MODELING MULTIPLE VEHICLE OCCUPANT INJURY SEVERITY: A COPULA-BASED MULTIVARIATE APPROACH

Naveen Eluru
The University of Texas at Austin
Dept of Civil, Architectural & Environmental Engineering
1 University Station C1761, Austin TX 78712-0278
Phone: 512-471-4535, Fax: 512-475-8744
E-mail: naveeneluru@mail.utexas.edu

Rajesh Paleti
The University of Texas at Austin
Dept of Civil, Architectural & Environmental Engineering
1 University Station C1761, Austin TX 78712-0278
Phone: 512-471-4535, Fax: 512-475-8744
E-mail: rajeshp@mail.utexas.edu

Ram M. Pendyala
Arizona State University
School of Sustainable Engineering and the Built Environment
Room ECG252, Tempe, AZ 85287-5306
Tel: (480) 727-9164; Fax: (480) 965-0557
Email: ram.pendyala@asu.edu

Chandra R. Bhat *(corresponding author)*
The University of Texas at Austin
Dept of Civil, Architectural & Environmental Engineering
1 University Station C1761, Austin TX 78712-0278
Phone: 512-471-4535, Fax: 512-475-8744
E-mail: bhat@mail.utexas.edu

July 2009
Revised November 2009

ABSTRACT
Previous research in crash injury severity analysis has largely focused on level of injury severity sustained by the driver of the vehicle or the most severely injured occupant of the vehicle. While such studies are undoubtedly useful, they do not provide a comprehensive picture of the injury profile of all vehicular occupants in crash-involved vehicles. This limits the ability to devise safety measures that enhance the safety and reduce the injury severity associated with all vehicular occupants. Moreover, such studies ignore the possible presence of correlated unobserved factors that may simultaneously influence and impact the injury severity levels of multiple occupants in the vehicle. This paper aims to fill this gap by presenting a simultaneous model of injury severity that can be applied to crashes involving any number of occupants. A copula-based methodology, that can be effectively used to estimate such complex model systems, is presented and applied to a data set of crashes drawn from the 2007 General Estimates System (GES) in the United States. The model estimation results provide strong evidence of the presence of correlated unobserved factors that affect injury severity levels among vehicle occupants. The correlation exhibits heterogeneity across vehicle types with greater level of inter-occupant dependency in heavier sport utility vehicles and pickup trucks. The study also sheds light on how numerous exogenous factors including occupant characteristics, vehicle characteristics, environmental factors, roadway attributes, and crash characteristics affect injury severity levels of occupants in different seat positions. The findings confirm that rear seat passengers are less vulnerable to severe injuries than front row passengers pointing to the need to enhance vehicular design features that promote front row occupant safety.

Keywords: statistical methodology, copula-based approach, simultaneous equations model, injury severity modeling, vehicle crash analysis

1. INTRODUCTION

The *Global Status Report on Road Safety* published recently by the World Health Organization (*1*) paints a grim picture of safety statistics on the world's highways. Using data derived from a 2008 survey of 178 countries around the world, the report notes that nearly 1.3 million people are killed and between 20 and 50 million people get injured every year around the globe in roadway crashes. The estimated cost of highway crashes to governments worldwide is estimated to be 518 billion US dollars. In the United States, about 40,000 fatalities and 2.3 million injuries occur on the nation's highways every year (*2*). While the World Health Organization (WHO) notes that enforcement of traffic rules, strict licensing standards, enhanced driver training, and community safety education campaigns would enhance roadway safety, it also identifies the need for a greater understanding of crash causation, injury severity, and risky road user behavior as one of the keys to reducing roadway fatalities and injuries. This paper aims to directly address this need by identifying both observed and unobserved factors that contribute to injury severity of multiple occupants in a vehicle, a topic that hitherto has received little attention in the literature.

In vehicular crashes where there are multiple occupants in a vehicle, the different occupants may experience varying levels of injury severity depending on a wide array of factors. Some factors may be observed (and therefore measured and reported in crash data sets), for example, seat belt use, alcohol involvement, vehicle type, and position of the occupant in the vehicle. Other factors, however, may be unobserved (and therefore go unmeasured and unreported in crash data sets). These factors may include such variables as vehicle condition and maintenance record, vehicle speed at the time of crash, condition and effectiveness of the vehicle safety equipment, and mental and physical state of the vehicle occupant. Given that there is potentially a wide array of factors, both observed and unobserved, that may affect injury severity *and* that injury severity may vary across occupants in a vehicle, the field would benefit from a study that models injury severity of multiple vehicle occupants while accounting for common observed and unobserved factors that may contribute to injury severity levels experienced by different occupants. This paper aims to present such a model system so that safety counter-measures can be devised to reduce injury severity levels for all vehicle occupants simultaneously.

The study of injury severity resulting from crashes has been of much interest in the profession. There is a large body of literature devoted to modeling injury severity, usually adopting some form of ordered response model specification. These studies typically examine the crash injury severity of the driver or the most severely injured vehicle occupant (*3-6*). However, not much attention has been paid to simultaneously modeling injury severity of multiple occupants in a vehicle. A couple of studies that have attempted to model injury severity of two occupants of the vehicle (usually the driver and the most severely injured passenger) include those by Hutchinson (*7*) and Yamamoto and Shankar (*8*). In both of these studies, a bivariate probit model specification is adopted to model injury severity for two vehicle occupants. The bivariate probit model specification incorporates the ability to account for the presence of common unobserved factors that influence injury severity across two vehicle occupants. Modeling injury severity simultaneously for more than two vehicle occupants presents a methodological challenge, however, due to the computational complexity associated with specifying, identifying, and estimating a multivariate probit model with more than two dimensions. This paper overcomes this challenge by presenting a simple and practical modeling approach and specification that accommodates the simultaneous analysis of injury severity of any number of vehicle occupants by seat position. The focus of this paper on injury severity as related to seat position is motivated by the considerable attention that has been devoted to this issue in the literature. There are numerous studies that examine the injury severity levels sustained by children seated in different positions in vehicles (*9-12*). Virtually all studies report findings that children seated in the front are more likely to sustain fatal or severe injuries than children seated in the rear.

The analysis of injury severity of multiple occupants in a vehicle has been limited by the methodological challenges associated with modeling such phenomena in a simultaneous (or joint) equations framework. Several studies have employed descriptive statistical analysis techniques, logistic regression approaches, or ordered response structures to model injury severity of occupants with explicit consideration of seat position, but as an explanatory variable. Evans and Frick (*13*), Smith and Cummings (*14,15*), Wang and Kockelman (*16*) Claret *et al.* (*17*), and Mayrose and Priya (*18*) constitute examples of such studies. All of these studies report that passengers seated in the rear seat sustain less severe injuries than those seated in the front, with those seated in the rear middle position generally sustaining the least severe injuries among all occupants. On the other hand, O'Donnell and Connor (*3*) undertake a comprehensive analysis of occupant injury severity using ordered logit and probit models and report that the driver seat position is the safest among all seat positions.

Although the previous literature has shed light on the influence of seat position on occupant injury severity, there is very little work on the joint modeling of multiple occupant injury severity that accounts for both observed and unobserved factors that simultaneously impact injury severity of multiple vehicle occupants. While the studies of Hutchinson (*7*) and Yamamoto and Shankar (*8*) provided an initial impetus to such simultaneous injury severity modeling, further work has been hampered by methodological challenges associated with specifying, identifying, and estimating such simultaneous equations models. This paper aims to contribute substantively to this arena by presenting a copula-based methodology that can be applied to estimate models of injury severity of any number of occupants in a vehicle simultaneously. The methodology is applied to the 2007 General Estimates System (GES) data set from the United States, a database of a sample of crashes from jurisdictions across the country.

The remainder of this paper is organized as follows. The next section presents the copula-based methodology adopted in this paper. The third section presents a detailed description of the data set while the fourth section presents the model estimation and validation results. Concluding thoughts are offered in the fifth and final section.

2. METHODOLOGY

Consistent with the literature on injury severity analysis, this paper adopts an ordered response modeling approach with an implicit assumption that there is an underlying continuous latent variable whose horizontal partitioning maps into the observed injury severity level. The issue that receives explicit consideration in this paper is that there is a potential inter-dependence in injury severity among different occupants of the same vehicle due to both observed and unobserved exogenous factors. If there are no common unobserved factors affecting injury severity across multiple vehicle occupants, then one can estimate independent ordered response models of injury severity separately for each vehicle occupant. However, if there are common unobserved factors, then a simultaneous ordered response model of vehicle occupant injury severity that accommodates error correlations needs to be specified and estimated. Common unobserved factors may include such variables as vehicle speed at the time of crash, vehicle condition and maintenance record, condition of vehicle safety equipment, vehicle safety features, and state of passengers prior to crash. The simultaneous equations modeling of occupant injury severity is a classic case of analyzing clusters of dependent random variables that has widely been considered in transportation and other fields (see, for example, *19-21*). However, these earlier studies *a priori* place restrictions on the dependency surface characterizing the relationship between the dependent random variables (mostly through what amounts to a symmetric multivariate normal dependency surface). However, it may be the case that the dependence among the injury propensities of vehicle occupants is asymmetric; for instance, one may observe vehicle occupants having a simultaneously high propensity for high injury severity levels, but not necessarily a propensity for simultaneously low injury severity levels. Alternatively, even if symmetric, the specific parametric functional form of the dependency may take one of several profiles. In the current paper, we use an approach that enables us to test the appropriateness of different parametric dependency surfaces to select the one that empirically fits the data best.

Specifically, this paper adopts a copula-based approach to accommodate the dependence in injury severity propensity among multiple vehicle occupants. In particular, this paper uses the Archimedean group of copulas to implement a computationally feasible maximum likelihood procedure for parameter estimation. The copula-based approach offers the ability to formulate a closed form likelihood function that eliminates the need to adopt the more computationally intensive simulation-based procedures for parameter estimation. Other advantages associated with adopting the Archimedean group of copulas for model estimation include the following:

- The Archimedean copulas can be used to obtain the joint multivariate cumulative distribution function of any number of individuals belonging to a cluster. Further, these copulas retain the same form regardless of cluster size, thus accommodating clusters of varying sizes in a straightforward manner.
- The Archimedean group of copulas allows testing a variety of radially symmetric and asymmetric joint distributions, as well as testing the assumption of within-cluster independence.
- The approach enables the specification of a variety of parametric marginal distributions for individual members in a cluster and preserves these marginal distributions when developing the joint probability distribution of the cluster. Further, the approach separates the marginal distributions from the dependence structure so that the dependence structure is entirely unaffected by the marginal distributions assumed.
- Finally, the approach allows the level of dependence within a cluster to vary based on cluster type. For example, the level of dependence of injury severity across vehicle occupants may be influenced by vehicle type and other vehicle characteristics. In fact, it is possible to allow the dependency structure to be different across cluster types (say, vehicle types) by using different copulas for different cluster types.

The remainder of this section presents the mathematical formulation of the modeling methodology.

2.1 Copula-Based Approaches

A copula is a device or function that generates a stochastic dependence relationship (*i.e.*, a multivariate distribution) among random variables with pre-specified marginal distributions. Bhat and Eluru (*22*) and Trivedi and Zimmer (*23*) offer detailed descriptions of the copula-based approaches to statistical model estimation and the types of copulas available for generating multivariate distribution functions with given marginals [see also Genest and MacKay (*24*)]. The precise definition of a copula is that it is a multivariate distribution function defined over the unit cube linking uniformly distributed marginals. Let C be an I-dimensional copula of uniformly distributed random variables $U_1, U_2, U_3, \ldots, U_I$ with support contained in $[0,1]^I$. Then,

$$C_\theta(u_1, u_2, \ldots, u_I) = \Pr(U_1 < u_1, U_2 < u_2, \ldots, U_I < u_I), \tag{1}$$

where θ is a parameter vector of the copula commonly referred to as the dependence parameter vector. Consider I random variables $\varepsilon_1, \varepsilon_2, \varepsilon_3, \ldots, \varepsilon_I$, each with univariate continuous marginal distribution function $F(z_i) = \Pr(\varepsilon_i < z_i)$.[1] Then, a joint I-dimensional distribution function of the random variables with the continuous marginal distribution functions $F(z_i)$ can be generated as follows (*25*):

$$\begin{aligned} F(z_1, z_2, \ldots, z_I) &= \Pr(\varepsilon_1 < z_1, \varepsilon_2 < z_2, \ldots, \varepsilon_I < z_I) = \Pr[U_1 < F(z_1), U_2 < F(z_2), \ldots, U_I < F(z_I)] \\ &= C_\theta[u_1 = F(z_1), u_2 = F(z_2), \ldots u_I = F(z_I)]. \end{aligned} \tag{2}$$

The above equation offers an approach to develop different dependency patterns for the random variables $\varepsilon_1, \varepsilon_2, \varepsilon_3, \ldots, \varepsilon_I$ based on the copula that is used as the underlying basis of construction. In the current paper, a class of copulas referred to as the Archimedean copulas is used to generate the dependency between the random variables. The Archimedean class of copulas is popular in empirical applications, and includes a whole suite of closed-form copulas that cover a wide range of dependency structures, including comprehensive and non-comprehensive copulas, radial symmetry and asymmetry, and asymptotic tail independence and dependence [see Nelsen (*26*) and Bhat and Eluru (*22*) for a detailed discussion]. This class of copulas is very flexible, and easy to construct.

Archimedean copulas are constructed based on an underlying continuous convex decreasing generator function (see Bhat and Eluru (*22*) for a discussion on the generation of Archimedean copulas). A whole variety of Archimedean copulas have been identified based on different forms of this generator function. In this paper, four different and most popular Archimedean copulas that span the spectrum of different kinds of dependency structures are considered. These are the Clayton, Gumbel, Frank, and Joe copulas (see Bhat and Eluru (*22*) for graphical descriptions of the implied dependency structures). All of these copulas allow only positive associations and equal dependencies among pairs of random variables in their multivariate forms, which is well-suited for cluster analysis where one would expect positive and equal dependencies among elements within a cluster. The Clayton copula (*27*) is best suited for strong left tail dependence and weak right tail dependence; that is, it is suitable for the case when, after controlling for observed covariates, vehicle occupants tend to have a simultaneously high propensity for low injury severity levels, but not a simultaneously high propensity for high injury severity levels. The Gumbel (*28*) copula (also referred to as the Gumbel-Hougaard copula)

[1] Note that the univariate marginal distribution functions of the random variables do not have to be identical. However, such a specification is often used when developing econometric models where the random terms represent individual-level idiosyncratic effects.

is well suited for the case when there is strong right tail dependence (strong correlation at high values) but weak left tail dependence (weak correlation at low values); that is, it is suitable for the case when, after controlling for observed covariates, vehicle occupants tend to have a simultaneously high propensity for high injury severity levels, but not a simultaneously high propensity for low injury severity levels. The Frank (*29*) copula is radially symmetric in its dependence structure like the Gaussian (normal) copula. This copula is suitable for equal levels of dependency in the left and right tails, with very strong clustering in the middle (much stronger than the Gaussian copula); that is, it is suitable for the case when vehicle occupants tend to have a simultaneously high propensity for high injury severity levels or a simultaneously high propensity for low injury severity levels.. The Joe (*30,31*) is similar to the Clayton copula discussed earlier, but the right tail positive dependence is stronger.

2.2 Model Formulation and Estimation

Let q be an index for clusters (vehicle in the current empirical context) (q = 1, 2, …, Q), and let i be the index for occupants (i = 1, 2, …, I_q, where I_q denotes the total number of occupants in vehicle q; in the current study I_q varies between 1 and 5). Also, let k be an index for the discrete outcomes corresponding to the injury severity level. The index k, for example, may take values of "no injury" (k = 1), "possible injury" (k = 2), "non-incapacitating injury" (k = 3), "incapacitating injury" (k = 4), and "fatal injury" (k = 5). In the usual ordered response framework notation, one can write the latent propensity (y^*_{qi}) of occupant i in vehicle q to sustain an injury severity level as a function of relevant covariates, and then relate this latent propensity to the severity outcome (y_{qi}) representing the injury severity sustained by occupant i in vehicle q through threshold bounds (see *32*):

$$y^*_{qi} = \beta' x_{qi} + \varepsilon_{qi}\,,\; y_{qi} = k \text{ if } \psi_k < y^*_{qi} \le \psi_{k+1}\,, \tag{3}$$

where x_{qi} is a ($L\times 1$) vector of exogenous variables for occupant i in vehicle q (not including a constant), β is a corresponding ($L\times 1$) vector of coefficients to be estimated, and ψ_k is the lower bound threshold for injury severity level k ($\psi_0 < \psi_1 < \psi_2 ... < \psi_K < \psi_{K+1}$; $\psi_0 = -\infty$, $\psi_{K+1} = +\infty$). The ε_{qi} terms capture the idiosyncratic effect of all omitted variables for occupant i in vehicle q, and are assumed to be independent of β and x_{qi}. The ε_{qi} terms are assumed identical across occupants, each with a univariate continuous marginal distribution function $F(z_{qi}) = \Pr(\varepsilon_{qi} < z_{qi})$. The error terms can take any parametric marginal distribution, although only the normal and logistic distributions are considered in the current paper. Due to identification considerations in the ordered-response model, the univariate distribution functions are standardized, so that they are standard normal or standard logistic distributed.

Dependence in the ε_{qi} terms across occupants i in the same vehicle q is accommodated to allow unobserved cluster effects. This dependency is generated through the use of an Archimedean copula based on Equation (2), where the only difference now is the introduction of the index q to reflect that the dependence is confined to occupants of the same vehicle:

$$\begin{aligned}\Pr(\varepsilon_{q1} < z_{q1}, \varepsilon_2 < z_{q2}, \ldots, \varepsilon_{qI_q} < z_{qI_q}) &= \Pr[U_{q1} < F(z_{q1}), U_{q2} < F(z_{q2}), \ldots, U_{qI_q} < F(z_{qI_q})] \\ &= C_{\theta_q}\ [u_{q1} = F(z_{q1}), u_{q2} = F(z_{q2}), \ldots u_{qI_q} = F(z_{qI_q})].\end{aligned} \tag{4}$$

It is important to note above that the level of dependence among occupants of a vehicle can vary across vehicles, as reflected by the θ_q notation for the dependence parameter. This dependence parameter is parameterized in this study as a function of observed vehicle characteristics[2].

Let m_{qi} be the actual observed categorical response for y_{qi} in the sample. Then, the probability of the observed vector of injury severity levels across occupants in vehicle q $(m_{q1}, m_{q2}, m_{q3}, ..., m_{qI_q})$ can be written as:

$$P(y_{q1} = m_{q1}, y_{q2} = m_{q2}, ..., y_{qI_q} = m_{qI_q}) = \int_{M_q} c_{\theta_q}\left(F(y^*_{q1}), F\ (y^*_{q2}), ..., F(y^*_{qI_q})\right) dy^*_{q1} dy^*_{q2} ... dy^*_{qI_q}, \tag{5}$$

where $M_q = \{y^*_{q1}, y^*_{q2}, ..., y^*_{qI_q} : \psi_{(m_{qi})} < y^*_{qi} < \psi_{(m_{qi}+1)}$ for all $i = 1, 2, ..., I_q\}$ and c_{θ_q} is the copula density. The integration domain M_q is simply the multivariate region of the y^*_{qi} variables $(i = 1, 2, ..., I_q)$ determined by the observed vector of injury outcomes $(m_{q1}, m_{q2}, ..., m_{qI_q})$. The dimensionality of the integration, in general, is equal to the number of occupants I_q in the vehicle. Thus, if one uses a Gaussian copula, one ends up with integrals of the order of the number of occupants in the vehicle for the joint probability of the observed combination of the injury severity levels across occupants in the vehicle. This will necessitate the use of simulation techniques when I_q is greater than three. However, in the case of a vehicle-level cluster with identical dependencies between pairs of occupants in the vehicle, one can gainfully employ the Archimedean copulas since they provide closed-form multivariate cumulative distribution functions. In particular, the probability in Equation (5) can be written in terms of 2^{I_q} closed-form multivariate cumulative distribution functions as follows:

$$\begin{aligned}&P(y_{q1} = m_{q1}, y_{q2} = m_{q2}, ..., y_{qI_q} = m_{qI_q}) = P(\psi_{m_{q1}} < y^*_{q1} < \psi_{m_{q1+1}}, \psi_{m2} < y^*_{q2} < \psi_{m_{q2+1}}, \ldots \psi_{m_{qI_q}} < y^*_{qI_q} < \psi_{m_{qI_q}+1}) \\ &= \sum_{a_1=1}^{2}\sum_{a_2=1}^{2}\cdots\sum_{a_{I_q}=1}^{2}(-1)^{a_1+a_2+\ldots a_{I_q}}\left[P(y^*_{q1} < \psi_{m_{q1}+a_1-1}, y^*_{q2} < \psi_{m_{q2}+a_2-1}, \ldots y^*_{qI_q} < \psi_{m_{qI_q}+a_{I_q}-1})\right] \\ &= \sum_{a_1=1}^{2}\sum_{a_2=1}^{2}\cdots\sum_{a_{I_q}=1}^{2}(-1)^{a_1+a_2+\ldots a_{I_q}}\left[C_{\theta_q}(u_{m_{q1}+a_1-1}, u_{m_{q2}+a_2-1}, \ldots u_{m_{qI_q}+a_{I_q}-1})\right]\end{aligned} \tag{6}$$

where C_{θ_q} is one of the four Archimedean copulas discussed previously with an association parameter θ_q, and $u_{m_{qi}+a_i-1} = F(\psi_{m_{qi}+a_i-1} - \beta' x_{qi})$. The number of cumulative distribution function

[2] It is possible to use different copula forms (*i.e.*, dependency surfaces) for different vehicles; however, in this paper, the same copula form is maintained across all vehicles to keep the estimation tractable.

computations increases rapidly with the number of individuals I_q in vehicle q, but this is not much of a problem when the cluster sizes are six or less because of the closed form structures of the cumulative distribution functions. In the current empirical context, $I_q \leq 5$, thus lending itself to the use of the copula-based approach for model estimation.

The association parameter θ_q is allowed to vary across vehicles. However, it is not possible to estimate a separate dependence term for each vehicle. Therefore, θ_q is parameterized as a function of a vector s_q of observed vehicle variables, while also choosing a functional form that ensures that θ_q for any vehicle q is within the allowable range for each copula. Thus, $\theta_q = \exp(\delta' s_q)$ for the Frank and Clayton copulas, and $\theta_q = 1 + \exp(\delta' s_q)$ for the Gumbel and Joe copulas.

The parameters to be estimated in the model may be gathered in a vector $\Omega = (\beta', \delta', \psi')'$, where the vector ψ is the vector of threshold bounds: $\psi = (\psi_1, \psi_2, \ldots \psi_K)$. The likelihood function for vehicle q may be constructed based on the probability expression in Equation (6) as:

$$L_q(\Omega) = P(y_{q1} = m_{q1}, y_{q2} = m_{q2}, \ldots, y_{qI_q} = m_{qI_q}). \tag{7}$$

The likelihood function to be maximized is then given by $L(\Omega) = \prod_q L_q(\Omega)$.

3. DATA

The crash data used in this study is derived from the 2007 General Estimates System (GES) obtained from the National Highway Traffic Safety Administration (NHTSA) in the United States (*2*). The GES consists of crash data compiled from a sample of police-reported accidents that involve at least one motor vehicle traveling on a roadway and resulting in property damage, injury, or death. The GES crash data are drawn from crashes in about 60 areas across the United States that reflect the geography, population, and traffic patterns of the country. A detailed discussion on the sampling and compilation of crash data for the GES is provided in the GES documentation available at the NHTSA website (http://www.nhtsa.gov). The 2007 GES includes information on 60,000 crashes involving about 150,000 individuals and 100,000 vehicles. A number of crash-related attributes are collected for each record in the GES including, for example, driver and vehicle characteristics, roadway design attributes, environmental conditions, and crash characteristics. The injury severity of each individual involved in a crash is coded on a five-point ordinal scale: (1) No injury, (2) Possible injury, (3) Non-incapacitating injury, (4) Incapacitating injury, and (5) Fatal injury.

In this study, the analysis is confined to examining injury severity of vehicle occupants in non-commercial (private) passenger vehicles involved in collisions, *i.e.*, where a vehicle collided with a stationary object or another vehicle. A vast majority of crashes in the database involved one or two vehicles, and therefore, records in which three or more vehicles were involved in a crash were not included in the analysis. Through an extensive data checking and cleaning effort, only those cases in which complete information was available for all occupants involved in the crash were selected for analysis. The cleaned data set used in this study consists of 35,978 vehicles and 48,004 occupants.

Due to the large sample nature of the data set, a random sample of records was drawn for the model estimation process. The random sample used for model estimation includes 5,297 occupants (4,000 drivers and 1,297 passengers) in 4,000 vehicles. The sample includes 77.3% single-occupant, 15.9% two-occupant, 4.3% three-occupant, 2.0% four-occupant, and 0.5% five-occupant crashes. Within each of the multiple occupant vehicles there are different seat positions possible for the occupants. There are 16 possible seat position configurations for up to five occupants. The distribution of the occupants based on the seat position in this estimation sample is as follows:

- Driver: 75.5%
- Front seat passenger: 15.0%
- Rear left position passenger: 3.5%
- Rear center position passenger: 1.3%
- Rear right position passenger: 4.7%

Table 1 presents a summary of the sample characteristics of the occupants of the vehicles involved in the crashes. More than three-quarters of the crashes involve two vehicles. The sample has a slightly higher fraction of males. Nearly one-half of the occupants are aged 21-45 years; children aged 15 years or less comprise nearly nine percent of the sample. Seatbelt use is quite high with 92 percent of the occupants reporting being buckled in. About 60 percent of the vehicles involved are sedans. Most of the crashes occur in the midday and evening, presumably due to the higher level of travel during these periods. Each of these periods accounts for more than 30 percent of the crashes in the sample. About 90 percent of the crashes took place on roadways with speed limits 55 mph or lower. Head-on collisions account for only five percent of the crashes while rear-end and angle collisions each account for more than 30 percent of the crashes.

The distributions of injury severity levels for the vehicle occupants show that nearly two-thirds of occupants (whether driver or passenger) report no injury. A little over 10 percent of the occupants report a possible injury or non-incapacitating injury in both the driver and passenger samples. The percent of individuals sustaining fatal injuries is extremely small in this random estimation sample (at about 0.6 percent). In order to ensure a reasonable share for each alternative outcome, the incapacitating and fatal injury categories were merged to generate a single “serious injury” category. This category accounts for about 8.5 percent of the outcomes reported in the sample.

4. MODEL ESTIMATION RESULTS

This section presents model estimation results in detail. First, the section presents the overall model specification considerations and model performance in terms of goodness-of-fit. Second, the section presents a detailed discussion of the actual factors affecting injury severity of multiple occupants in vehicles involved in crashes.

4.1 Model Specification and Overall Performance

The model specification included a range of variables covering five broad categories of factors. These are:

- Occupant characteristics including age, sex, alcohol state, and seat belt use
- Vehicle characteristics including vehicle type

- Environmental characteristics including day of week, time of day, lighting conditions, and weather conditions
- Roadway design attributes including speed limit, type of roadway, roadway alignment, and number of lanes
- Crash characteristics including whether person was ejected from vehicle, whether vehicle rolled over, whether single-vehicle or two-vehicle crash, collision type, and role of the driver's vehicle in a two-vehicle crash

The final model specification was derived based on a systematic process of considering variables for inclusion based on statistical significance, intuitive interpretation, parsimony in specification, and consistency with results reported in prior studies of injury severity. Several different combinations of variables, functional forms, and interaction terms were considered.

In this research effort, and as discussed earlier, we examined four different copula structures (Clayton, Gumbel, Frank and Joe) for specifying the dependency between the ε_{qi} terms across vehicle occupants to represent the vehicle cluster effect, and two different univariate distribution assumptions (normal and logistic) for the random error term ε_{qi}. For the sake of brevity, and also due to space considerations, we present the results for the best copula model and the best independent model (from the logistic and the normal distributions for the ε_{qi} terms). To determine the best model among the copula models, we employ the Bayesian Information Criterion (BIC) [for details, see Quinn (*33*), Trivedi and Zimmer (*23*)]. The BIC for a given copula model is equal to $-2\ln(L)+K\ln(Q)$, where $\ln(L)$ is the log-likelihood value at convergence, K is the number of parameters, and Q is the number of observations. The copula that results in the lowest BIC value is the preferred copula. The BIC based selection procedure in our research effort is equivalent to selection based on the largest value of the log-likelihood function at convergence because all the competing models have the same exogenous variables and the same number of thresholds.

Among the copula models, the results indicated that the Probit-Frank (PF) model provides the best data fit with a likelihood value of -4677.9. However, in all of the copula models, the dependency parameters were highly statistically significant, with the vehicle-level dependency in unobserved factors varying based on vehicle type. Specifically, the vehicle-level dependency was different across four vehicle types – sedan, SUV, pickup truck, and van. Between the two independent models, the normal error term distribution for the marginals (*i.e.*, the ordered-response probit or ORP) provides a slightly better fit than the logistic error term distribution for the marginal (*i.e.*, the ordered-response logit). The likelihood ratio test comparing the PF model in this paper with the independent ORP model yields a test statistic value of 373.0 which is substantially larger than the critical χ^2 value with 4 degrees of freedom (corresponding to the four dependency parameters) at any reasonable level of significance, confirming the importance of accommodating dependence in injury severity propensity among vehicle occupants.

4.2 Key Findings

Model estimation was undertaken for all occupants together while accommodating unobserved dependencies in the latent injury propensities of occupants within a vehicle. Specifically, separate coefficients were estimated for the driver, front seat passenger, and rear seat passengers. Coefficients for all rear seat passengers were restricted to be equal to accommodate the small sample of rear seat passengers; however, indicator variables were included in the model specification to accommodate potential differences across different rear seat passengers. Model estimation results are presented in Table 2.

The coefficients presented in the table indicate the effects of variables on the latent injury severity propensity of an occupant. A positive coefficient associated with a variable indicates that the variable contributes positively to a higher injury severity propensity. The first set of values in the table present the thresholds of the ordered response model that simply serve to translate the latent propensity into the observed ordered categories of injury severity. For the dummy variables (including variables with multiple levels), we have a reference category that may vary by seat position category. In Table 2, when all the levels of a dummy variable are present, the coefficients with "---" represent the reference category. For other dummy variables, we have explicitly identified the reference category in the table.

Among occupant characteristics, it is found that males have a lower propensity to sustain severe injuries when compared with females when seated in the front row (either as driver or passenger). When the driver is male, the front passenger (who is more likely to be a female) has a higher propensity to experience a severe injury. This finding confirms previous research [see Ulfarsson and Mannering (*5*)] indicating significant gender differences (say related to weight, body structure, or other factors) in injury severity outcomes after controlling for the factors usually available in injury severity analyses. Children aged 0-5 years are less likely to be severely injured when they are seated in the rear. In comparison to older drivers (65+ years of age), younger drivers are less likely to sustain a severe injury. This is particularly so for the youngest group of drivers 16-20 years of age. Older passengers (65+ years) seated in the front have a higher propensity to be severely injured compared to front row passengers of other age groups. With an increasing number of elderly people in virtually every country of the world, many of whom are going to depend on others for a ride, this finding merits serious consideration for the implementation of counter-measures. This finding may call for the installation of special safety devices and equipment in vehicles that would protect the elderly whose physical condition may be more fragile in comparison to other groups.

As expected, seat belt use consistently results in a lower injury severity propensity. When the driver is under the influence of alcohol and behaves "recklessly", the propensity of severe injuries rises. A driver in a full vehicle appears less likely to be severely injured, presumably because the driver is careful in light of having the responsibility to transport a full vehicle of passengers and the passengers in turn ensure that the driver is careful in operating the vehicle.[3] In comparison to the larger vehicle types of SUV, pickup truck, and van, driving or riding in a sedan is associated with a propensity to experience a higher injury severity level. This finding is intuitive, as one would expect to be better protected from severe injury when driving or riding in a heavier and larger vehicle type. Rear passengers seated in the middle have a propensity to experience a lower level of injury level, suggesting that the middle position in the rear is likely to be the safest position in the event of a crash. This finding is consistent with that reported in the literature (*13*).

---

[3] As pointed out by a reviewer, careful driving would be more reasonably discovered in a crash rate analysis. However, it is reasonable to assume that a person who drivers defensively will incur less severe injuries if in a crash [a recent study by Paleti *et al*. (*34*) has found this to be the case]. In addition to the safe driving habits (which unfortunately did not prevent a crash) that may lead to less severe injuries, it is plausible that careful drivers see a crash developing earlier and take evasive actions to reduce the severity of a crash. However, it is important to note here that other explanations are also possible for this result.

Among environmental factors, it was found that time of day significantly impacts injury severity levels. Occupants seated in the front row (whether driver or passenger) are likely to sustain more severe injuries in crashes during the overnight and early morning hours (12 midnight to 6 am) than during other time periods of the day, a finding consistent with expectations that night-time driving may be more challenging and associated with potential alcohol involvement. Conditions in which lighting is present contributes to a lower injury severity propensity. In wet and snowy conditions, both drivers and front row passengers have a lower injury propensity, presumably because vehicles are proceeding at slower speeds and drivers are more cautious under these conditions.

As expected, in comparison to roadways where the speed limit is low, crashes on higher speed limit roadways are associated with a higher injury severity propensity for occupants in all positions. The absence of a median dividing the roadway and the presence of curves in the roadway contribute positively to injury severity propensity for the driver and the front seat passenger. These findings have important implications for roadway design and alignment. Crashes in which the vehicle rolls over, the occupant is ejected from the vehicle, or the vehicle collides with a stationary object are associated with higher levels of injury severity, particularly for front row occupants. When occupants are in a vehicle that both strikes another vehicle and is struck itself by the other vehicle, then the likelihood of a severe injury rises substantially as indicated by the higher positive coefficient than that associated with a vehicle that was only struck by another vehicle. Consistent with expectations, it was found that head-on collisions and angle crashes showed high injury severity propensities for all vehicle occupants. Rear end collisions were associated with a higher injury propensity for a front passenger, while side-swipe collisions were associated with a lower injury severity levels. Counter-measures that aim to reduce the occurrence of head-on and angle crashes are likely to be most effective in reducing injury severity levels associated with crashes.

Overall, the findings are consistent with expectations and speak to the important role played by observed factors in affecting injury severity levels. The findings confirm results reported previously in the literature and offer some insights into the types of safety counter-measures that can reduce injury severity levels of drivers and passengers seated in different positions. However, this paper goes beyond what has been done previously to also examine for possible unobserved dependence effects that can substantially impact values of model coefficients, if indeed such effects are present. The estimated copula-based clustered ordered response model incorporates the jointness in injury severity across vehicle occupants that may be caused by the presence of common unobserved factors. Ignoring such dependencies completely or pre-imposing specific functional forms of the dependency can, and in general will, lead to inappropriate covariate influence estimates on injury severity levels.

4.3 Model Assessment and Validation

As indicated earlier, the Frank copula model form provided the best fit. The association parameter is parameterized in the Frank copula as $\theta_q = \exp(\delta' s_q)$ to accommodate potential heterogeneity in the dependence effects across clusters. In this paper, explicit recognition is given to the possibility that dependence effects in injury severity across vehicle occupants vary by vehicle type. Therefore, in this study, the s_q vector includes four dummy vehicle type variables, *i.e.*, sedan, SUV, pickup truck, and van. The implied Frank association parameter θ_q for these four vehicle types and their corresponding standard errors [computed using the delta method; see Greene (*35*] are as follows: Sedan: 5.2651 (2.023), SUV: 7.3068 (2.985), Pick up truck: 7.6156 (3.244) and Van: 4.3462 (1.152). All of these parameters are highly statistically significant (relative to the value of zero, which corresponds to independence), indicating the strong dependence among the unobserved injury severity determinants of vehicle occupants. Another common way to quantify the dependence in the copula literature is to compute the Kendall's measure of dependence [see Bhat and Eluru (*22*) for a detailed description of this measure]. The Kendall's measure of dependence takes the place of a traditional correlation coefficient when one is dealing with asymmetric distributions. For the estimated association parameters, θ_q, the values of the Kendall's measures are: Sedan: 0.473, SUV: 0.575, Pick-up truck: 0.588 and Van: 0.413. These measures of concordance coupled with the dependence form of the Frank copula imply that the dependency in unobserved components across occupants in the propensity to sustain an injury severity level is very strong. In particular, the highest level of dependence in injury severity due to unobserved factors is for occupants of SUVs and pick-up trucks[4].

In an effort to further assess the Probit-Frank (PF) model, a model validation effort was undertaken. The performance of the PF model is compared against that of the ordered response probit (ORP) model of independence for a validation sample that was not part of the estimation data set. The validation sample consisted of 1,000 vehicles and 1,365 occupants. To perform the validation, the predictive log-likelihood measure is computed for both models for various subsamples. The results of the validation effort are presented in Table 3.

An examination of the results of the validation exercise confirms that the PF model clearly offers a superior statistical fit and predictive power than the ORP model of independence. The likelihood ratio test presented in the last column offers a statistical basis to compare the performance of the PF model against that of the ORP model. For the full sample, and most subsamples considered in the table, the PF model is statistically significantly better than the ORP model. The PF model performs substantially better than the ORP model when there are multiple occupants (particularly when there are three or four occupants in the vehicle). This finding is consistent with expectations as one would expect the correlation across dimensions to be substantive only for vehicles that have multiple occupants. The PF model shows a statistically significant superior data fit for one-vehicle and two-vehicle crashes, sedan and SUV crashes, crashes on roadways with speed limit between 35 and 55 mph, and rear-end collisions. For other subsamples, such as head-on collisions and crashes on roadways with speed limit less than or equal to 35 mph or greater than 55 mph, the PF model and the ORP model do not provide significantly different fit measures.

5. CONCLUSIONS

There is substantial interest in the profession to understand and identify factors contributing to the severity of injuries of vehicle occupants in crashes. Past research has generally focused on the injury severity for that occupant who was most severely injured in a crash as opposed to examining the injury severity levels of all occupants in a crash-involved vehicle. While there have been studies examining the injury severity level of vehicle occupants by seat position, there has been virtually no study that simultaneously analyzes the injury severity levels of all vehicle

[4] We generated scatter plots to illustrate the relationship between the unobserved components ε_{qi} of injury severity propensity for any two occupants in the same vehicle q, based on vehicle type. However, due to space constraints, these plots are not presented. The plots are available from the authors.

occupants in a crash. Initial attempts at doing so have been limited in their scope to examining injury severity of two vehicle occupants using bivariate probit models, such as that by Yamamoto and Shankar (*8*). The ability to examine additional vehicle occupants simultaneously has been seriously hampered by methodological challenges associated with jointly modeling multidimensional phenomena with complex error correlation structures.

In this paper, an ordered probit Frank copula model is specified and estimated to allow for the joint modeling of injury severity outcomes for all vehicle occupants in a vehicle while accomplishing three major objectives:

1. Accounting for the presence of common unobserved factors (error correlations) that simultaneously affect the injury severity outcomes of multiple occupants in a crash-involved vehicle

2. Accounting for the differential effects of various exogenous factors on injury severity according to the seat position of the vehicle occupant

3. Accounting for the heterogeneity in injury severity dependency effects among vehicle occupants across vehicle types by parameterizing the copula association parameter as a function of vehicle body type

The specification and estimation of such a model system constitutes the major contribution of this paper in the context of earlier research in this topic area. A random sample of the 2007 General Estimates System (GES) data set in the United States is used to estimate the ordered probit Frank copula model. The performance of the Frank copula based model was compared against that of the ordered probit/logit model of independence and it was found that the copula-based model that accommodated unobserved common determinants consistently outperformed the model of independence for virtually every subsample of crashes considered in this paper. The findings clearly point to the presence of correlated unobserved factors that determine the crash injury severity outcomes of multiple vehicle occupants in a vehicle and that the degree of correlation varies by vehicle body type. Models that ignore or neglect the presence of such common unobserved factors provide poorer fit, and therefore inferior predictive power.

From a safety perspective, the findings of this paper have important implications. First and foremost, the findings suggest that crash studies that model or predict injury severity levels associated with a transportation facility should consider adopting approaches wherein the injury severity levels of all vehicle occupants are modeled simultaneously or jointly. In this way, one can obtain a more complete picture of the injury severity profile associated with a facility and devise counter-measures that address the entire profile of crash-related injuries. The use of a joint equations model such as that presented in this paper would allow one to do this while accounting for correlated unobserved factors that are often not present in safety data sets (*e.g.*, speed of travel prior to crash, vehicle condition).

In addition, model estimation results presented in the paper offer key insights into factors that affect crash severity levels for multiple occupants by seat position. The results have important implications for passenger safety and vehicle design. For example, consider that females in the front row are more likely to be severely injured than males. Females and older individuals in the front passenger seat are more likely to suffer severe injuries in comparison to other demographic groups. It would behoove the profession and vehicle manufacturers to consider enhancing safety devices and vehicular designs/ergonomics to better accommodate the physical characteristics of females and older individuals. The finding that higher injury severity levels are associated with riding in a sedan implies that vehicle manufacturers need to enhance safety features in smaller cars. Vehicle designs need to be enhanced to mimic the safety of the rear center seat position in other seat positions as well. Designs that can minimize vehicle rollover and occupant ejection from a vehicle would result in a decreased propensity to experience severe injury levels. Similarly, findings regarding the benefits of seatbelt use and the dangers associated with alcohol involvement confirm many of the previous findings reported in the literature. Safety and education campaigns aimed at raising awareness of these issues are likely to have a beneficial impact on reducing injury severity levels.

ACKNOWLEDGEMENTS
The authors acknowledge the helpful comments of five anonymous reviewers on an earlier version of the paper. The authors are grateful to Lisa Macias for her help in formatting this document.

REFERENCES

1. World Health Organization. *Global Status Report on Road Safety*. World Health Organization, Geneva, Switzerland, 2009.
2. NHTSA. 2008 Traffic Safety Annual Assessment – Highlights. *Traffic Safety Facts*. National Center for Statistics and Analysis, National Highway Traffic Safety Administration, Washington, D.C., 2009.
3. O'Donnell, C. J., and D. H. Connor. Predicting the Severity of Motor Vehicle Accident Injuries using Models of Ordered Multiple Choice. *Accident Analysis and Prevention*, Vol. 28, No. 6, 1996, pp. 739-753.
4. Khattak, A. J., P. Kantor, and F. M. Council. Role of Adverse Weather in Key Crash Types on Limited-Access Roadways. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1621, Transportation Research Board of the National Academies, Washington, D.C., 1998, pp. 10-19.
5. Ulfarsson, G. F., and F. L. Mannering. Differences in Male and Female Injury Severities in Sport-Utility Vehicle, Minivan, Pickup and Passenger Car Accidents. *Accident Analysis and Prevention*, Vol. 36, No. 2, 2004, pp. 135-147.
6. Eluru, N., and C. R. Bhat. A Joint Econometric Analysis of Seat Belt Use and Crash-Related Injury Severity. *Accident Analysis and Prevention*, Vol. 39, No. 5, 2007, pp. 1037-1049.
7. Hutchinson, T. P. Statistical Modelling of Injury Severity, with Special Reference to Driver and Front Seat Passenger in Single-Vehicle Crashes. *Accident Analysis and Prevention*, Vol. 18, No. 2, 1986, pp. 157-167.
8. Yamamoto, T., and V. N. Shankar. Bivariate Ordered-Response Probit Model of Driver's and Passenger's Injury Severities in Collisions with Fixed Objects. *Accident Analysis and Prevention*, Vol. 36, No. 5, 2004, pp. 869-876.
9. Lennon, A., V. Siskind, and N. Haworth. Rear Seat Safer: Seating Position, Restraint Use and Injuries in Children in Traffic Crashes in Victoria, Australia. *Accident Analysis and Prevention*, Vol. 40, No. 2, 2008, pp. 829-834.
10. Lund, U. J. The Effect of Seating Location on the Injury of Properly Restrained Children in Child Safety Seats. *Accident Analysis and Prevention*, Vol. 37, No. 3, 2005, pp. 435-439.
11. Berg, M. D., L. Cook, H. Corneli, D. D. Vernon, and J. M. Dean. Effect of Seat Position and Restraint Use on Injuries to Children in Motor Vehicle Crashes. *Pediatrics*, Vol. 105, No. 4, 2000, pp. 831-835.
12. Braver, E. R., R. Whitfield, and S. A. Ferguson. Seating Position and Children's Risk of Dying in Motor Vehicle Crashes. *Injury Prevention*, Vol. 4, No. 3, 1998, pp. 181-187.
13. Evans, L., and M. Frick. Seating Position in Cars and Fatality Risk. *American Journal of Public Health*, Vol. 78, No. 11, 1988, pp. 1456-1458.
14. Smith, K. M., and P. Cummings. Passenger Seating Position and the Risk of Passenger Death or Injury in Traffic Crashes. *Accident Analysis and Prevention*, Vol. 36, No. 2, 2004, pp. 257-260.

15. Smith, K. M., and P. Cummings. Passenger Seating Position and the Risk of Passenger Death in Traffic Crashes: A Matched Cohort Study. *Injury Prevention*, Vol. 12, No. 2, 2006, pp. 83-86.

16. Wang, X., and K. M. Kockelman. Occupant Injury Severity Using a Heteroscedastic Ordered Logit Model: Distinguishing the Effects of Vehicle Weight and Type. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1908, Transportation Research Board of the National Academies, Washington, D.C., 2005, pp. 195-204.

17. Claret, P. L, J. J. Jimenez-Moleon, J. de D. Luna-Del-Castillo, and A. Bueno-Cavanillas. Individual Factors Affecting the Risk of Death for Rear-Seated Passengers in Road Crashes. *Accident Analysis and Prevention*, Vol. 38, No. 3, 2006, pp. 563-566.

18. Mayrose, J., and A. Priya. The Safest Seat: Effect of Seating Position on Occupant Mortality. *Journal of Safety Research*, Vol. 39, No. 4, 2008, pp. 433-436.

19. Bhat, C. R. A Multi-Level Cross-Classified Model for Discrete Response Variables. *Transportation Research Part B*, Vol. 34, No. 7, 2000, pp. 567-582.

20. Bottai, M., N. Salvati, and N. Orsini. Multilevel Models for Analyzing People's Daily Movement Behavior. *Journal of Geographical Systems*, Vol. 8, No. 1, 2006, pp. 97-108.

21. Czado, C., and S. Prokopenko. Modeling Transport Mode Decisions Using Hierarchical Binary Spatial Regression Models with Cluster Effects. *Statistical Modeling*, Vol. 8, No. 4, 2008, pp. 315-345.

22. Bhat, C. R., and N. Eluru. A Copula-Based Approach to Accommodate Residential Self-Selection Effects in Travel Behavior Modeling. *Transportation Research Part B*, Vol. 43, No. 7, 2009, pp. 749-765.

23. Trivedi, P. K., and D. M. Zimmer. Copula Modeling: An Introduction for Practitioners. *Foundations and Trends in Econometrics*, Vol. 1, No. 1, 2007, pp. 1-110.

24. Genest, C., and R. J. MacKay. Copules Archimediennes et Familles de Lois Bidimensionnelles Dont les Marges Sont Donnees. *The Canadian Journal of Statistics*, Vol. 14, No. 2, 1986, pp. 145-159.

25. Sklar, A. Random Variables, Joint Distribution Functions, and Copulas. *Kybernetika*, Vol. 9, 1973, pp. 449-460.

26. Nelsen, R. B. *An Introduction to Copulas, 2nd ed*. Springer-Verlag, New York, 2006.

27. Clayton, D. G. A Model for Association in Bivariate Life Tables and its Application in Epidemiological Studies of Family Tendency in Chronic Disease Incidence. *Biometrika*, Vol. 65, No. 1, 1978, pp. 141-151.

28. Gumbel, E. J. Bivariate Exponential Distributions. *Journal of the American Statistical Association*, Vol. 55, No. 292, 1960, pp. 698-707.

29. Frank, M. J. On the Simultaneous Associativity of F(x, y) and x + y - F(x, y). *Aequationes Mathematicae*, Vol. 19, No. 1, 1979, pp. 194-226.

30. Joe, H. Parametric Families of Multivariate Distributions with Given Marginals. *Journal of Multivariate Analysis*, Vol. 46, No. 2, 1993, pp. 262-282.

31. Joe, H. *Multivariate Models and Dependence Concepts*. Chapman and Hall, London, UK, 1997.

32. McKelvey, R. D., and W. Zavoina. A Statistical Model for the Analysis of Ordinal Level Dependent Variables. *Journal of Mathematical Sociology*, Vol. 4, 1975, pp. 103-120.

33. Quinn, C. The Health-Economic Applications of Copulas: Methods in Applied Econometric Research. Health, Econometrics and Data Group (HEDG) Working Paper 07/22, Department of Economics, University of York, 2007.

34. Paleti, R., N. Eluru, and C. R. Bhat. Examining the Influence of Aggressive Driving Behavior on Driver Injury Severity in Traffic Crashes. Technical paper, Department of Civil, Architectural & Environmental Engineering, The University of Texas at Austin, August 2009.

35. Greene, W. *Econometric Analysis, 6th ed.* Prentice Hall, Upper Saddle River, 2008.

LIST OF TABLES

TABLE 1 Sample Characteristics

Occupant characteristics			
Male			52.1%
0-15 years			8.6%
16-20 years			17.1%
21-45 years			46.3%
46-65 years			20.2%
65+ years			7.8%
Wearing seat belt			92.0%
Vehicle Characteristics			
Vehicle type			
Sedan			59.2%
SUV			17.2%
Pickup truck			16.5%
Van			7.1%
Environment characteristics			
Time of crash			
12 AM - 6 AM			6.9%
6 AM - 9 AM			13.3%
9 AM - 3 PM			32.3%
3 PM - 7 PM			32.2%
7 PM - 12 AM			15.3%
Roadway Attributes			
Speed limit			
≤ 35 mph			41.2%
35 - 55mph			48.1%
55 + mph			10.7%
Crash Characteristics			
Number of Vehicles involved			
1 vehicle			22.2%
2 vehicles			77.8%
Crash Type			
Head-on			5.2%
Rear-end			30.5%
Side-swipe			5.9%
Angle			36.2%
Other (single vehicle/fixed object crashes)			22.2%
Injury Outcome	Driver	Passenger	Overall
No injury	66.0%	65.8%	65.9%
Possible injury	13.0%	15.0%	13.5%
Non-incapacitating injury	12.2%	11.6%	12.1%
Incapacitating + fatal injury	8.8%	7.6%	8.5%

TABLE 2 Vehicle Occupant Injury Severity Estimation Results

	Driver		Front passenger		Rear passenger	
Variable	**Parameter**	**t-stat**	**Parameter**	**t-stat**	**Parameter**	**t-stat**
Threshold parameters						
Threshold 1	0.0195	0.154	0.4646	1.466	0.8537	5.086
Threshold 2	0.4660	3.669	1.0116	3.199	1.3423	7.414
Threshold 3	1.1114	8.617	1.6895	5.247	1.9510	10.245
Occupant characteristics						
Male	-0.1644	-4.078	-0.1923	-2.483	---	---
Driver is Male	---	---	0.1499	1.905	---	---
Occupant age						
0-5 years	---	---	---	---	-0.6065	-4.453
16-20 years	-0.2710	-3.222	---	---	---	---
21-44 years	-0.1364	-1.955	---	---	---	---
45-64 years	-0.1364	-1.955	---	---	---	---
≥65 years			0.2649	1.944		
Driver age (base is “<45 years”)						
≥45 years	---	---	0.2873	3.222	---	---
Seat Belt used (base is seat belt not used)	-0.8629	-10.235	-0.5189	-3.016	-0.3072	-2.673
Driver under the influence of alcohol (base is “No alcohol influence”)	0.4067	4.797	---	---	---	---
Driver behavior characterized as “reckless” (base is “Not reckless”)	0.5432	1.939	---	---	---	---
Number of occupants = 5	-0.7578	-3.010	---	---	---	---
Vehicle characteristics						
Vehicle type (base is “all other vehicle types”)						
Sedan	0.2051	4.853	0.1885	2.407	0.4104	3.409
Vehicle type of colliding vehicle (base is “all other vehicle types”)						
Sedan	-0.1266	-2.761	---	---	---	---
Rear center indicator (base is “other rear seat configuration”)	---	---	---	---	-0.4146	-2.431
Environment factors						
Time of crash (base is “7:00 pm to 12:00 am”)						
12:00 am to 6:00 am	0.2405	3.373	---	---	---	---
6:00 am to 9:00 am	0.1320	2.260	-0.3378	-2.228	---	---
9:00 am to 3:00 pm	---	---	-0.2734	-2.321	---	---
3:00 pm to 7:00 pm	---	---	-0.2929	-2.530	---	---

Lighting condition (base is "normal lighting")						
Dark with lighting	---	---	-0.3535	-2.883	---	---
Dark	---	---	---	---	---	---
Adverse weather and road condition (base is "no adverse weather condition")						
Wet	-0.0992	-2.003	---	---	---	---
Snow	-0.0992	-2.003	-0.3154	-2.867	---	---
Ice	-0.0992	-2.003	---	---	---	---
Rain	---	---	-0.3154	-2.867	-0.2497	-1.621
Roadway attributes						
Speed limit (base is "≤ 35 mph")						
35 - 55mph	0.1779	4.097	0.2254	2.681	0.2136	1.669
>55mph	0.3120	4.298	0.2254	2.681	0.6242	3.201
Traffic way without median	0.0920	1.975	0.1893	2.261		
Roadway alignment						
Curved road	0.1814	2.878	0.3905	3.062	---	---
Crash characteristics						
Vehicle rolled over (base is "No rollover")	0.8271	7.612	0.8150	4.150	0.9008	3.387
Occupant ejected from the vehicle	1.4659	2.772	0.6602	0.998	---	---
Crash with a stationary object (base is "crash with another vehicle")	0.3919	5.572	0.5840	2.233	---	---
Role of vehicle in two vehicle crashes (base is "vehicle strikes the other vehicle")						
Contacted vehicle	0.1452	2.943	0.2782	3.022	---	---
Both striking & contacted vehicle	0.6340	4.447	0.5867	2.663	---	---
Type of Collision (base is "other" type of crashes)						
Head on	1.1335	12.029	1.3135	4.501	1.1508	4.169
Rear-end	---	---	0.3072	1.194	---	---
Side-swipe	-0.2848	-2.546	---	---	---	---
Angle	0.4230	8.014	0.7260	2.921	0.4169	3.315

TABLE 3 Disaggregate Measures of Fit in Validation Sample

Sample details	Number of observations	ORP Predictive likelihood	PF Predictive likelihood	Predictive likelihood ratio test ($\chi^2_{4,0.05} = 9.49$)
Full validation sample	1000	-1213.02	-1191.53	42.98
Number of occupants				
One	759	-643.70	-645.29	-3.17
Two	154	-301.70	-298.54	6.31
Three	54	-133.63	-126.10	15.06
Four	27	-101.17	-85.99	30.36
Five	6	-32.83	-35.61	-5.57
Number of vehicles				
One	198	-266.81	-256.87	19.86
Two	902	-946.21	-934.66	23.12
Vehicle type				
Sedan	574	-725.14	-717.17	15.94
SUV	176	-238.77	-227.46	22.61
Speed limit				
≤ 35 mph	394	-424.35	-420.40	7.90
35-55 mph	474	-593.95	-572.69	42.52
> 55 mph	132	-194.72	-198.44	-7.42
Collision type				
Head-on	49	-85.89	-84.77	2.24
Rear-end	319	-323.03	-308.34	29.37