Title
Influential Factors on Level of Injury in Pedestrian Crashes: Applications of Ordered Probit Model with Robust Standard Errors

Authors
Jang, Kitae
Park, Shin Hyoung
Chung, Sungbong
et al.

Publication Date
2010

Peer reviewed

SafeTREC Safe Transportation Research & Education Center

Influential Factors on Level of Injury in Pedestrian Crashes: Applications of Ordered Probit Model with Robust Standard Errors

Kitae Jang, SafeTREC,
Shin Hyoung Park, Seoul National University,
Sungbong Chung and Ki Han Song, Korea Transport Institute

2010
RR-2010-4
http://www.safetrec.berkeley.edu

ABSTRACT

Pedestrian-involved crashes that occurred in the city of San Francisco over a six-year period, 2002–2007, were analyzed to examine various influential factors on the injury severity of pedestrian crashes. The crash data extracted from the Statewide Integrated Traffic Records System (SWITRS) include five categorical levels of injury severity in traffic crashes also in addition to detailed information about the features of each crash. This study applied an ordered probit model for injury severity analysis to specify the ordinal nature of injury categories. To draw unbiased implications from the estimated parameters, statistical tests were performed on the parameters based on robust standard errors. Then, the marginal effects of each variable on the likelihood of each injury level were computed. The variables that significantly increased the probability of severe injury and fatality were: i) age (under age 15 and over age 65), alcohol consumption and cell phone use among pedestrian characteristics; ii) nighttime, weekends and rainy weather among environmental characteristics; and iii) influence of alcohol, larger vehicles (pickups, buses and trucks) and vehicle proceeding straight when striking a pedestrian among crash characteristics. Crash characteristics were found to influence significantly on the level of pedestrian injury. Based on the findings of this analysis, policy implications and countermeasures are also discussed.

1. INTRODUCTION

Walking is the most basic and common form of transportation mode associated with daily life, and it offers many health benefits—provided that injuries caused by traffic crashes are avoided. In 2007 in the United States, 4,652 pedestrians were killed and approximately 70,000 were injured in traffic crashes; accounting for 11% of total traffic fatalities and 3% of total traffic injuries (*1*). Though continuously diminishing in number over the last decade, pedestrian crashes remain a serious public health problem. Because the human body is directly exposed to the collision force, pedestrians are more vulnerable in traffic crashes than people using other transportation modes. Pucher and Dijkstra (*2*) reported that pedestrians were 23 times more likely to be killed than vehicle users when fatality rates for these two transportation modes were compared.

This high risk of pedestrian injury and fatality in the U.S. has garnered increased attention in recent years and extensive research efforts have been devoted to enhance the level of pedestrian safety via various approaches. However, some important questions regarding pedestrian safety still remain unanswered. Since pedestrians are likely to be severely injured when exposed to traffic crashes, research focusing on various types of—yet unexamined—risk factors for injury severity of pedestrian crashes is essential. Therefore, understanding the relationship between these risk factors and injury severity will provide background for developing safety countermeasures against pedestrian crashes and lay the groundwork for planning a walkable environment. In the present study, ordered probit models were specified to evaluate various risk factors for pedestrian injuries, using pedestrian crashes that occurred in the city of San Francisco from 2002 to 2007.

The objectives of this research are: i) to investigate the relationship between the level of injury in a pedestrian crash and the various characteristics associated with that crash; ii) to quantify the effects of these characteristics on the level of pedestrian injury; and iii) to provide policy and planning implications to improve traffic safety for pedestrians. The remainder of this paper is organized as follows: Section 2 reviews relevant previous research, Section 3 describes the data used in this study, Section 4 summarizes methodology for specification and estimation of the ordered probit model, Section 5 applies an ordered probit model to the data collected in Section 2 and reports the estimation results, and Section 6 discusses implications based on the estimated model.

2. LITERATURE REVIEW

In recent decades, one of the mainstreams in pedestrian safety research has been regression analysis to model the relationship between pedestrian crash count and a variety of explanatory variables (e.g., geometric features of the site, traffic volume and other environmental features). The Poisson, Negative Binomial models have been widely used for this purpose (e.g., *3*, *4*, and *5*). In the meantime, another approach has been to model the severity of pedestrian injury in the occurrence of a traffic crash. Since the objective of the present research is to investigate the effects of various risk factors on severity of pedestrian injury in traffic crashes, this section focuses on the literature examining these risk factors.

Roudsari et al. (*6*) and Sze and Wong (*7*) conducted multivariate binary logistic regression analysis to evaluate the injury risk of pedestrian casualties in traffic crashes in relation to contributory factors to severe injuries and fatalities. This research reported that light truck vehicles (LTVs) were associated with a two times higher risk of pedestrian fatalities and a three times higher risk of pedestrian severe injuries, compared with pedestrian crashes involving passenger vehicles. In addition to vehicle type, two other factors were included in the regression model: pedestrian age and impact speed. However, due to aggregation, binary measurement of injury severity cannot properly reflect variations in injury severity. Additionally, the model used by Roudsari et al. (*6*) did not include other factors which may have had a potential influence on the injury severity and possibly induced confounding in the outcomes of the model.

Davis (*8*) used both logistic and ordered probit models to relate the injury severity of a pedestrian to the impact speed of the vehicle for three different age groups; children (ages 0 to 14), adults (ages 15 to 59) and elderly pedestrians (ages 60+). The results indicate that elderly pedestrians were more likely to experience severe injury than other age groups, when exposed to the same level of impact speed. Again, the model in this study only considered two variables, age and impact speed, while the effects of other characteristics, which are likely to influence injury severity, were overlooked.

Eluru et al. (*9*) developed the mixed generalized ordered response logit model and applied it to the level of injuries among pedestrians and bicyclists involved in traffic crashes. This study reported that the most important variables influencing non-motorist injury severity were age of non-motorist, roadway speed limit, crash location, and time of crash. Since the study attempted to estimate risk factors for two transportation modes (with different characteristics) together, only characteristics common to both modes were included in the analysis.

Zajac and Ivan (*10*) and Lee and Abdel-Aty (*11*) estimated the ordered probit model to investigate the impact of various features on level of injury in pedestrian crashes in rural two-lane highways and at intersections, respectively. Both studies identified some common features significantly influencing pedestrian injury including type of vehicle, driver and pedestrian alcohol involvement and pedestrian older than 65. The models of these studies are limited to specific roadway conditions.

Siddiqui et al. (*12*) specified the ordered probit model to assess the impacts of crossing locations and lighting conditions on pedestrian injury severity, while controlling for other factors that may also impact pedestrian injury severity, including pedestrian attributes (age, race, disability and alcohol involvement), driver attributes (age, race, disability and alcohol involvement), and environmental attributes (roadway geometry features, speed limit, location and year). Though this study estimated the effect of various features on pedestrian injury severity, there was a lack of consideration of the characteristics regarding the crash itself.

Despite these extensive research efforts to determine the impacts of risk factors on the severity of pedestrian injury, there still remain undiscovered risk factors (e.g. crash characteristics). To enhance the understanding of those undiscovered risk factors, the present study categorizes pedestrian injury into five levels to accurately reflect the variations in pedestrian injury severity and the large number of potential risk factors (including those addressed in previous research).

3. DATA DESCRIPTION

In California, the California Highway Patrol (CHP) enters data from CHP-generated reports, as well as those from local law enforcement agencies, into the Statewide Integrated Traffic Records System (SWITRS). Each year data from approximately 4,000 fatal and 190,000 non-fatal injury crashes are added to the system. In addition, data from more than 200,000 Property Damage Only (PDO) crashes are added (*13*).

In this research, to investigate the influence of risk factors on injury severity in pedestrian crashes, data on all levels of pedestrian injury (including PDO) crashes on public roadways in the city of San Francisco from 2002 to 2007 were obtained from SWITRS. Each record contains detailed information on when and where the crash occurred, the road and weather conditions, how many people were killed or injured, and whether the crash involved pedestrians, bicycles, motorcycles or trucks. Other useful information such as crash type and primary crash factors can also be obtained from the record. As shown in Table 1, a total of 5,084 pedestrian crashes including PDO were recorded in San Francisco over the six-year period (2002–2007). The dependent variables are the five levels of injury associated with pedestrian crashes: PDO, slight injury (complaint of pain), visible injury (other visible), severe injury (extended hospitalization), and fatal (see Table 1 (a)). Slight and visible injuries comprised over 85% of total crashes: slight injuries 53.42%, and visible injuries 32.83%. PDO and fatal crashes comprised only 1.91% and 2.85%, respectively. Additional 25 explanatory variables were classified into four categories describing the characteristics of pedestrian, driver, environment and crash for each recorded crash, as summarized in Table 1 (b), (c), (d) and (e), respectively. The reference case was shown in *italics* in Table 1.

Table 1 (a) Dependent Variables in the Models

Variables	Description	Number	Percentage
Level of Pedestrian Injury	Property Damage Only (PDO)	97	1.91%
	Slight Injury (complaint of pain)	2716	53.42%
	Visible Injury (other visible)	1669	32.83%
	Severe Injury (extended hospitalization)	457	8.99%
	Fatal	145	2.85%

Table 1 (b) Pedestrian Characteristics in the Models

Variables	Description	Number	Percentage
PFAULT	*Pedestrian at Fault*	1,652	32.49%
	Otherwise	3,432	67.51%
PSEX	*Female*	2,374	46.70%
	Male	2,653	52.18%
	Unknown	57	1.12%
PAGE	*Younger Than 15*	407	8.01%
	Older Than 65	653	12.84%
	Between Ages 15 and 65	3,831	75.35%
	Unknown	193	3.80%
PUI	*Pedestrian Alcohol Use*	186	3.66%
	Otherwise	4,898	96.34%
PCELL	*Pedestrian Cell Phone Use*	31	0.61%
	Otherwise	5,053	99.39%
PRACE	*Asian*	1,146	22.54%
	African American	755	14.85%
	White	1,942	38.20%
	Hispanic	702	13.81%
	Others	539	10.60%

Table 1 (c) Driver Characteristics in the Models

Variables	Description	Number	Percentage
DFAULT	*Driver at Fault*	3,113	61.23%
	Otherwise	1,971	38.77%
DSEX	*Female*	1,371	26.97%
	Male	3,369	66.27%
	Unknown	344	6.77%
DAGE	*Younger Than 15*	7	0.14%
	Older Than 65	365	7.18%
	Between Ages 15 and 65	3,979	78.27%
	Unknown	733	14.42%
DUI	*Pedestrian Alcohol Use*	81	1.59%
	Otherwise	5,003	98.41%
DCELL	*Pedestrian Cell Phone Use*	24	0.47%
	Otherwise	5,060	99.53%
DRACE	*Asian*	898	17.66%
	African American	621	12.21%
	White	2,098	41.27%
	Hispanic	580	11.41%
	Others	887	17.45%

Table 1 (d) Environmental Characteristics in the Models			
Variables	**Description**	**Number**	**Percentage**
YEAR	*2002*	951	18.71%
	2003	892	17.55%
	2004	784	15.42%
	2005	806	15.85%
	2006	781	15.36%
	2007	870	17.11%
TIME	*Midnight to 6:00 AM*	380	7.47%
	6:00 AM to Noon	1,308	25.73%
	Noon to 6:00 PM	1,975	38.85%
	6:00 PM to Midnight	1,421	27.95%
WEEKEND	*Weekdays*	3918	77.06%
	Weekends	1166	22.94%
INTERSECT	Intersection Crash	1,548	30.45%
	Otherwise	3,536	69.55%
WEATHER	*CLEAR*	3,978	78.25%
	Raining	591	11.62%
	Others	515	10.13%
CROSSWALK	Crash While Pedestrian Crossing a Crosswalk	2,891	56.86%
	Otherwise	2,193	43.14%
NCROSSWALK	Crash While Pedestrian Crossing a Non-Crosswalk	1,184	23.29%
	Otherwise	3,900	76.71%
LIGHTING	*Daylight*	3,215	63.24%
	Dusk-Dawn	190	3.74%
	Dark-Light	1,587	31.22%
	Dark-No Light	65	1.28%
	Unknown	27	0.53%

Table 1 (e) Crash Characteristics in the Models			
Variables	**Description**	**Number**	**Percentage**
Primary Crash Factor	*Influence of Alcohol*	59	1.16%
	Unsafe Speed	286	5.63%
	Improper Passing	79	1.55%
	Improper Turning	56	1.10%
	Automobile Right-of-Way	68	1.34%
	Pedestrian Right-of-Way	1,870	36.78%
	Pedestrian Violation	1,657	32.59%
	Traffic Signals and Signs	236	4.64%
	Other Hazardous Violation	91	1.79%
	Unsafe Starting or Backing	241	4.74%
	Others	145	2.85%
	Unknown	296	5.82%
HITRUN	*Hit and Run Crash*	694	13.65%
	Otherwise	4,390	86.35%
DMOVE	*Proceeding Straight*	1,741	34.24%
	Making Right Turn	379	7.45%
	Making Left Turn	731	14.38%
	Backing	186	3.66%
	Others	366	7.20%
	N/A	1,681	33.07%
DVEHTYPE	*Passenger Car*	3,173	62.41%
	Motorcycle/Scooter	94	1.85%

(continued on next page)

Table 1 (e) Continued			
DVEHTYPE	Pickup	421	8.28%
	Truck	66	1.30%
	Bus	192	3.78%
	Bicycle	108	2.12%
	Others	1,030	20.26%
PARTIES (Other Than Pedestrian)	*0*	4	0.08%
	1	4,717	92.78%
	2	279	5.49%
	3	49	0.96%
	4	23	0.45%
	5	8	0.16%
	6	2	0.04%
	7	2	0.04%

4. MODEL SPECIFICATION

Pedestrian injuries in traffic crashes are categorized into discrete and ordinal levels according to injury severity, given a latent and continuous injury descriptor underlying the categories. Though the level of pedestrian injury is categorical, multinomial logit and probit models do not account for the ordinal nature[1] inherent to the level of injury and therefore, these models are not appropriate in evaluating pedestrian injuries. Since the ordered probit model was developed to estimate the latent descriptor for categories with an ordinal nature, the ordered model was adapted for the specification of the level of injury in the present research. The ordered logit model is also suitable for analyzing the level of injury. The difference between ordered logit and probit is the assumption for the distribution of error, ε_p : the ordered logit model uses logistic distribution while the ordered probit uses standard normal distribution. In previous research, both models resulted in comparable outcomes (e.g., *14*, *15*, and *16*).

4.1 Model Specification (Ordered Probit Model)

The ordered probit model is specified as follows:

$$I_p^* = \beta' X_p + \varepsilon_p$$

Where, I_p^* = a latent and continuous variable measuring injury severity of p^{th} pedestrian; β is a vector of unknown parameters to be estimated; X_p is a vector of observed variables describing the pedestrian, driver, environment and crash involved with p^{th} pedestrian; and ε_p is a random error term, which is assumed to be normally distributed with zero mean and unit variance (i.e., a standard normal distribution).

I_p^* cannot be directly observed in any given pedestrian crash but only a discrete level of injury severity, I_p, is observed and determined from the model in a form of censoring:

[1] Ordinal nature indicates that the discrete (categorical) dependent variable is ranked in a certain order and the differences between ranks are not necessarily equivalent. The levels of injury, the dependent variable of the present research, have ordered ranks (PDO, slight injury, visible injury, severe injury and fatal). Also, the differences between any pairs of two consecutive injury ranks can signify unequal magnitude.

$$I_p = \begin{cases} 1 & \text{if} \quad -\infty < I_p^* \le \psi_1 & \text{Property Damage Only (PDO)} \\ 2 & \text{if} \quad \psi_1 < I_p^* \le \psi_2 & \text{Slight injury (complaint of pain) (Injury}_1\text{)} \\ 3 & \text{if} \quad \psi_2 < I_p^* \le \psi_3 & \text{Visible injury (other visible) (Injury}_2\text{)} \\ 4 & \text{if} \quad \psi_3 < I_p^* \le \psi_4 & \text{Severe injury (extended hospitalization) (Injury}_3\text{)} \\ 5 & \text{if} \quad \psi_4 < I_p^* \le \infty & \text{Fatal} \end{cases}$$

Where, thresholds ψ_i's are unknown parameters to be estimated along with β.

$$\psi_{i-1} < I_p^* \le \psi_i \Leftrightarrow \psi_{i-1} < \beta' X_p + \varepsilon_p \le \psi_i \Leftrightarrow \psi_{i-1} - \beta' X_p < \varepsilon_p \le \psi_i - \beta' X_p$$

Since ε_p is assumed to follow a standard normal distribution,

$$\Pr(I_p = i) = \Phi\left(I_p^* < \psi_i - \beta' X_p\right) - \Phi\left(I_p^* \le \psi_{i-1} - \beta' X_p\right)$$

Where, $\Pr(I_p = i)$ is the probability that p^{th} pedestrian experiences i level of injury (i=1, 2, …, 5); $\Phi\left(I_p^* < \psi_0 - \beta' X_p\right) = 0$; $\Phi\left(I_p^* \le \psi_5 - \beta' X_p\right) = 1$; and $\Phi(\ )$ is the standard normal cumulative distribution function.

The maximum likelihood estimation (MLE) was used to obtain estimators of parameters in the model: $\psi_1, \psi_2, \psi_3, \psi_4, \beta_0, \beta_1, \ldots, \beta_n$. Then, the likelihood function, L, can be formulated as:

$$L = L\left(I_p \mid \psi_1, \psi_2, \psi_3, \psi_4, \beta_0, \beta_1, \ldots, \beta_n\right) = \prod_{j=1}^{p} \prod_{i=1}^{5} \left\{\Phi\left(\psi_i - \beta' X_j\right) - \Phi\left(\psi_{i-1} - \beta' X_j\right)\right\}^{I_{p,n+5}}$$

$$ln\, L = \sum_{j=1}^{p} \sum_{i=1}^{5} I_{p,n+5} \cdot log\left(\Phi\left(\psi_i - \beta' X_j\right) - \Phi\left(\psi_{i-1} - \beta' X_j\right)\right)$$

Since this log-likelihood function, ln L, is a function of $\psi_1, \psi_2, \psi_3, \psi_4, \beta_0, \beta_1, \ldots, \beta_n$ and can be maximized subject to $\psi_1 \le \psi_2 \le \psi_3 \le \psi_4$, then it is in turn, a convex maximization problem with a single constraint, which can be solved by taking the (first and second) partial derivatives of *ln L* for all the parameters and setting them equal to zero or by using commercially available statistical software (e.g., Limdep and STATA). Since the full derivation of maximization procedure is outside of our research scope, detailed derivation is not described in this paper. For the full derivation, please see McKelvey and Zavoina (*17*).

4.2 Robust Standard Errors

Like other regression models, ordered probit models also estimate standard errors to provide information about the precision of parameter estimates and to draw inferences about the covariates' marginal effects. The standard errors can be estimated by taking the square roots of the diagonal elements in the inverse of the so-called "Fisher information matrix" as shown in the following equation:

$$\hat{V} = \left[-E_\theta\left[\frac{\partial^2}{\partial \theta^2} ln\, L\right]\right]^{-1} \text{ where, } \theta \text{ is a parameter vector, } \left\{\psi_1, \psi_2, \psi_3, \psi_4, \beta_0, \beta_1, \ldots, \beta_n\right\}$$

Given that some conditions—a model is properly specified and the error term generally satisfies the independently and identically distributed (IID) condition—are met, consistent estimates of standard errors of the estimated parameters can be obtained. However, these conditions are often violated, thus the obtained standard errors are invalid, making it difficult to draw conclusive inferences. For reliable inferences, thus, robust estimators of the variance-covariance matrix (so-called "Huber Sandwich estimator") were developed (Huber (*18*) and White (*19*)):

$$\hat{V}_R = \hat{V}\left\{\left(\frac{\partial}{\partial\theta} ln\, L(\hat{\theta})\right)^T \left(\frac{\partial}{\partial\theta} ln\, L(\hat{\theta})\right)\right\}\hat{V}$$ where, superscript T indicates transposition

The square roots of the diagonal elements of $\hat{V}_R$ are robust standard errors. Intuitively, $\hat{V}_R$ weighs the contribution of each observation to variance-covariance estimate by the amount of that observation's actual residual variability such that the variance estimate can be empirically corrected. It is also known as the sandwich estimator since the correction segment is sandwiched by $\hat{V}$.[2] Robust standard errors consistently estimate the true standard errors and provide a basis for valid inferences about the parameter estimates, even when the conditions are not satisfied. However, if the model is nearly correct, the usual standard errors are likely to be valid and equivalent to the robust standard errors (*20*). Therefore, the differences between usual and robust standard errors can be used to validate the parameter estimates.

4.3 Measures of Fit

Likelihood Ratio Index

The likelihood ratio index measures goodness of fit of the estimated model based on the log-likelihood value at the convergence. Likelihood ratio index is defined as:

$$\rho^2 = 1 - \left(\frac{\ln L(I_p \mid \theta_A)}{\ln L(I_p \mid \theta = 0)}\right)$$

Where, $\ln L(I_p \mid \theta = 0)$ is the log-likelihood computed with only a constant term (i.e., the vector of coefficients was set to zero) and $\ln L(I_p \mid \theta_A)$ is the log-likelihood value at convergence. The value of ρ^2 has a value between one and zero. The measure equals zero when all the coefficients are zero. As the estimated model improves its goodness of fit, the value of ρ^2 increases and becomes close to one (although it cannot be equal to one).

Likelihood Ratio Test

Complementing the likelihood ratio index, likelihood ratio tests were performed to determine the most appropriate model because the outcome of likelihood ratio test indicates whether the selected model explains the pedestrian injury significantly better than another model does. In the present research, likelihood ratio tests were conducted to test whether the addition of characteristics involved in pedestrian crashes on the base model significantly improved the overall model performance.

Under the null and alternative hypothesis:

[2] Options estimating robust standard errors are available in many statistical packages (e.g. sandwich package in R and robust option in STATA).

H_0: $\theta = \hat{\theta}_0$

H_A: $\theta = \hat{\theta}_A$

Where, $\hat{\theta}_0$ is the vector of estimated parameters in the null model, and $\hat{\theta}_A$ is the vector of estimated parameters in the alternative model, the likelihood ratio test statistic is:

$$\Lambda = -2 \cdot \ln\left(\frac{L(I_p \mid \theta_0)}{L(I_p \mid \theta_A)}\right) = -2 \cdot \left[\ln\{L(I_p \mid \theta_0)\} - \ln\{L(I_p \mid \theta_A)\}\right]$$

Where, $L(I_p \mid \theta_0)$ is the likelihood of the null model and $L(I_p \mid \theta_A)$ is the likelihood of the alternative model. The test rejects the null hypothesis—that the alternative model performs better than the null model—if $\Lambda > \chi^2_{\alpha,\beta}$, where α is degree of freedom and β is significance level. The degree of freedom is the number of additional parameters in the alternative model with respect to the null model, while the significance level is set at 5%.

4.4 Marginal Effects

In the ordered probit model, the parameters are not directly interpreted in terms of the marginal effects of $x_{p,n}$ on the probabilities. Since the level of injury in the pedestrian crash has five categories, the model has four unknown threshold parameters, $\psi_1, \psi_2, \psi_3, \psi_4$. As specified, the probabilities are:

$$\Pr(I_p = 1) = \Phi\left(I_p^* < \psi_1 - \beta' X_p\right) - \Phi\left(I_p^* < \psi_0 - \beta' X_p\right) = \Phi\left(\psi_1 - \beta' X_p\right)$$

$$\Pr(I_p = 2) = \Phi\left(I_p^* < \psi_2 - \beta' X_p\right) - \Phi\left(I_p^* < \psi_1 - \beta' X_p\right) = \Phi\left(\psi_2 - \beta' X_p\right) - \Phi\left(\psi_1 - \beta' X_p\right)$$

$$\Pr(I_p = 3) = \Phi\left(I_p^* < \psi_3 - \beta' X_p\right) - \Phi\left(I_p^* < \psi_2 - \beta' X_p\right) = \Phi\left(\psi_3 - \beta' X_p\right) - \Phi\left(\psi_2 - \beta' X_p\right)$$

$$\Pr(I_p = 4) = \Phi\left(I_p^* < \psi_4 - \beta' X_p\right) - \Phi\left(I_p^* < \psi_3 - \beta' X_p\right) = \Phi\left(\psi_4 - \beta' X_p\right) - \Phi\left(\psi_3 - \beta' X_p\right)$$

$$\Pr(I_p = 5) = \Phi\left(I_p^* < \psi_5 - \beta' X_p\right) - \Phi\left(I_p^* < \psi_4 - \beta' X_p\right) = 1 - \Phi\left(\psi_4 - \beta' X_p\right)$$

Since all the variables in the model are binary (dummy) variables, the effect of a variable is analyzed by comparing the probabilities when the variable takes one value with those when the variable takes zero value while all other variables remain constant. Therefore, the marginal effect of a variable, $x_{n,p}$, on each ordinal categories can be computed as follows:

$$\Delta(I_p = 1 \mid x_{n,p}) = \Pr(I_p = 1 \mid x_{n,p} = 1) - \Pr(I_p = 1 \mid x_{n,p} = 0)$$

$$\Delta(I_p = 2 \mid x_{n,p}) = \Pr(I_p = 2 \mid x_{n,p} = 1) - \Pr(I_p = 2 \mid x_{n,p} = 0)$$

$$\Delta(I_p = 3 \mid x_{n,p}) = \Pr(I_p = 3 \mid x_{n,p} = 1) - \Pr(I_p = 3 \mid x_{n,p} = 0)$$

$$\Delta(I_p = 4 \mid x_{n,p}) = \Pr(I_p = 4 \mid x_{n,p} = 1) - \Pr(I_p = 4 \mid x_{n,p} = 0)$$

$$\Delta(I_p = 5 \mid x_{n,p}) = \Pr(I_p = 5 \mid x_{n,p} = 1) - \Pr(I_p = 5 \mid x_{n,p} = 0)$$

While holding all others constant, one unit change in variable, $x_{p,n}$, shifts the distribution toward the direction of the sign β. The increase in variable, $x_{p,n}$, associated with the parameter, β, with positive sign shifts the distribution toward the right. Thus, this shift results in an increase in the probability of the rightmost category (i.e., $I_p = 5$, Fatal) and diminishing the probability of the leftmost category (i.e., $I_p = 1$, PDO). Meanwhile, the negative signs are conversely interpreted. However, the marginal effects for the categories in between depend on the shifted amount of densities.

5. ESTIMATION RESULT

5.1 Model Selection

Since the pedestrian is the subject directly exposed to the crashes—as well as of our primary interest—the ordered probit model with every combination of characteristics including pedestrian characteristics was estimated and the log likelihood value for each model was computed. With these log likelihood values, a log likelihood ratio index was calculated to measure the model's goodness of fit, and a log likelihood ratio test was performed for each pair of models to test whether the difference was statistically significant. Table 2 summarizes the log likelihood values and indices for models with different input characteristics, and Table 3 presents the outcomes of log likelihood ratio tests at a 5% significance level conducted for the notable differences in log likelihood values.

Following the inclusion of additional variables, the likelihood ratio index indicated the continuous improvement in goodness of fit. Compared with the model of pedestrian characteristics only (model 1), models with two characteristics (models 2, 3 and 4) show higher values of likelihood ratio index. Among those, however, only the difference between models 1 and 4 (test 4 in Table 3) was statistically significant. In other words, including only crash characteristics significantly improved the model. This comparison indicates that crash characteristics have a more significant influence on the level of pedestrian injury severity than driver and environmental characteristics.

Similarly, models with three characteristics (models 5, 6 and 7) also return higher values of likelihood ratio index than those with two characteristics (models 2, 3 and 4). Again, the statistical tests were conducted between models to examine how the model improved after adding either driver or environmental characteristics. First, model 5 and models 2 and 3 were tested and both tests (tests 4 and 5 in Table 3) were statistically significant. Then, statistical tests were performed for the differences between model 4 and models 6 and 7, and the test results indicated that the differences were all statistically significant. Though the model improved (with statistical significance) by adding one more characteristic to the models with two characteristics, models showed greater improvement following the addition of environmental characteristics rather than driver characteristics.

The model with all four characteristics (model 8) was finally selected because it outperformed all the other models in terms of likelihood index ratio and test. The estimation results of model 8 are summarized in Table 4.

Table 2 Log Likelihood of Alternative Models

Classification	Selected Characteristics				Number of Independent Variables	Log Likelihood	Likelihood Ratio Index (ρ^2)
	Pedestrian	Driver	Environmental	Crash			
Model 0	X	X	X	X	0	-5562.63	-
Model 1	O	X	X	X	6	-5415.68	0.0264
Model 2	O	O	X	X	12	-5414.50	0.0266
Model 3	O	X	O	X	15	-5413.46	0.0268
Model 4	O	X	X	O	11	-5357.30	0.0369
Model 5	O	O	O	X	21	-5386.83	0.0316
Model 6	O	O	X	O	17	-5343.36	0.0394
Model 7	O	X	O	O	20	-5327.72	0.0422
Model 8	O	O	O	O	26	-5316.33	0.0443

Table 3 Log Likelihood Ratio Tests between Alternative Models

	Null Model	Alternative Model	Likelihood Ratio Test Statistics (Λ)	Chi-squared ($\chi^2_{\alpha,\beta}$)	Significance
Test 1	Model 1	Model 2	2.36	12.59	Not Significant
Test 2	Model 1	Model 3	4.44	16.92	Not Significant
Test 3	Model 1	Model 4	116.76	11.07	Significant
Test 4	Model 2	Model 5	55.34	16.92	Significant
Test 5	Model 3	Model 5	53.26	12.59	Significant
Test 6	Model 4	Model 6	27.88	12.59	Significant
Test 7	Model 4	Model 7	59.16	16.92	Significant
Test 8	Model 7	Model 8	22.78	12.59	Significant

5.2 Model Estimates

As shown in Table 4, values of robust standard errors were comparable to those of usual standard errors, signifying that the parameters were properly estimated. To draw valid interpretation, robust standard errors were used to calculate the p-value for each estimated parameter. Table 4 summarizes only coefficients for which p-values indicate statistical significance (at the level of 10%). P-values show noticeable results: i) in pedestrian characteristics, the parameters of pedestrian age, pedestrian alcohol involvement and pedestrian cell phone use were statistically significant; ii) most of the parameters in drivers' characteristics were not significant; iii) time of crashes, weekends, and rainy weather were statistically significant, iv) among crash characteristics, almost all of the parameters for primary crash factors, vehicle movement and vehicle types were statistically significant. The results of p-values for the estimated parameters also indicate that the parameters for pedestrian and crash characteristics significantly affected the level of pedestrian injury severity as indicated in log likelihood ratio indices and tests presented in Table 2 and 3.

Table 4 Ordered Probit Estimates for Pedestrian Injuries (Model 8)

Variable Categories		Variables	Coef.	Std. Err.	Robust Std. Err.	p-value
Pedestrian Characteristics	PAGE	Older Than 65	0.203^{***}	0.072	0.074	0.006
		Between Ages 15 and 65	-0.193^{***}	0.06	0.06	0.001
		Unknown	0.354^{***}	0.103	0.13	0.006
	PUI	Pedestrian Alcohol Use	0.400^{***}	0.087	0.097	0
	PCELL	Pedestrian Cell Phone Use	0.422^{**}	0.2	0.167	0.011
	PRACE	African American	-0.341^{***}	0.057	0.056	0
Environment Characteristics	TIME	6:00 AM to Noon	-0.258^{***}	0.087	0.094	0.006
		Noon to 6:00 PM	-0.270^{***}	0.083	0.091	0.003
		6:00 PM to Midnight	-0.254^{***}	0.068	0.076	0.001
	WEEKEND	Saturday and Sunday	0.083^{**}	0.039	0.039	0.035
	RAINING	Raining	0.173^{**}	0.071	0.071	0.014
	NCROSSWALK	Crash While Pedestrian Crossing a Non-Crosswalk	0.103^{*}	0.056	0.057	0.072
	LIGHTING	Unknown	0.374^{*}	0.218	0.194	0.054
Crash Characteristics	PCF	Unsafe Speed	-0.739^{***}	0.275	0.284	0.009
		Improper Passing	-0.668^{**}	0.298	0.309	0.031
		Improper Turning	-0.812^{**}	0.311	0.334	0.015
		Pedestrian Right-of-Way	-0.705^{**}	0.272	0.28	0.012
		Pedestrian Violation	-0.809^{***}	0.281	0.291	0.006

(continued on next page)

Table 4 Continued						
Crash Characteristics	PCF	Traffic Signals and Signs	-0.552*	0.278	0.29	0.057
		Other Hazardous Violation	-0.875***	0.296	0.307	0.004
		Unsafe Starting or Backing	-0.784***	0.284	0.288	0.007
		Others	-0.760***	0.285	0.293	0.009
		Unknown	-0.728**	0.281	0.289	0.012
	DMOVE	Making Right Turn	-0.242***	0.062	0.058	0
		Making Left Turn	-0.167***	0.05	0.049	0.001
		Backing	-0.312***	0.099	0.089	0
		Others	-0.161**	0.062	0.067	0.017
	DVEHTYPE	Pickup	0.145**	0.06	0.059	0.014
		Truck	0.444***	0.141	0.169	0.009
		Bus	0.299***	0.085	0.093	0.001
		Bicycle	0.284***	0.11	0.096	0.003
		Others	-0.079*	0.043	0.043	0.069
	PARTIES	4	1.482*	0.609	0.9	0.1
		5	1.866*	0.683	0.96	0.052
		7	1.983**	1.044	0.941	0.035
Severity			Coefficient		Standard Error	
ψ_1 (between PDO and Injury$_1$)			-2.111		0.7281	
Ψ_2 (between Injury$_1$ and Injury$_2$)			0.181		0.7277	
Ψ_3 (between Injury$_2$ and Injury$_3$)			1.301		0.7277	
ψ_4 (between Injury$_3$ and Fatal)			2.108		0.7283	

*: Significance level: 10%, **: Significance level: 5%, ***: Significance level: 1%

5.3 Marginal Effects of the Estimates

In assessing the marginal effects of the estimates, variables with higher statistical significance (identified in Section 5.2) were interpreted because these variables were more likely to have statistically significant effects on pedestrian injury severity.

The marginal effects for each coefficient included in Table 4 were calculated and are shown in Table 5. The marginal effects are the substantive effects of the explanatory variables on the changes in the probability of a certain level of pedestrian injury severity in the occurrence of a pedestrian traffic crash. These are a relative measure to a reference, which, in this case, was the model with all dummy variables set to equal zero (the reference case was presented as *italics* in Table 1).

Pedestrian Characteristics

Compared with young pedestrians (younger than age 15), older pedestrians (older than age 65) tend to experience increased injury levels, while pedestrians between ages 15 and 65 were more likely to experience diminished injury levels. A higher probability of severe injury and fatality in older and younger age groups can be explained by the fact that pedestrians in those age groups are more vulnerable to the impacts, less responsive to the risks, and exhibit slower perception and reaction times. As expected, pedestrian alcohol consumption increased the level of injury severity risk, although the primary crash factor was not directly related to alcohol consumption. Alcohol consumption is associated with diminished physical abilities (e.g., slower reaction time, blurred vision, inaccurate motion tracking and lack of concentration), leading to the increased risk of severe injury in pedestrian crashes. Additionally, pedestrians engaged in cell phone use appeared to have an increased risk of injury severity, possibly due to lack of concentration. Interestingly, there are risk differences across races: African Americans experienced decreased levels of injury severity compared with pedestrians of other races.

Driver Characteristics

Variables in driver characteristics were included in the model to examine their influences on the severity of pedestrian injury under all types of pedestrian-involved traffic crashes. Among driver characteristics, none of the coefficients were statistically significant. In pedestrian crashes, drivers are not directly exposed to the traffic crash, and thus, driver characteristics are less influential on the level of pedestrian injury.

Environmental Characteristics

Compared with crashes that occurred between midnight and 6 a.m., pedestrian crashes that occurred in other time periods appeared to present a diminished risk of severe injury and fatality. Shorter visible range at night, faster vehicle speeds under light traffic conditions, and other factors may contribute to the propensity of severe injury and fatality risk at night. Pedestrian crashes during weekends were associated with increased risk of severe injury and fatality, probably due to the difference in travel patterns between weekends and weekdays (since travel during weekdays is more likely to be work-related and along familiar routes than weekend travel).

As reported in previous research, precipitation is also shown as a factor for higher risk of severe injury and fatality among pedestrians. When and where a pedestrian crash occurred (e.g., intersection, crossing a crosswalk or crossing a non-crosswalk) are not statistically significant or are marginally significant. In previous research (e.g., *5* and *21*), crosswalks appeared to be associated with a higher risk of pedestrian crashes (not the level of injury, but the frequency), which has drawn great attention to crosswalk design. However, since the model in the present research estimated the level of pedestrian injury given that the crash is already occurring, based on the current model, it cannot be determined whether these variables contribute to the level of pedestrian injury.

Crash Characteristics

Among crash characteristics, coefficients of primary crash factor, movement of vehicle and vehicle type are statistically significant. Since the primary crash factor was identified and recorded in the database based on a police officer's direct observation of a crash scene, it delivers information (in the form of categorical data) about the qualitative measures of primary causes associated with the crash. Compared with crashes due to the influence of alcohol, other primary crash factors were associated with lower probability of severe pedestrian injury and fatality. In other words, crashes caused by the influence of alcohol are most likely to result in higher levels of pedestrian injury severity. Pedestrian crashes related to automobile right-of-way and traffic signals and signs were also found to result in more severe pedestrian injuries. This can be explained by unexpected situations experienced while driving or walking. While driving in automobile right-of-ways, drivers do not expect pedestrians on the road and focus more on other vehicles rather than pedestrians. Similarly, when drivers and pedestrians follow traffic signals and signs, they tend to heed the guidance of these signals and signs without considering other circumstances, leading to a higher risk of a traffic crash occurring.

Coefficients in vehicle movement at the time that the crash occurred indicate that proceeding straight was found to result in a higher probability of severe injury and fatality. Pedestrians struck by larger vehicles (e.g., pickups, trucks and buses) were more likely to be severely injured or killed. This may be explained primarily by the heavier weight of larger vehicles. Unexpectedly, bicycle-pedestrian crashes appeared to result in higher level of pedestrian injury. However, this finding should be cautiously interpreted because samples of bicycle-involved crashes might be over-representing the population due to a small sample size and, additionally, distribution of bicycle-involved injury risk might differ from that of other motorized vehicle-involved injuries. Future studies using a larger sample size and focusing on groups of pedestrian-non-motorized vehicle crashes could shed further light on this issue.

Table 5 Marginal Effects

Variable Categories		Variables	PDO	Injury$_1$	Injury$_2$	Injury$_3$	Fatal
Pedestrian Characteristics	PAGE	Older than 65	-0.007	-0.074	0.04	0.029	0.011
		Between Ages 15 and 65	0.007	0.07	-0.039	-0.027	-0.01
		Unknown	-0.01	-0.131	0.062	0.054	0.024
	PUI	Pedestrian Alcohol Use	-0.011	-0.148	0.068	0.062	0.028
	PCELL	Pedestrian Cell Phone Use	-0.011	-0.156	0.069	0.067	0.031
	PRACE	African American	0.017	0.114	-0.078	-0.04	-0.013
Environment Characteristics	TIME	6:00 AM to Noon	0.012	0.089	-0.057	-0.032	-0.011
		Noon to 6:00 PM	0.011	0.095	-0.058	-0.035	-0.012
		6:00 PM to Midnight	0.011	0.088	-0.056	-0.032	-0.011
	WEEKEND	Saturday and Sunday	-0.003	-0.03	0.017	0.011	0.004
	RAINING	Raining	-0.006	-0.063	0.035	0.025	0.01
	NCROSSWALK	Crash While Pedestrian Crossing Non-Crosswalk	-0.004	-0.037	0.021	0.014	0.005
	LIGHTING	Unknown	-0.01	-0.138	0.064	0.058	0.026
Crash Characteristics	PCF	Unsafe Speed	0.058	0.203	-0.173	-0.069	-0.019
		Improper Passing	0.051	0.185	-0.158	-0.062	-0.017
		Improper Turning	0.072	0.204	-0.189	-0.069	-0.018
		Automobile Right-of-Way	0.024	0.127	-0.094	-0.044	-0.013
		Pedestrian Right-of-Way	0.035	0.233	-0.153	-0.086	-0.03
		Pedestrian Violation	0.045	0.258	-0.178	-0.094	-0.032
		Traffic Signals and Signs	0.037	0.166	-0.13	-0.056	-0.016
		Other Hazardous Violation	0.082	0.211	-0.202	-0.072	-0.019
		Unsafe Starting or Backing	0.065	0.208	-0.183	-0.07	-0.019
		Others	0.063	0.201	-0.178	-0.068	-0.018
		Unknown	0.056	0.201	-0.171	-0.068	-0.019
	DMOVE	Making Right Turn	0.012	0.082	-0.055	-0.029	-0.01
		Making Left Turn	0.007	0.058	-0.037	-0.021	-0.007
		Backing	0.016	0.103	-0.072	-0.036	-0.011
		Others	0.007	0.056	-0.035	-0.021	-0.007
	DVEHTYPE	Pickup	-0.005	-0.053	0.029	0.021	0.008
		Truck	-0.011	-0.164	0.071	0.07	0.033
		Bus	-0.009	-0.11	0.055	0.045	0.019
		Bicycle	-0.008	-0.105	0.052	0.043	0.018
		Others	0.003	0.028	-0.017	-0.01	-0.004
	PARTIES	4	-0.016	-0.45	-0.019	0.224	0.261
		5	-0.015	-0.496	-0.113	0.222	0.403
		7	-0.015	-0.505	-0.142	0.212	0.45

Injury$_1$: Injury (slight injury), Injury$_2$: Injury (visible injury), Injury$_3$: Injury (severe injury)

6. CONCLUSION

The present research evaluated the impact of various risk factors on the severity of pedestrian injury in traffic crashes. Pedestrian-involved crashes that occurred in the city of San Francisco from 2002 to 2007 were extracted from the SWITRS database. Variables in the database were categorized into four groups of characteristics—pedestrian, driver, environment and crash—and entered into the model as explanatory variables. Using these variables, an ordered probit model for levels of pedestrian injury severity (dependent variable) was estimated. Statistical tests were performed to select proper sets of characteristics involved in traffic crashes and the significant parameters were identified based on robust (unbiased) standard errors. The parameters and marginal effects of significant variables were interpreted to examine the influence of various characteristics on pedestrian injury severity.

In the course of model selection, it was determined that the model improved most significantly when it included crash characteristics. Meanwhile, variables in the group of driver characteristics appeared to be statistically insignificant and, in turn, improved the model least significantly when included. Pedestrian characteristics that increased pedestrian injury severity included alcohol involvement (even when not a primary crash factor), cell phone use, and age—younger than age 15 and older than age 65. Environmental characteristics including nighttime, weekends and rainy weather were associated with increased probability of severe injury and fatality.

Among crash characteristics, primary crash factors, vehicle movement and type of vehicle were shown to be significant. The primary crash factor resulting in the most severe injuries appeared to be the influence of alcohol. Among vehicle movements at the time of a pedestrian crash, the probability of severe injury and fatality was increased when a pedestrian was hit by a vehicle that was proceeding straight. Compared with passenger vehicles, larger vehicles including pickups, trucks and buses were associated with more severe injury.

The results of this study are useful in understanding which risk factors have a greater impact on severe injury in pedestrian crashes and, thus, effective policy implications and countermeasures of pedestrian injuries can be recommended. For example, since the findings from the estimated model indicate that lack of awareness of the crash situation are likely to increase injury risk, pedestrians should be informed of the increased risk factors associated with walking: i) pedestrians under the age 15 or over the age of 65, ii) walking while using a cell phone, and iii) walking after drinking alcohol. On roadways, countermeasures should be implemented to address nighttime and rainy weather crashes, including improving light conditions and maintaining proper pavement conditions during wet weather. Providing traffic information for drivers traveling during weekends may be also helpful, as they are less likely to drive familiar routes (to work or other daily destinations) compared with weekday travelers. Moreover, countermeasures can be developed based on crash characteristics and further (quantitatively) evaluated using the estimated model—however, the countermeasures may vary across locations in accordance with the most frequently observed primary crash factors: vehicle movement type and vehicle type.

The present research specified the ordered probit model, incorporating a set of variables which had rarely been examined in previous research, including primary crash factors and vehicle movement at the time of a crash. Since the ordered probit model is conditioned on the occurrence of crashes, however, the model may be limited to providing overall crash risk associated with walking. It should be noted that the factors identified as increasing injury severity might be contributing the occurrence of crashes or *vice versa*. Thus, the model estimated in the present research, if combined with count models for occurrence of pedestrian crashes, can further enhance the understanding of risk factors on overall pedestrian traffic safety.

The SWITRS database (similar to other police-reported crash database) may underreport PDO and minor injury crashes (e.g., *22* and *13*) and, in turn, may contain some biased estimates toward severe injury. Though this study incorporated a large number of samples (5,084 pedestrian crashes), it focused only on one jurisdiction and therefore the results might not be directly applicable to other jurisdictions. Thus, other econometric models or statistical techniques (e.g., analysis of censored, truncated and missing data) should be considered to overcome these possible limitations. In addition, crash data from additional jurisdictions should be included to confirm the relationship between risk factors and pedestrian injury severity, taking into account potential differences across jurisdictions. These remain the topics of future research.

ACKNOWLEDGEMENTS

The authors are very grateful to Professor David Ragland for his support and constructive comments on the study. The second author would like to acknowledge the support of Safe and Sustainable Infrastructure Research Group at Seoul National University, as part of the Brain Korea (BK) 21 research Program funded by the Ministry of Education, Science and Technology in South Korea.

REFERENCE

1. National Highway Traffic Safety Administration (2008) Traffic safety facts: 2007 data pedestrian, Washington, D.C.
2. Pucher, J, Dijkstra, L. (2003) Promoting safe walking and cycling to improve public health: lessons from The Netherlands and Germany. *American Journal of Public Health,* 93, pp. 1509–1516
3. Fridstrøm L. and Ingebrigtsen S. (1991) An aggregated accident model based on pooled, regional time-series data, *Accident Analysis and Prevention*, 23, pp. 363–378
4. Shankara, V.N., Ulfarssona, G.F., Pendyalab, R.M. and Nebergallc, M.B., (2003) Modeling crashes involving pedestrians and motorized traffic, *Safety Science*, 41, pp. 627–640
5. Zegeer, C.V., Stewart, J.R., Huang, H. and Lagerwey, P. (2001) Safety effects of marked versus unmarked crosswalks at uncontrolled locations analysis of pedestrian crashes in 30 cities, *Transportation Research Record: Journal of the Transportation Research Board*, 1773, pp. 56–68
6. Roudsari, B. S., Mock C.N., Kaufman R., Grossman D., Henary B.Y. and Crandall, J. (2004) Pedestrian crashes: higher injury severity and mortality rate for light truck vehicles compared with passenger vehicles, *Injury Prevention,* 10, pp. 154–158
7. Sze, N.N. and Wong, S.C. (2007) Diagnostic analysis of the logistic model for pedestrian injury severity in traffic crashes, *Accident Analysis and Prevention*, 39, pp. 1267–1278
8. Davis, G.A. (2001) Relating Severity of Pedestrian Injury to Impact Speed in Vehicle-Pedestrian Crashes, *Transportation Research Record,* 1773, pp. 108–113
9. Eluru, N., Bhat, C.R. and Hensher, D.A. (2008) A mixed generalized ordered response model for examining pedestrian and bicyclist injury severity level in traffic crashes, *Accident Analysis and Prevention,* 40, pp. 1033–1054
10. Zajac, S.S. and Ivan, J.N. (2003) Factors influencing injury severity of motor vehicle–crossing pedestrian crashes in rural Connecticut, *Accident Analysis and Prevention,* 35, pp. 369–379
11. Lee, C. and Abdel-Aty, M. (2005) Comprehensive analysis of vehicle–pedestrian crashes at intersections in Florida, *Accident Analysis and Prevention,* 37, pp. 775–786
12. Siddiqui, N.A., Chu, X. and Guttenplan, M. (2006) Crossing Locations, Light Conditions, and Pedestrian Injury Severity, *Transportation Research Record: Journal of the Transportation Research Board*, 1982, pp 141-149
13. Bigham, J.M., Rice, T.M., Pande, S., Lee, J., Park, S.H., Gutierrez, N.B. and Ragland, D.R. (2009) A comprehensive approach to geocoding police collision report data in California, *Accident Analysis and Prevention (submitted)*
14. Kockelman, K. and Kweon, Y.J. (2002) Driver injury severity: An application of ordered probit models. Accident Analysis and Prevention, 34, pp. 313-321
15. Quddus, M.A., Noland, R.B. and Chin, H.C. (2002) An analysis of motorcycle injury and vehicle damage severity using ordered probit models, *Journal of Safety Research*, 33, pp.445–462
16. Greene, W.H. (2003) Econometric Analysis, Fifth Edition. Prentice Hall, New Jersey.
17. McKelvey, R.D. and Zavoina, W. (1975) A statistical model for the analysis of ordinal level dependent variables, *Journal of Mathematical Sociology*, 4, pp. 103–120
18. Huber, P.J. (1967) The Behavior of Maximum Likelihood Estimates under Nonstandard Conditions, *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, vol. I, pp. 221–233
19. White, H. (1980) A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity, *Econometrica*, 48, pp. 817–838
20. Freedman, D.A. (2006) On the so-called "Huber sandwich estimator" and "robust standard errors", *American Statistician,* 60, pp. 299–302
21. Koepsell, T., McCloskey, L., Wolf, M., Moudon, A.V., Buchner, D., Kraus, J. and Patterson, M. (2002) Crosswalk Markings and the Risk of Pedestrian–Motor Vehicle Collisions in Older Pedestrians, *Journal of the American Medical Association* , 288, pp. 2136–2143

22. Blincoe, L.J., Seay, A.G., Zalonshnja, E., Miller, T.R., Romano, E.O. Luchter, S. Spicer, R.S. (2002) The economic impact of motor vehicle crashes 2000. Washington, DC: National Highway Traffic Safety Administration, US Department of Transportation.