their prevalence on the nation's roadways – both in sheer volume and in vehicle miles traveled.

It is not possible to provide a comprehensive review of the transportation safety literature within the scope of this paper. The key aspect of the prior research that this paper attempts to address is that the literature has generally dealt with modeling and explaining the influence of various factors on crashes of a *specific type*, involving *specific classes* of vehicles, or occurring at *specific locations*. Neyens and Boyle (2007) examine the effects of distractions on crash occurrence among teenage drivers; they consider crash types (angular, rear-end, and collision with fixed object), but do not consider the types of vehicles involved in the crash. A study by Ghazizadeh and Boyle is another example of such a study examining the effects of distracted driving with consideration of crash type, but no consideration of vehicle type. Bham et al (2011) estimated a multinomial logit model of collision type, and included consideration of the number of vehicles involved in the collision (single-vehicle versus multi-vehicle collisions), but did not consider vehicle size/body/weight in their characterization of crashes. There are other studies that have explicitly considered vehicle type in crash analysis. Abdel-aty and Abdelwahab (2004) modeled rear-end collisions involving light trucks using a nested logit structure; thus their analysis is focused on a very specific collision and vehicle type. Yan et al (2005) used a logistic regression modeling approach to identify factors influencing rear-end collisions at signalized intersections, and included consideration of vehicle types (passenger car, passenger van, pickup/light truck, and large size vehicle) in their analysis. However, their analysis is limited to rear-end collisions at signalized intersections. Pai et al (2009) used the mixed logit modeling methodology to examine factors contributing to motorcycle accidents at priority T-junctions, while Haque and Chin (2010) focused their analysis on motorcycle accidents at signalized intersections. Schneider et al (2012) examined factors contributing to collisions involving an automobile and a motorcycle using crash record database for the State of Ohio. Crashes involving heavy vehicles have been studied quite extensively, given the concerns associated with injury severity when such vehicles are involved. Romo et al (2014) and Stevenson et al (2013) are examples of studies that focus on heavy vehicle crashes under specific circumstances. A study by Mitchell et al (2015) compares factors contributing to crashes involving novice and mature drivers in New South Wales, Australia; once again, while the study considers different collision types, crashes are not distinguished by vehicle type.

A common aspect that is pervasive in the safety literature is that crashes of specific types or involving certain vehicle types or occurring at specific locations are generally analyzed and modeled in isolation. This methodology has proven effective at identifying factors and counter-measures that influence specific crash types. However, this approach does not provide a holistic view of how factors and associated countermeasures can simultaneously and differentially affect crashes of diverse types and involving diverse vehicle types. This paper aims to build on the accumulated knowledge in the literature about factors that affect crashes of different types to provide a more holistic model of crash probability with explicit consideration of collision and vehicle types in the definition of the crash types considered. Moreover, the paper considers crashes that occur at any and all locations and times of the day, and does not focus on a specific subpopulation of transport system users. The comprehensive model system presented in this paper considers eight different collision types and three different vehicle types as follows:

- Collision Types
  - Collision with a stationary object
  - Collision with a parked vehicle
  - Collision with a pedestrian
  - Collision with a bicyclist
  - Head-on collision (includes both front-to-front and opposite direction sideswipe collisions)
  - Angle collision (vehicles that are not traveling in the same direction collide at an angle to one another)
  - Rear-end collision (includes both front-to-rear and same direction sideswipe collisions)
  - Rear-to-side collision (rear of one vehicle collides with the side of another vehicle)
- Vehicle Types
  - Light vehicles (automobiles, utility vehicles, and light trucks $\leq$ 4,536 kg Gross Vehicle Weight Rating)
  - Heavy vehicles (medium/heavy trucks > 4,536 kg Gross Vehicle Weight Rating)
  - Motorcycles, including motorcycles, mopeds, three wheeled motorcycle or mopeds, minibikes, and motor scooters

Prior research has shown that light vehicles, heavy vehicles, and motorcycles are each more prone to different types of crashes; by modeling all crashes comprehensively while explicitly accounting for collision and vehicle type, it will be possible to identify how explanatory factors affect different types of crashes within a unified holistic framework. For example, suppose there is a roadway characteristic that contributes to fewer angle collisions but increased rear-end collisions; the comprehensive model system presented in this paper will be able to identify this competing influence, and thus help identify countermeasures that may help reduce crashes without resulting in unintended consequences. This paper is intended to offer a comprehensive model of crash occurrence so that such a holistic perspective can be obtained when assessing the potential effectiveness of safety measures.

3. DATA

This study utilizes crash records from the 2011-2013 GES crash database. The crash records system is maintained by the NHTSA in the United States. The GES database contains a nationally representative sample of crashes reported to and recorded by the police. The crashes involve at least one motor vehicle traveling on a roadway resulting in death, injury, or property damage. The accident reports included in the sample are chosen from 60 areas that reflect the geography, roadway mileage, population, and traffic conditions of the United States. GES data collectors make weekly visits to approximately 400 police jurisdictions in the 60 areas across the United States, where they randomly sample about 50,000 police accident reports each year (NHTSA, 2015). It should be noted that, because GES data are estimates, differences across years may be attributed at least partly to the sampling process (and may not be reflective of an actual trend).

The database compiled for this research effort included 151,557 motor vehicle crashes reported over the three year period. As mentioned earlier, this study focuses on the four most common types of collisions that involve motor vehicles in transport (MVIT): head-on, angle, rear-end, and rear-to-side. The study also focuses on four non-MVIT collision types: collision with a stationary object, collision with a parked vehicle, collision with a pedestrian, and collision with a bicyclist. The three distinct vehicle body types are light vehicles, heavy vehicles, and motorcycles. Crashes that did not fall within any of these categories were excluded from the analysis. Crashes with incomplete or missing data on variables of interest were also excluded. Buses and vehicles in other category (farm equipment, golf carts, and construction equipment) were excluded from consideration as well. The final data set for use in this study includes 71,481 crashes.

Table 1 summarizes the distribution of crashes by crash type or alternative. Each record in the database corresponds to one reported motor vehicle crash, irrespective of the number of vehicles involved. A total of 20 crash alternatives are considered because some crash types that had very few observations had to be aggregated into a single alternative. For example, all collisions involving two heavy vehicles were combined into a single alternative. Also, collisions between a heavy vehicle and a motorcycle were excluded from the final data set due to a paucity of observations (even across the three years of observation).

TABLE 1 Distribution of Crash Types in Study Data Set

Crash Type	Frequency	Percentage
Collision between a light vehicle and a stationary object	21109	29.5
Collision between a light vehicle and a parked vehicle	2638	3.7
Collision between a light vehicle and a pedestrian	1772	2.5
Collision between a light vehicle and a bicyclist	1194	1.7
Collision between a heavy vehicle and a stationary object	1540	2.2
Collision between a heavy vehicle and a parked vehicle	143	0.2
Collision between a heavy vehicle and pedestrian/bicyclist	94	0.1
Collision between a motorcycle and a stationary object, parked vehicle, pedestrian, bicyclist, or another motorcycle	3013	4.2
Head-on collision between two light vehicles	2905	4.1
Angle collision between two light vehicles	9459	13.2
Rear-end collision between two light vehicles	17471	24.4
Rear-to-side collision between two light vehicles	189	0.3
Collision between two heavy vehicles	369	0.5
Head-on collision between a light and a heavy vehicle	361	0.5
Angle collision between a light vehicle and a heavy vehicle	1498	2.1
Rear-end or rear-to-side collision between a light vehicle and a heavy vehicle	4536	6.3
Head-on or angle collision between a light vehicle and a motorcycle	782	1.1
Rear-end or rear-to-side collision between a light vehicle and a motorcycle	816	1.1
Head-on or angle collision between multiple vehicles	615	0.9
Rear-end or rear-to-side collision between multiple vehicles	977	1.4
Total	**71481**	**100.0**

The dataset includes a host of explanatory variables that may be used in model specifications. Factors related to weather, time of day, roadway characteristics, and driver behavior are available in the dataset. Weather and temporal attributes include season, day of week, time of day, and weather conditions at time of crash. Roadway characteristics include intersection type, roadway alignment, traffic control devices, and trafficway description. Driver behavior variables include the violation type(s) charged to the driver. A very limited set of demographic variables such as age and gender are available, but were not included in the model specification of this paper as safety countermeasures are frequently designed to address all drivers regardless of age and gender. While it is certainly plausible that some interventions are targeted towards certain demographics (such as the elderly or teenage drivers), this paper focuses on the influence of non-personal factors on occurrence of crashes by type.

It is not possible to offer an exhaustive descriptive analysis of crash type by explanatory factor within the scope of this paper. The study involved an extensive exploratory and descriptive analysis of the data to understand how crash occurrence may be associated with the variables available in the dataset. An illustrative example of the descriptive statistics is presented in Table 2. This table shows the distribution of collisions of light vehicles with a non-MVIT by explanatory factor. Similar tables were constructed for all other collision types to see how the distributions varied by explanatory factor. These distributions helped inform the model specifications tested and adopted in this paper.

TABLE 2 Descriptive Statistics of Light Vehicle Collisions with a Non-Motor Vehicle in Transport

	Light Vehicle Collision with Non-Motor Vehicle in Transport			
	Stationary Object 21109	Parked Vehicle 2638	Pedestrian 1772	Bicyclist 1194
Weather/Temporal Attributes				
Season				
Autumn (base)	25.8%	25.3%	27.7%	28.0%
Winter	28.7%	25.9%	28.6%	17.2%
Spring	22.9%	23.8%	24.9%	24.0%
Summer	22.6%	25.1%	18.7%	30.8%
Weather				
Clear (base)	62.2%	74.1%	72.6%	78.6%
Cloudy	16.9%	14.9%	14.5%	15.5%
Rain or Drizzle	14.4%	7.6%	11.7%	5.4%
Snow	5.1%	2.7%	0.8%	0.1%
Fog or Smog	0.8%	0.3%	0.3%	0.2%
Severe Wind/Sand/Other	0.5%	0.3%	0.1%	0.3%
Day of Week				
Weekday (base)	66.7%	64.1%	75.3%	74.5%
Weekend	33.3%	35.9%	24.7%	25.5%
Time of Day				
12am-7am	29.1%	28.1%	14.6%	8.1%
7am-10am	16.0%	13.4%	14.3%	20.9%
10am-4pm (base)	23.3%	25.7%	24.5%	34.0%
4pm-8pm	18.2%	17.9%	31.8%	27.8%
8pm-12am	13.4%	14.9%	14.7%	9.1%
Roadway Characteristics				
Intersection Type				
Non-intersection (base)	93.1%	94.8%	55.0%	39.9%
Four-way Intersection	3.3%	2.3%	34.4%	42.9%
T-Intersection	3.1%	2.7%	9.7%	15.8%
Y-Intersection	0.3%	0.1%	0.4%	0.3%
Traffic circle, Roundabout or L-Intersection	0.2%	0.2%	0.6%	1.1%
Roadway Alignment				
Straight (base)	73.2%	91.2%	96.4%	97.4%
Curved	26.8%	8.8%	3.6%	2.6%
Traffic Control Device				
No Controls (base)	97.4%	98.8%	69.2%	62.1%
Traffic Signal	2.5%	1.1%	30.5%	37.4%
Flashing Signal	0.1%	0.0%	0.2%	0.5%
Other	0.0%	0.1%	0.1%	0.1%
Trafficway Description				
Two-way, Not Divided (base)	54.0%	75.2%	55.0%	57.2%
Two-way, Divided, Unprotected Median	11.5%	7.8%	16.1%	15.7%
Two-way, Divided, Positive Median Barrier	25.9%	8.0%	13.8%	14.8%
One-way Traffic	2.2%	6.3%	6.8%	4.4%
Two-way, Undivided, Continuous Left-Turn Lane	2.0%	2.0%	7.6%	7.4%
Entrance/Exit Ramp	4.5%	0.7%	0.6%	0.6%
Violation Charged to Driver in Vehicles				
None	65.9%	55.4%	77.8%	72.3%
Reckless Offense	4.0%	6.2%	3.0%	5.6%
Impairment Offense	4.3%	7.2%	0.6%	0.3%
Speed-related Offense	4.5%	3.0%	0.6%	0.2%
Rules of the Road	2.5%	2.7%	8.4%	12.1%
License, Registration, Equipment Violations	8.5%	9.9%	4.7%	4.7%
Multiple Violations Charged to Driver	10.4%	15.5%	5.0%	4.9%

Table 2 provides an initial glimpse into how collisions between a light vehicle and a non-MVIT may be associated with various explanatory factors. A majority of such collisions occur on weekdays, presumably because weekdays account for a larger portion of light vehicle travel.

However, what is interesting to note is that the proportion of such crashes is even higher for collisions involving pedestrians and bicyclists (on weekdays), perhaps because of the greater prevalence of such non-motorized mode users on weekdays, the higher levels of traffic congestion on weekdays, and the rush of commute traffic when individuals may be in a rush to get to work, school, or home in a timely manner. Pedestrians and bicyclist involved collisions appear to occur more in the 4-8 PM hours, presumably due to diminished visibility during these hours and the prevalence of pedestrians and bicyclists (for recreational and utilitarian travel purposes) in the evening hours. The weather-related statistics indicate that a large proportion of crashes occur on clear days, although collisions with a stationary object show a higher percent (relative to other collision types) during rain and snow – an observation that is consistent with expectations.

Within the intersection-related crashes, bicyclists and pedestrians are quite vulnerable at four-way intersections and T-intersections, possibly due to the multiple conflict points prevalent at such intersections. Curved roads and curved intersection approaches are associated with light vehicle crashes involving a stationary object (26.8 percent is considerably higher than other percentages in that row). A rather large percent of pedestrian- and bicyclist-involved crashes occur when a traffic signal is present (note that traffic signals can occur at non-intersection locations too, such as a crosswalk at a mid-block of a roadway); again, the presence of multiple conflict points and non-adherence to traffic signal indications may contribute to these high percentages (30.5 percent for pedestrians and 37.4 percent for cyclists). Two-way divided roadways (that are likely to be wider to cross and operate at higher speeds) are associated with a higher prevalence of pedestrian and bicyclist involved crashes. In most crashes, drivers have not been charged or cited. In the case of collisions involving a stationary object or a parked vehicle, however, drivers are cited more than in collisions involving pedestrians and bicyclists.

The dataset was analyzed and descriptive statistics such as those in Table 2 were studied carefully to help identify trends in the data that could help inform the model specifications adopted in this paper.

4. METHODOLOGY

In this research study, a mixed random parameter multinomial logit (MMNL) approach is adopted for modeling crash types, which are categorized by both the manner of collision and the vehicle type(s) involved. Each case in the dataset represents a reported motor vehicle crash that occurred between the years of 2011 to 2013 in the United States. Therefore, the decision maker, or in this case, the crash type involving the driver(s) of the motor vehicle(s) can be observed to be one among 20 distinct crash types (i.e., 20 alternatives). Each alternative represents one combination of crash type (e.g., rear-end or with a stationary object) and vehicle type (i.e., light vehicle, heavy vehicle, or motorcycle).

The likelihood of crash type $j (j = 1,2,...,J)$ for driver $q (q = 1,2,...,Q)$ can be specified as:

$$\begin{aligned} U_{qj} &= \boldsymbol{\beta}'_q \boldsymbol{x}_{qj} + \varepsilon_j \\ &= (\boldsymbol{b} + \tilde{\boldsymbol{\beta}})' \boldsymbol{x}_{qj} + \varepsilon_j \end{aligned} \tag{1}$$

where $\boldsymbol{x}_{qj}$ is a column vector of explanatory variables that is related to weather/temporal attributes, roadway characteristics, and driver behavior factors. $\boldsymbol{b}$ is a column vector of coefficients representing the mean effect of explanatory variables. $\tilde{\boldsymbol{\beta}}$ is a column vector of coefficients representing the random effect. Further, $\tilde{\boldsymbol{\beta}}$ is assumed to be distributed normal and uncorrelated across parameters, i.e., $\tilde{\boldsymbol{\beta}} \sim \text{MVN}(\mathbf{0}, \mathbf{v})$. Finally, ε_j is the random error term which is distributed independently and identically and has an extreme value distribution.

Thus the probability of observing a crash of type j for driver q can be written as (Revelt and Train, 1998):

$$P_{qj} = \int_{\tilde{\boldsymbol{\beta}}} \frac{\exp[(\boldsymbol{b} + \tilde{\boldsymbol{\beta}})' \boldsymbol{x}_{qj}]}{\sum_{\forall j} \exp[(\boldsymbol{b} + \tilde{\boldsymbol{\beta}})' \boldsymbol{x}_{qj}]} f(\tilde{\boldsymbol{\beta}} \mid \mathbf{0}, \mathbf{v}) d\tilde{\boldsymbol{\beta}} \tag{2}$$

As the integral in equation (2) does not have a closed form solution, a maximum simulated likelihood approach is used to obtain the probability of a crash. In the simulated likelihood approach, equation (2) may be written as:

$$P_{qj} = \frac{1}{R} \sum_{r=1}^{R} \frac{\exp[(\boldsymbol{b} + \mathbf{v}\boldsymbol{w}_r)' \boldsymbol{x}_{qj}]}{\sum_{\forall j} \exp[(\boldsymbol{b} + \mathbf{v}\boldsymbol{w}_r)' \boldsymbol{x}_{qj}]} \tag{3}$$

where $\boldsymbol{w}_r$ is a column vector of Halton draws. In this study, 250 Halton draws are used in the maximum simulated likelihood estimation approach. Details about the maximum simulated likelihood estimation approach, and the use of Halton draws to compute choice probabilities, may be obtained in Bhat (2000) and (2001).

5. MODEL ESTIMATION RESULTS

A mixed multinomial logit model (MMNL) was estimated on the data set of 71,481 crash records considering 20 distinct crash alternatives. The mixed logit model was adopted to account for potential heterogeneity in the effects of certain variables on crash types by nature of collision and vehicle involvement. The methodology was coded in the GAUSS programming language and model estimation was accomplished using the simulated maximum likelihood approach. For convenience, the base category for model estimation was set as the collision between a light vehicle and a stationary object. It should be noted that the model indicates the probability of involvement in one of the crash types, given that a crash has occurred. The model does not purport to explain the propensity of crash occurrence, crash frequency, or crash/injury severity. The sole purpose of the model is to determine the influence of various factors on the likelihood of different crash types (conditional on a crash event) under various conditions.

This section presents a summary of key illustrative findings based on the model estimation results. Due to the size of the model estimation results table, it is not possible to provide the entire set of estimation results within the scope of this paper. For illustrative purposes, the model estimation results are furnished in their entirety for two specific alternatives in Table 3. The complete estimation results are available from the authors upon request. The descriptive write-up highlights results seen in Table 3 as well as results associated with other crash alternatives not shown in Table 3. The write-up is organized with respect to the various sets of attributes considered in the model specification. The base alternative in the mixed logit model estimation results in Table 3 corresponds to light vehicle collisions with a stationary object.

The constants in the model reflect that the probability of virtually all crash types is lower than that of the base alternative – namely, the collision of a light vehicle with a stationary object. The one exception, where a positive constant is noted, is the rear-end collision between two light vehicles. It appears that collisions are likely to be more of the rear-end type (involving two light vehicles) than other types of collisions (note that the constants have a meaningful interpretation because all exogenous variables are categorical).

TABLE 3 Illustrative Mixed Logit Model Estimation Results for Two Crash Type Alternatives

	Light Vehicle Collision with Non-MVIT			Collision Between Two Light Vehicles			
Variable	Parked Veh	Pedestrian	Bicyclist	Head-on	Angle	Rear-End	Rear-Side
Constant	-1.4210	-2.7800	-3.2800	-1.8180	-1.2340	0.1010	-4.6920
Weather/Temporal Attributes							
Season (Base: Autumn)							
Winter	--	--	--	--	0.0490	--	--
Spring	--	--	--	--	0.0790	0.0790	--
Summer	--	--	--	--	--	--	--
Weather (Base: Clear, Fog or Smog[1], Severe Crosswind or Blowing Sand or Other Weather[1])							
Cloudy	-0.2240	-0.2240	-0.2240	-0.1660	-0.2240	-0.2240	--
Rain/Drizzle	-0.5790	--	--	-0.1660	--	-0.2880	--
Snow	-0.5790	-0.9240	-0.9240	-0.1660	-0.9240	-0.9240	--
Day of Week (Base: Weekday)							
Weekend	0.1770	-0.4130	--	-0.4130	-0.4130	-0.4130	--
Time of Day (Base: 10am-4pm)							
12am-7am	--	-0.9150	-1.9350	-0.9150	-1.9350	-1.9350	-1.9350
7am-10am	--	--	--	-0.3280	-0.3280	-0.3280	--
4pm-8pm	--	0.4060	0.4060	--	-0.1430	-0.1430	--
8pm-12am	--	--	-0.8640	-0.8640	-0.8640	-1.4200	--
Roadway Characteristics							
Intersection Type (Base: Non-Intersection)							
4-way Intersection							
Mean	--	2.2810	2.2810	1.0270	2.2810	1.0270	1.0270
Std Dev	--	--	--	0.2340	--	0.2340	0.2340
T-Intersection							
Mean	--	0.7580	1.2190	--	1.2190	--	--
Std Dev	--	--	0.8690	--	0.8690	--	--
Y-Intersection	--	0.7580	--	--	--	--	--
Traffic Circle	0.2470	0.7580	0.2470	0.2470	--	--	0.2470
Roadway Alignment (Base: Straight)							
Curved							
Mean	-1.6780	-1.6780	-1.6780	--	-1.6780	-1.6780	--
Std Dev	--	--	--	--	--	--	--
Traffic Control Device (Base: No Controls, Other[1])							
Traffic Signal							
Mean	-1.5500	1.5910	1.5910	1.5910	1.5910	1.5910	1.5910
Std Dev	--	--	--	--	--	--	--
Flashing Signal	--	--	--	0.6990	0.6990	0.6990	0.6990
Trafficway Description (Base: Two-way, Not Divided)							
Two-way, Divided, Unprotected Median							
Mean	-0.6710	0.0850	0.0850	-0.6710	0.0850	0.0850	--
Std Dev	--	0.9810	0.9810	--	0.9810	0.9810	--
Two-way, Divided, Positive Median Barrier							
Mean	-1.5180	-0.4830	--	-1.5180	-0.4830	0.0700	-0.4830
Std Dev	--	--	--	--	--	1.4670	--
One-way Traffic							
Mean	0.7420	0.7420	0.7420	-0.9380	--	0.7420	0.7420
Std Dev	0.3080	0.3080	0.3080	--	--	0.3080	0.3080
Two-way, Undivided, Left-Turn Lane							
Mean	--	0.8520	0.8520	0.8520	0.8520	0.8520	0.0490
Std Dev	--	0.6340	0.6340	0.6340	0.6340	0.6340	--
Entrance/Exit Ramp							
Mean	-1.5110	-1.5110	-1.5110	-1.5110	-1.5110	0.6280	--
Std Dev	--	--	--	--	--	0.3110	--
Driver Behavior (Base: None)							
Reckless	--	--	--	0.5870	0.5870	0.5870	--
Impairment	--	--	--	0.5480	-0.1830	-0.1830	--
Speed-related	--	--	--	-0.6520	-0.6520	1.2400	--
Rules of Road	--	--	--	1.5920	1.5920	--	--
Lic/Regn/Equip	--	--	--	0.4920	0.4920	0.4920	--
Multiple Violat	--	--	--	0.4540	0.4540	0.4540	--

[1]Estimated coefficients statistically insignificant at 95 percent confidence level

Season and Weather

Estimation results (not shown in Table 3) suggest that crashes involving motorcycles are less likely to occur in the winter compared to other seasons of the year. This finding, consistent with that reported by Branas and Knudson (2001), may be attributed to the fact that motorcycle riding is largely a fair-weather activity; fewer motorcycles on the roads during winter months will naturally lead to fewer crashes involving motorcycles. Likewise, it was found that the probability of crashes involving motorcycles is higher in the spring and summer months. The likelihood of two light vehicles getting into an angle collision was found to be higher in the winter, compared to fall and summer (see Table 3) – a finding that may be attributed to the more adverse driving conditions in winter that could contribute more to angle crashes. Some coefficients are statistically significant, but not necessarily easily explained. For example, the higher propensity for angle and rear-end collisions in the spring (as signified by the positive coefficient of 0.0790) warrants further research.

An interesting finding is that all crash types are less likely to occur under adverse weather conditions (such as rain/drizzle, cloudy, and snow) *in comparison to the collision involving a vehicle striking a stationary object.* This finding is consistent with expectations. During inclement weather, roads are slippery and visibility is diminished; given a crash occurs, it is more likely to be of the type where a driver skids off the road and strikes an object. That is the most likely crash type under such conditions. Another possible explanation is that there are more vehicles on the roadways during clear days, thus making it plausible to expect other types of crashes (for example, two light vehicles striking one another) to be more likely to occur on such fair weather days. Previous research (Kilpeläinen and Summala, 2007) has found that drivers are more cautious when driving under adverse weather conditions; this is another reason why negative coefficients are associated with weather conditions in Table 3.

Temporal Attributes

The weekend days are associated with a lower likelihood of collisions between two light vehicles and collisions between a light vehicle and pedestrians. On weekend days, travelers are likely to be more relaxed, traffic congestion is likely to be less severe, and travelers are likely pursuing more leisure-type activities. For these reasons, the lower propensity for such crash types is quite reasonable. The propensity for a crash type where a light vehicle strikes a parked vehicle is higher, however, on weekends; this is likely due to the larger number of social recreational and shopping trips on weekends where travelers are undertaking parking maneuvers to a larger degree than on weekdays. This finding is consistent with that reported by Bham et al (2011) who found that the risk of single-vehicle collisions is higher on weekends, whereas the risk of multi-vehicle collisions is higher on weekdays. They believe that this is due to lower traffic volumes on weekends. The lower prevalence of trucks and heavy vehicles on the roadways during weekends contributes to a lower propensity for crash types involving heavy vehicles on weekend days (not seen in Table 3). Crashes involving motorcycles were found to be more likely on weekends, a finding that is consistent with the higher level of recreational motorcycle riding on weekends.

In general, all crash types are least likely to occur between 12 midnight and 7 AM when there is less traffic on the roadways. Between 4 PM and 8 PM, there is a higher likelihood of crashes involving a light vehicle colliding with a pedestrian or bicyclist. These positive coefficients (0.4060) reflect the higher propensity for pedestrian and bicycle crashes to occur in the afternoon and evening peak hours when individuals are pursuing a variety of activities, traffic volumes are high, and children may be pursuing after school activities. Cinnamon et al (2011) conducted research on several high-incident intersections in Vancouver, Canada and found that pedestrian and motor vehicle violations occurred to a higher degree in the afternoon peak period of 4 PM to 6 PM. Collisions involving two light vehicles are most likely to occur in the midday (10 AM to 4 PM) as evidenced by the negative coefficients on all other time of day variables.

Roadway Characteristics

A variety of roadway characteristics were included in the model specification and tested for random effects to capture potential unobserved heterogeneity that may be present in the way in which a roadway characteristic affects crash type probabilities. In general, it is seen that crashes of various types are more likely to occur at intersections, including four-way intersections, T-intersections, Y-intersections, and roundabouts. The larger number of conflict points and approaches at intersections increases the propensity for crashes that involve pedestrians and bicyclists, and collisions of various types that involve two vehicles. Unobserved heterogeneity is significant at four-way intersections for the head-on and rear-end collisions involving two light vehicles. This is likely due to the presence of unobserved factors (not contained in the data set) that affect crash type propensity; for example, traffic volume, geometric configuration of approach and turning lanes, adjoining land uses, and turning movements affect crash type probabilities. All of these factors remain unmeasured and hence the effect of a four-way intersection on crash type probabilities exhibits a significant amount of heterogeneity. Findings in this paper corroborate results reported by Niewoehner and Berg (2005) and Pai and Saleh (2008), who note that crashes of various types are more prevalent at intersections. The absence of clear pedestrian crosswalks and the inability to adequately assess vehicular maneuvers at Y-intersections contributes to greater pedestrian-involved crashes at such intersections (California Department of Transportation, 2010).

An examination of roadway alignment effects suggests that crashes of various types are less likely to occur on curved roads and curved intersection approaches, *in comparison to the base alternative of crashes that involve a vehicle striking a stationary object*. Wang et al (2013) note that road curvature may be beneficial from a safety standpoint as drivers slow down when maneuvering around curves, tend to be more alert and careful when navigating curves, and are less likely to be bored and sleepy when their path involves curves. If a crash does occur, then it is more likely to be one where the driver runs off the road and strikes a stationary object (Bham et al, 2011). Motorcycle collisions, on the other hand, are more likely to occur on curved roads, a finding that is consistent with expectations (not shown in Table 3).

Traffic signals are likely to be present at intersections and locations where traffic volumes are high, the number of conflict points is high, and safety hazards exist. As such, it is not surprising that the presence of a traffic signal is associated with a higher crash probability for crashes of various types, except for the crash type where a light vehicle collides with a parked vehicle. As parked vehicles are not likely to be in the vicinity of a signal, this finding is consistent with expectations.

Descriptors of the trafficway influence crash type probabilities significantly. Consider a two-way roadway with an unprotected median. The propensity for angle, rear-end, and pedestrian/bicyclist involved crashes is higher, as evidenced by the positive (mean) coefficient. The standard deviation is also statistically significant, indicating the presence of unobserved heterogeneity. Bicyclists and pedestrians may cross such a trafficway mid-block because of the presence of the median. When they do this, they are more likely to be involved in a crash as vehicular drivers are not expecting such road users to be encountered mid-block. This phenomenon may contribute to an *increase* in crash propensity for bicyclists and pedestrians on such trafficways. On the other hand, a median may serve as a protective shelter for pedestrians and bicyclists, thus *decreasing* the crash propensity. In other words, there may be considerable variation in how this particular trafficway configuration affects crash propensity for bicyclists and pedestrians; the mixed logit model offers a way to capture the unobserved heterogeneity or variation in the impacts of this trafficway configuration variable. Rear-end collisions show a greater propensity to occur on divided highways; this may appear counter-intuitive at first, but is consistent with results reported by Yan et al (2005) who note that such roadways may see higher rear-end collisions because of higher traffic volumes. Rear-end collisions are also more likely at entrance and exit ramps (compared to two-way undivided roads), presumably because of the speed variability on ramps; the significant heterogeneity for this variable suggests that such unobserved factors influence rear-end crash propensity.

As expected, head-on collisions are less likely to occur on one-way streets. However, crashes involving a non-MVIT and rear-end crashes are more likely to occur on one-way streets. As one-way streets are more likely to be encountered in dense central city areas, it is not surprising to see higher crash propensities involving a non-MVIT (parked vehicles on the side of the road, pedestrians, and bicyclists are likely to be present in larger numbers in such locations). The significant heterogeneity term suggests that the configuration of the one-way street, the surrounding land use, and the provision of sidewalks, crosswalks, and bicycle lanes may constitute unobserved factors that contribute to variation in how one-way streets affect crash type probabilities.

Driver Behavior

Compared to the base case involving a collision of a vehicle with a non-MVIT object, drivers are more likely to be charged in collisions that involve multiple vehicles (collisions between two light vehicles in Table 3). When a vehicle strikes a non-MVIT object, it may be difficult to identify the individual who is at fault. On the other hand, when two vehicles are involved in a collision, one or more of the drivers may be at fault thus resulting in a citation. Reckless driver behavior, not following rules of the road, drivers with faulty equipment and expired license/registration, and drivers with multiple infractions are likely to contribute to all types of collisions involving two light vehicles (except for rear-to-side collisions). Speed related infractions contribute less to head-on and angle crashes, and more to rear-end collisions – a finding consistent with the notion that higher speeds require longer stopping distances and hence the higher likelihood of rear-end collisions. Impaired driving contributes positively to head-on collisions (because drivers are not able to maintain their path), but negatively to angle and rear-end collisions – a finding that is somewhat counterintuitive and worthy of additional investigation. It may be that angle and rear-end collisions are associated more with other driver infractions than impaired driving.

Goodness of Fit Measures

The mixed logit model offers a superior goodness of fit to the simpler multinomial logit model that does not account for unobserved heterogeneity. The log-likelihood value for the model with constants-only is -153472.9825 with 19 parameters. The multinomial logit model has a log-likelihood value of -99021.4303 with 90 parameters, while the mixed logit model has a log-likelihood value of -93364.7000 with 100 parameters. The improvement in the log-likelihood due to the inclusion of explanatory variables with heterogeneity terms results in the following:

$$\bar{\rho}_C^2 = 1 - \frac{-93364.70 - 100}{-153472.98 - 19} = 0.34$$

In addition, the improved fit offered by the mixed logit over the multinomial logit may be assessed by computing the likelihood ratio χ^2 statistic as:

$$-2[-99021.4303 - (-93364.7000)] = 11313.50$$

This value is far greater than the critical χ^2 statistic of 28.30 at 12 degrees of freedom. This implies that the additional parameters introduced in the mixed logit specification offer significant explanatory power and capture unobserved heterogeneity that is not adequately accounted for in the multinomial logit model.

6. CONCLUSIONS

This paper presents a comprehensive model of highway safety considering the full range of crash types defined by the nature of collision and vehicles involved. Previous research in the transportation safety arena has largely focused on analyzing and modeling crash occurrence, crash frequency, or injury severity for a subgroup of transport system users along specific dimensions. While such literature has offered rich insights into the factors that contribute to crashes and injury severity of different types, it does not provide a holistic view of the influence of various explanatory factors on a multitude of crash types *simultaneously*. How does a certain roadway attribute affect the probability of a rear-end collision involving two vehicles and the probability of a crash involving a light vehicle striking a pedestrian? The answer to such a question can be obtained by modeling all crash type outcomes in a single comprehensive model. More importantly, by examining how a factor affects multiple crash type outcomes simultaneously, it is possible to devise countermeasures, improvements to roadway geometry, and traffic control strategies while minimizing unintended consequences.

In this paper, a comprehensive model of roadway crash type is presented. The model considers 20 different crash type alternatives, considering eight different collision types and three different vehicle types. A mixed logit model of crash type is estimated using the 2011-2013 GES crash database. Roadway characteristics, weather and seasonal attributes, temporal attributes are explanatory factors included in the model. In addition, the mixed logit model specification accommodates for the presence of unobserved heterogeneity in the effects of various factors on crash type propensity. In general, it is found that several roadway attributes exhibit such unobserved heterogeneity; this is not surprising given that the data set does not include detailed information about traffic volumes and congestion levels, lane configurations, bicycle and pedestrian facilities, and adjoining land uses. The mixed logit model specification is able to account for variations in impacts due to such unobserved factors and is found to offer a statistically superior goodness-of-fit in comparison to the regular multinomial logit model.

The importance of modeling safety in a comprehensive framework is evident in the model estimation results. For example, the model estimation results show that the introduction of an unprotected median in a two-way roadway could reduce head-on collisions between two light vehicles. Similarly, converting a street to a one-way street will result in reduced likelihood of head-on collisions. However, these strategies alone contribute positively to the probability of other crash types, unless the strategies are implemented in a way that minimizes unintended consequences. Both of these variables exhibit considerable unobserved heterogeneity in the manner in which they impact crash type probabilities. Through careful consideration of such unobserved factors, it will be possible to design effective safety measures that produce the intended and desired outcomes without increasing a different type of crash risk. The introduction of a positive median barrier appears to decrease the probability of several crash types, thus suggesting it is an effective safety measure; however, it also increases the probability of specific crash types including a heavy vehicle striking a stationary object, rear-end collision between two light vehicles, and collision between two heavy vehicles. It is important to understand how and why median barriers contribute positively to such crash types; the provision of median barriers can then be combined with other safety measures that reduce or eliminate the increase in probability of certain crash types. For example, restrictions on the passage of heavy vehicles during certain high traffic periods of the day may be a strategy that can be combined with the provision of median barriers.

This research offers insights into factors affecting the probability of crashes of various types by comprehensively considering all crash types simultaneously. The results may be of value in the design of automotive safety systems; for example, the results in this paper suggest that pedestrian and bicyclist safety is compromised when a larger heavy vehicle approaches an intersection, presumably because heavy vehicle drivers are not able to see pedestrians and bicyclists easily and are distracted by the presence of other vehicles and conflicting movements at intersections. Heavy vehicles can be equipped with sensors alerting drivers to the presence of such non-motorized road users. Comprehensive models of safety will be of considerable value in the march towards vehicle connectivity and automation.

ACKNOWLEDGEMENTS

The authors are grateful to Lisa Macias for her help in formatting this document, and to Subodh Dubey for help with coding and running specifications.

REFERENCES

Abdel-Aty, M., and 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 & Prevention*, *36*(3), 447-456.

Bham, G.H., Javvadi, B.S., and Manepalli, U.R. (2011). Multinomial logistic regression model for Single-Vehicle and multivehicle collisions on urban U.S Highways in Arkansas. *Journal of Transportation Engineering*, *138*(6), 786-797.

Bhat, C.R. (2000). A multi-level cross-classified model for discrete response variables. *Transportation Research Part B*, *34*(7), 567-582.

Bhat, C.R. (2001). Quasi-random maximum simulated likelihood estimation of the mixed multinomial logit model. *Transportation Research Part B*, *35*(7), 677-693.

Branas, C.C., and Knudson, M.M. (2001). Helmet laws and motorcycle rider death rates. *Accident Analysis & Prevention*, *33*(5), 641-648.

California Department of Transportation (2010). Complete Intersections: A Guide to Reconstructing Intersections and Interchanges for Bicyclists and Pedestrians. Retrieved from http://nacto.org/docs/usdg/complete_intersections_caltrans.pdf, July 28, 2015.

Cinnamon, J., Schuurman, N., and Hameed, S.M. (2011). Pedestrian injury and human behaviour: observing road-rule violations at high-incident intersections. PLoS ONE 6(6): e21063. doi:10.1371/journal.pone.0021063.

Ghazizadeh, M., and Boyle, L. (2009). Influence of driver distractions on the likelihood of rear-end, angular, and single-vehicle crashes in Missouri. *Transportation Research Record: Journal of the Transportation Research Board*, 2138, 1-5.

Haque, M., and Chin, H. (2010). Right-angle crash vulnerability of motorcycles at signalized intersections: mixed logit analysis. *Transportation Research Record: Journal of the Transportation Research Board*, 2194, 82-90.

Kilpeläinen, M., and Summala, H. (2007). Effects of weather and weather forecasts on driver behaviour. *Transportation Research Part F*, *10*(4), 288-299.

Mitchell, R.J., Senserrick, T., Bambach, M.R., and Mattos, G. (2015). Comparison of novice and full-licenced driver common crash types in New South Wales, Australia, 2001–2011. *Accident Analysis & Prevention*, *81*, 204-210.

Neyens, D.M., and Boyle, L.N. (2007). The effect of distractions on the crash types of teenage drivers. *Accident Analysis & Prevention*, *39*(1), 206-212.

National Highway Traffic Safety Administration, 2015. Traffic Safety Facts 2013: A Compilation of Motor Vehicle Crash Data from the Fatality Analysis Reporting System and the General Estimates System. Retrieved from http://www-nrd.nhtsa.dot.gov/Pubs/812139.pdf, July 28, 2015.

Niewoehner, W., and Berg, F.A. (2005). Endangerment of pedestrians and bicyclists at intersections by right turning trucks. *Statistics*, 1-15.

Pai, C.W., Hwang, K.P., and Saleh, W. (2009). A mixed logit analysis of motorists' right-of-way violation in motorcycle accidents at priority T-junctions. *Accident Analysis & Prevention*, *41*(3), 565-573.

Pai, C.W., and Saleh, W. (2008). Exploring motorcyclist injury severity in approach-turn collisions at T-junctions: Focusing on the effects of driver's failure to yield and junction control measures. *Accident Analysis & Prevention*, *40*(2), 479-486.

Revelt, D., and Train, K. (1998). Mixed logit with repeated choices: households' choices of appliance efficiency level. *Review of Economics and Statistics*, *80*(4), 647-657.

Romo, A., Hernandez, S., and Cheu, R.L. (2014). Identifying precrash factors for cars and trucks on interstate highways: mixed logit model approach. *Journal of Transportation Engineering*, *140*(3), 04013016.

Schneider, W.H., Savolainen, P.T., Van Boxel, D., and Beverley, R. (2012). Examination of factors determining fault in two-vehicle motorcycle crashes. *Accident Analysis & Prevention*, *45*, 669-676.

Stevenson, M.R., Elkington, J., Sharwood, L., Meuleners, L., Ivers, R., Boufous, S., Williamson, A., Haworth, N., Quinlan, M., Grunstein, R., Norton, R., and Wong, K. (2013). The role of sleepiness, sleep disorders, and the work environment on heavy-vehicle crashes in 2 Australian states. *American Journal of Epidemiology*, DOI: 10.1093/aje/kwt305.

Wang, C., Quddus, M.A., and Ison, S.G. (2013). The effect of traffic and road characteristics on road safety: A review and future research direction. *Safety Science*, *57*, 264-275.

Yan, X., Radwan, E., and Abdel-Aty, M. (2005). Characteristics of rear-end accidents at signalized intersections using multiple logistic regression model. *Accident Analysis & Prevention*, *37*(6), 983-995.