

**Pooling Data from Fatality Analysis Reporting System (FARS) and Generalized Estimates System (GES) to Explore the Continuum of Injury Severity Spectrum**

**Shamsunnahar Yasmin**

Department of Civil Engineering & Applied Mechanics  
McGill University

Ph: 514 398 6823, Fax: 514 398 7361

Email: [shamsunnahar.yasmin@mail.mcgill.ca](mailto:shamsunnahar.yasmin@mail.mcgill.ca)

**Naveen Eluru\***

Department of Civil, Environmental and Construction Engineering  
University of Central Florida

Ph: 407 823 4815; Fax: 407 823 3315

Email: [naveen.eluru@ucf.edu](mailto:naveen.eluru@ucf.edu)

**Abdul R. Pinjari**

Department of Civil and Environmental Engineering  
University of South Florida

Ph: 813 974 9671, Fax: 813 974 2957

Email: [apinjari@eng.usf.edu](mailto:apinjari@eng.usf.edu)

\*Corresponding author

## ABSTRACT

Fatality Analysis Reporting System (FARS) and Generalized Estimates System (GES) data are most commonly used datasets to examine motor vehicle occupant injury severity in the United States (US). The FARS dataset focuses exclusively on fatal crashes, but provides detailed information on the continuum of fatality (a spectrum ranging from a death occurring within thirty days of the crash up to instantaneous death). While such data is beneficial for understanding fatal crashes, it inherently excludes crashes without fatalities. Hence, the exogenous factors identified as critical in contributing (or reducing) to fatality in the FARS data might possibly offer different effects on non-fatal crash severity levels when a truly random sample of crashes is considered. The GES data fills this gap by compiling data on a sample of roadway crashes involving all possible severity consequences providing a more representative sample of traffic crashes in the US. FARS data provides a continuous timeline of the fatal occurrences from the time to crash – as opposed to considering all fatalities to be the same. This allows an analysis of the survival time of victims before their death. The GES, on the other hand, does not offer such detailed information except identifying who died in the crash. The challenge in obtaining representative estimates for the crash population is the lack of readily available “appropriate” data that contains information available in both GES and FARS datasets. One way to address this issue is to replace the fatal crashes in the GES data with fatal crashes from FARS data thus augmenting the GES data sample with a very refined categorization of fatal crashes. The sample thus formed, *if statistically valid*, will provide us with a reasonable representation of the crash population.

This paper focuses on developing a framework for pooling of data from FARS and GES data. The validation of the pooled sample against the original GES sample (unpooled sample) is carried out through two methods: (1) univariate sample comparison and (2) econometric model parameter estimate comparison. The validation exercise indicates that parameter estimates obtained using the pooled data model closely resemble the parameter estimates obtained using the unpooled data. After we confirm that the differences in model estimates obtained using the pooled and unpooled data are within an acceptable margin, we also simultaneously examine the whole spectrum of injury severity on an eleven point ordinal severity scale – no injury, minor injury, severe injury, incapacitating injury, and 7 refined categories of fatalities ranging from fatality after 30 days to instant death – using a nationally representative pooled dataset. The model estimates are augmented by conducting elasticity analysis to illustrate the applicability of the proposed framework.

Keywords: Fatality, Fatality Analysis Reporting System (FARS), Generalized Estimates System (GES), Data Pooling, Generalized ordered logit model

## 1. INTRODUCTION

Traffic crashes result in physical and emotional trauma as well as huge financial losses for the individuals involved, their families and the society at large. Across the world, these crashes account for 18 deaths and 1,136 disability-adjusted life years (DALY) lost per 100,000 individuals annually (WHO, 2013a; WHO, 2013b). Researchers and practitioners are constantly seeking remedial measures to reduce the burden of these unfortunate events. Towards this end, literature in transportation safety has evolved along two major streams: the first stream of research is focused on identifying attributes that result in traffic crashes and propose means to reduce the occurrence of traffic crashes (see Lord and Mannering (2010) for a review of these studies); the second stream of work examines crash events and identifies factors that impact the crash outcome and suggests countermeasures to reduce crash related consequences (injuries and fatalities) (see Savolainen et al. (2011) and Yasmin and Eluru (2013) for a review). The current research study contributes to the second stream of literature with a specific focus on driver injury severity analysis.

A number of studies have explored the impact of various factors on vehicle occupant injury severity at disaggregate level (see Bédard et al., 2002; Fredette et al., 2008; and Yasmin and Eluru, 2013 for a detailed review). These studies can broadly be categorized as: a) studies that focus exclusively on crashes involving only fatalities (employing a sample of crashes involving fatalities) and b) studies that examine crashes that involve all levels of injury severity – ranging from no injury to fatality (employing a random sample of traffic crashes that compile different levels of injury severity). In the United States (US), the former category of studies predominantly use the Fatality Analysis Reporting System (FARS) database (see Evans and Frick, 1988; Preusser et al., 1998; Zador et al., 2000; Gates et al., 2013) while the latter group of studies primarily employ the General Estimates System (GES) database (see Kockelman and Kweon, 2002; Eluru and Bhat, 2007; Yasmin and Eluru, 2013).

The FARS database is a census (not a sample) of all fatal crashes in the US; i.e., crashes that lead to at least one fatality within thirty consecutive days from the time of crash. The GES database, on the other hand, comprises a sample of road crashes across the US involving at least one motor vehicle travelling on a roadway and resulting in property damage, injury or death to the road users. The two datasets employed in the safety literature have their own advantages and limitations. The FARS focuses exclusively on fatal crashes. Therefore, one cannot reliably use this data to analyze the factors that increase or decrease the probability of fatality (because the data does not include crashes that do not lead to fatalities). The GES fills this gap by compiling data on a sample of roadway crashes involving all possible severity consequences (no injury, possible injury, non-incapacitating injury, incapacitating injury and fatality) providing a more representative sample of traffic crashes in the US. One of the advantages of FARS, however, is that the collected information includes the date and time of occurrence of the fatalities resulting within a 30-day time period from the crash. This detailed information provides us a continuous timeline of the fatal occurrences from the time to crash (instead of considering all fatalities to be the same). This allows for an analysis of the survival time of victims before their death. The GES, on the other hand, does not offer such detailed information except identifying who died in the crash.

Examining the impact of various exogenous factors on all levels of injury severity as well as on the survival time of fatalities can potentially play a critical role in field triage - screening process to determine the more severe cases. Preclinical trauma care is one of the most important factors affecting the outcome of motor vehicle crash (MVC) victims (Chalya et al., 2012; Palanca

et al., 2003). In prehospital setting, along with the anatomic and physiological conditions of MVC victims, different mechanism-of-injuries (vehicle intrusion, occupant ejection, vehicle telemetry and death in same passenger compartment) are also considered by emergency medical service (EMS) personnel as conditions for trauma triage of victims (Sasser et al., 2012; Isenberg et al., 2011). In fact, it is evident from previous studies (Stewart, 1990) that prolonging survival beyond the first hour can potentially help avoid fatality with proper preclinical care. Hence, a refined specification of fatality might allow us to identify potential survivors that might benefit by providing pre- and post-hospital treatment.

In an effort to identifying exogenous factors that help in prolonging survival time, using detailed information available in FARS data, Yasmin et al. (2015) examined fatal crashes from a new perspective. The authors recognize that fatality is an aggregation of a continuous spectrum ranging from dying instantly to dying within thirty days of crash (as reported in the FARS data). Keeping all else same, a fatal crash that results in an immediate fatality is clearly much more severe than another crash that leads to fatality after several days. Therefore, it is useful to explicitly recognize the different levels of severity among fatal crashes. Such refined definition of fatal crashes, as opposed to lumping all fatal crashes into a single category, allows one to differentiate fatal crashes based on the survival time and to derive insights on factors that can prolong survival time. A disadvantage of the study by Yasmin et al. (2015) is that, as discussed before, the FARS dataset focuses exclusively on fatal crashes. While using the FARS data is very helpful for understanding the differences across different fatal crashes, it inherently excludes crashes with other possible, non-fatal injury severity outcomes. This makes it difficult to generalize the findings to the overall crash population. Besides, while analyzing the survival time of only fatal crash victims (using FARS data) helps in deriving the influence of various exogenous factors on survival time conditional upon the occurrence of a fatality, it doesn't allow the analyst to derive the influence of those factors in increasing the chances of survival. This is because the FARS data doesn't provide a representative sample of non-fatal crashes.

One way to address this issue is combining information from both the FARS and GES datasets into a single, disaggregate crash-level database<sup>1</sup>. This will bring together the strengths of both datasets – the representativeness of crashes with all injury severity outcomes from the GES data and the detailed information on fatal crashes from the FARS data. The challenge, however, lies in combining the two datasets in a statistically appropriate way. Since FARS is a census of all fatal traffic crashes in the US, all fatal crashes in the GES sample for a year should be available in the FARS data for that year. Now, if one could identify these crashes directly, it would be easy to augment the fatal crash records in GES with the detailed information from FARS. However, there is no mechanism to easily link crashes across these two databases because the datasets do not have a common identifier. Hence, an alternative, statistically valid method needs to be used for fusing information from both the datasets.

The approach is a proof of concept investigation of data pooling from two datasets while ensuring statistical validity. While, there could be various other alternative datasets for such investigation, given the extensive use of GES and FARS datasets in safety literature, they serve as good candidates for the research exercise. In this context, this paper is geared towards addressing the challenge of pooling data from GES and FARS databases. While several approaches exist in

---

<sup>1</sup> To be sure, the reader would note that there have been compilation of GES and FARS datasets to obtain the Annual Traffic Safety Facts (see NHTSA, 2012). However, in these efforts, there is no attempt to pool disaggregate level data from the two sources. The report provides trends separately for FARS and GES datasets. Further, in our research, we examine the effect of exogenous variables on severity in pooled and unpooled data.

the literature to fuse information from different data sources without a common identifier (Konduri et al., 2011; Sivakumar and Polak, 2013), a simple approach is to replace fatal crashes from the GES sample by a random sample from the FARS census of crashes. We conduct statistical tests to assess if this approach suffices for the purpose of developing a database that allows us to examine the whole spectrum of injury severity ranging from no injury to fatality, along with differentiating fatal crashes based on survival time. Moreover, the simultaneous interpretation of information would allow researchers to provide recommendations using a single modeling framework, rather than making inferences from the results of separate econometric models from different datasets.

In summary, the current research makes a fourfold contribution to the literature on vehicle occupant injury severity analysis. *First*, we propose and test the efficacy of a simple yet statistically valid approach to fuse the FARS and GES datasets into a single, disaggregate crash level database that combines information from both the datasets. *Second*, we employ a sampling design approach for generating a nationally representative pooled sample of all crashes. *Third*, the Generalized Ordered Logit (GOL) model (also referred to as Partial Proportional Odds model) is employed on the pooled dataset to analyze the influence of a variety of exogenous factors on traffic crash injury severity, while considering a very refined characterization of fatal crashes along with other, non-fatal injury severity outcomes. *Finally*, we compute elasticity measures to identify important factors affecting driver injury severity outcomes.

The rest of the paper is organized as follows. The data source and sample formation are presented in Section 2. Section 3 provides details of the approach used for pooling data from FARS and GES. Section 4 presents the empirical analysis along with a statistical assessment of the proposed approach to fuse information from both data sources. The estimation results of the GOL model are described in Section 5. The elasticity effects are presented in section 6 and section 7 concludes the paper.

## 2. DATA SOURCE AND SAMPLE FORMATION

The data for the current study is sourced from the FARS and GES databases for the year 2010. FARS data is a census of all fatal crashes in the US and compiles crashes if at least one person involved in the crash dies within thirty consecutive days from the time of crash. The FARS database has a record of 30,196 fatal crashes with 32,885 numbers of fatalities involving 74,863 road users for the year 2010. The GES database is a nationally representative weighted stratified sample of road crashes collected and compiled from about 60 jurisdictions across the US. It includes information of reports compiled by police officers for crashes involving at least one motor vehicle travelling on a roadway and resulting in property damage, injury or death. The GES crash database has a record of 46,391 crashes involving 81,406 motor vehicles and 116,020 individuals for the year of 2010. These databases are obtained from the US Department of Transportation, National Highway Traffic Safety Administration's National Center for Statistics and Analysis (<ftp://ftp.nhtsa.dot.gov>) and provide information on a multitude of factors (driver characteristics, vehicle characteristics, roadway design and operational attributes, environmental factors, crash characteristics and situational variables) representing the crash situation and events. The reader would note that the exogenous variable information available in FARS and GES datasets are very similar making it relatively easier to fuse the fatality information from FARS into the GES data.

This study is focused on injury severity outcome of passenger vehicles' drivers who were involved in either a single or two vehicle crashes. The crashes that involve more than two vehicles are excluded from both FARS and GES datasets. Commercial vehicles involved collisions are also

excluded in order to avoid the potential systematic differences between commercial and non-commercial driver groups. In order to prepare the final FARS dataset, crash records involving non-motorized road users (19,670 records), commercial vehicles (17,795 records), records of passengers and crashes involving more than two vehicles (18,073 records), non-fatal crash records of drivers (8,012 records) and records with missing information for essential attributes (2,468 records) are deleted. Thus, the final FARS dataset consisted of 8,845 records. From the continuous timeline of the fatal occurrences, a seven point discrete ordinal variable is created to represent the scale of fatal injury severity of drivers involved in these crashes - from least severe to most severe fatal crashes (and their proportions): 1) Died between 6 and 30 days of crash (6.0%), 2) Died between 2 and 5 days of crash (5.2%), 3) Died between 7 and 24 hours of crash (4.4%), 4) Died between 2 and 6 hours of crash (21.6%), 5) Died between 31 and 60 minutes of crash (14.5%), 6) Died between 1 and 30 minutes of crash (20.1%) and 7) Died instantly (28.3%) (see Yasmin et al. (2015) for a similar fatality continuum representation).

In order to prepare the final GES dataset, crash records involving non-motorized road users and commercial vehicles (34,808 records), records of passengers and crashes involving more than two vehicles (32,824 records), and records with missing information for essential attributes (23,094 records) are deleted. Thus, the final GES dataset consisted of about 25,294 records. From this dataset, a sample of 6,062 records is randomly sampled out for the purpose of validating pooled models. The reader would note that the simple random sampling process was employed for the validation exercise to reduce the computational time necessary to validate and compare the models described subsequently. A five point ordinal scale is used in the database to represent the injury severity of individuals involved in these crashes. In the validation sample, the distributions of driver injury severities are as follows: No injury 63.7%, Possible injury 14.0%, Non-incapacitating injury 13.1%, Incapacitating injury 8.2% and Fatal injury 1.0%. However, GES is a probability sample of police reported traffic crashes. A weight variable is associated with each record of this stratified sample to represent the national crash trend. Therefore, we also select a sample of 19,181 records from the final dataset with 25,294 records by using proportionate sampling method for the purpose of estimating models to produce national estimates. In the weighted estimation sample, the distribution of driver injury severity are as follows: No injury 83.7%, Possible injury 10.0%, Non-incapacitating injury 5.0%, Incapacitating injury 1.2% and Fatal injury 0.1%.

### **3. RESEARCH FRAMEWORK**

In the current research effort, we employ the Generalized Ordered Logit (GOL) or the partial proportional odds logit model (see Eluru and Yasmin, 2015; Yasmin and Eluru, 2013; Eluru, 2013 and Mooradian et al., 2013 for a detailed description of the econometric framework) to examine the driver injury severity by using pooled dataset from FARS and GES. The injury severity variable is analyzed using the ordered outcome framework to recognize the inherent ordinality of the injury severity levels. The traditional ordered outcome models (ordered logit and ordered probit) restrict the impact of exogenous variables on the outcome process to be same across all alternatives (Eluru et al., 2008). Recent research (Eluru, 2013; Eluru et al., 2008) has addressed this limitation by allowing the analyst to estimate individual level thresholds as function of exogenous variables as opposed to retaining the same thresholds across the population (as is the case in standard ordered logit). However, the prerequisites for any data pooling exercise are that different sources to be pooled are comparable (Verma et al., 2009) and share a common data generation process (Louviere et al., 1999). This section presents a roadmap to pool information from both the data sources and

the tests used to assess if the pooled data represents a common data generation process for the individual data sources. A conceptual diagram of the research methodology employed in validating the pooled estimates is provided in Figure 1. Further, this section also presents a sampling design of the pooled dataset to produce representative estimates of driver injury severity levels.

### 3.1 Testing Data Pooling Exercise

The GES dataset has a five point ordinal scale to represent injury severity while a seven point ordinal scale is defined to distinguish the severity of different fatal crashes based on the survival time. In this study, we form the pooled dataset by replacing the fatal crash records in GES with a random sample of crashes in FARS. In the pooled dataset we can generate an eleven point ordinal representation of injury severity, with 4 categories for non-fatal crashes and 7 categories for fatal crashes ( $5 + 7 - 1$ ). Prior to developing models to analyze the newly generated injury severity scale, it is imperative that we validate the pooled dataset. As the actual data generation process is latent we have to resort to comparing the pooled dataset with the unpooled dataset. In our pooling exercise, the records from FARS are being added to the GES data, the evaluation would be geared towards comparing the pooled data with the original GES data (unpooled data). Specifically, we undertake comparison of the pooled sample with the unpooled sample in two ways: (1) univariate sample comparison, by simply comparing the distributions of the variables in the two samples and (2) econometric model estimate comparison. The validation of pooling exercise is done by using the GES validation sample with 6,062 records<sup>2</sup>.

While the descriptive comparison of pooled and original samples is relatively straight forward, the more challenging task is to perform a more statistically rigorous analysis to examine if the crash records from FARS can replace those in the GES data. For this purpose, as a *first step*, we estimate the injury severity model using the original GES validation data and compare the model estimates with the injury severity model estimated from the pooled dataset – while maintaining the same number of injury severity categories in the GES and pooled datasets. To do so, all the fatal records pooled from FARS into the GES sample were categorized as fatal (i.e., a single category) regardless of the survival time of the victims. The pooled data sample is obtained by removing the 59 fatal records in the GES sample of 6,062 records.

To statistically ensure the validity of our comparison results and to ensure that the statistical results obtained from the pooled samples are stable, we consider multiple samples of fatal crash data from FARS to replace fatalities in GES. Specifically, for testing the validity of the pooled data, 15 data samples – 5 samples of about 2,000 records; 5 samples of about 3,000 records and 5 samples of about 5,000 records – are randomly generated from the 8,845 records of FARS database and combined with the GES data to form pooled data. These 15 data samples along with the full sample (of 8,845 records) from FARS dataset are used to generate 16 different sets of pooled databases. The fatal records replaced in GES by the FARS fatal records in these 16 samples are presented in Table 1. GOL models of injury severity are estimated for these 16 pooled samples under the five point ordinal scale system and compared with the GOL model parameters obtained

---

<sup>2</sup> At this juncture, it is important to highlight that the validation exercise does not consider the weights available in the GES dataset. However, as the records are being added to GES from FARS to create the pooled sample and validated against the unpooled sample, not considering weights does not affect the findings of the validation exercise. The econometric models are estimated for the pooled and unpooled models with the same weight distribution. Hence, the comparison is valid. However, the model estimates from these pooled models are not nationally representative and hence econometric models for weighted datasets are also estimated.

using unpooled GES data to ensure that the estimates have not been altered significantly due to the newly added records.

### 3.2 Weight Variable for Pooling

The reader would note, from Table 1, that the GES (unpooled) database of validation sample has a very small percentage of fatalities. This is because the percentage of fatal crashes is small compared to all other crashes. As our primary objective is examining the impact of exogenous variables on seven categories of the fatality spectrum (based on survival time) it is useful to oversample the fatal crashes from FARS. Otherwise, we are likely to have very small number of records for each of the fatal injury severity alternatives. Of course, the oversampling of fatalities from FARS to replace GES fatalities necessitates creating an appropriate weight variable to weight the pooled data. This approach ensures that the distribution of the injury severity variable in the pooled data is the same as that in the GES data. Therefore, to generate the pooled sample, we remove the fatal crashes ( $m_i$ ) from the GES sample and replace it with fatal cases ( $n_i$ ) from the FARS along with a specific weight  $\omega_{FG}$  computed as  $\frac{m_i}{n_i}$ . Specifically, a weight of  $\omega_{FG}$  is assigned to the FARS crash records (that replace the GES fatalities) in the pooled samples while the other non-fatal crash records (from GES validation sample) were weighted by 1. The associated weights for 16 different pooled samples are shown in Table 1.

### 3.3 Severity Parameter Comparison Exercise

The 16 pooled data samples created with appropriate weights are employed to generate injury severity parameter estimates. The parameter estimates obtained using the pooled data are compared with that of the original GES parameter estimates obtained using unpooled data (i.e., the original GES validation data) by computing the percentage error (considering parameter estimates from unpooled data as the base case). Then, a hypothesis test that the parameters are obtained from the same distribution (i.e.,  $\beta_P = \beta_{UP}$ <sup>3</sup> where  $P$ =Pooled and  $UP$ =Unpooled) is carried out to examine the differences between parameter estimates. If this hypothesis is rejected, the estimates from pooled model represent estimates from a dissimilar latent data generation process (Bass and Wittink, 1975). On the contrary, if the hypothesis is not rejected, it will provide support that the proposed pooling of GES and FARS datasets has not altered the distribution of the parameters and that the pooling process is statistically valid. The percentage error in parameter estimates and the hypothesis tests are first computed separately for each of the 16 pooled data samples. Subsequently, for ease of presentation, we present and discuss the average measures from each sample type – 1 pooled sample with 8,845 records from FARS (sample 16<sup>th</sup> of Table 1); 5 pooled samples with about 5,000 records from FARS (samples 11-15<sup>th</sup> of Table 1); 5 pooled samples with about 3,000 records from FARS (samples 6-10<sup>th</sup> of Table 1); and 5 pooled samples with about 2,000 records from FARS (samples 1-5<sup>th</sup> of Table 1).

---

<sup>3</sup> To test the hypothesis that  $\beta_P = \beta_{UP}$ , we need to obtain the distribution of  $(\beta_P - \beta_{UP})$ . The standard error for the distribution is obtained as  $\sqrt{SE_P^2 + SE_{UP}^2}$  where  $SE_P$  and  $SE_{UP}$  represent standard errors of the parameters obtained using pooled and unpooled data respectively. Then, one can simply do a t-test on  $(\beta_P - \beta_{UP})$ . That is, if the ratio of the estimate of  $(\beta_P - \beta_{UP})$  to its standard error is less than the critical t-value at a chosen confidence level, then one cannot reject the hypothesis that  $\beta_P = \beta_{UP}$ .

### **3.4 Performance Evaluation of Eleven Point Pooled Model**

After we confirm that the differences in model estimates from the five point ordinal models are within an acceptable margin, we can employ the pooled data to estimate an injury severity model with an eleven point severity scale with 4 categories of injury severity for non-fatal crashes and 7 categories for fatal crashes. However, another issue that needs to be addressed before estimating the eleven point scale ordinal model is developing a statistical approach to determine if the eleven point ordinal model is an improvement on the five point ordinal model (with all fatal crashes lumped into a single category). Due to the nature of the log-likelihood measure employed in model estimation, increasing the resolution will lead to deterioration of model log-likelihood. Hence, comparing log-likelihoods between a five alternative model and eleven alternative model is not statistically valid. Interestingly, we could not find any method in literature to make a meaningful comparison of models with different resolutions of dependent variable definitions. Hence, we developed an approach based on first principles to address this issue and compare the performance of eleven point ordinal model with that of five point ordinal model.

### **3.5 Sampling design for Population Representative Estimates**

The pooled estimates using an unweighted GES sample is not representative of the population. Thus, it is important to incorporate the associated “weight” variable of GES data sample in the pooled dataset in order to produce nationally representative estimates. Using the weight of GES in the non-fatal categories of pooled data is straightforward as these records are directly drawn from the GES database. However, it is a challenge to incorporate the weight for the fatal crashes in the pooled sample after replacement from FARS. To address this issue, we employ a two-step approach in designing a nationally representative pooled sample of all crashes. First, we generate a nationally representative GES sample, which is yet unpooled, by employing proportional sampling strategy<sup>4</sup> (see Paleti et al. (2010) for a similar approach). This approach through choice based sampling obtains a sample of GES records that closely match the weighted shares of the GES sample. The approach ensures that the estimates obtained from the weighted sample are not different from the estimates obtained from the choice based sample. The approach allows us to avoid the consideration of weight in the model estimation process and thus makes replacement relatively straight forward.

Once we ensure appropriate representation of the national crash profile, in the second step, we replace the fatal crashes of the proportionate GES sample (26 records) with a random sample of fatal crashes from FARS along with weight variable for pooling as described in Section 3.2. The proposed method allows us to circumvent the need to operate with two weight variables and offers a statistically easier alternative to generating a nationally representative dataset. The pooled data sample generated by using the above steps is then used for the purpose of estimating the final driver injury severity model. For our analysis we chose one sample from the 16 different pooled data samples. The chosen sample has 2,967 randomly sampled records from the FARS data to replace the 26 fatal records from GES and the remaining 19,155 records from the GES data.

---

<sup>4</sup> The reader is referred to Schutt et al. (2011) for a detailed discussion on proportional stratified sampling approach.

## 4. EMPIRICAL ANALYSIS

### 4.1. Variables Considered

In our analysis, to estimate models using pooled data, we prepared the datasets such that both GES and FARS datasets have exactly the same set of independent variables. We selected a host of variables from six broad categories: Driver characteristics (including driver gender, driver age, restraint system use, alcohol consumption and physical impairment), Vehicle characteristics (including vehicle type and vehicle age), Roadway design and operational attributes (including speed limit, types of intersection and traffic control device), Environmental factors (including time of day, lighting condition, day of week and road surface condition), Crash characteristics (including collision object, manner of collision, collision location and trajectory of vehicle's motion), and Situational variables (including number of passengers and driver ejection). It should be noted here that several variables such as presence of shoulder, shoulder width, point of impact, roadway class and number of lanes could not be considered in our analysis because either the information was entirely unavailable or there was a large fraction of missing data for these attributes in the dataset. To be sure, we employ the manner of collision and speed limit variables as surrogates for point of impact and roadway class, respectively. In the final specification of the model, statistically insignificant variables were removed. The reader would note that the pooling exercise was undertaken using the variables that are common to both datasets. Hence, variables such as emergency crew arrival times were not considered in our models as they are unavailable in GES data.

### 4.2. Validation Exercise of Pooled Data

The first step in the validation exercise was to examine the similarities and dissimilarities in independent variables across the pooled and unpooled samples (to be sure, the validation exercise is done by using the validation GES sample). In the comparison, we found that the exogenous factor distributions of all pooled datasets (16 datasets) are almost the same. For the sake of brevity we chose to present the results for one sample only. The sample characteristics of the exogenous factors of unpooled and one pooled (weighted) dataset are presented in Table 2. Overall, we find that the characteristics of the pooled and unpooled samples across the entire sample (in columns 2 and 3) and across fatal crashes (columns 4 and 5) are very similar. We observe that there are slightly higher proportions of driving under the influence of alcohol and negotiating curves among the fatal crashes in the pooled data than those in the unpooled data. Also, the proportions for fatal crashes in the pooled dataset are marginally lower for two way traffic-with median and for vehicle age 6-10 years. It is not unanticipated that pooling would introduce such minor differences between the datasets.

In the second step of our validation, a comparison exercise between the parameter estimates obtained using unpooled and pooled data is also carried out by using 16 different pooled samples. The reader would note that a direct comparison of parameter estimates is considered only for illustrative purposes. A more rigorous statistical approach is also undertaken. The percentage errors in injury severity parameter estimates obtained using pooled datasets compared to parameter estimates obtained using unpooled data are presented in Figure 2 for all the variables (variable numbers are defined in Appendix A along with the injury severity estimates obtained using the unpooled model). From this plot, we can see that, among 44 variables in the final models, 32

variables have an error percentage lower than 10%, 8 variables have an error percentage between 10 and 25% and 4 variables have an error percentage higher than 25%. Overall, for such highly non-linear models such as GOL, estimated using two datasets, these are reasonably small differences.

To undertake a more rigorous statistical comparison, we test the hypothesis that the parameter estimates obtained using the pooled and unpooled datasets are not systematically different and the observed numerical differences can be accounted by the randomness in data samples. The test values of the homogeneity hypothesis test ( $\beta_P = \beta_{UP}$ ) between parameter estimates obtained using unpooled and pooled datasets are plotted against the variable numbers and is presented in Figure 3. From this plot, we can clearly see that the test statistics lie within the bounds +1.96 and -1.96 (critical t-stats at 95% confidence level). In fact, the largest difference is less than 1 indicating that there is no systematic difference in the estimates from pooled and unpooled models. This same trend can be observed for all types of pooled data samples with different numbers of FARS records in the pooled data. Thus, we can find no evidence to reject the hypothesis that the severity parameter estimates obtained using pooled data and the severity parameter estimates obtained using unpooled data follow different distribution. Based on our comparison of descriptive statistics and severity parameter estimates, we can argue that there is no evidence to suggest that the data pooled from GES and FARS results from a distinct latent data generation process than that in GES.

### 4.3. Metric for Comparing Eleven Point Model with Five Point Model

The second step in the validation exercise was to develop a statistical approach to determine if the eleven point ordinal model is an improvement on the five point ordinal model. In a five point ordinal scale model all fatalities are treated equally i.e. there is no distinction across fatal crashes. So in a five alternative model, we implicitly assume that the seven fatality groups considered in the eleven alternative model are all equally likely. Recognizing this assumption, one could generate an equivalent eleven alternative log-likelihood based on the five alternative model log-likelihood value. This can be compared with the log-likelihood of the eleven alternative model that differentiates between the various fatality classes.

The exact equation for the computation of log-likelihood takes the following form:

$$L = \sum_{i=1}^N \left[ \left( \sum_{j=1}^4 (\log P_i(j))^{d_{ij}} \right) + \frac{1}{7} * \omega_{FG} * (\log P_i(J))^{d_{i5}} \right]$$

where,  $\omega_{FG}$  is the weight,  $i$  be the index for drivers ( $i = 1, 2, \dots, N$ ),  $j$  be the index for driver injury severity levels ( $j = 1, 2, \dots, J$ ),  $P_i(j)$  represents the probability of injury severity level  $j$ , and  $d_{ij}$  is a dummy variable taking the value 1 if the driver  $i$  sustains an injury of level  $j$  and 0 otherwise. Once the equivalent log-likelihood is generated based on the above equation, one could easily employ the likelihood ratio (LR) test to check if the eleven point ordinal scale model offers additional improvement. The LR test statistic is defined as  $2 * (LL_{11} - LL_5)$  where  $LL_{11}$  and  $LL_5$  represent log-likelihood values at convergence of the eleven point and equivalent five point ordinal models, respectively. The LR test statistic thus computed is compared with the chi-square distribution value of  $k$  degrees of freedom where  $k$  corresponds to the additional parameters in the unrestricted model. In our case, for all samples, the additional number of parameters is 6. Hence,

if the LR test statistic is larger than the  $\psi^2$  value for 6 degrees of freedom, we can conclude that the considering fatality as multiple states enhances the data fit.

The log-likelihood values along with the LR test statistic for the equivalent and the actual eleven point models for various samples are presented in Table 3. The resulting LR test values for the comparison of equivalent/actual eleven point models for all sample types are more than 23 indicating the actual eleven point model outperforms the equivalent eleven point model at any reasonable level of statistical significance. The consistent improvement offered by the pooled model clearly indicates that the refined categorization of fatal injury crashes improves the model fit and provides more information to the model for examining the injury severity outcome. This is of particular relevance to this empirical exercise because fatal crashes comprise a very small portion of our sample (only 1%) – thus by introducing further disaggregation of an alternative with such a small sample share, there was a risk of worsening the model.

## 5. ESTIMATION RESULTS

The driver injury severity model of the nationally representative pooled data sample for the eleven point ordinal injury severity categorization is discussed in this section. To reiterate, the dependent variable under consideration is the eleven point ordinal variable defined as: no injury, possible injury, non-incapacitating injury, incapacitating injury, and 7 categories within fatal crashes - died between 6 and 30 days of crash, died between 2 and 5 days of crash, died between 7 and 24 hours of crash, died between 2 and 6 hours of crash, died between 31 and 60 minutes of crash, died between 1 and 30 minutes of crash and died instantly. The estimation results are presented in Table 4. In GOL model, when the threshold parameter is positive (negative), the result implies that the threshold is bound to increase (decrease); the actual effect on the probability is quite non-linear and can only be judged in conjunction with the influence of the variable on propensity and other thresholds. In the following sections, the estimation results are discussed by variable groups.

Driver Characteristics: In the category of driver characteristics, the result for driver gender indicates higher injury risk propensity for female drivers compared to male drivers. The effect of this variable is also significant for the threshold demarcating possible and non-incapacitating injury. The positive sign of the coefficient in the threshold indicates higher likelihood of possible injury for the female drivers. The result perhaps is indicative of the lower physiological strength of female drivers (compared to male drivers) in withstanding the impact of a crash (Xie et al., 2009; Chen and Chen, 2011). The age of drivers involved in the collision also has a significant influence on injury severity. As found in previous studies (Xie et al., 2012; O'Donnell and Connor, 1996; Castro et al., 2013), the parameter characterizing the effect of young driver (age<25) suggests a reduction in the likelihood of severe injuries compared to middle-aged drivers (age 25 to 64). However, the estimation result indicates that compared to the middle aged driver, the latent injury propensity is higher for older drivers (age≥65).

As expected, injury risk propensity is higher for the drivers not wearing seat belts relative to the drivers using seat belts (see Obeng, 2008; Yau, 2004; Yasmin et al., 2012; Eluru and Bhat, 2007 for a similar result). At the same time, the negative value of the threshold demarcating the possible and non-incapacitating injury of unrestrained driver reflects lower likelihood of possible injuries and, in general, higher likelihood of dying instantly for those drivers. The result related to drunk driving indicates that alcohol impairment leads to higher injury risk propensity of drivers compared to sober drivers. The negative effect of this variable on the threshold separating non-

incapacitating and incapacitating injury level indicates a lower likelihood of non-incapacitating injury for the alcohol impaired drivers. The net implications of these effects is that alcohol impaired drivers have a lower likelihood of no injury and a higher likelihood of dying instantly in a crash compared to sober drivers. A crash involving physically impaired drivers is associated with an overall higher injury risk propensity. The result may be reflecting increased reaction times for physically impaired drivers.

Vehicle Characteristics: With respect to driver's vehicle type, the estimation results show that latent injury risk propensities are lower for the drivers of sports utility vehicle (SUV), pickups and vans compared to the drivers of passenger car, presumably because SUV, pickups and vans have huge mass which offer more protection to the occupants of these vehicles (Kockelman and Kweon, 2002; Xie et al., 2009; Eluru et al., 2010; Fredette et al., 2008). The effect of SUV is also significant in second threshold and indicates increased probability of possible injury. The vehicle age results demonstrate that latent injury propensities are higher for drivers in older vehicles (vehicle age 6-10 years and vehicle age  $\geq 11$  years) relative to drivers in newer vehicles (vehicle age  $\leq 5$  years). As is expected, within the vehicle age categories considered the oldest vehicle age category has a larger impact relative to the moderately older vehicle age category. The higher injury risk of older vehicle's driver may be attributable to the absence of advanced safety features and/or the involvement of suspended and unlicensed drivers in older vehicles (Lécuyer and Chouinard, 2006, Kim et al., 2013; Islam and Mannering, 2006).

Roadway Design Attributes and Operational Attributes: Several roadway design attributes considered are found to be significant determinants of driver injury severity. Among those, the injury risk propensities are higher with overall increased likelihoods of dying instantly (as indicated by positive signs of thresholds demarcating possible and non-incapacitating injury) for crashes occurring on medium (26 to 50 mph) and high (above 50 mph) speed limit locations (with larger impact for high speed limit locations) compared to lower (less than 26 mph) speed limit locations (see Eluru et al., 2010; Chen et al., 2012; Tay and Rifaat, 2007 for similar results). The presence of traffic control device is also found to have significant effect on the severity of crashes. Crashes at traffic controlled and stop-sign controlled intersections seem to decrease the likelihood of serious crashes. However, the effect of stop-sign on threshold parameterization also indicates increased likelihood of incapacitation injury, possibly suggesting non-compliance with this traffic control device and judgment problems (Chipman, 2004; Retting et al., 2003). The influence of traffic control device also reveals that the presence of other traffic control devices (such as warning sign, regulatory sign, railway crossing sign) increases the likelihood of injury risk propensity of the drivers.

Environmental Factors: Several environmental factors considered are found to be significant determinants of driver injury severity in the final model specification. With respect to time of day, the latent propensity for evening peak period (related to morning peak, off peak and late evening) is found to be negative, indicating lower likelihood of serious injury, and is may be a result of traffic congestion and slow driving speeds during this period. The likelihood of injury risk propensity is found to be higher for late night (12.00 a.m. to 5.59 a.m.) period. This finding is consistent with several previous studies; attributable to reduced visibility, fatigue, longer emergency response times, higher driver reaction time and/or increased traffic speed (Plainis et al., 2006; Helai et al., 2008; Hu and Donnell, 2010; Kockelman and Kweon, 2002; de Lapparent,

2008). The findings of the lighting condition indicate that if collisions occur during dusk, the consequence is likely to be more injurious as compared to the crashes during other lighting condition (daylight, dawn and darkness). The sunglare during dusk period might pose such risk on drivers (Jurado-Piña et al., 2010; Gray and Regan, 2007). As found in previous studies (Kockelman and Kweon, 2002; Quddus et al., 2002), our study also found that the likelihood of driver injury risk propensity is higher during weekend compared to weekdays. The surface condition effects are simplified to a simple binary representation of presence/absence of snowy road surface condition. The result for the variables indicates that if collisions occur on a snowy road surface (relative to those on other surface conditions), the drivers are more likely to evade injury, perhaps due to reduced speeding possibility and/or could be related to more cautious driving (Edwards, 1998; Mao et al., 1997; Eluru and Bhat, 2007).

Crash Characteristics: Collision with large object (building, concrete traffic barrier, wall, tree, bridge, snow bank) result does not have any effect on the propensity of injury severity, but demonstrates a higher likelihood of non-incapacitating injury and in general, a higher probability of instant death in a crash (related to collision with small object and moving vehicle). The result is in line with several previous studies (Yamamoto et. al., 2004; Holdridge et al., 2005). The result also suggests that collision with other object (animal, non-fixed object) has a lower injury risk propensity. The results related to collision type reflect the anticipated higher injury risk propensity for head-on collision compared to other collision types. This is perhaps a consequence of greater dissipation of kinetic energy. The results in Table 4 related to sideswipe (both same and opposing direction) collisions underscore lower injury risk propensities relative to other collision types. The negative sign of propensity associated with front to rear collision reflects lower injury risk propensity. On the other hand, the impacts of front to rear collision on both of the first two thresholds are positive, which implies that the effects of front to rear collision on different injury categories are crash and driver-specific. However, the results suggest an increased probability of no injury category and, in general, a decreased possibility of instant death category. Crashes in driveway access location lead to an overall reduced injury risk propensity (relative to collision at other location) perhaps indicating driving at lower speed or more watchful driving at these locations (Rifaat and Tay, 2009).

The effects of the trajectory of vehicle's motions underscore an overall higher injury risk propensity for the driver whose vehicle was stopped in a traffic lane compared to the one who was going straight at the time of collision. Both turning manoeuvres (left and right) of drivers have lower injury risk propensities compared to going straight. This may be reflecting more watchful driving as well as lower speeds while turning. Changing traffic lane has a lower impact on the risk propensity, while the indicator variable has a negative impact on the threshold between non-incapacitating and incapacitating injury. This effect implies a lower probability of non-incapacitating injury and an overall higher probability for instant death (relative to going straight).

Situational Variables: Among different situational variables, number of passenger and driver ejection are found to affect driver injury severity. A higher injury risk propensity is observed for the presence of one passenger in the vehicle relative to presence of more than one or no passenger. Finally, the coefficient corresponding to driver ejection reveals that drivers who are ejected out of their vehicle during a crash have a high probability of sustaining serious injuries compared to those who were not ejected out. The result concurs with several previous studies (Palanca et al., 2003; Eluru and Bhat, 2007).

## 6. ELASTICITY EFFECTS AND IMPLICATIONS

The pooling exercise as presented in the paper shows that the pooled model (eleven point) provides a superior fit over unpooled (five point) model in examining driver injury severity outcomes. Therefore, we can expect that the pooled model provides more information on injury severity process relative to the unpooled model. However, both the pooled and unpooled models are estimated by using the same set of exogenous variables. Hence, for further policy analysis, it is beneficial to identify the differences between the pooled and unpooled models along with the additional information that the pooled model has to offer over unpooled model.

The parameter effects of the exogenous variables in Table 4 do not provide the magnitude of the variable effects on the injury severity of drivers. To quantify the effects of these variables and to identify the differences between pooled and unpooled models on driver injury severity outcomes, we compute the aggregate level “elasticity effects” (see Eluru and Bhat (2007) for a discussion on the methodology for computing elasticities) for a selected set of independent variables – driver age $\geq 65$ , other physical impairment, vehicle age 6-10 years, medium speed limit road, late night, weekend, head-on collision, changing lane and presence of one passenger. The elasticity estimates are calculated for both the pooled and unpooled models of the nationally representative pooled and unpooled samples, respectively.

In order to identify the differences between the pooled and unpooled models, we compute the differences in elasticity effects of variables for the non-fatal crash categories as: [Elasticity (Pooled) - Elasticity (Unpooled)]. These differences are presented in Figure 4. The following observations can be made based on the plot presented in Figure 4. First, there are considerable differences in elasticity effects between the pooled and unpooled models. The differences increase with increasing non-fatal crash severity levels i.e. the severe is the crash, the larger is the under-estimation of elasticity for the five alternative model. To illustrate the difference in estimates for fatal categories, we plot the elasticity effects from the unpooled (one category) and the pooled model (seven categories). This plot is presented in Figure 5. The following observations can be made based on the elasticity effects presented in Figure 5. First, the results in Figure 5 indicate that there are considerable differences in the elasticity effects of unpooled and pooled fatal crash categories. Second, there are also substantial differences across different fatal crash categories of pooled model. Specifically, the differences for collision on medium speed limit road, other physical impairment and head-on collision are significant. These findings support our hypothesis that the severity of fatal crashes is not a single, un-separable category but rather is a continuum ranging from dying instantly to dying within thirty days of crash. These results also suggest that considering a fine resolution categorization of fatal crashes in examining the crash injury severity outcome offers the potential to provide additional information on injury severity mechanism. This information has important implication for policy makers in developing the EMS system and trauma triage. Third, the most important variables in terms of early death are collision on a medium speed limit road, head-on collision and driving under other physical impairment. These variable effects have important implications in terms of enforcement, engineering and educational strategies. In terms of engineering measures, a forgiving road environment should be designed for a higher speed limit road location to allow the drivers more space to recover from a driving error. Head-on collisions are often caused by drivers violating traffic rules, driving across the centerline, driving too fast for the roadway conditions and thus by losing control of their vehicles (Zhang and Ivan, 2005). Therefore, policies concerning the enforcement in reducing the traffic violation have the

potential to reduce this type of collision. With respect to enforcement and education, our results endorse a continuous education program and stricter enforcement to prevent impaired-driving. Public health effort and education campaigns against intoxicated driving are needed for this group of drivers.

Finally, the elasticity analysis conducted provides an illustration of how the proposed pooled model can be applied to determine the critical factors contributing to reducing the survival time. For example, based on crash characteristic elasticities computed, if EMS services can identify *critical* crashes with likelihood for survival on the field it might assist in determining the appropriate mode of patient transfer (by road or air lifting depending on the crash characteristics) and also providing appropriate medical supervision at the hospital.

## 7. CONCLUSIONS

The focus of this paper was to develop a framework for pooling of data from Fatality Analysis Reporting System (FARS) and Generalized Estimates System (GES) data. The current research makes four important contributions to literature on driver injury severity analysis. *First*, we developed and tested a simple approach to combine information from FARS and GES databases toward a pooled database that brings together the strengths of individual databases. *Second*, we employed a sampling design approach for generating a nationally representative pooled sample of all crashes. *Third*, after demonstrating the validity of the approach, the nationally represented pooled data set was employed to undertake injury severity analysis with a very refined characterization of fatality along with other injury severity levels. Specifically, a Generalized Ordered Logit model (also referred to as Partial Proportional Odds model) was estimated on an eleven-alternative ordinal categorization of injury severity – no injury, minor injury, severe injury, incapacitating injury, and 7 categories of fatal injury ranging from fatality after 30 days of crash to instant death. *Finally*, using the empirical model results, we identified important factors affecting driver severity levels by evaluating elasticities of a selected set of exogenous variables.

The empirical analysis involved the validation of the five point ordinal (no injury, possible injury, non-incapacitating injury, incapacitating injury and fatal injury) pooled sample against the validation GES sample (unpooled sample) through two methods: (1) univariate sample comparison and (2) econometric model estimate comparison. The validation exercise confirmed that there was no evidence to suggest that the data pooled from GES and FARS resulted from distinct latent data generation process than the GES sample - the severity parameter estimates obtained using the pooled data closely resembled the severity parameter estimates obtained using the unpooled GES data. After we confirmed that the differences in parameter estimates obtained using pooled and unpooled data from the five point ordinal models were within the acceptable margins, we employed the pooled data to estimate models of fine resolution of injury severity with an eleven point ordinal scale defined as: no injury, possible injury, non-incapacitating injury, incapacitating injury, died between 6 and 30 days of crash, died between 2 and 5 days of crash, died between 7th-24 hours of crash, died between 2 and 6 hours of crash, died between 31 and 60 minutes of crash, died between 1 and 30 minutes of crash and died instantly. To compare the model with the five-alternative model estimated using the unpooled data, we generated an equivalent eleven alternative log-likelihood based on the five alternative model. The consistent improvement offered by the model estimated using the pooled data clearly indicated that inclusion of multiple discrete states of fatal injury category improves the model fit and provides more information in examining the injury severity outcome. Finally, a nationally representative pooled data sample was generated by

using the two- step sampling design approach which was then used for the purpose of estimating a nationally representative eleven point driver injury severity model.

In our research, to further understand the impact of various exogenous factors and to identify the differences between pooled and unpooled models, elasticity effects were estimated for a selected set of exogenous variables. The elasticity effects indicated that there were considerable differences in the elasticity effects across different crash categories of pooled and unpooled estimates. The substantial differences in elasticity effects across different fatal crash categories of pooled dataset signify the importance of considering the fine resolution of fatal crashes in examining the crash injury severity outcome. The most important variables in terms of early death were collision on the medium speed limit road, head-on collision and driving under other physical impairment. In summary, the pooling of fatal crashes with high resolution information from FARS dataset and replacing the fatal crashes in GES data allowed us to examine the impact of various attributes on all levels of injury severity and in turn allowed us to draw on the strengths of FARS and GES datasets to generate a single, potentially more beneficial sample for analysis. Finally, through the elasticity exercise, we demonstrate how our approach can be employed to identify factors affecting potentially fatal crashes (non-instantaneous) and improving the chances of survival of motor vehicle occupants involved.

The study is not without limitations. The datasets employed in our analysis are not perfect. For example, there are clear documented evidence on underreporting problems in relation to less severe crashes (see Elvik and Mysen, 1999; Yamamoto et al., 2008). The injury reporting data is fraught with police error (see Tsui et al., 2009; Schiff and Cummings, 2004; Loo and Tsui, 2007). However, our study is an attempt to bridge the two datasets and their strengths. Any enhancements or improvements to the datasets themselves will further enhance the value of our proposed approach. For example, augmenting the police reported data with hospital recorded data would allow us to better capture the interaction of transportation crashes and treatment on severity and fatality analysis. This would allow us to not be restricted by the questionable 30 day limit for fatal records to be considered in FARS. Further, in our research effort to keep the estimation time of the validation exercise within a reasonable limit, we have considered a random sample of 6,602 crashes from GES dataset. Another aspect of interest is the categorization of the fatality spectrum - we categorized the spectrum of fatal crashes in seven refined categories of fatalities ranging from fatality after thirty days to instant death. There has been earlier work on characterizing the distribution of survival times (Trunkey, 1983; Clark et al., 2012). Exploring these characterizations is an avenue for future research. Finally, the pooling exercise considered in our analysis is based on replacing GES fatal records with FARS fatal records without any exogenous variable specific controls. There is scope for considering more advanced pooling approaches where the replacement is undertaken by controlling for select exogenous variables such as crash type or vehicle type.

## **ACKNOWLEDGEMENTS**

The authors would like to acknowledge insightful feedback from two anonymous reviewers on an earlier version of the manuscript.

## **References**

Bass, F. M., and D. R. Wittink, Pooling Issues and Methods in Regression Analysis with Examples in Marketing Research. *Journal of Marketing Research*, Vol. 12, No. 4, 1975, pp. 414-425.

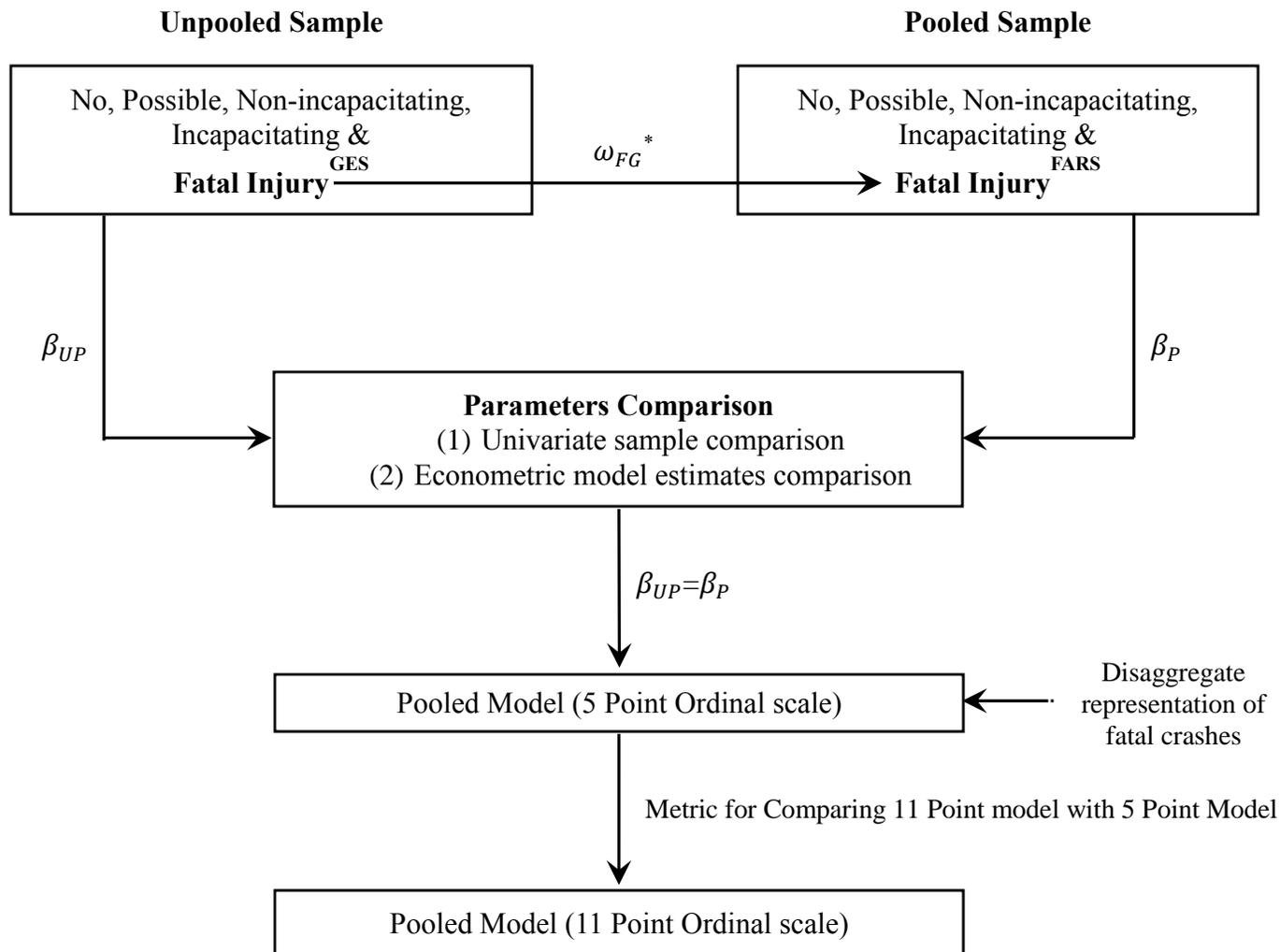
- Bédard, M., G. H. Guyatt, M. J. Stones, and J. P. Hirdes. The Independent Contribution of Driver, Crash, and Vehicle Characteristics to Driver Fatalities. *Accident Analysis and Prevention*, Vol. 34, No. 6, 2002, pp. 717-727.
- Castro, M., R. Paleti, and C. R. Bhat. A Spatial Generalized Ordered Response Model to Examine Highway Crash Injury Severity. *Accident Analysis and Prevention*, Vol. 52, 2013, pp. 188-203.
- Chalya, P. L., J. B. Mabula, R. M. Dass, N. Mbelenge, I. H. Ngayomela, A. B. Chandika, and J. M. Gilyoma. Injury Characteristics and Outcome of Road Traffic Crash Victims at Bugando Medical Centre in Northwestern Tanzania. *Journal of Trauma Management and Outcomes*, Vol. 6, No.1, 2012, pp. 1-8.
- Chen, F., and S. Chen. Injury Severities of Truck Drivers in Single- and Multi-Vehicle Accidents on Rural Highways. *Accident Analysis and Prevention*, Vol. 43, No. 5, 2011, pp. 1677-1688.
- Chen, H., L. Cao, and D. B. Logan. Analysis of Risk Factors Affecting the Severity of Intersection Crashes by Logistic Regression. *Traffic Injury Prevention*, Vol. 13, No. 3, 2012, pp. 300-307.
- Chipman, M. L. Side Impact Crashes - Factors Affecting Incidence and Severity: Review of the Literature. *Traffic Injury Prevention*, Vol. 5, No. 1, 2004, pp. 67-75.
- Clark, D.E., J. Qian, K. C. Sihler, L. D. Hallagan, and R. A. Betensky. The distribution of survival times after injury. *World journal of surgery*, Vol. 36, 2012, pp. 1562-1570.
- de Lapparent, M. Willingness to Use Safety Belt and Levels of Injury in Car Accidents. *Accident Analysis and Prevention*, Vol. 40, No. 3, 2008, pp. 1023-1032.
- Edwards, J. B. The Relationship between Road Accident Severity and Recorded Weather. *Journal of Safety Research*, Vol. 29, No. 4, 1998, pp. 249-262.
- Eluru N., and S. Yasmin. A note on Generalized Ordered Outcome Models. *Analytic Methods in Accident Research*, 2015 (forthcoming).
- Eluru, N. Evaluating Alternate Discrete Choice Frameworks for Modeling Ordinal Discrete Variables. *Accident Analysis and Prevention*, Vol. 55, 2013, pp. 1-11.
- 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.
- Eluru, N., C. R. Bhat, and D. A. Hensher. A Mixed Generalized Ordered Response Model for Examining Pedestrian and Bicyclist Injury Severity Level in Traffic Crashes. *Accident Analysis and Prevention*, Vol. 40, No. 3, 2008, pp.1033-1054.
- Eluru, N., R. Paleti, R. M. Pendyala, and C. R. Bhat. Modeling Injury Severity of Multiple Occupants of Vehicles: Copula-Based Multivariate Approach. *Transportation Research Record*, No. 2165, 2010, pp. 1-11.
- Elvik, R., and A. B. Mysen. Incomplete Accident Reporting: Meta-analysis of Studies Made in 13 Countries. *Transportation Research Board*, No. 1665, 1999, pp. 133-140.
- Evans, L., and M. C. Frick. Seating Position in Cars and Fatality Risk. *American Journal of Public Health*, Vol. 78, No. 11, 1988, pp. 1456-1458.
- Fredette, M., L. S. Mambu, A. Chouinard, and F. Bellavance. Safety Impacts Due to the Incompatibility of SUVs, Minivans, and Pickup Trucks in Two-Vehicle Collisions, *Accident Analysis and Prevention*, Vol. 40, No. 6, 2008, pp. 1987-95.

- Gates, J., S. Dubois, N. Mullen, B. Weaver, and M. Bédard. The Influence of Stimulants on Truck Driver Crash Responsibility in Fatal Crashes. *Forensic Science International*, Vol. 228, No. 1-3, 2013, pp. 15-20.
- Gray, R., and D. Regan. Glare susceptibility test results correlate with temporal safety margin when executing turns across approaching vehicles in simulated low-sun conditions. *Ophthalmic and Physiological Optics*, Vol. 27, No. 5, 2007, pp. 440-450.
- Helai, H., H. Chin, and M. Haque. Severity of Driver Injury and Vehicle Damage in Traffic Crashes at Intersections: A Bayesian Hierarchical Analysis. *Accident Analysis and Prevention*, Vol. 40, No. 1, 2008, pp. 45-54.
- Holdridge, J. M., V. N. Shankar, and G. F. Ulfarsson. The Crash Severity Impacts of Fixed Roadside Objects. *Journal of Safety Research*, Vol. 36, No. 2, 2005, pp. 139-47.
- Hu, W., and E. Donnell. Median Barrier Crash Severity: Some New Insights. *Accident Analysis and Prevention*, Vol. 42, No.6, 2010, pp. 1697-1704.
- Isenberg, D., D. C. Cone, and F. E. Vaca. Motor Vehicle Intrusion Alone does not Predict Trauma Center Admission or Use of Trauma Center Resources. *Prehospital Emergency Care*, Vol. 15, No. 2, 2011, pp. 203-207.
- Islam, S., and F. L. Mannering. Driver Aging and Its Effect on Male and Female Single-Vehicle Accident Injuries: Some Additional Evidence. *Journal of Safety Research*, Vol. 37, No. 3, 2006, pp. 267-276.
- Jurado-Piña, R., J. M. Pardillo-Mayora, and R. Jiménez. Methodology to analyze sun glare related safety problems at highway tunnel exits. *Journal of Transportation Engineering*, Vol. 136, No. 6, 2010, pp. 545-553.
- Kim, J.-K., G. F. Ulfarsson, S. Kim, and V. N. Shankar. Driver-Injury Severity in Single-Vehicle Crashes in California: A Mixed Logit Analysis of Heterogeneity Due to Age and Gender. *Accident Analysis and Prevention*, Vol. 50, 2013, pp. 1073-1081.
- Kockelman, K. M., and Y.-J. Kweon,. Driver Injury Severity: An Application of Ordered Probit Models. *Accident Analysis and Prevention*, Vol. 34, No. 3, 2002, pp. 313-321.
- Konduri, K., S. Astroza, B. Sana, R. Pendyala, and S. Jara-Díaz. Joint Analysis of Time Use and Consumer Expenditure Data. *Transportation Research Record*, No. 2231, 2011, pp. 53-60.
- Lécuyer, J. F., and A. Chouinard. Study on the Effect of Vehicle Age and the Importation of Vehicles 15 Years and Older on the Number of Fatalities, Serious Injuries and Collisions in Canada. *Proceedings of the Canadian Multidisciplinary Road Safety Conference XVI*, 2006.
- Loo, B. P. Y. and K. L. Tsui. Factors Affecting the Likelihood of Reporting Road Crashes Resulting in Medical Treatment to the Police. *Injury Prevention*, Vol. 13, No. 3, 2007, pp. 186-189.
- Lord, D., and F. L. Mannering. The Statistical Analysis of Crash-Frequency Data: A Review and Assessment of Methodological Alternatives. *Transportation Research Part A: Policy and Practice*. Vol. 44, No. 5, 2010, pp. 291-305.

- Louviere, J., R. Meyer, D. Bunch, R. Carson, B. Dellaert, W. M. Hanemann, D. Hensher, and J. Irwin. Combining Sources of Preference Data for Modeling Complex Decision Processes. *Marketing Letters*, Vol. 10, No. 3, 1999, pp. 205-217.
- Mao, Y., J. Zhang, G. Robbins, K. Clarke, M. Lam, and W. Pickett. Factors Affecting the Severity of Motor Vehicle Traffic Crashes Involving Young Drivers in Ontario. *Injury Prevention: Journal of the International Society for Child and Adolescent Injury Prevention*, Vol. 3, No. 3, 1997, pp. 183-189.
- Mooradian, J., J. N. Ivan, N. Ravishanker, and S. Hu. Analysis of Driver and Passenger Crash Injury Severity Using Partial Proportional Odds Models. *Accident Analysis and Prevention*, Vol. 58, 2013, pp. 53-58.
- NHTSA 2010, Traffic Safety Facts 2012, A Compilation of Motor Vehicle Crash Data from the Fatality Analysis Reporting System and the General Estimates System, accessed online at <http://www-nrd.nhtsa.dot.gov/Pubs/812032.pdf> on September 9th 2014.
- 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.
- Obeng, K. Injury Severity, Vehicle Safety Features, and Intersection Crashes. *Traffic Injury Prevention*, Vol. 9, No. 3, 2008, pp. 268-76.
- Palanca, S., D. M. Taylor, M. Bailey, and P. A. Cameron. Mechanisms of Motor Vehicle Accidents that Predict Major Injury. *Emergency Medicine*, Vol. 15, No. 5-6, 2003, 423-428.
- Paleti, R., N. Eluru, and C. R. Bhat. Examining the influence of aggressive driving behavior on driver injury severity in traffic crashes. *Accident Analysis and Prevention*, Vol. 42, No. 6, 2010, pp. 1839-1854.
- Plainis, S., I. Murray, and I. Pallikaris. Road Traffic Casualties: Understanding the Night-time Death Toll. *Injury Prevention*, Vol. 12, No. 2, 2006, pp. 125-128.
- Preusser, D. F., A. F. Williams, S. A. Ferguson, R. G. Ulmer, and H. B. Weinstein. Fatal Crash Risk for Older Drivers at Intersections. *Accident Analysis and Prevention*, Vol. 30, No. 2, 1998, pp. 151-59.
- Quddus, M., R. Noland, and H. Chin. An Analysis of Motorcycle Injury and Vehicle Damage Severity Using Ordered Probit Models. *Journal of Safety Research*, Vol. 33, No. 4, 2002, pp. 445-462.
- Retting, R. A., H. B. Weinstein, and M. G. Solomon. Analysis of motor-vehicle crashes at stop signs in four U.S. cities. *Journal of Safety Research*, Vol. 34, No. 5, 2003, pp. 485-489.
- Rifaat, S., and R. Tay. Effects of Street Patterns on Injury Risks in Two-Vehicle Crashes. *Transportation Research Record*, No. 2012, 2009, pp. 61-67.
- Sasser, S. M., R. C. Hunt, M. Faul, D. Sugeran, W. S. Pearson, T. Dulski, M. M. Wald, G. J. Jurkovich, C. D. Newgard, E. B. Lerner, A. Cooper, S. C. Wang, M. C. Henry, J. P. Salomone, R. L. Galli. Guidelines for Field Triage of Injured Patients Recommendations of the National Expert Panel on Field Triage, 2011. *Morbidity and Mortality Weekly Report*, No. 61(RR-1), 2012, pp. 1-23.

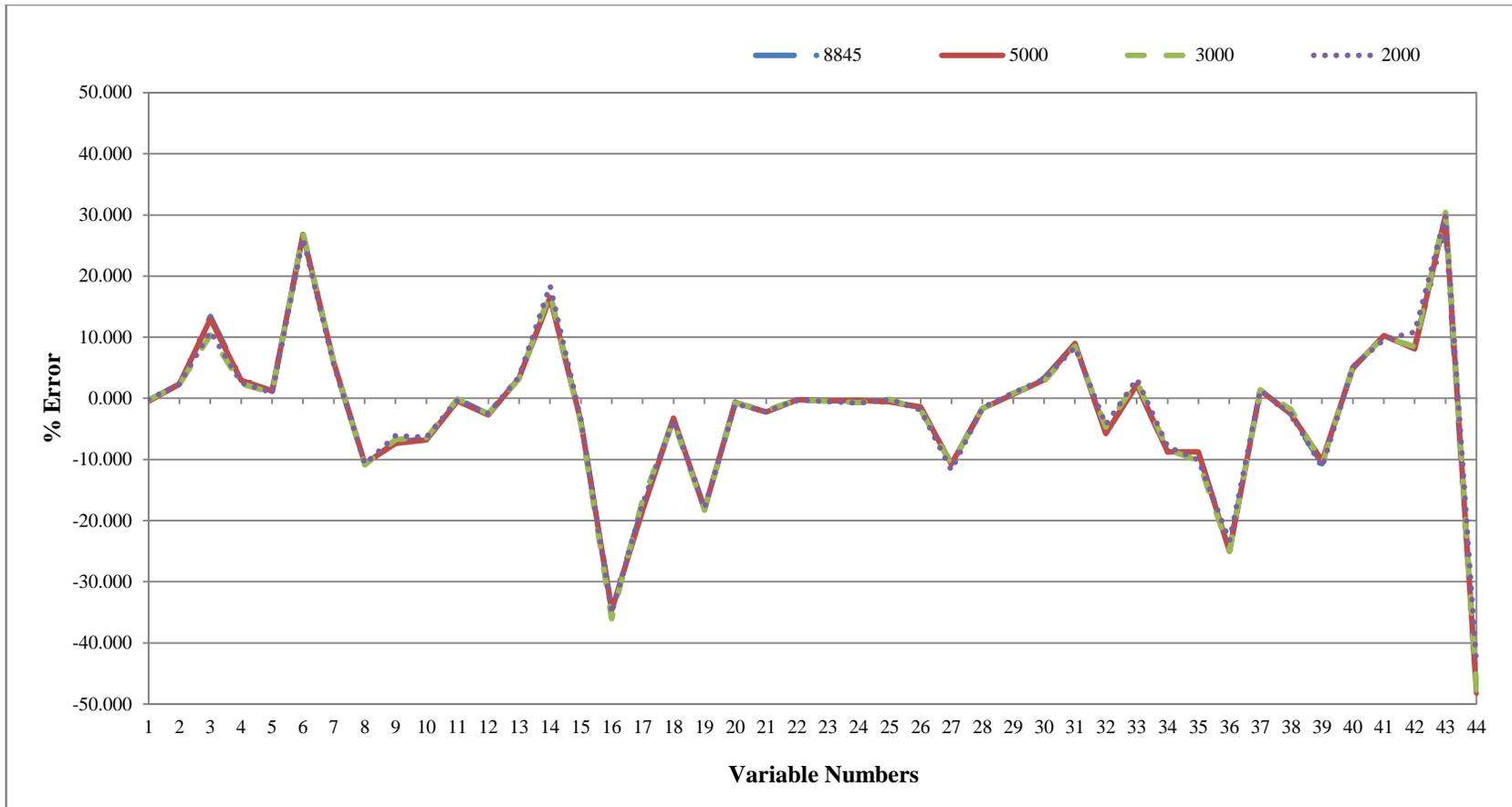
- Savolainen, P. T., F. L. Mannering, D. Lord, and M. A. Quddus. The Statistical Analysis of Highway Crash-Injury Severities: A Review and Assessment of Methodological Alternatives. *Accident Analysis and Prevention*. Vol. 43, No. 5, 2011, pp. 1666-76.
- Schiff, M. A., and P. Cummings. Comparison of Reporting of Seat Belt Use by Police and Crash Investigators: Variation in Agreement by Injury Severity. *Accident Analysis and Prevention*, Vol. 36, No. 6, 2004, pp. 961-965.
- Schutt R. K. *Investigating the social world: The process and practice of research*, Pine Forge Press, 2011.
- Sivakumar, A. and Polak, J.W. (2013). Exploration of Data-Pooling Techniques: Modeling Activity Participation and Household Technology Holdings. Presented at the 92<sup>nd</sup> Annual Meeting of the Transportation Research Board, January 2013, Washington D.C., USA.
- Stewart, R. D. Prehospital Care of Trauma. In: *Management of Blunt Trauma*. Baltimore, MD: Williams and Wilkins, 1990, pp. 23–29.
- Tay, R., and S. M. Rifaat. Factors Contributing to the Severity of Intersection Crashes. *Journal of Advanced Transportation*, Vol. 41, No. 3, 2007, pp. 245-265.
- Trunkey DD., 1983Trauma. *Sci Am* 249, 28–35.
- Tsui, K. L., F. L. So, N. N. Sze, S. C. Wong, and T. F. Leung. Misclassification of Injury Severity among Road Casualties in Police Reports. *Accident Analysis and Prevention*, Vol. 41, No. 1, 2009, pp. 84-89.
- Verma, V., F. Gagliardi, and C. Ferretti. On Pooling of Data and Measures. Working Paper no<sup>o</sup> 84/2009, DMQ, Università di Siena, 2009.
- World Health Organization (WHO). *Global Health Estimates Summary Tables: DALYs by Cause, Age and Sex*. World Health Organization, Geneva, Switzerland, 2013b.
- World Health Organization (WHO). *Global Status Report on Road Safety 2013: Supporting a Decade of Action*. World Health Organization, Geneva, Switzerland, 2013a.
- Xie, Y., K. Zhao, and N. Huynh. Analysis of Driver Injury Severity in Rural Single-Vehicle Crashes. *Accident Analysis and Prevention*, Vol. 47, 2012, pp. 36-44.
- Xie, Y., Y. Zhang, and F. Liang. Crash Injury Severity Analysis Using Bayesian Ordered Probit Models. *Journal of Transportation Engineering*, Vol. 135, No. 1, 2009, pp. 18-25.
- 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.
- Yamamoto, T., J. Hashiji, and V. N. Shankar. Underreporting in Traffic Accident Data, Bias in Parameters and the Structure of Injury Severity Models. *Accident Analysis and Prevention*, Vol. 40, No. 4, 2008, pp. 1320-1329.
- Yasmin, S., S. Anowar, and R. Tay. Effects of Drivers' Actions on Severity of Emergency Vehicle Collisions. *Transportation Research Record*, No. 2318, 2012, pp. 90-97.
- Yasmin. S., and N. Eluru. Evaluating Alternate Discrete Outcome Frameworks for Modeling Crash Injury Severity. *Accident Analysis and Prevention*, Vol. 59, 2013, pp. 506-21.

- Yasmin, S., N. Eluru, and A. R. Pinjari. Analyzing the continuum of fatal crashes: A generalized ordered approach. *Analytic Methods in Accident Research*, Vol. 7, 2015, pp. 1-15.
- Yau, K. K. W. Risk Factors Affecting the Severity of Single Vehicle Traffic Accidents in Hong Kong. *Accident Analysis and Prevention*, Vol. 36, No. 3, 2004, pp. 333-40.
- Zador, P. L., S. A. Krawchuk, and R. B. Voas. Alcohol-Related Relative Risk of Driver Fatalities and Driver Involvement in Fatal Crashes in Relation to Driver Age and Gender: An Update Using 1996 Data. *Journal of Studies on Alcohol*, Vol. 61, No. 3, 2000, pp. 387-95.
- Zhang, C., and J. N. Ivan. Effects of Geometric Characteristics on Head-on Crash Incidence on Two-Lane Roads in Connecticut. *Transportation Research Record*, No. 1908, 2005, pp. 159-164.



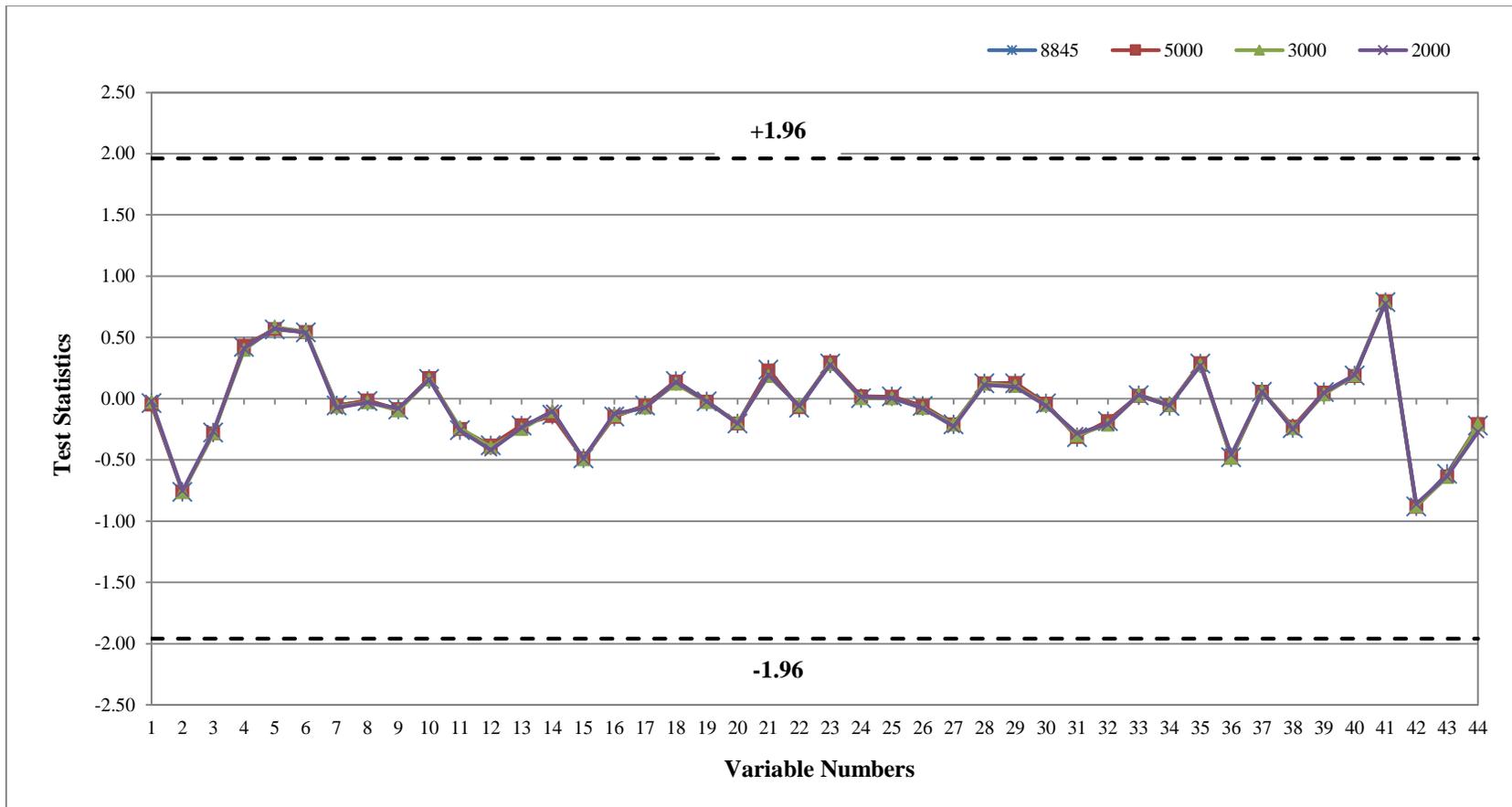
\*Specific weights for FARS crash records in pooled dataset

**FIGURE 1 Flow Chart Showing Research Framework for Validating Pooled dataset**



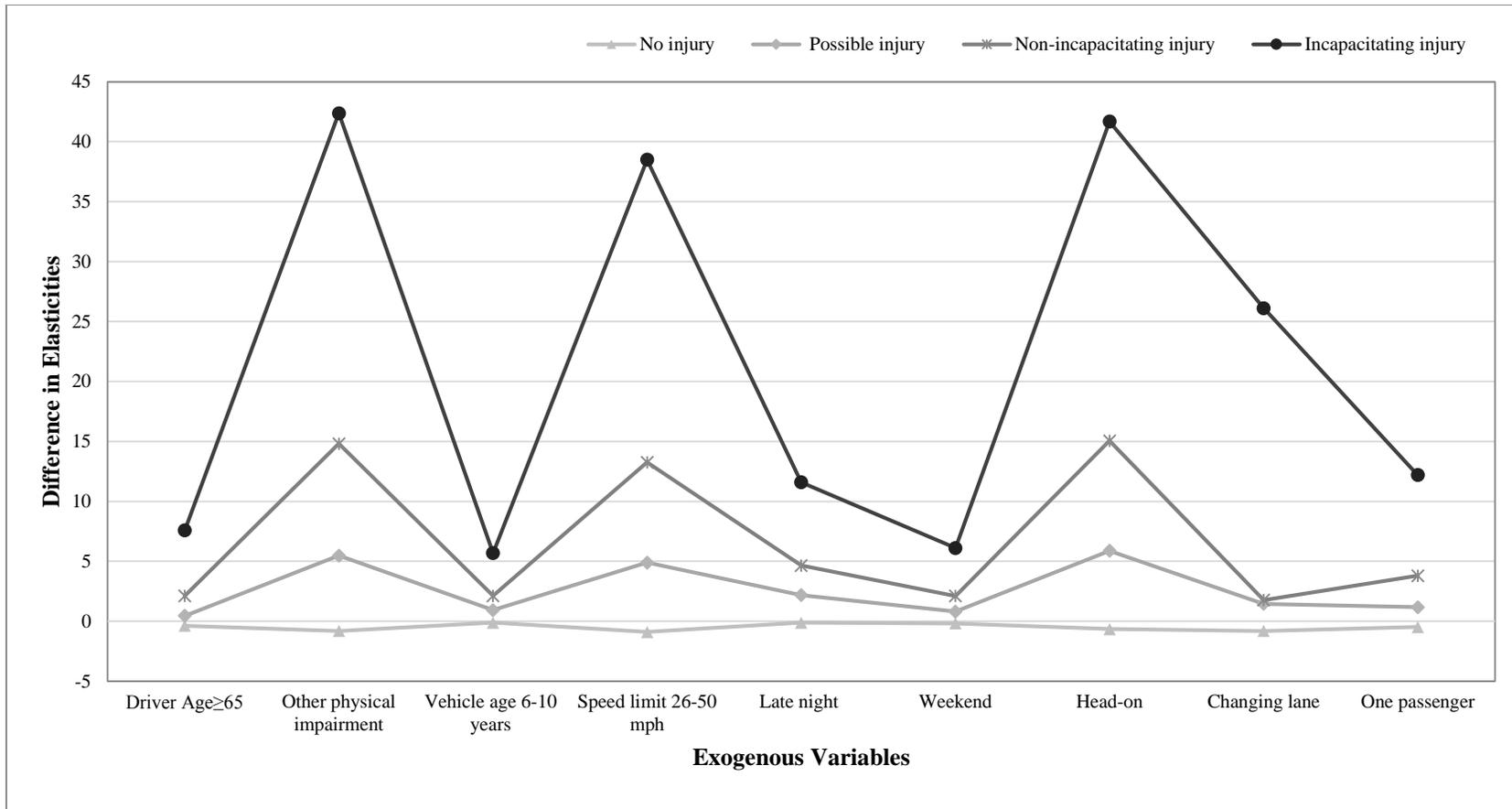
**Note** - Pooled data is obtained by replacing 59 fatality records from GES with 8,845 records from the FARS data for the 8,845 sample. The same process is applied to other sample sizes.

**FIGURE 2 % Error in Parameter Estimates obtained using Pooled model Plotted against Variable Numbers**



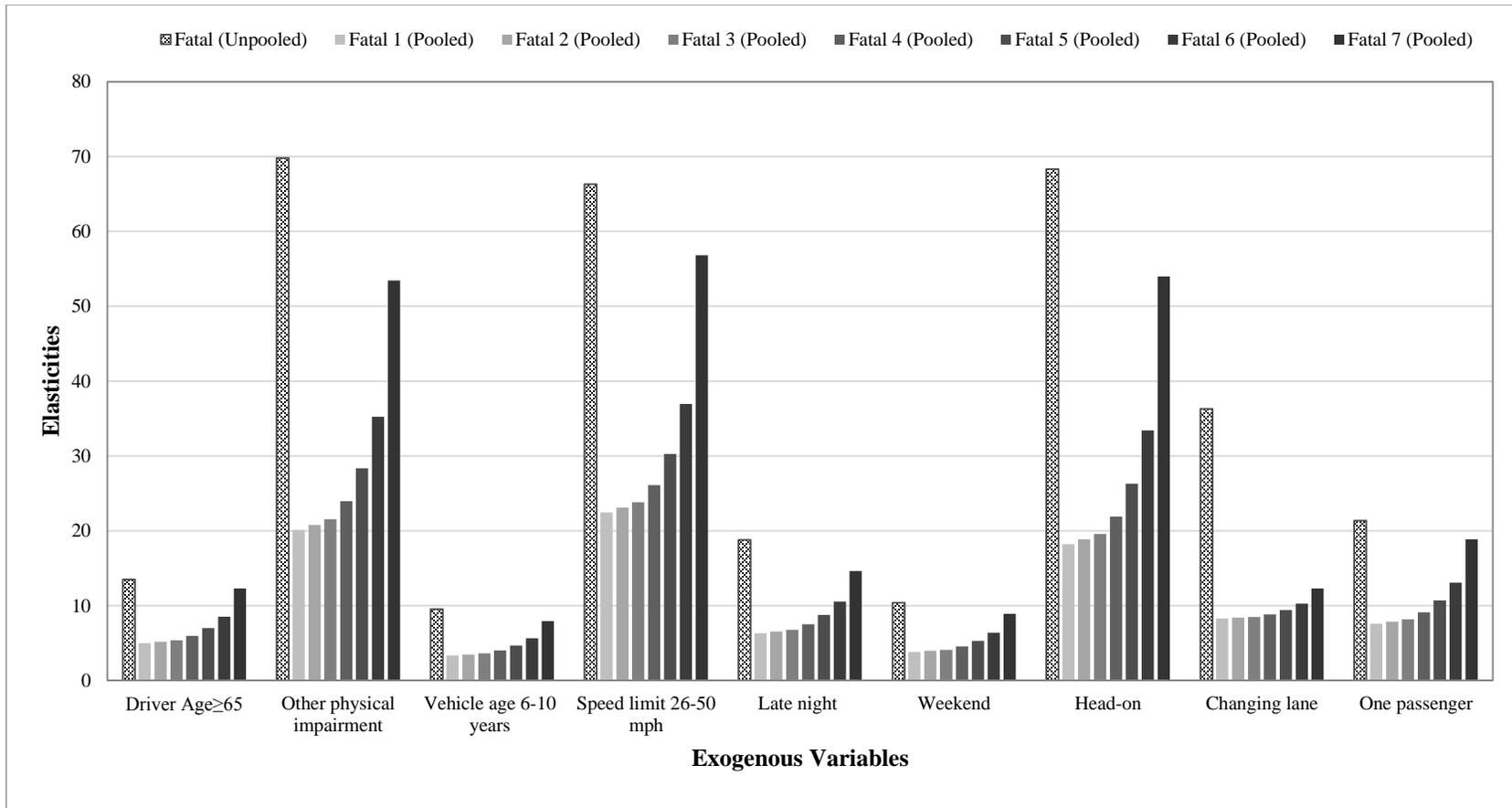
**Note** - Pooled data is obtained by replacing 59 fatality records from GES with 8,845 records from the FARS data for the 8,845 sample. The same process is applied to other sample sizes.

**FIGURE 3 Test Statistics for Parameter Estimates Plotted against Variable Numbers**



**Note** - Differences in elasticity effects are calculated as: [Elasticity (Pooled) - Elasticity (Unpooled)].

**FIGURE 4 Differences in Elasticity Effects of Non-Fatal Crash Categories for Pooled and Unpooled Models**



**Note** – Fatal 1 (Pooled) = Died between 6 and 30 days of crash, Fatal 2 (Pooled) = Died between 2 and 5 days of crash, Fatal 3 (Pooled) = Died between 7 and 24 hours of crash, Fatal 4 (Pooled) = Died between 2 and 6 hours of crash, Fatal 5 (Pooled) = Died between 31 and 60 minutes of crash, Fatal 6 (Pooled) = Died between 1 and 30 minutes of crash and Fatal 7 (Pooled) = Died instantly.

**FIGURE 4 Elasticity Effects of Fatal Crash Categories for Pooled and Unpooled Models**

**TABLE 1 Fatal Cases and Weight of Data Samples**

<b>Datasets</b>	<b>Samples</b>	<b>Fatal Cases</b>	<b>Weight</b>
<b>Unpooled</b>	---	59	---
	1	1956	59/1956
	2	1945	59/1945
	3	2010	59/2010
	4	1921	59/1921
	5	1983	59/1983
	6	2967	59/2967
	7	3101	59/3101
	8	3062	59/3062
<b>Pooled</b>	9	2980	59/2980
	10	2983	59/2983
	11	4976	59/4976
	12	4939	59/4939
	13	4921	59/4921
	14	4931	59/4931
	15	5004	59/5004
	16	8845	59/8845

**TABLE 2 Sample Characteristics of “Driver Injury Severity”**

Variables	Sample		Fatal Crashes	
	Unpooled Data	Pooled Data (With Weight)	Unpooled Data	Pooled Data (With Weight)
Frequency				
<b>Driver Characteristics</b>				
<i>Driver gender (Base: Male)</i>				
Female	2786	2786	18	18
<i>Driver age (Base: Age 25 to 64)</i>				
Age less than 25	1671	1666	19	14
Age above 65 & above	514	514	11	11
<i>Restraint system use (Base: Restrained)</i>				
Unrestrained	230	233	28	31
<i>Under the influence of alcohol</i>	312	325	11	24
<i>Other physical impairment</i>	195	197	6	8
<b>Vehicle Characteristics</b>				
<i>Vehicle Type (Base: SUV, Passenger car)</i>				
Pickup	1010	1019	4	13
Vans	413	414	2	3
<i>Vehicle age (Base: Vehicle age ≤ 5 years)</i>				
Vehicle age 6-10 years	2077	2068	28	19
Vehicle age ≥ 11 years	1897	1903	21	27
<b>Roadway Design and Operational Attributes</b>				
<i>Speed limit (Base: Speed limit less than 26 mph)</i>				
Speed limit 26-50 mph	3948	3940	32	24
Speed limit >50mph	1445	1452	25	32
<i>Traffic Control Device</i>				
Other traffic control device	145	148	1	4

<i>Type of intersection</i>				
T intersection	729	731	2	4
<i>Traffic-way description</i>				
Two way-with median	1398	1387	17	6
<b>Environmental Factor</b>				
<i>Time of Day (Base: 6.00 a.m. to 11.59 p.m. )</i>				
Late night (12.00 a.m. to 5.59 a.m.)	473	472	16	15
<i>Surface condition</i>				
Snowy	262	263	1	2
<b>Crash Characteristics</b>				
<i>Collision object (Base: Another moving vehicle)</i>				
Collision with large stationary object	525	517	25	17
Collision with other object	205	206	0	1
<i>Manner of collision (Base: Angular collision)</i>				
Head-on	347	346	10	9
Side swipe-same direction	342	342	1	1
Front to rear	1858	1858	1	1
<i>Collision location (Base: Non-intersection)</i>				
Driveway access	625	626	0	1
Intersection	2641	2646	6	11
<i>Trajectory of vehicle's motions (Base: Going straight)</i>				
Stopped in Traffic Lane	584	583	1	0
Turning right	155	155	0	0
Turning Left	680	683	0	3
Negotiating a curve	318	325	10	17

**TABLE 3 Log-likelihood values for Equivalent and Actual Eleven Point ordinal “Driver Injury Severity” Models**

Samples	Average log-likelihood		Log-likelihood Ratio Test Statistic
	Equivalent eleven point ordinal model	Actual eleven point ordinal model	
2000 (5 random samples)	-5976.897	-5964.741	24.312
3000 (5 random samples)	-5977.325	-5965.546	23.558
5000 (5 random samples)	-5977.418	-5965.788	23.260
8845 (1 sample)	-5977.790	-5966.164	23.252

**Note** - Pooled data is obtained by replacing 59 fatality records from GES with 8,845 records from the FARS data for the 8,845 sample. The same process is applied to other sample sizes.

**TABLE 4 Estimation Results of “Driver Injury Severity” by using the Population Representative Pooled Dataset**

Variables	Latent Propensity	$\tau_2$	$\tau_3$	$\tau_4$	$\tau_5$	$\tau_6$	$\tau_7$	$\tau_8$	$\tau_9$	$\tau_{10}$
Constant	2.054 (0.090) ‡	0.127 (0.092)	0.539 (0.057)	0.916 (0.342)	-2.692 (8.409)	-2.712 (9.416)	-2.669 (9.873)	-1.148 (4.508)	-1.282 (5.460)	-0.554 (4.619)
<b>Driver Characteristics</b>										
<i>Driver gender (Base: Male)</i>										
Female	0.546 (0.043)	0.122 (0.046)	--	--	--	--	--	--	--	--
<i>Driver age (Base: Age 25 to 64)</i>										
Age less than 25	-0.238 (0.048)	--	--	--	--	--	--	--	--	--
Age above 65 & above	0.145 (0.072)	--	--	--	--	--	--	--	--	--
<i>Restraint system use (Base: Restrained)</i>										
Unrestrained	1.392 (0.116)	-0.355 (0.126)	--	--	--	--	--	--	--	--
<i>Under the influence of alcohol</i>										
	0.523 (0.091)	--	-0.370 (0.115)	--	--	--	--	--	--	--
<i>Other physical impairment</i>										
	0.577 (0.114)	--	--	--	--	--	--	--	--	--
<b>Vehicle Characteristics</b>										
<i>Vehicle Type (Base: Passenger car)</i>										
SUV	-0.201 (0.055)	0.129 (0.058)	--	--	--	--	--	--	--	--
Pickup	-0.377 (0.064)	--	--	--	--	--	--	--	--	--
Vans	-0.341 (0.087)	--	--	--	--	--	--	--	--	--
<i>Vehicle age (Base: Vehicle age <math>\leq 5</math> years)</i>										
Vehicle age 6-10 years	0.097 (0.050)	--	--	--	--	--	--	--	--	--
Vehicle age $\geq 11$ years	0.277 (0.051)	--	--	--	--	--	--	--	--	--
<b>Roadway Design and Operational Attributes</b>										
<i>Speed limit (Base: Speed limit less than 26 mph)</i>										
Speed limit 26-50 mph	0.547 (0.073)	-0.168 (0.079)	--	--	--	--	--	--	--	--
Speed limit >50mph	0.859 (0.082)	-0.219 (0.089)	--	--	--	--	--	--	--	--
<i>Traffic Control Device (Base: No traffic control)</i>										
Traffic Signal	-0.175 (0.053)	--	--	--	--	--	--	--	--	--
Stop sign	-0.325 (0.094)	--	-0.320 (0.166)	--	--	--	--	--	--	--
Other traffic control device	0.331 (0.127)	--	--	--	--	--	--	--	--	--
<b>Environmental Factor</b>										
<i>Time of Day (Base: Morning peak and Off peak )</i>										

Evening peak (3.00 pm. to 5.59 pm.)	-0.111 (0.048)	--	--	-	-	-	-	-	-	-
Late night (12.00 a.m. to 5.59 a.m.)	0.177 (0.079)	--	--	-	-	-	-	-	-	-
<i>Lighting Condition (Base: Other lighting condition)</i>										
Dusk	0.229 (0.123)	--	--	-	-	-	-	-	-	-
<i>Day of Week (Base: Weekdays)</i>										
Weekend	0.109 (0.047)	--	--	-	-	-	-	-	-	-
<i>Surface condition (Base: Other surface condition)</i>										
Snowy	-0.587 (0.101)	--	--	-	-	-	-	-	-	-
<b>Crash Characteristics</b>										
<i>Collision object (Base: Another moving vehicle and collision with small object)</i>										
Collision with large stationary object	--	-0.234 (0.09)	--	-	-	-	-	-	-	-
Collision with other object	-2.079 (0.183)	--	--	-	-	-	-	-	-	-
<i>Manner of collision (Base: Other collision type)</i>										
Head-on	0.567 (0.08)	--	--	-	-	-	-	-	-	-
Side swipe-same direction	-1.519 (0.123)	--	--	-	-	-	-	-	-	-
Side swipe-opposite direction	-0.538 (0.182)	--	--	-	-	-	-	-	-	-
Front to rear	-0.911 (0.058)	0.239 (0.064)	0.516 (0.104)	-	-	-	-	-	-	-
Angular		0.174 (0.058)								
<i>Collision location (Base: Non-intersection and other location)</i>										
Driveway access	-0.334 (0.073)	--	--	-	-	-	-	-	-	-
<i>Trajectory of vehicle's motions (Base: Going straight)</i>										
Stopped in Traffic Lane	0.346 (0.078)	--	--	-	-	-	-	-	-	-
Turning right	-0.610 (0.168)	--	--	-	-	-	-	-	-	-
Turning Left	-0.226 (0.068)	--	--	-	-	-	-	-	-	-
Changing lane	-0.373 (0.169)	--	-0.491 (0.302)	-	-	-	-	-	-	-
<b>Situational Variable</b>										
<i>Number of Passengers (Base: No passenger and more than one passenger)</i>										
One passenger	0.221 (0.048)	--	--	-	-	-	-	-	-	-
<i>Driver Ejection (Base: Non ejected out)</i>										
Ejected Out	3.469 (0.552)	--	--	-	-	-	-	-	-	-

---

$\tau_2$  = Threshold between possible injury/non-incapacitating injury;  $\tau_3$ = Threshold between non-incapacitating injury/incapacitating injury;  $\tau_4$ = Threshold between incapacitating injury/6to30 days;  $\tau_5$ = Threshold between 6to30 days/ 1 to 5 days;  $\tau_6$ = Threshold between 1 to 5 days/ 7 to 24 hours;  $\tau_7$ = Threshold between 7 to 24 hours/ 1 to 6 hours;  $\tau_8$  = Threshold between 1 to 6 hours/ 31 to 60 minutes;  $\tau_9$  = Threshold between 31 to 60 minutes/ 1 to 30 minutes;  $\tau_{10}$  = Threshold between 1 to 30 minutes/ Died Instantly

---

‡ Standard errors are presented in parenthesis

**APPENDIX A Estimation Results of “Driver Injury Severity” by using Unpooled (GES) Data**

<b>Variables</b>	<b>Latent Propensity</b>	<b><math>\tau_2</math></b>	<b><math>\tau_3</math></b>	<b><math>\tau_4</math></b>
Constant	1.253 <sup>1*</sup> (0.122) <sup>‡</sup>	-0.534 <sup>28</sup> (0.071)	0.166 <sup>37</sup> (0.038)	0.971 <sup>41</sup> (0.075)
<b>Driver Characteristics</b>				
<i>Driver gender (Base: Male)</i>				
Female	0.537 (0.060) <sup>2</sup>	0.245 (0.066) <sup>29</sup>	--	--
<i>Driver age (Base: Age 25 to 64)</i>				
Age less than 25	-0.117 (0.064) <sup>3</sup>	--	--	--
Age 65 & above	0.237 (0.101) <sup>4</sup>	-0.213 (0.121) <sup>30</sup>	--	-0.471 (0.178) <sup>42</sup>
<i>Restraint system use (Base: Restrained)</i>				
Unrestrained	1.617 (0.142) <sup>5</sup>	--	--	-0.361 (0.132) <sup>43</sup>
<i>Under the influence of alcohol</i>	0.570 (0.131) <sup>6</sup>	--	--	--
<i>Other physical impairment</i>	0.708 (0.140) <sup>7</sup>	--	--	--
<b>Vehicle Characteristics</b>				
<i>Vehicle Type (Base: SUV, Passenger car)</i>				
Pickup	-0.423 (0.081) <sup>8</sup>	--	--	--
Vans	-0.239 (0.118) <sup>9</sup>	--	--	--
<i>Vehicle age (Base: Vehicle age ≤ 5 years)</i>				
Vehicle age 6-10 years	0.315 (0.069) <sup>10</sup>	0.218 (0.065) <sup>31</sup>	--	--
Vehicle age ≥ 11 years	0.328 (0.071) <sup>11</sup>	--	--	--
<b>Roadway Design and Operational Attributes</b>				
<i>Speed limit (Base: Speed limit less than 26 mph)</i>				
Speed limit 26-50 mph	0.642 (0.102) <sup>12</sup>	--	--	--
Speed limit >50mph	0.896 (0.117) <sup>13</sup>	--	--	--
<i>Traffic Control Device</i>				
Other traffic control device	0.465 (0.185) <sup>14</sup>	--	--	--
<i>Type of intersection</i>				
T intersection	-0.205 (0.088) <sup>15</sup>	--	--	--
<i>Traffic way description</i>				
Two way-with median	0.138 (0.069) <sup>16</sup>	--	--	--
<b>Environmental Factor</b>				
<i>Time of Day (Base: 6.00 a.m. to 11.59 p.m. )</i>				
Late night (12.00 a.m. to 5.59 a.m.)	0.281 (0.107) <sup>17</sup>	--	--	--

<i>Surface condition</i>				
Snowy	-1.040 (0.153) <sup>18</sup>	--	--	--
<b>Crash Characteristics</b>				
<i>Collision object (Base: Another moving vehicle)</i>				
Collision with large stationary object	0.379 (0.106) <sup>19</sup>	-0.289 (0.129) <sup>32</sup>	--	--
Collision with other object	-1.827 (0.217) <sup>20</sup>	-0.637 (0.349) <sup>33</sup>	--	--
<i>Manner of collision (Base: Angular collision)</i>				
Head-on	0.669 (0.106) <sup>21</sup>	--	-0.234 (0.124) <sup>38</sup>	--
Side swipe-same direction	-1.603 (0.157) <sup>22</sup>	--	--	--
Front to rear	-1.159 (0.079) <sup>23</sup>	--	--	--
<i>Collision location (Base: Non-intersection)</i>				
Driveway access	-0.440 (0.108) <sup>24</sup>	0.252 (0.119) <sup>34</sup>	0.303 (0.132) <sup>39</sup>	--
Intersection	--	0.266 (0.071) <sup>35</sup>	--	0.515 (0.133) <sup>44</sup>
<i>Trajectory of vehicle's motions (Base: Going straight)</i>				
Stopped in Traffic Lane	0.324 (0.111) <sup>25</sup>	--	0.529 (0.139) <sup>40</sup>	--
Turning right	-0.991 (0.241) <sup>26</sup>	--	--	--
Turning Left	-0.188 (0.094) <sup>27</sup>	--	--	--
Negotiating a curve	--	0.266 (0.120) <sup>36</sup>	--	--

$\tau_2$  = Threshold between possible injury/non-incapacitating injury;  $\tau_3$  = Threshold between non-incapacitating injury/incapacitating injury;  $\tau_4$  = Threshold between incapacitating injury/fatal injury

† Standard errors are presented in parenthesis

\* Variable Numbers