



## Alternative Ordered Response Frameworks for Examining Pedestrian Injury Severity in New York City

Shamsunnahar Yasmin, Naveen Eluru & Satish V. Ukkusuri

To cite this article: Shamsunnahar Yasmin, Naveen Eluru & Satish V. Ukkusuri (2014) Alternative Ordered Response Frameworks for Examining Pedestrian Injury Severity in New York City, Journal of Transportation Safety & Security, 6:4, 275-300, DOI: [10.1080/19439962.2013.839590](https://doi.org/10.1080/19439962.2013.839590)

To link to this article: <https://doi.org/10.1080/19439962.2013.839590>



Accepted author version posted online: 12 Sep 2013.  
Published online: 12 Sep 2013.



Submit your article to this journal [↗](#)



Article views: 195



View Crossmark data [↗](#)



Citing articles: 21 View citing articles [↗](#)

## Alternative Ordered Response Frameworks for Examining Pedestrian Injury Severity in New York City

SHAMSUNNAHAR YASMIN,<sup>1</sup> NAVEEN ELURU,<sup>1</sup>  
AND SATISH V. UKKUSURI<sup>2</sup>

<sup>1</sup>Department of Civil Engineering and Applied Mechanics, McGill University, Montréal, Québec, Canada

<sup>2</sup>School of Civil Engineering, Purdue University, West Lafayette, Indiana, USA

*This article focuses on identifying the appropriate ordered response structure for modeling pedestrian injury severity. The alternative ordered response approaches considered for the empirical analysis include ordered logit model (OL), generalized ordered logit model (GOL), and latent segmentation based ordered logit model (LSOL). The GOL and LSOL models enhance the traditional OL model in different ways. The GOL model relaxes the restrictive thresholds in the OL model by allowing for individual-level exogenous variable impacts on the threshold parameters. On the other hand, the LSOL model allows for differential impact on the alternatives by segmenting the pedestrian crash population into various segments with segment-specific OL parameters. In this study, the authors focus on examining the performance of these two model structures relative to the traditional OL model in the context of pedestrian injury severity. The performance of the formulated injury severity models are tested based on the New York City (NYC) Pedestrian Research Data Base for the years 2002 through 2006. To the authors' knowledge, the study provides a first of its kind comparison exercise among OL, GOL, and LSOL models for examining pedestrian injury severity. The model estimation results clearly highlight the presence of segmentation based on the crash location attributes of pedestrian accidents. The crash location attributes that affect the allocation of pedestrians into these segments include regional county, functional classification of roadway, pedestrian location on roadway, number of travel lanes, and number of parking lanes in the roadway system. The key factors influencing pedestrian injury severity are weather condition, lighting condition, vehicle types, pedestrian age, and season. Overall, the results of the empirical analysis provide credence to the hypothesis that LSOL model is a promising ordered framework to accommodate population heterogeneity in the context of pedestrian injury severity.*

**Keywords** pedestrian safety, comparison of ordered response models, generalized ordered logit model, latent segmentation based ordered logit model, validation

Address correspondence to Naveen Eluru, Department of Civil Engineering and Applied Mechanics, McGill University, Suite 483, 817 Sherbrooke St. W., Montréal, Québec H3A 2K6, Canada. E-mail: naveen.eluru@mcgill.ca

## 1. Introduction

Pedestrian safety is a global health concern, and the United States of America is no exception. Earlier research reveals that the risk of being injured while walking could be about 4 times higher than driving an auto (Elvik, 2009). In 2008, about 4,378 pedestrian fatalities and 69,000 other pedestrian injuries from road traffic crashes were recorded in the United States (National Highway Traffic Safety Administration [NHTSA], 2009). More alarmingly, about 72% of these fatalities occurred in urban areas underscoring the need for examining urban pedestrian safety. To study pedestrian safety issues in urban regions, we focus on pedestrian crash injury severity data from the largest and most populous urbanized area of the United States, New York City (NYC). In NYC, approximately 10% of the journey-to-work trips are conducted on foot (Ukkusuri *et al.*, 2012), whereas the national average for journey-to-work trips is approximately 6% (Loukaitou-Sideris *et al.*, 2007). However, the city has a poor pedestrian safety record compared to rest of the nation. The pedestrian fatality rate is almost fivefold higher in NYC compared to the whole nation (pedestrian fatality rate is 11% in the United States and 52% in NYC) (NHTSA, 2009; New York City Department of Transportation [NYCDOT], 2010). On the NYC street, pedestrians are 10 times more likely to die than a motor vehicle occupant, and the annual cost of these pedestrian crashes is as high as \$1.38 billion (NYCDOT, 2010)—approximately 32% of total traffic crash costs of the city. These statistics clearly indicate that pedestrian safety is of great concern for NYC region. Any effort to reduce the social burden of these crashes and enhance pedestrian safety would necessitate the examination of factors that contribute significantly to crash likelihood and/or pedestrian injury severity in the event of a crash. This study endeavors to identify the various factors that affect pedestrian crash severity in a pedestrian–motor vehicle crash in NYC by developing econometric models that quantify the impact of these factors on injury severity.

A critical component in the process of identifying the contributing factors is the application of appropriate econometric model. Pedestrian injury severity is often reported as an ordered variable resulting in the application of ordered response models for analyzing risk factors. The most commonly employed approach to pedestrian injury severity is the ordered logit/probit (OL/OP) models. These models recognize the inherent ordering in the severity variable while apportioning the probability of injury severity to various alternatives based on population specific thresholds. The traditional OL/OP model has been the workhorse for examining injury severity in safety literature. However, the traditional ordered response model imposes strong restrictions on the structure of threshold parameters. Specifically, the traditional OL/OP model assumes the thresholds to remain fixed across crashes (Eluru *et al.*, 2008). However, it is possible that depending on the individual and crash attributes the thresholds could vary across crashes.

In recent years, two enhanced econometric frameworks have been proposed and implemented in transportation safety literature allowing us to tackle the restriction on the threshold parameters in the traditional OL model. These advanced frameworks include (1) generalized ordered logit (GOL) (Eluru *et al.*, 2008) and (2) latent segmentation based ordered logit (LSOL) model (Eluru *et al.*, 2012). The GOL model relaxes the restrictive assumption by allowing the threshold in the traditional OL model to vary based on individual and crash attributes. On the other hand, the LSOL model relaxes the assumption of the traditional OL model by categorizing pedestrian crashes into different segments based on crash location attributes while estimating segment specific OL parameters. The pedestrians involved in crashes are assigned to the various segments probabilistically as a

function of individual-level exogenous variables. The two approaches represent crash injury severity more realistically compared to the traditional OL model and improve the accuracy in quantifying the impact of exogenous variables on pedestrian injury severity (Eluru et al., 2008, 2012). However, there has been no research conducted on examining which of these two approaches offers a better framework for injury severity analysis. In this study, we undertake a comparison exercise to identify the preferred ordered model for examining pedestrian injury severity. The comparison exercise is conducted based on the performance of the various ordered response frameworks in model estimation as well as in model validation (using a hold-out sample).

The rest of the article is organized as follows. Section 2 provides a discussion of earlier research on pedestrian injury severity modeling while positioning the current study. Section 3 provides details of the various econometric model frameworks used in the analyses. In Section 4, the data source and sample formation procedures are described. The model estimation results, elasticity effects and validation measures are presented in Section 5. Section 6 concludes the article and presents directions for future research along with the recommendations.

## 2. Earlier Research

A number of research efforts have examined pedestrian crashes to gain a comprehensive understanding of the factors that affect crash frequency and severity. A category of these studies has focused on pedestrian crash frequencies (Gårder, 2004), or on the exposure measure of pedestrian crash risk (Keall, 1995) whereas another category of studies has examined the determinants of injury severity in the event of a pedestrian crash. The review in this study is restricted to the latter category. A summary of earlier research on pedestrian injury severity analysis is provided in Table 1. The table provides information on the analysis framework employed, pedestrian injury severity representation, and characteristics/factors considered in the analysis (including crash characteristics, vehicular characteristics, roadway design and land use attributes, environmental factors, pedestrian characteristics, and driver characteristics).

The following observations may be made from the review presented in Table 1. First, the dependent variable examined in most of the earlier studies is pedestrian injury severity ranging from two (killed/severe injury to slight injury/property damage only [PDO]) to five (fatality, disabling injury, not disabling injury, probable injury, to no injury) severity levels. Second, the most prevalent mechanism to study pedestrian injury severity is logistic regression and OL/OP models (14 out of 19). Third, only two earlier research efforts (Eluru et al., 2008; Kim et al., 2008) have considered variables from all factor categories. Fourth, the unordered response structure predominantly considered is the multinomial logit model (or its mixed version). Fifth, only two of the studies (Clifton et al., 2009; Eluru et al., 2008) considered the generalized version of ordered response model, and notably none of the earlier studies has considered latent segmentation based models for examining pedestrian injury severity. Finally, two studies (Eluru et al., 2008; Rifaat et al., 2011) examined the pedestrian and bicyclist crash severities together in their empirical analysis.

The overall findings from earlier research efforts are usually consistent. The most common factors that increase pedestrian crash severities include older pedestrians, male pedestrians, intoxicated pedestrians and/or drivers, occurrence of crash in darkness (with or without lighting), vehicle speeding, crash location is in a commercial area or on highways, and crash with bus or truck. On the other hand, exogenous factors that reduce the pedestrian

**Table 1**  
Summary of existing pedestrian injury severity studies

Article	Analysis Framework Employed	Pedestrian Injury Severity Representation	Crash	Vehicle	Characteristics/Factors Considered				
					Roadway Design & Land Use	Environment	Pedestrian	Driver	
Moudon et al. (2011)	Binary logistic regression	Severely injured/dying, Suffering minor/no injury	—	—	Yes	Yes	Yes	Yes	
Tefft (2013)	Logistic regression	Severe injury, nonsevere injury/fatal injury, nonfatal injury	Yes	Yes	Yes	—	Yes	—	
Ballesteros et al. (2004)	Logistic regression	Mortality, nonmortality/injury severity score <16 and $\geq 16$	—	Yes	Yes	—	—	—	
Sze and Wong (2007)	Binary logistic regression	Killed or severe injury, slight injury	Yes	—	Yes	Yes	Yes	—	
Roudsari et al. (2004)	Multivariate logistic regression	Severe injury, nonsevere injury	—	Yes	Yes	—	Yes	—	
Kim et al. (2008a)	Logistic regression	Serious injury, noninjury	—	—	Yes	Yes	Yes	—	
Zajac and Ivan (2003)	Ordered probit	Fatality, disabling injury, not disabling injury, probable injury, no injury	—	Yes	Yes	Yes	Yes	—	

Zahabi et al. (2011)	Ordered logit	No injury, minor injury, fatal/major injury	Yes	Yes	Yes	—	—
Lee and Abdel-Aty (2005)	Ordered probit	No injury, possible injury, nonincapacitating injury, incapacitating injury, fatal injury	—	Yes	Yes	Yes	—
Clifton et al. (2009)	Generalized ordered probit	No injury, injury, fatality	Yes	—	Yes	Yes	—
Eluru et al. (2008)	Mixed generalized ordered logit, Ordered logit	No injury, nonincapacitating injury, incapacitating injury, fatal injury	Yes	Yes	Yes	Yes	Yes
Tay et al. (2011)	Multinomial logit	Minor injury, serious injury, fatal injury	—	Yes	Yes	Yes	Yes
Rifaat et al. (2011)	Multinomial logit	No injury, injury, fatality	—	—	Yes	—	Yes
Kim et al. (2008)	Heteroskedastic generalized extreme value logit	Fatal, incapacitating injury, nonincapacitating injury, possible or no injury	Yes	Yes	Yes	Yes	Yes
Aziz et al. (2013)	Random-parameter multinomial logit	Property damage and possible injury, severe injury, fatality	Yes	Yes	Yes	Yes	—

(Continued on next page)

**Table 1**

Summary of existing pedestrian injury severity studies (*continued*)

Article	Analysis Framework Employed	Pedestrian Injury Severity Representation	Characteristics/Factors Considered					
			Crash	Vehicle	Roadway Design & Land Use	Environment	Pedestrian	Driver
Kim et al. (2010)	Mixed logit model	Fatal injury, incapacitating injury, nonincapacitating injury, possible/no injury	—	Yes	Yes	Yes	Yes	Yes
Kwigizile et al. (2011)	Ordered probit, Multinomial logit	No/possible injury, nonincapacitating injury, incapacitating injury, fatal injury	Yes	Yes	Yes	—	Yes	Yes
Abay (2013)	Ordered logit, Mixed ordered logit, Multinomial logit, Mixed multinomial logit	Slight/no injury, serious injury, fatal injury	Yes	Yes	Yes	Yes	Yes	Yes
Mohamed et al. (2013)	Latent Class Clustering: Ordered probit, K-Means: Multinomial logit	Injury and fatal injury, no injury, minor injury and fatal injury	—	Yes	Yes	Yes	Yes	Yes

crash severities include older drivers, presence of traffic signal control, snowy weather, and crash occurrence during the day.

## **2.1. This Study in Context**

As is evident from the review presented in preceding section, many studies in earlier literature have employed traditional ordered and unordered outcome models. The traditional ordered response model imposes a restrictive assumption on the impact of exogenous variables by constraining their impact to be the same for all alternatives. Toward addressing this limitation, researchers have employed the unordered response models that allow the impact of exogenous variables to vary across the injury severity levels. A sample of recent research efforts used the ordered and unordered response models for pedestrian crash severity analysis and find that the implications obtained from the two model systems differ for some variables (Abay, 2013; Kwigizile et al., 2011). In both studies, the unordered systems fit the data better than the ordered systems indicating that the additional flexibility of estimating alternative specific parameters is resulting in an improved fit. However, the increased flexibility from the unordered models is obtained at a cost. The unordered response models neglect the inherent ordering of crash severity outcome.

Ideally, it would be beneficial to incorporate the influence of alternative specific effects of exogenous variable in the ordered response framework. The generalized ordered response model (or proportional odds logit) (Castro et al., 2013; Clifton et al., 2009; Eluru, 2013; Eluru et al., 2008; Quddus et al., 2010; Mooradian et al., 2013; Yasmin et al., 2012; Yasmin & Eluru, 2013; Wang & Abdel-Aty, 2008) relaxes the restrictive assumption of the traditional ordered response model by allowing for differential impact of exogenous variables on injury severity levels. Recent evidence on the comparison of the unordered models and generalized ordered response systems in the context of injury severity highlight how the generalized ordered response model offers the same flexibility (if not better) as that of the unordered systems (Eluru, 2013; Yasmin & Eluru, 2013). Hence it is important that pedestrian injury severity models consider the generalized ordered response framework.

In recent years, there has also been a revival of latent segmentation models in safety and transportation. The traditional discrete outcome models restrict the impact of exogenous variables to be same across all crash locations: homogeneity assumption. But the impact of control variables on pedestrian crash severity might vary across individuals based on crash location attributes. To illustrate this, let us consider the pedestrian crash severity outcomes at two different crash locations (L1 and L2). For the ease of comparison, let us also assume that all crash attributes are identical with the only difference between the two scenarios being the location; suppose L1 is an intersection and L2 is a midblock location. Now let us consider the influence of dark road - lighted variable in these crash locations. In the first crash at L1, the driver might be travelling at a lower speed while approaching an intersection thus requiring a smaller reaction time to reduce the impact of crash with a pedestrian at this location. In this case, the illumination of road light might help the drivers and pedestrians to be more heedful in their movements and thus result in a less severe injury for pedestrians. On the other hand, at location L2, the driver would not have stopped and possibly would be travelling at a higher speed as he or she would not expect any pedestrian at midblock location. In this event, the advantages of having illumination at the dark period would be reduced and the higher impact force would increase the pedestrian crash severity at L2. This is an example of how pedestrian location at the time of crash moderates the influence of one variable (dark road - lighted) in determining pedestrian crash severity outcomes. However, it is possible that crash locations might serve as a moderating influence on multiple control



variables in the context of pedestrian injury severity. Ignoring such heterogeneous impact of variable might result in incorrect coefficient estimates.

A common approach employed to accommodate heterogeneity, in discrete outcome models, is the estimation of mixed/random coefficients version of the discrete outcome models (Eluru & Bhat, 2007; Kim *et al.*, 2013; Morgan & Mannering, 2011; Paleti *et al.*, 2010; Srinivasan, 2002; Xiong & Mannering, 2013).<sup>1</sup> However, in this approach the focus is on incorporating unobserved heterogeneity through the error term while also necessitating extensive amount of simulation for model estimation. In an attempt to accommodate for systematic heterogeneity researchers have considered segmenting the population based on exogenous variables (such as gender, age, location) and estimate separate models for each segment (see Aziz *et al.*, 2013, for segmentation based on location). These approaches divide the population into groups and are meaningful only for one or two exogenous variables. When segmentation by many variables is considered simultaneously it results in a large number of data samples necessitating a huge number of model estimations and also might lead to very few records in the various samples. To address this limitation, more advanced approaches such as clustering techniques that allow us to segment based on a multivariate set of factors have been suggested. The approach has been successfully employed (Mohamed *et al.*, 2013) in examining the pedestrian crash severity. However, the approach still requires allocating data records exclusively to a particular segment. So it is possible that some clusters end up with very little records (as was the case for one or two clusters in Mohamed *et al.* [2013]).

An alternative approach to accommodate for population heterogeneity is to undertake an endogenous segmentation approach. In this approach crash records are allocated probabilistically to different segments and a segment specific model is estimated for each segment. The approach does not allocate data records exclusively and hence allows the segment specific models to be estimated on the entire population (as opposed to data records allocated to the segment in clustering or simple segmentation approaches). Also, unlike the mixed models, it does not require any specific distributional assumption for the parameters (Greene & Hensher, 2003). The probabilistic allocation improves model estimation efficiency and is appropriate irrespective of the sample size of the data set, that is, it works efficiently for large and small data sets. On the other hand, the clustering approach is suitable only when we have a large data set. The endogenous segmentation approach, also referred to as the latent segmentation approach, has been employed in safety literature recently (Eluru *et al.*, 2012; Xie *et al.*, 2012).

The current research draws on these advances to contribute to pedestrian injury severity modeling along two directions. First, the study contributes to econometric modeling by identifying the appropriate framework for examining pedestrian injury severity through an exhaustive comparison exercise across three ordered response models (OL, GOL, and LSOL models). The models estimated are rigorously compared employing various comparison metrics for estimation and validation. Second, the study contributes substantively by identifying the various determinants of pedestrian injury severity in NYC allowing us to suggest remedial measures to improve pedestrian safety. The ordered response models are estimated using an exhaustive set of exogenous variables (crash characteristics, environmental factors, vehicle characteristics, roadway design and operational attributes, land use characteristics, and pedestrian characteristics).

<sup>1</sup>The readers are encouraged to review Yasmin and Eluru (2013) for an extensive review of discrete outcome modeling approaches employed in transportation safety literature (not just pedestrian literature).

### 3. Econometric Framework

In this section, we provide the brief discussion of the methodology of all the models considered for examining pedestrian injury severity in our research.

#### 3.1. Standard Ordered Logit Model

In the traditional ordered response model, the discrete injury severity levels ( $y_i$ ) are assumed to be associated with an underlying continuous latent variable ( $y_i^*$ ). This latent variable is typically specified as the following linear function:

$$y_i^* = X_i\beta + \varepsilon_i, \text{ for } i = 1, 2, \dots, N \quad (1)$$

where,

- $i$  ( $i = 1, 2, \dots, N$ ) represents the pedestrians
- $X_i$  is a vector of exogenous variables (excluding a constant)
- $\beta$  is a vector of unknown parameters to be estimated
- $\varepsilon$  is the random disturbance term assumed to be standard logistic.

Let  $j$  ( $j = 1, 2, \dots, J$ ) denotes the injury severity levels and  $\tau_j$  represents the thresholds associated with these severity levels. These unknown  $\tau_j$ s are assumed to partition the propensity into  $J - 1$  intervals. The unobservable latent variable  $y_i^*$  is related to the observable ordinal variable  $y_i$  by the  $\tau_j$  with a response mechanism of the following form:

$$y_i = j, \text{ if } \tau_{j-1} < y_i^* < \tau_j \text{ for } j = 1, 2, \dots, J \quad (2)$$

To ensure the well-defined intervals and natural ordering of observed severity, the thresholds are assumed to be ascending in order, such that  $\tau_0 < \tau_1 < \dots < \tau_J$  where  $\tau_0 = -\infty$  and  $\tau_J = +\infty$ . Given these relationships across the different parameters, the resulting probability expressions for individual  $i$  and alternative  $j$  for the OL model take the following form:

$$\pi_{ij} = \Pr(y_i = j | X_i) = \Lambda(\tau_j - X_i\beta) - \Lambda(\tau_{j-1} - X_i\beta) \quad (3)$$

where  $\Lambda(\cdot)$  represents the standard logistic cumulative distribution function.

#### 3.2. Generalized Ordered Logit Model

The GOL model relaxes the constant threshold across population restriction to provide a flexible form of the traditional OL model. The basic idea of the GOL is to represent the threshold parameters as a linear function of exogenous variables (Eluru et al., 2008; Maddala, 1983; Srinivasan, 2002; Terza, 1985). Thus the thresholds are expressed as:

$$\tau_j = fn(Z_{ij}) \quad (4)$$

where,  $Z_{ij}$  is a set of exogenous variable (including a constant) associated with  $j$ th threshold. Further, to ensure the accepted ordering of observed discrete severity ( $-\infty < \tau_1 < \tau_2 < \dots < \tau_{J-1} < +\infty$ ), we employ the following parametric form as employed by Eluru et al. (2008):

$$\tau_j = \tau_{j-1} + \exp(\delta_j Z_{ij}) \quad (5)$$

where,  $\delta_j$  is a vector of parameters to be estimated. The remaining structure and probability expressions are similar to the OL model. For identification reasons, we need to restrict one of the  $\delta_j$  vectors to zero.

### 3.3. Latent Segmentation Based Ordered Logit Model

Standard ordered response model restricts the impact of crash-related explanatory variables to be identical for all individuals (Eluru et al., 2008). The latent segmentation model relaxes this homogeneity assumption of the standard ordered response model by classifying the crashes based on crash location attributes and subsequently model the effect of crash attributes within each segment separately (Eluru et al., 2012). Let us consider  $S$  homogenous segments of crash locations. The multinomial logit structure is used to assign the pedestrians to these segments based on the attributes of crash locations. The utility for assigning a pedestrian  $i$ 's crash location to segment  $s$  is defined as:

$$U_{is}^* = \alpha'_s Z_i + \xi_{is} \quad (6)$$

Where

$\alpha'_s$  is a vector of coefficients

$Z_i$  is a vector of attributes that influences the propensity of belonging to segment  $s$ .

$\xi_{is}$  is an idiosyncratic random error term assumed to be identically and independently Type 1 extreme value distributed across individual  $i$  and segment  $s$ .

Then the probability that pedestrian  $i$ 's crash location belongs to segment  $s$  is given as:

$$P_{is} = \frac{\exp[\alpha'_s Z_i]}{\sum_s \exp[\alpha'_s Z_i]} \quad (7)$$

Within the latent segmentation approach, the unconditional probability of individual sustaining injury severity level  $j$  is given as:

$$P_i(j) = \sum_{s=1}^S (P_i(j|s)(P_{is})) \quad (8)$$

where,  $P_i(j|S)$  represents the probability of pedestrian  $i$  sustaining injury severity level  $j$  within the segment  $s$ . Now, if we consider the injury severity sustain by pedestrian  $i$  to be ordered and if the accident belongs to segment  $s$ , we can represent the latent propensity function as follows as in standard OL model:

$$y_{is}^* = X_i \beta'_s + \varepsilon_{is}, \text{ for } i = 1, 2, \dots, N \quad (9)$$

Thus, the probability expressions take the form:

$$P_i(j|s) = \Lambda(\tau_{sj} - X_i \beta'_s) - \Lambda(\tau_{s,j-1} - X_i \beta'_s) \quad (10)$$

where  $\Lambda(\cdot)$  represents the standard logistic cumulative distribution function.

## 4. Data

### 4.1. Data Source

The pedestrian crash data for NYC used in the research effort is extracted from NYC Pedestrian Research Data Base for the year of 2002 through 2006. The coded information of this database is compiled from three different data sources: (1) New York State Department of Transportation-Safety Management System (NYSDOT SMS) data, (2) New York City Department of Transportation (NYCDOT) data, and (3) U.S. Census data. The NYC pedestrian research database consists of a total of 7,354 crashes involving at least one pedestrian during this 5-year period. These crashes involve about 7,647 pedestrians and 7,712 motor vehicles resulting in 739 fatalities and 6,710 serious injuries to the pedestrians. A number of crash-related factors are extracted from these databases to explore the effect of various exogenous variables on pedestrian injury severity.

### 4.2. Sample Formation and Description

In an effort to clearly examine the influence of exogenous variables on pedestrian injury severity, we limit our attention to crashes involving one pedestrian and one vehicle. Further, records with missing information for essential attributes were also removed. The final compiled data set consisted of 4,701 records. In this final sample of accidents, the proportions of injury severity for the reported categories were as follows: (a) PDO – 0.4%, (b) minor injury – 1.0%, (c) serious injury – 89.0%, and (d) fatal injury – 9.7%. To obtain a reasonable sample share for all alternatives the PDO and minor injury categories were merged for our analysis. From the final data set, 4,258 records are sampled out for the purpose of model estimation and the remaining 443 records are set aside for validation. Table 2 offers a summary of the sample characteristics of the dataset. From the descriptive analysis, we observe that a large portion of crashes occur at an intersection (72.3%), in clear weather (71.9%), in the presence of daylight (56.3%), and on urban streets (96.9%). The majority of pedestrians are adult (59.2%) and mostly struck by a passenger vehicle (74.1%).

## 5. Empirical Analysis

### 5.1. Variables Considered

For our analysis, we selected variables from six broad categories: crash characteristics (crash location), environmental factors (weather condition, season, and lighting condition) vehicle characteristics (type of vehicle), roadway design and operational attributes (roadway class,<sup>2</sup> travel lane, and parking lane), land use characteristics (boroughs), and pedestrian characteristics (pedestrian age). The impact of several variables such as presence of shoulder, shoulder width, vehicle weight, point of impact, pedestrian condition, driver condition, at fault-status, and time of day could not be explored, because the information for these variables was either entirely unavailable or there was a large fraction of missing data for these attributes in the data set. To be sure, we employed lighting condition and vehicle type to act as surrogates for time of day and vehicle weight, respectively. In the final specification of model, statistically insignificant variables were removed through a systematic process based on statistical significance (90% significance) from the universal variable set. The insignificant variables from our analysis include day-of-week and trajectory of vehicle's

<sup>2</sup>Roadway class served as a surrogate for speed limit.

**Table 2**  
Crash database sample statistics

Categorical Explanatory Variables	Sample Share	
	Frequency	Percentage
Crash location		
At intersection	3079	72.31
At midblock	1179	27.69
Weather condition		
Snowy/foggy	47	1.10
Clear	3061	71.89
Cloudy	555	13.03
Rain	595	13.97
Season		
Winter	1041	24.45
Spring	1024	24.05
Summer	1060	24.89
Autumn	1133	26.61
Light conditions		
Daylight	2396	56.27
Dawn	100	2.35
Dusk	195	4.58
Dark road - lighted	1518	35.65
Dark road - unlighted	49	1.15
Vehicle type		
Other vehicle type	827	19.42
Car/van/pickup	3155	74.10
Truck	140	3.29
Bus	106	2.49
Motorcycle	30	0.70
Roadway class		
Town	15	0.35
Urban street	4125	96.88
Parkway	35	0.82
Parking lot & other nontraffic	52	1.22
Highway	31	0.73
Land use		
Family residential	729	17.12
Mixed residential and commercial	403	9.46
Commercial and office	752	17.66
Industrial/manufacturing	953	22.38
Open space & vacant land	426	10.00

**Table 2**  
Crash database sample statistics (*Continued*)

Categorical Explanatory Variables	Sample Share	
	Frequency	Percentage
Parking facilities and transportation utility	750	17.61
Public facilities and institutions	153	3.59
Miscellaneous lots	92	2.16
Boroughs		
Bronx	661	15.52
Brooklyn	1408	33.07
Manhattan	1178	27.67
Queens	848	19.92
Staten island	163	3.83
Pedestrian age		
Children (Pedestrian age 12 and younger)	612	14.37
Teenager (Pedestrian age 13–18)	378	8.88
Adult (Pedestrian age 19–65)	2521	59.21
Older (Pedestrian age 65+)	747	17.54
Ordinal explanatory variables	Mean	
Travel lane	2.030	
Parking lane	1.420	

motion. Further, in cases where the variable effects were not significantly different across different variable levels the coefficients were restricted to be the same.

### 5.2. Overall Measures of Fit

We estimated four models: (1) OL, (2) GOL, (3) LSOL with two segments (LSOL II), and (4) LSOL with three segments (LSOL III). After extensively testing for three segments in latent segmentation approach we found that the model collapses to the two segment model. Hence, from here on, the entire comparison exercise is focused on three models: OL, GOL, and LSOL II. Prior to discussing the estimation results, we compare the performance of these models. The GOL is a generalized version of OL. Thus, we can compare these two models by using likelihood ratio test for selecting the preferred model. However, the comparison with the LSOL II model using the likelihood ratio is not possible because these structures are not nested within one another. Hence, we employ different measures that are routinely applied in comparing econometric models including (1) Bayesian information

**Table 3**  
Measures of fit in estimation sample

Summary Statistic	OL	GOL	LSOL II
Log-likelihood at zero	−4677.9	−4677.9	−4677.9
Log-likelihood at sample shares	−1654.5	−1654.5	−1654.5
Log-likelihood at convergence	−1481.6	−1454.5	−1456.19
Number of parameters	20	22	16
Number of observations	4258	4258	4258
BIC	3120.1	3092.8	3046.1
AICc	3007.5	2953.2	2944.5
Adjusted likelihood ration index ( $\bar{\rho}_2^2$ )	0.092	0.108	0.110

OL = Ordered logit model; GOL = generalized ordered logit model; LSOLII = latent segmentation based ordered logit model with two segment; BIC = Bayesian information criterion; AICc = Akaike information criterion corrected.

criterion (BIC), (2) Akaike information criterion corrected (AICc),<sup>3</sup> and (3) Ben-Akiva and Lerman's adjusted likelihood ratio (BL) test. These estimates are presented in Table 3.

The BIC for a given empirical model is equal to  $-2\ln(L) + K \ln(Q)$  and the AICc for an empirical model is given by  $AIC + [2 K(K+1)/(Q - K - 1)]$ , where  $\ln(L)$  is the log-likelihood value at convergence,  $K$  is the number of parameters, and  $Q$  is the number of observations. The model with the lower BIC and AICc values is the preferred model. The numbers in Table 3 indicate that the LSOL II model has lower values for both measures.

The BL test statistic (Ben-Akiva & Lerman, 1985) is computed as:  $\lambda = \Phi\{-[\sqrt{-2(\bar{\rho}_2^2 - \bar{\rho}_1^2)L(C)} + (M_2 - M_1)]\}$ , where  $\bar{\rho}^2$  represents the McFadden's adjusted rho-square value for the model. It is defined as  $\bar{\rho}_i^2 = 1 - \frac{L_i(\beta) - M_i}{L(C)}$ , where  $L_i(\beta)$  represents log-likelihood at convergence for the  $i$ th model,  $L(C)$  represents log-likelihood at sample shares and  $M_i$  is the number of parameters in the model (Windmeijer, 1995). The  $\Phi(\cdot)$  represents the cumulative standard normal distribution function. This BL test statistic identifies if the adjusted likelihood ratio indices of the two non-nested models are significantly different. It compares two models by computing the probability ( $\lambda$ ) that we could have obtained the higher  $\bar{\rho}^2$  value for the "best" model even though this is not the case. The resulting  $\lambda$  values (presented in Table 3) thus clearly indicate that LSOL II offers superior fit compared to OL and GOL models at any significance level (LSOL II has the higher  $\bar{\rho}^2$  with lowest number of parameters compared to OL and GOL). The comparison exercise, therefore, highlights the superiority of the LSOL II model in terms of data fit compared to the OL and GOL models. In the following discussion, we always denote LSOL II as LSOL for simplicity.

<sup>3</sup>AICc is a more stringent version of the AIC [ $AIC = 2K - 2\ln(L)$ ] in penalizing for additional parameters.

### 5.3. Estimation Results

To conserve on space, the subsequent discussions of exogenous variable impacts are restricted to LSOL estimates. Table 4 presents the estimation results of the LSOL model. An intuitive discussion of the LSOL model is presented followed by the discussion of segmentation component parameters and severity component parameters specific to Segments 1 and 2.

*5.3.1. Intuitive Interpretation of LSOL Model.* To delve into the segmentation characteristics, the model estimates are used to generate information on (1) population share across the two segments and (2) overall injury severity shares within each segment. These estimates are presented in Table 4. From the estimates, it is evident that the probability of pedestrians being assigned to Segment 2 is substantially higher than the probability of being assigned to Segment 1. Further, the likelihood of injury severity for pedestrians conditional on their belonging to a particular segment offer contrasting results indicating that two segments exhibit distinct injury severity profiles in the current research context. It is clear that pedestrians assigned to Segment 1 are more likely to sustain fatalities whereas those assigned to Segment 2 are more likely to sustain serious injury. To facilitate the discussion from here on, we label segment 1 as the “fatality segment” and segment 2 as the “serious injury segment,”

*5.3.2. Latent Segmentation Component.* The latent segmentation component determines the probability that a pedestrian is assigned to one of the two latent segments based on the crash location attributes. The latent segmentation component coefficients correspond to the likelihood of assigning pedestrians to serious injury segment. The positive sign of the constant term indicates a larger likelihood for pedestrians being assigned to serious injury segment. Crash location attributes that affect the allocation of pedestrians, into either of the segment, include regional county, functional classification of roadway, pedestrian location on roadway, number of travel lanes, and number of parking lanes in the roadway system.

NYC is composed of five boroughs: Manhattan, Bronx, Brooklyn, Queens, and Staten Island. From the estimated results, we find that the likelihood of pedestrian being assigned to serious injury segment is higher for Manhattan compared to other boroughs. On the other hand, pedestrian crashes occurring in Staten Island are likely to be assigned to fatality segment. The coefficients for functional classification of roadway indicate that highways and parkways increase the likelihood of pedestrians being assigned to fatality segment. It is not surprising that these variables increase the likelihood of a pedestrian being assigned to fatality segment, because highways and parkways represent roadway facilities with the highest speeds. The result associated with the crash at midblock location reflects an increase likelihood of pedestrian being assigned to fatality segment. Motorists may not expect pedestrians at locations not designated for them, and they tend to drive at a higher speed at midblock compared to intersection (Kim et al., 2010). Consequently, they might fail to notice the pedestrian in time leading to increased propensity of fatality. An increase in total number of travel lanes in roadway increases the likelihood of assigning the pedestrians to fatality segment. Generally, higher the number of travel lanes, higher is the width of the roadway. Wider roads imply longer exposure time for pedestrians to vehicular traffic, and higher vehicular speed, increasing the likelihood of pedestrian fatalities (Sze & Wong, 2007; Tay et al., 2011). The result associated with parking lanes indicates that an increase in number of parking lanes on roadway increase the likelihood of assigning the pedestrians to serious injury segment. The presence of on-street parking, in general, is expected to force



**Table 4**

Latent segmentation based ordered logit model with two segments (LSOL II) estimates

	Segments			
	Segment 1		Segment 2	
Pedestrian population share	0.11		0.89	
Injury severity				
Property damage only (PDO)/minor injury	6.07		0.84	
Serious injury	49.13		93.82	
Fatal	44.80		5.34	
Parameter Estimates				
Explanatory variables	Segment 1		Segment 2	
	Estimate	t-stat	Estimate	t-stat
Segmentation components				
Constant	—	—	3.0813	7.4560
Regional country (Reference: Bronx, Brooklyn, Queens)				
Manhattan	—	—	0.5560	2.1420
Staten island	—	—	−0.8659	−1.8810
Functional class of roadway (Reference: Urban street)				
Highway and parkway	—	—	−2.6484	−3.9940
Pedestrian location (Reference: Pedestrian at intersection)				
Pedestrian at midblock	—	—	−1.2367	−5.0650
Travel lane	—	—	−0.4107	−4.6530
Parking lane	—	—	0.2940	2.4460
Injury severity components				
Threshold parameters				
Threshold 1	−2.8668	—	−4.8333	—
Threshold 2	0.4893	—	3.9953	—
Weather condition (Reference: Clear and rain)				
Snowy/foggy	−3.9723	−2.6350	—	—
Cloudy	—	—	0.5210	2.1540
Light conditions (Reference: Daylight)				
Dark road - lighted	1.4309	2.3580	−0.4675	−1.7630
Vehicle type (Reference: Car/van/pickup)				
Truck	—	—	2.5982	8.0080
Bus	—	—	2.1061	5.7110
Pedestrian age (Reference: Adult)				
Children	−0.7475	−1.6760	—	—
Teenager	−1.3881	−2.6140	—	—
Older	—	—	2.4296	7.5770
Season (Reference: Winter, Summer, Autumn)				
Spring	—	—	−0.5052	−2.2530

the drivers to be more watchful for parked cars entering the roadway or sporadic pedestrian movement as people get in and out of their vehicles (Zajac & Ivan, 2003). Overall, fatality segment is characterized by high-speed facilities in the roadway system, crashes occurring at midblock location of road, increased number of travel lanes and lower number of parking lanes in roadway system.

*5.3.3. Injury Severity Component: Segment 1.* The injury severity component within the fatality segment is discussed in this section. The interpretation of the coefficients follows the usual ordered response framework.

Weather is considered to be one of the most important environmental components that affect driving. The results presented in Table 4 indicate that snowy/foggy weather condition results in less severe pedestrian crashes compared to the clear and rainy weather. The reduced probability of severe pedestrian crashes during snowy/foggy period perhaps can be attributed to the reduced driving speed (Eluru et al., 2008; Kim et al., 2010) and more cautious pedestrian activities. The unfavorable driving conditions during adverse weather perhaps also result in increased driver attention thereby reducing the severity of pedestrian crashes.

Visibility significantly affects driver and pedestrian activity. In this study, the lighting condition has significant association with the pedestrian injury severity propensity. It is very interesting to note that crashes of fatality segment occurring in the presence of artificial illumination (street lights) during dark periods increases the likelihood of severe pedestrian injury compared to other lighting conditions (daylight, dawn, and dusk). Problems associated with darkness at night time could be attributed to poor visual conditions, higher vehicular speed, fatigue, possible negligence, inattentiveness, and driving and walking under the influence of alcohol (Lee & Abdel-Aty, 2005; Kim et al., 2008, 2010; Montella et al., 2011; Rifaat et al., 2011; Sze & Wong, 2007; Tay et al., 2011). These conditions perhaps increase the reaction time and braking distance of vehicles and lead to greater impact at the time of a crash. Further, longer response time by emergency crews at the crash location during darkness (Klop & Khattak, 1999) could also exacerbate the resulting pedestrian crash severity.

The relevance of pedestrian age has long been recognized as an important contributory factor in pedestrian crash severity studies. The model results reveal a reduction in the risk propensity for child and teenager pedestrian groups compared to the adult group perhaps because these pedestrian groups are more physically fit (Lee & Abdel-Aty, 2005) compared to other pedestrians. Moreover, child pedestrians are less likely to be transgressors of traffic rules (Kim et al., 2008b) and thus unlikely to be involved in severe crashes. Thresholds in the ordered response model form the boundary points for different levels of injury severities. In the first segment, when the latent propensity of the individual is less than  $-2.87$  the pedestrian sustains PDO/minor injury. The pedestrian sustains a serious injury when the propensity is between  $-2.87$  and  $0.50$ . The pedestrian is fatally injured when the propensity value is greater than  $0.50$ .

*5.3.4. Injury Severity Component: Segment 2.* The OL model corresponding to serious injury segment provides variable impacts that are significantly different, in magnitude as well as in sign (for a few variables), from the impacts offered by the exogenous variables in fatality segment. Further, we also notice that the number of variables that moderate the influence of injury severity is significantly higher for serious injury segment.

In the second segment, the results indicate that cloudy weather affects the injury sustained by a pedestrian in a crash. In particular, cloudy weather results in more severe

crashes compared to the clear and rainy weather perhaps because of the reduced visibility, which presumably results in reduced perception-reaction and reduced ability to take evasive actions at the crash incident (Tay *et al.*, 2011). Cloudy days perhaps also adversely affect human psychology. For example, Kim *et al.* (2008b) observed less compliance with traffic rules on behalf of the drivers and pedestrians on a cloudy day.

The influence of lighting condition has a strikingly different influence on the pedestrians compared to the effect in fatality segment. For the dark-lighted condition, the latent propensity is found negative, which indicates a lower injury risk propensity of pedestrian in serious injury segment. This might indicate more heedful movement of pedestrians and more effective avoidance maneuver of drivers during darkness. The result also highlights how the same variable can have distinct influence on injury severity based on the segment to which the crash is allocated. The LSOL approach allows for capturing such complex interactions. With respect to vehicle type involved in a crash with pedestrian, these study results show that a pedestrian struck by a truck or a bus has a higher injury risk propensity. The reason may be attributed to the heavier vehicle mass, greater stiffness, greater momentum, large area of impact for pedestrians, higher bumper height, blunter geometry, and longer stopping distances of bus and trucks compared to other vehicles (Eluru *et al.*, 2008; Kim *et al.*, 2008; Lee & Abdel-Aty, 2005; Montella *et al.*, 2011; Tay *et al.*, 2011). For bus, the weaving pattern of movement in traffic due to frequent on-street bus stops could also impose higher fatality risk on pedestrians.

The influence of pedestrian age on crash severity is along expected lines. It is found that the older pedestrians are associated with the higher likelihood of severe crashes compared to the adult pedestrian groups. Older pedestrians might be physically weak and may be medically unfit with problems related to hearing, vision, and contrast sensitivity (Lee & Abdel-Aty, 2005; Kim *et al.*, 2010). Moreover, older pedestrians tend to be slow in reacting to hazardous situations, walk slower, can withstand low impact force of crash impact, and may select insufficient gaps while crossing the road (Anowar *et al.*, 2010; Sze & Wong, 2007); all of which contribute to their higher severity risk. The result associated with the season reflect that injury risk propensity of pedestrian crashes is lower in spring compared to any other season. The result is quite interesting, and the reasons for the effect are not very clear. It is possibly a manifestation of some unobserved variables that are not considered in our analysis.

In the serious injury segment, when the latent propensity of the individual is less than  $-4.83$  the pedestrian sustains PDO/minor injury. The pedestrian sustains a serious injury when the propensity is between  $-4.83$  and  $3.99$ . The pedestrian is fatally injured when the propensity value is greater than  $3.99$ . This again highlights the difference between the two segments.

#### **5.4. Elasticity Effects**

The parameter effects of the exogenous variables do not provide the magnitude of the effects on injury severity. For this purpose, we compute the aggregate-level “elasticity effects” for all independent variables (Eluru & Bhat, 2007) for a discussion on the methodology for computing elasticities) and present the computed elasticities in Table 5. The effects are computed for the GOL and LSOL models.

The following observations can be made based on the results presented in Table 5. First, the results from the elasticity effects indicate that there are substantial differences in the elasticity effects of these two models. For instance, the LSOL model predicts a reduction in PDO/minor injury for pedestrians in Manhattan region whereas the GOL model predicts an

**Table 5**  
Elasticity effects

Variables	LSOL II			GOL		
	PDO/Minor Injury	Serious Injury	Fatal Injury	PDO/Minor Injury	Serious Injury	Fatal Injury
Manhattan	−16.05	2.22	−18.25	34.67	2.24	−25.31
Staten Island	36.22	−4.83	39.51	−36.54	−4.03	42.04
Highway	167.058	−22.153	181.128	628.96	−24.28	134.67
Parkway	167.058	−22.153	181.128	−75.58	−17.78	173.39
Pedestrian at midblock	46.58	−6.39	52.48	56.72	−7.08	56.95
Travel lane	15.25	−2.03	16.61	0.00	−2.31	21.13
Parking lane	−8.58	1.15	−9.42	15.88	1.02	−11.52
Snowy/foggy	466.79	−2.08	−48.48	173.41	3.89	−59.79
Cloudy	−22.98	−2.07	22.53	−37.81	−3.86	40.65
Dark-lighted	−18.20	−1.81	19.46	0.00	−4.14	37.94
Dark-unlighted	—	—	—	343.97	−17.92	116.14
Truck	−51.23	−22.09	212.26	−86.55	−26.15	251.52
Bus	−47.82	−15.46	150.28	0.00	−18.60	170.27
Pedestrian age 12 and younger	35.60	1.28	−17.02	39.60	2.27	−26.32
Pedestrian age 13–18	83.70	1.85	−29.32	89.59	3.64	−45.81
Pedestrian age 65+	−57.75	−14.57	143.44	−83.96	−14.87	147.89
Spring	30.52	1.42	−17.61	24.51	1.62	−18.22
Industrial and vacant land	—	—	—	73.27	0.17	−11.75

LSOL II = Latent segmentation based ordered logit model with two segments; GOL = generalized ordered logit model; PDO = property damage only.

increase in the PDO/minor injury category. In a similar vein, differences can be observed for travel lane and dark-lighted variables. Second, the most significant variables in terms of fatal injury (from both models) for pedestrians are pedestrian age 65 and older, crash with a truck or bus, crash occurred on a highway or parkway. In terms of fatal injury reduction the important factors are pedestrian age between age 13 and 18 and snowy conditions. Finally, the dark unlighted variable exhibits substantially different impacts in the LSOL and GOL models. The reason for the difference could be attributed to differences in the model structures of the LSOL and GOL models.

### 5.5. Validation Analysis

We also evaluate the performance of these models on a validation sample. We use the aggregate and disaggregate measures of fit for the validation exercise. At the disaggregate level we calculate predictive log-likelihood (computed by calculating the log-likelihood for the predicted probabilities of the sample), probability of correct prediction (computed as the

probability that the chosen alternative has the highest predicted probability), and probability of correct prediction  $>0.7$  (computed as the probability that the chosen alternative has the highest predicted probability and it is also  $>0.7$ ). At the aggregate level, root mean square error (RMSE) and mean absolute percentage error (MAPE) are computed by comparing the predicted and actual (observed) shares of injuries for each injury severity level. We compute these measures for the complete validation sample and specific sub-samples in the population: Manhattan, cloudy weather, dark road-lighted, pedestrian age 12 and younger, and spring. The results for computation of these measures are presented in Table 6.

At the disaggregate level the fit measures show that GOL performs marginally better than LSOL. At the aggregate level, the RMSE and MAPE values are very close for the two model systems. It is important to note that with six fewer parameters the LSOL model performance only marginally poorly compared to the performance of the GOL model. The close performance of the two model systems is further illustrated through the computation of the validation measures for various subsamples of the population. The results indicate that LSOL and GOL models offer very similar prediction for the various subsamples at the aggregate and disaggregate level. The results reinforce the strengths of the LSOL model to perform very close to the GOL model.

## 6. Discussion and Recommendations

The results obtained in our analysis have important implications for improving the safety of pedestrians in terms of enforcement, engineering, and educational strategies (3 “E” approach). With respect to enforcement and education, our results endorse a continuous education program and stricter enforcement to prevent unsafe road crossing behavior of pedestrians. Most importantly, public education campaigns are needed to encourage the pedestrians to wear “reflectors” to increase their conspicuity in darkness (European Commission, 2007). Retroreflective materials (lamps, flashing lights, or retroreflective vest) enhance the drivers’ detection of pedestrians at nighttime (Kwan & Mapstone, 2009; Luoma *et al.*, 1996; Muttart, 2000; Wood, 2005) and hence in turn reduce the pedestrian crash severity.

Policies concerning the visual enhancement of pedestrians during night time—in the form of artificial illumination, increased intensity of roadway lighting, adaptive headlighting solution, illuminated crosswalk, illuminated warning sign and smart lighting—could improve pedestrian safety (Polus & Katz, 1978; Ragland *et al.*, 2003; Sullivan & Flannagan, 1999, 2007). Enhanced visibility is identified in several previous studies to be the most cost-effective measures in improving pedestrian safety and the greatest safety benefit from it is observed in reducing fatal pedestrian crashes (Lalani and Institute of Transportation Engineers Pedestrian and Bicycle Task Force, 2001; Ragland *et al.*, 2003; Sullivan & Flannagan, 1999). Campaigns should also address the issue of increased fatality risk of older pedestrians and advise seniors of the potential risk and suggest avoiding walking during night time and on high-speed corridors. However, there is lack of evidence regarding the impact of safety education or training for this group of pedestrians (Duperrex *et al.*, 2002).

In terms of engineering measures, pedestrian crossing should be designed either to be space or time separated from vehicular movement. In terms of space separation, pedestrian crossings should be separated by barrier and fences or off-road or grade-separated facility on high-speed corridors and on the roadways with more travel lanes. Several previous studies have demonstrated the safety benefits of such facilities (Berger, 1975; Elvik *et al.*, 2009; Retting *et al.*, 2003). The greater injury severity of pedestrian in the event of crashes with bus and truck further endorse the importance of separating pedestrian traffic from

**Table 6**  
Measures of fit in validation sample

Disaggregate Measure of Fit in Validation Sample			
Summary Statistic		GOL	LSOL II
Number of observations		443	443
Predictive log-likelihood		−168.43	−168.94
Average probability of correct prediction		88.49	83.30
Average probability for chosen probability > 0.70		83.97	85.78
Aggregate Measure of Fit in Validation Sample			
Injury Categories/	Actual	GOL	LSOL II
Measures of Fit	Shares	Predictions	Predictions
PDO/minor injury	1.80	1.39	1.47
Serious injury	88.90	88.86	88.12
Fatal injury	9.30	9.75	10.41
RMSE	—	0.30	0.70
MPAE	—	6.89	7.82
Manhattan			
PDO/minor injury	1.60	1.59	1.39
Serious injury	87.70	89.55	88.56
Fatal injury	10.70	8.86	10.05
RMSE	—	10.78	10.20
MPAE	—	11.73	10.98
Predictive log-likelihood	—	−52.98	−52.65
Cloudy			
PDO/minor injury	3.2	0.99	1.22
Serious injury	88.9	88.13	89.12
Fatal injury	7.9	10.87	9.66
RMSE	—	16.75	16.45
MPAE	—	32.82	33.95
Predictive log-likelihood	—	−27.76	−28.80
Dark road – lighted			
PDO/minor injury	3.1	1.42	1.48
Serious injury	86.3	87.81	87.44
Fatal injury	10.6	10.77	11.08
RMSE	—	27.30	27.04
MPAE	—	37.96	38.26
Predictive log-likelihood	—	−79.29	−77.10
Pedestrian age 12 and less			
PDO/minor injury	1.5	1.76	1.86
Serious injury	95.5	93.64	93.35
Fatal injury	3.0	4.61	4.80
RMSE	—	16.38	16.48
MPAE	—	14.70	14.68
Predictive log-likelihood	—	−13.36	−14.15

(Continued on next page)

**Table 6**  
Measures of fit in validation sample (*Continued*)

Injury Categories/ Measures of Fit	Actual Shares	GOL Predictions	LSOL II Predictions
Spring			
PDO/minor injury	0.9	1.66	1.65
Serious injury	92.0	89.86	89.69
Fatal injury	7.1	8.48	8.65
RMSE	—	4.95	4.89
MPAE	—	38.40	38.89
Predictive log-likelihood	—	−32.41	−32.76

GOL = Generalized ordered logit model; LSOLII = latent segmentation based ordered logit model with two segments; PDO = property damage only; RMSE = root mean square error; MPAE = mean absolute percentage error.

heavy-vehicle traffic flow on roadways. With respect to time separation, pedestrians might be separated from vehicular movement by installation of traffic signal, by implementing exclusive pedestrian signal phase or by installing pedestrian prompting devices (Retting et al., 2002). The implementation of these recommendations will enhance pedestrian safety.

## 7. Conclusions

This article focuses on identifying the appropriate ordered response structure for modeling pedestrian injury severity. Pedestrian injury severity is often reported as an ordered variable resulting in the application of ordered response models for analyzing risk factors. However, the traditional ordered response model (OL/OP) restrict the impact of risk factors to be the same across all alternatives, thus the model cannot identify risk factors that specifically influence a particular injury category. On the contrary, the GOL model relaxes the restrictive assumption by allowing for exogenous variable impacts on the threshold parameters in the OL structure. Another ordered response approach, the LSOL model, allows for differential impact on the alternatives by segmenting the pedestrian crash population into various segments with segment specific OL parameters. Earlier research efforts have concluded that these approaches are substantially better than the traditional OL model. However, a comparison of these two model frameworks has not been undertaken. Toward this end we undertake a comparison exercise involving four ordered response frameworks: OL, GOL, LSOL with two segments (LSOL II), and LSOL with three segments (LSOL III) to identify the preferred ordered model for examining pedestrian injury severity. The comparison exercise is conducted to identify the various risk factors affecting pedestrian injury severity in terms of model estimation as well as model validation (using a hold-out sample). The empirical analysis is conducted using the NYC Pedestrian Research Data Base for the years of 2002 through 2006. The ordered response models are estimated using a comprehensive set of exogenous variables including crash characteristics, environmental factors, vehicle characteristics, roadway design and operational attributes, land use characteristics, and pedestrian characteristics. The comparison exercise highlights the superiority of the LSOL II model on the estimation sample in terms of data fit compared to the OL and GOL models. In the LSOL approach, pedestrian crashes are assigned probabilistically to two segments—fatality segment and serious injury segment—based on a host of crash

location attributes. The fatality segment is characterized by crashes on high speed roadway facility, midblock location of road, increased number of travel lanes, and decreased number of parking lanes in the roadway system. In the fatality segment, crash during dark-lighted period contribute to increasing the injury severity while snowy/foggy weather, pedestrian age 12 and younger, and pedestrian age 13 to 18 reduce the injury severity. On the other hand, for the serious injury segment the results indicate that cloudy weather, crash with truck or bus and pedestrian age 65+ increase injury severities whereas crash during dark-lighted period and spring season are likely to reduce injury severity. Overall, the number of variables moderating the effect of injury severity is higher in the serious injury segment compared to the fatality segment. Further, some variables provide different impacts, in magnitude as well as in sign, for two segments highlighting how the same variable can have distinct influence on injury severity based on the segment to which the pedestrian is allocated. Thus, target specific countermeasures for “fatality segment” crash locations might ensure an effective reduction of crash related pedestrian fatalities of NYC.

In our research, to further understand the impact of various exogenous factors, elasticity effects for the exogenous variables for the LSOL and GOL models are computed. The elasticity effects indicate that the most significant variables in terms of fatal injury (from both models) for pedestrians are pedestrian age 65 and older, crash with a truck or bus and crash occurred on a highway or parkway.

The results from the elasticity effects also indicate that there are substantial differences in the elasticity effects of these two models, and these differences might be attributed to the implicit structural differences of these two frameworks. The performance evaluation of these models on a validation sample reveals that GOL model performs marginally better than the LSOL model. The differences in the validation measures at aggregate and disaggregate level are very small for the two frameworks. The improvement in validation predictions for the GOL model is obtained at the cost of six additional parameters. Overall, the comparison exercise supports the hypothesis that the GOL and LSOL models clearly provide a better fit than standard OL model. The empirical results also show that LSOL model (with six fewer parameters compared to GOL) performance is satisfactory relative to the GOL model performance in the current research context. Further, it is important to reemphasize the fact that LSOL model provides valuable insights on how the explanatory variables affect segmentation of pedestrians into fatality and serious injury segments. In conclusion, the results from our analysis identify LSOL model as a promising ordered response framework for accommodating population heterogeneity in the context of pedestrian injury severity.

## Acknowledgments

The authors would like to acknowledge insightful feedback from three anonymous reviewers on a previous version of the article.

## Funding

Naveen Eluru would like to acknowledge financial support from Fonds de recherche du Québec - Nature et technologies (FQRNT) Team Grant Program for undertaking the research.

## References

- Abay, K. A. (2013). Examining pedestrian-injury severity using alternative disaggregate models. *Research in Transportation Economics*, 43(1), 123–136.



- Anowar, S., Yasmin, S., & Tay, R. (2010). A logistic model for examining pedestrian fatality risk in traffic crashes. In *Proceedings of the 20th Canadian Multidisciplinary Road Safety Conference (CMRSC)*, June 6–9, Niagara Falls, Ontario, Canada, pp. 1–16.
- Aziz, H. M. A., Ukkusuri, V. S., & Hasan, S. (2013). Exploring the, in New York City. *Accident Analysis & Prevention*, 50, 1298–1309.
- Ballesteros, M. F., Dischinger, P. C., & Langenberg, P. (2004). Pedestrian injuries and vehicle type in Maryland, 1995–1999. *Accident Analysis & Prevention*, 36(1), 73–81.
- Ben-Akiva, M. E., & Lerman, S. R. (1985). *Discrete choice analysis: Theory and application to travel demand*. Cambridge, MA: MIT Press.
- Berger, W. G. (1975). *Urban pedestrian accident countermeasures experimental evaluation: Volume I- Behavioral studies*. Washington, DC: U.S. Dept of Transportation.
- Castro, M., Paletti, R., & Bhat, C. R. (2013). A spatial generalized ordered response model to examine highway crash injury severity. *Accident Analysis & Prevention*, 52, 188–203.
- Clifton, K. J., Burnier, C. V., & Akar, G. (2009). Severity of injury resulting from pedestrian-vehicle crashes: What can we learn from examining the built environment? *Transportation Research Part D: Transport and Environment*, 14(6), 425–436.
- Duperrex, O., Bunn, F., & Roberts, I. (2002). Safety education of pedestrians for injury prevention: A systematic review of randomised controlled trials. *British Medical Journal*, 324(7346), 1129–1131.
- Eluru, N. (2013). Evaluating alternate discrete choice frameworks for modeling ordinal discrete variables. *Accident Analysis & Prevention*, 55(1), 1–11.
- Eluru, N., Bagheri, M., Miranda-Moreno, L. F., & Fu, L. (2012). A latent class modeling approach for identifying vehicle driver injury severity factors at highway-railway crossings. *Accident Analysis & Prevention*, 47, 119–127.
- Eluru, N., & Bhat, C. R. (2007). A joint econometric analysis of seat belt use and crash-related injury severity. *Accident Analysis & Prevention*, 39(5), 1037–1049.
- Eluru, N., Bhat, C. R., & Hensher, D. A. (2008). A mixed generalized ordered response model for examining pedestrian and bicyclist injury severity level in traffic crashes. *Accident Analysis & Prevention*, 40(3), 1033–1054.
- Elvik, R. (2009). The non-linearity of risk and the promotion of environmentally sustainable transport. *Accident Analysis & Prevention*, 41(4), 849–855.
- Elvik, R., Vaa, T., & Sørensen, M. (2009). *The handbook of road safety measures*. Bingley, UK: Emerald Group Publishing Ltd.
- European Commission. (2007). Best Practices in Road Safety - Handbook for Measures at the Country Level. Summary and Publication of Best Practices in Road Safety in the European Member States (SUPREME), European Commission.
- Gårder, P. E. (2004). The impact of speed and other variables on pedestrian safety in Maine. *Accident Analysis & Prevention*, 36(4), 533–542.
- Greene, W. H., & Hensher, D. A. (2003). A latent class model for discrete choice analysis: Contrasts with mixed logit. *Transportation Research Part B: Methodological*, 37(8), 681–698.
- Keall, M. D. (1995). Pedestrian exposure to risk of road accident in New Zealand. *Accident Analysis & Prevention*, 27(5), 729–740.
- Kim, J.-K., Ulfarsson, G. F., Kim, S., & Shankar, V. N. (2013). Driver-injury severity in single-vehicle crashes in California: A mixed logit analysis of heterogeneity due to age and gender. *Accident Analysis & Prevention*, 50, 1073–1081.
- Kim, J. K., Ulfarsson, G. F., Shankar, V. N., & Kim, S. (2008). Age and pedestrian injury severity in motor-vehicle crashes: A heteroskedastic logit analysis. *Accident Analysis & Prevention*, 40(5), 1695–1702.
- Kim, J.-K., Ulfarsson, G. F., Shankar, V. N., & Mannering, F. L. (2010). A note on modeling pedestrian-injury severity in motor-vehicle crashes with the mixed logit model. *Accident Analysis & Prevention*, 42(6), 1751–1758.
- Kim, K., Brunner, I. M., & Yamashita, E. (2008a). Modeling fault among accident-involved pedestrians and motorists in Hawaii. *Accident Analysis & Prevention*, 40(6), 2043–2049.

- Kim, K., Brunner, I. M., & Yamashita, E. (2008b). Modeling violation of Hawaii's crosswalk law. *Accident Analysis & Prevention*, 40(3), 894–904.
- Klop, J. R., & Khattak, A. J. (1999). Factors influencing bicycle crash severity on two-lane, undivided roadways in North Carolina. *Transportation Research Record: Journal of the Transportation Research Board*, 1674, 78–85.
- Kwan, I., & Mapstone, J. (2009). Interventions for increasing pedestrian and cyclist visibility for the prevention of death and injuries. *Cochrane Database of Systematic Reviews* (4), CD003438.
- Kwigizile, V., Sando, T., & Chimba, D. (2011). Inconsistencies of ordered and unordered probability models for pedestrian injury severity. *Transportation Research Record: Journal of the Transportation Research Board*, 2264, 110–118.
- Lalani, N., and Institute of Transportation Engineers Pedestrian and Bicycle Task Force. (2001). Alternative Treatments for At-Grade Pedestrian Crossings. Publication LP-629. Institute of Transportation Engineers, Washington, D.C.
- Lee, C., & Abdel-Aty, M. (2005). Comprehensive analysis of vehicle-pedestrian crashes at intersections in Florida. *Accident Analysis & Prevention*, 37(4), 775–786.
- Loukaitou-Sideris, A., Liggett, R., & Sung, H. (2007). Death on the crosswalk: A study of pedestrian-automobile collisions in Los Angeles. *Journal of Planning Education and Research*, 26(3), 338–351.
- Luoma, J., Schumann, J., & Traube, E. C. (1996). Effects of retroreflector positioning on nighttime recognition of pedestrians. *Accident Analysis & Prevention*, 28(3), 377–383.
- Maddala, G. S. (1983). *Limited-dependent and qualitative variables in econometrics*. New York, NY: Cambridge University Press.
- Mohamed, M. G., Saunier, N., Miranda-Moreno, L. F., & Ukkusuri, S. V. (2013). A clustering regression approach: A comprehensive injury severity analysis of pedestrian-vehicle crashes in New York, US and Montreal, Canada. *Safety Science*, 54, 27–37.
- Montella, A., Aria, M., D'Ambrosio, A., & Mauriello, F. (2011). Data-mining techniques for exploratory analysis of pedestrian crashes. *Transportation Research Record: Journal of the Transportation Research Board*, 2237, 107–116.
- Mooradian, J., Ivan J. N., Ravishanker, N., & Hu, S. (2013). Analysis of driver and passenger crash injury severity using partial proportional odds models. *Accident Analysis & Prevention*, 58, 53–58.
- Morgan, A., & Mannering, F. L. (2011). The effects of road-surface conditions, age, and gender on driver-injury severities. *Accident Analysis & Prevention*, 43(5), 1852–1863.
- Moudon, A. V., Lin, L., Jiao, J., Hurvitz, P., & Reeves, P. (2011). The risk of pedestrian injury and fatality in collisions with motor vehicles, a social ecological study of state routes and city streets in King County, Washington. *Accident Analysis & Prevention*, 43(1), 11–24.
- Muttart, J. (2000). Effects of retroreflective material on pedestrian identification at night. *Accident Reconstruction Journal*, 11(1), 51–57.
- National Highway Traffic Safety Administration. (2009). *2008 pedestrians traffic safety fact sheet*. Washington, DC: National Highway Traffic Safety Administration, National Center for Statistics and Analysis.
- New York City Department of Transportation. (2010). *The New York City Pedestrian Safety Study and action plan*. New York, NY: New York City Department of Transportation.
- Paleti, R., Eluru, N., & Bhat, C. R. (2010). Examining the influence of aggressive driving behavior on driver injury severity in traffic crashes. *Accident Analysis & Prevention*, 42(6), 1839–1854.
- Polus, A., & Katz, A. (1978). An analysis of nighttime pedestrian accidents at specially illuminated crosswalks. *Accident Analysis & Prevention*, 10(3), 223–228.
- Quddus, M., Wang, C., & Ison, S. (2009). Road traffic congestion and crash severity: econometric analysis using ordered response models. *Journal of Transportation Engineering*, 136(5), 424–435.
- Ragland, D. R., Markowitz, F., & MacLeod, K. E. (2003). *An intensive pedestrian safety engineering study using computerized crash analysis*. Berkeley, CA: University of California. Berkeley: Safe Transportation Research & Education Center.

- Retting, R. A., Chapline, J. F., & Williams, A. F. (2002). Changes in crash risk following re-timing of traffic signal change intervals. *Accident Analysis & Prevention*, 34(2), 215–220.
- Retting, R. A., Ferguson, S. A., & McCartt, A. T. (2003). A review of evidence-based traffic engineering measures designed to reduce pedestrian-motor vehicle crashes. *American Journal of Public Health*, 93(9), 1456–1463.
- Rifaat, S. M., Tay, R., & de-Barros, A. (2011). Effect of street pattern on the severity of crashes involving vulnerable road users. *Accident Analysis & Prevention*, 43(1), 276–283.
- Roudsari, B. S., Mock, C. N., Kaufman, R., Grossman, D., Henary, B. Y., & Crandall, J. (2004). Pedestrian crashes: Higher injury severity and mortality rate for light truck vehicles compared with passenger vehicles. *Injury Prevention*, 10(3), 154–158.
- Srinivasan, K. K. (2002). Injury severity analysis with variable and correlated thresholds: ordered mixed logit formulation. *Transportation Research Record: Journal of the Transportation Research Board*, 1784, 132–142.
- Sullivan, J. M., & Flannagan, M. J. (1999). *Assessing the potential benefit of adaptive headlighting using crash databases*. Ann Arbor, MI: University of Michigan, Transportation Research Institute.
- Sullivan, J. M., & Flannagan, M. J. (2007). Determining the potential safety benefit of improved lighting in three pedestrian crash scenarios. *Accident Analysis & Prevention*, 39(3), 638–647.
- Sze, N. N., & Wong, S. C. (2007). Diagnostic analysis of the logistic model for pedestrian injury severity in traffic crashes. *Accident Analysis & Prevention*, 39(6), 1267–1278.
- Tay, R., Choi, J., Kattan, L., & Khan, A. (2011). A multinomial logit model of pedestrian-vehicle crash severity. *International Journal of Sustainable Transportation*, 5(4) 233–249.
- Tefft, B. C. (2013). Impact speed and a pedestrian's risk of severe injury or death. *Accident Analysis & Prevention*, 50, 871–878.
- Terza, J. V. (1985). Ordinal probit: A generalization. *Communications in Statistics: Theory and Methods*, 14(1), 1–11.
- Ukkusuri, S. V., Miranda-Moreno, L. F., Ramadurai, G., & Isa-Tavarez, J. (2012). The role of built environment on pedestrian crash frequency. *Safety Science*, 50(4), 1141–1151.
- Wang, X., & Abdel-Aty, M. (2008). Analysis of left-turn crash injury severity by conflicting pattern using partial proportional odds models. *Accident Analysis & Prevention*, 40(5), 1674–1682.
- Windmeijer, F. A. G. (1995). Goodness-of-fit measures in binary choice models. *Econometric Reviews*, 14(1), 101–116.
- Wood, J. M. (2005). Limitations in drivers' ability to recognise pedestrians at night. *Human Factors*, 47(3), 644–653.
- Xie, Y., Zhao, K., & Huynh, N. (2012). Analysis of driver injury severity in rural single-vehicle crashes. *Accident Analysis & Prevention*, 47, 36–44.
- Xiong, Y., & Mannering, F. L. (2013). The heterogeneous effects of guardian supervision on adolescent driver-injury severities: A finite-mixture random-parameters approach. *Transportation Research Part B: Methodological*, 49, 39–54.
- Yasmin, S., Anowar, S., & Tay, R. (2012). Effects of drivers' actions on severity of emergency vehicle collisions. *Transportation Research Record: Journal of the Transportation Research Board*, 2318, 90–97.
- Yasmin, S., & Eluru, N. (2013). Evaluating alternate discrete outcome frameworks for modeling crash injury severity. *Accident Analysis & Prevention*, 59(1), 506–521.
- Zahabi, S. A. H., Strauss, J., Manaugh, K., & Miranda-Moreno, L. F. (2011). Estimating potential effect of speed limits, built environment, and other factors on severity of pedestrian and cyclist injuries in crashes. *Transportation Research Record: Journal of the Transportation Research Board*, 2247, 81–90.
- Zajac, S. S., & Ivan, J. N. (2003). Factors influencing injury severity of motor vehicle-crossing pedestrian crashes in rural Connecticut. *Accident Analysis & Prevention*, 35(3), 369–379.