sustainability

MDPI

Article

Injury Severity of Bus–Pedestrian Crashes in South Korea Considering the Effects of Regional and Company Factors

Ho-Chul Park [1], Yang-Jun Joo [2], Seung-Young Kho [2], Dong-Kyu Kim [2,*] and Byung-Jung Park [3]

1 Department of Civil, Environmental and Construction Engineering, University of Central Florida, Orlando, FL 32816-2450, USA; hochulpark@knights.ucf.edu

2 Department of Civil and Environmental Engineering, Seoul National University, Seoul 08826, Korea; yangjun5@snu.ac.kr (Y.-J.J.); sykho@snu.ac.kr (S.-Y.K.)

3 Department of Transportation Engineering, Myongji University, Yongin 17058, Korea; bjpark@mju.ac.kr

* Correspondence: dongkyukim@snu.ac.kr; Tel.: +82-2-880-7348

Received: 23 April 2019; Accepted: 31 May 2019; Published: 5 June 2019

Abstract: Bus–pedestrian crashes typically result in more severe injuries and deaths than any other type of bus crash. Thus, it is important to screen and improve the risk factors that affect bus–pedestrian crashes. However, bus–pedestrian crashes that are affected by a company's and regional characteristics have a cross-classified hierarchical structure, which is difficult to address properly using a single-level model or even a two-level multi-level model. In this study, we used a cross-classified, multi-level model to consider simultaneously the unobserved heterogeneities at these two distinct levels. Using bus–pedestrian crash data in South Korea from 2011 through to 2015, in this study, we investigated the factors related to the injury severity of the crashes, including crash level, regional and company level factors. The results indicate that the company and regional effects are 16.8% and 5.1%, respectively, which justified the use of a multi-level model. We confirm that type I errors may arise when the effects of upper-level groups are ignored. We also identified the factors that are statistically significant, including three regional-level factors, i.e., the elderly ratio, the ratio of the transportation infrastructure budget, and the number of doctors, and 13 crash-level factors. This study provides useful insights concerning bus–pedestrian crashes, and a safety policy is suggested to enhance bus–pedestrian safety.

Keywords: bus–pedestrian crash; injury severity; cross-classified multi-level model (CCMM); heterogeneity; type I statistical error

1. Introduction

Crashes that involve buses are known to result in serious injuries due in part to their physical characteristics, e.g., heavy weight, large size, and maneuvering restrictions, such as a large minimum turning radius [1]. In South Korea, according to the data on vehicle crashes in 2015, the fatalities that occurred in bus crashes represented 2.2% of the total fatalities, which was slightly more than double the number of fatalities that occurred in taxi crashes. Notably, the fatality rate of bus–pedestrian crashes was 5.9%, which was considerably higher than the 1.3% fatality rate of bus–vehicle crashes. Due to the access for passengers to get on and off buses, many conflicts occur between pedestrians and buses, resulting in a large number of casualties. In order to achieve a sustainable transportation system, the bus–pedestrian safety should be improved. Recently, despite the pedestrian-oriented policies that have been implemented to improve pedestrian safety in many countries, pedestrians continue to be exposed to the risk of crashes with buses [2,3]. Therefore, we should identify the factors that result in bus–pedestrian crashes and continue to make efforts to eliminate those factors.

Various factors affect bus–pedestrian crashes. Many studies have considered these factors, such as drivers, vehicles, roadways, and factors related to the environment. In crashes with buses, pedestrians are more vulnerable than they are in crashes with autos. Therefore, the severity of the injuries in bus–pedestrian crashes can depend to some extent on the condition of the bus driver and her/his ability to deal with dangerous situations. Some studies have focused on the characteristics of commercial vehicle companies, including the buses and the drivers' working conditions [4–6]. The results of these studies have indicated that safety-oriented work schedules and safety-related attitudes and behaviors of bus drivers are essential to safe driving. These factors may be affected by the safety management practices of companies. In addition, the characteristics of the region where the crash occurred also can affect bus–pedestrian crashes, and some studies have focused on the regional effects on bus–pedestrian crashes [1,7]. Therefore, bus–pedestrian crashes can be affected by the crash level factors, the characteristics of the company, and the characteristics of the region.

Bus–pedestrian crash data, which include both crash level and group level (region and company) factors, have a cross-classified hierarchical structure. Some research results have indicated that traffic crash data have a hierarchical structure, usually focused on geographical or regional effects, and multi-level modeling should be used to account for group-level factors [1,7–15]. Concerning the severity of the injuries in bus or pedestrian crashes, two-level models were used to account for the correlation of the crashes in a defined region [1,7]. These models set the crash level factors to a lower level (i.e., individual or level 1 factors) and the regional group factors to a higher level (i.e., group or level 2 factors). Unlike the indications of previous studies, bus–pedestrian crashes are affected by the companies that own and operate the buses involved and by the regions in which the crashes occurred. This means that bus–pedestrian crashes have two distinct hierarchical paths, and applying a two-level model to cross-classified hierarchical data may yield a biased estimate of the parameters [12]. Because cross-group heterogeneity is neglected in hierarchical data, a two-level model can underestimate the standard error, thereby resulting in a type I statistical error when testing statistical hypotheses [13]. To analyze these two higher group factors, a cross-classified, multi-level model (CCMM) is required, because it can consider non-nested groups simultaneously.

In this study, we identified the factors that influence the severity of injuries in bus–pedestrian crashes, including both the crash-level factors and the regional and company-group factors.

We used CCMM to take into account the heterogeneity in two distinct group level factors, i.e., the region and the company (cross-classified hierarchical structure). We also compared the results we obtained with the results of a single-level model and two-level multi-level models to show the need for a CCMM.

2. Literature Review

2.1. Research on the Injury Severity in Crashes Involving Buses and Pedestrians

Table 1 shows that recent studies on the injury severity in bus and pedestrian crashes have developed various models to identify the factors that influence such crashes, including temporal factors (season and time of day); individual factors (ages and genders of the drivers and victims); the type of vehicle and its age and mileage; and environmental and roadway characteristics (surface conditions and width of the road). Distinctive factors have also been considered to screen the risk factors for bus and pedestrian crashes. These factors can be classified into pedestrian, bus, regional, and social characteristics. Studies of the characteristics of pedestrians have taken into account the movement of pedestrians, the colors of the clothes, and the errors they make. Studies of the characteristics of buses have considered the type of buses (e.g., school buses and van buses). Regarding the regional characteristics, street pattern, land use, number of schools or shops, and network type have been considered. Also, for the social characteristics, land price, household income, and people's nationalities were considered. In addition, previous studies used various statistical models, such as ordered logit and average direct pseudo-elastics models, and more recent studies have used methods based on machine learning, such as clustering-based regression, classification and regression trees (CARTs), and association rules [16–18].

Table 1. Summary of studies concerning the severity of crashes involving buses and pedestrians.

Author	Type of Crash	Estimation Model	Distinctive Factors	Purpose of the Study
Rifaat et al. [19]	Pedestrian	Multinomial logit (ML)	Street pattern (grid, cul-de-sac, warped parallel, mixed)	Study of the effect of street patterns on the severity of crashes
Moudon et al. [20]	Pedestrian	Binary logit	Average annual daily traffic, home value (0.5 km buffer), land uses (Number of Schools and shops)	Study of the risk factors of crashes on state routes and city streets
Kaplan and Prato [21]	Bus	Generalized ordered logit (GOL)	Type of bus (school bus, van bus), collision type (front, rear, head, end)	Study of the risk factors associated with bus crashes in the in U.S.
Mohamed et al. [22]	Pedestrian (including bus crashes)	Clustering, ordered probit (OP), K-means, ML	Pedestrian's error, light condition, network (road, town, city street)	Clustering pedestrian crashes and investigating each cluster's risk factors
Aziz et al. [23]	Pedestrian (including bus crashes)	Random-parameter multinomial logit (RPML)	Region (Bronx, Brooklyn, Manhattan, Queens, Staten Island)	Analysis of unobserved heterogeneity in risk factors of pedestrian crashes
Tefft [24]	Pedestrian	Logistic regression	Pedestrian's height, weight, and body mass index (BMI)	Investigating impact speed of pedestrian crashes
Prato and Kaplan [25]	Bus	GOL, ADPE	Region (West coast, Canterbury, Southland)	Analyzing the severity of vehicle crashes at highway-rail grade crossings
Islam and Jones [26]	Pedestrian	Mixed logit (MXL), ADPE	Pedestrian not visible (dark clothing), pedestrian darting, sitting in roadway	Investigating risk factors of pedestrian at-fault crashes in urban and rural locations
Osman et al. [27]	Heavy vehicle (including bus crashes)	GOL	Access control (no control), peak, worker present, work zone type (lane closure, shoulder), location (transition area), on-bridge	Analysis of injury severity of large truck crashes in work zones
Pour-Rouholamin and Zhou [28]	Pedestrian (including bus crashes)	Partial proportional odds (PPO)	Urban, region (Chicago), traffic control device (traffic signal and sign, no control)	Analysis of the injury severity in pedestrian crashes in Illinois
Li et al. [29]	Pedestrian	Classification and regression trees (CARTs), random forest	Vehicle maneuver (standing), pedestrian movement (driver's nearside, offside)	Analyzing the injury severity to pedestrians in crashes under different weather conditions
Toran Pour et al. [30]	Pedestrian (including bus crashes)	CARTs, boosted decision tree, bagged decision tree	Social characteristics (white-, blue-, pink-collar worker, income)	Modeling the severity of pedestrian crashes at mid-block

2.2. Multi-Level Models in Traffic Safety

There is growing interest in the use of multi-level models to analyze hierarchically structured traffic crash data, and these models are often referred to as hierarchical models [1,7–15]. Two research studies have summarized the use of multi-level analysis of traffic crash data and suggested further studies [14,15]. Among these, Huang and Abdel-Aty (2010) proposed a five-level hierarchy that represented the spatial distribution, i.e., geographic region, traffic site, traffic crash, driver-vehicle unit, and occupant [14]. In addition, they mentioned that crash observation could be separated into two distinct, but not strictly-hierarchical, dimensions at the same time, which is known as cross-classification. If the reality that the observations simultaneously and independently belong to the two distinct groups is ignored (by only considering a single context), the results may be misleading and either overestimate or underestimate the effect of one group [31].

In a variety of areas, such as education, medicine, and crime, the CCMM has been used to analyze the data with a cross-classified structure. In education, CCMMs have been used to identify the effects of schools and neighborhoods on the smoking behaviors of adolescents [31]. Because an adolescent belongs to multiple groups (i.e., school and neighborhood) at the same time, biased estimation results can occur when ignoring non-hierarchical nesting structures. The study pointed out that the CCMM can reduce the limitations of existing models. In medicine, the study of nurses with depression disorders used the CCMM to deal with cross-classified data, including counties, organizational associations, and workplace groups [32]. In crime, CCMMs have been used to study the consequences of the cases involving suspected terrorists to consider the effects of terrorist organizations and criminal court environment [33]. These studies also showed that a better model can be obtained by considering the effects of higher groups simultaneously.

Similar to these studies, we adopted the CCMM to overcome the limitations of traditional single-level models and conventional two-level models. Bus–pedestrian crashes are inevitably affected by two higher groups, i.e., region and company, which violates the assumption of regression that the crashes are independent of each other. The consequence of failing to consider hierarchical structures is that standard errors of estimated coefficients may be underestimated, thereby resulting in an overstatement of statistical significance. In order to prevent unbiased results, we considered two higher groups simultaneously by using the CCMM. Also, the CCMM allows identification of the effects of higher groups on observations. By comparing the variances of higher groups in the model, we confirmed the effects of company and region on injury severities in bus–pedestrian crashes. Lastly, in the CCMM, we can estimate both effects due to observed and unobserved group factors. In traditional regression models, dummy variables for groups can be included, i.e., fixed effects models. However, in these models, it is not possible to distinguish the effects due to observed and unobserved group factors. The distinguishable results from the CCMM will be more helpful for a clear understanding of bus–pedestrian crashes.

2.3. Macro-Level Factors in Traffic Safety

The model developed in this study included company-level factors and regional-level factors, both of which are macro-level factors. Many previous studies have also developed models that include macro-level factors [34–38]. The factors in the previous studies included demographic, socioeconomic, land use, road network, and transit characteristics, and they were aggregated using various geographical units, e.g., traffic analysis zones, block groups, and census tracts [37]. These factors provide useful insights in terms of long-term improvements in traffic safety. Thus, by analyzing these factors, researchers can suggest implications for improving traffic safety in both traffic engineering and non-traffic engineering, which can lead to many aspects of policies and decisions related to traffic safety investments [37,38].

3. Data and Methods

3.1. Data Preparation

For this study, we combined three datasets, one each from the Traffic Accident Analysis System (TAAS), the Transport Workers Management System (TWMS), and the Korean Statistical Information Service (KOSIS), each of which covered the five-year period from 2011 through to 2015 in South Korea. The TAAS dataset contains all police-reported crash data, and it is maintained by the National Police Agency. The bus–pedestrian crash data in TAAS were used in this study. TWMS includes only bus crashes with information about drivers and companies, and it is maintained by the Korea Transportation Safety Authority, which operates and manages various transportation safety projects to prevent the crashes of commercial vehicles. Therefore, it is necessary to match each crash of the two different datasets to obtain the cross-classified crash structure of companies and regions, because TAAS has no information on the companies that employ the drivers and TWMS lacks crash-level information.

The TAAS and TWMS datasets were combined based on common variables. We matched 9913 bus–pedestrian crashes from the TAAS and TWMS datasets, using the time and location of each crash, the gender and age of the driver, the type of vehicle, and the severity of the injuries as matching variables. As a result of the matching, the bus–pedestrian crashes had both crash-level information and company-level information (1447 companies). In addition, based on the identification of the region from which the crash-level information was acquired, we added regional-level information that was provided by KOSIS for 222 municipalities.

In this database, the severity of injuries was classified into three levels, i.e., fatal injuries (4.5%), major injuries (50.7%), and minor injuries (44.8%). In the study, a crash identified as a "fatal injury crash" represents a crash where an injury caused the death of at least one person within 30 days of the crash. A major injury crash refers to a crash in which at least one person suffered an injury that required treatment for 3 weeks or more after the crash. A minor injury crash indicates a crash in which at least one person suffered an injury that required more than 5 days but less than 3 weeks of treatment after the crash.

The bus–pedestrian crash data comprise two cross-classified levels, i.e., the lower-level for crash-level factors, and the upper-level for company-level and regional-level factors. The crash-level factors include driver related factors, the vehicle type and its age, road conditions, and weather conditions. The company-level factors include information related to the driver and the company, such as the average age of the drivers, whether the buses are inter-city or intra-city buses, the status of each company's safety inspections, the number of vehicles and drivers reflecting the size of the company, and the number of violations of each company. The regional level factors include the road pavement ratio, elderly ratio, the ratio of the transportation budget, the number of doctors, population density, financial independence rate, and the number of vehicles divided by the total population. In the study, only statistically significant factors were used in the models.

We considered 15 crash-level factors, three company-level factors, and three regional-level factors. Table 2 provides the descriptive statistics of the factors considered in the model, which include the mean and the standard deviation. The number of minor injuries, major injuries, and fatal injuries were 4440, 5028, and 445, respectively. The crash-level factors in Table 2 show that the values of individual crashes were averaged by the level of severity. In the company-level and regional-level factors, the values of the upper groups, i.e., the groups to which individual crashes belong, were assigned to the individual crashes, and the values of individual crashes were averaged according to the level of severity. Table 3 shows the average of factors according to the year (2011 to 2015). About 2000 crashes have occurred every year, but we were unable to identify a specific trend in the factors over time. In addition, we conducted a variance inflation test to examine the possibility of multi-collinearity between the factors. The results of the variance inflation factor ranged from 1.01 to 1.16, indicating that the effects of multi-collinearities were minimal.

Table 2. Description of factors.

Factor	Coding Convention	Level of Severity (Mean (Standard Deviation))		
		Minor Injury	Major Injury	Fatal Injury
Crash Frequency		4440	5028	445
Company Related				
Intercity Bus	1 = if the company operates an intercity bus; 0 = otherwise	0.131(0.337)	0.146(0.354)	0.191(0.394)
Average age of drivers	Average age of drivers in company	49.56(2.003)	49.52(2.270)	49.60(2.148)
Number of violations	Average of the company's count of law violations	1.741(2.482)	1.618(2.360)	1.508(2.281)
Region Related				
Elderly ratio	Ratio of elderly in region	11.71(3.640)	12.08(4.240)	12.67(4.846)
Ratio of transportation infrastructure budget	Ratio of total municipal budget spent on transportation infrastructure	41.18(13.470)	39.74(13.331)	39.53(17.726)
Number of doctors	Number of doctors per thousand residents	3.508(3.419)	3.356(3.228)	3.036(2.818)
Crash Related				
Crash Location				
Intersection	1 = if crash occurred at intersection; 0 = otherwise	0.244(0.430)	0.293(0.455)	0.323(0.468)
Exclusive bus lane	1 = if crash occurred at exclusive bus lane; 0 = otherwise	0.014(0.118)	0.023(0.150)	0.040(0.195)
Road Alignment				
Right curve	1 = if crash occurred at right curve; 0 = otherwise	0.024(0.152)	0.033(0.180)	0.057(0.232)
Left curve	1 = if crash occurred at left curve; 0 = otherwise	0.018(0.131)	0.020(0.140)	0.018(0.132)
Road Type				
Rural road	1 = if crash occurred at rural road; 0 = otherwise	0.035(0.184)	0.049(0.215)	0.081(0.274)
Residential street	1 = if crash occurred at residential street; 0 = otherwise	0.048(0.214)	0.029(0.168)	0.020(0.139)
Road Width				
Medium sized road (13–20 m)	1 = if crash occurred at road which has 13–20 m width; 0 = otherwise	0.150(0.357)	0.160(0.367)	0.152(0.359)
Wide road (>20 m)	1 = if crash occurred at road with wider than 20 m; 0 = otherwise	0.119(0.324)	0.152(0.360)	0.207(0.405)
Driver Behavior				
Speeding	1 = if driver was speeding when the crash occurred; 0 = otherwise	0.001(0.026)	0.004(0.066)	0.037(0.190)
Signal violation	1 = if driver violated traffic signal when the crash occurred; 0 = otherwise	0.074(0.262)	0.098(0.297)	0.097(0.296)
Speed 30–60 km/h	1 = if driver's speed was 30 to 60 km/h just before the crash; 0 = otherwise	0.093(0.291)	0.187(0.390)	0.341(0.475)
Speed over 60 km/h	1 = if driver's speed was faster than 60 km/h just before the crash; 0 = otherwise	0.003(0.058)	0.017(0.360)	0.121(0.326)
Vehicle				
Vehicle age	1 = if driver's vehicle was older than 7 years; 0 = otherwise	0.139(0.346)	0.163(0.369)	0.198(0.399)
Roadway/Environment				
Weekend	1 = if crash occurred on weekend; 0 = otherwise	0.249(0.432)	0.222(0.415)	0.181(0.450)
Wet road	1 = if crash occurred on wet road; 0 = otherwise	0.113(0.317)	0.137(0.344)	0.191(0.394)

Note. Standard deviations appear in parentheses.

Table 3. Average of factors for each year.

Factor	Year				
	2011	**2012**	**2013**	**2014**	**2015**
Crash Frequency	1953	2059	2104	1906	1891
Company Related					
Intercity bus	0.127	0.136	0.156	0.145	0.141
Average age of drivers	49.584	49.524	49.619	49.463	49.525
Number of violations	1.762	1.862	1.629	1.491	1.582
Region Related					
Elderly ratio	12.132	11.880	11.911	11.891	11.899
Ratio of transportation infrastructure budget	39.766	39.939	40.316	39.978	41.965
Number of doctors	3.583	3.498	3.330	3.333	3.300
Crash Related					
Crash Location					
Intersection	0.227	0.257	0.287	0.275	0.317
Exclusive bus lane	0.030	0.023	0.018	0.020	0.006
Road Alignment					
Right curve	0.033	0.034	0.030	0.030	0.023
Left curve	0.019	0.022	0.021	0.019	0.012
Road Type					
Rural road	0.040	0.051	0.047	0.055	0.026
Residential street	0.050	0.022	0.035	0.046	0.086
Road Width					
Medium sized road (13–20 m)	0.153	0.156	0.163	0.159	0.144
Wide road (>20 m)	0.145	0.141	0.158	0.138	0.115
Driver Behavior					
Speeding	0.004	0.004	0.004	0.001	0.008
Signal violation	0.082	0.093	0.091	0.080	0.088
Speed 30–60 km/h	0.204	0.206	0.179	0.161	0.165
Speed over 60 km/h	0.023	0.021	0.019	0.015	0.014
Vehicle					
Vehicle age	0.165	0.191	0.184	0.221	0.181
Roadway/Environment					
Weekend	0.252	0.235	0.230	0.243	0.222
Wet road	0.130	0.154	0.126	0.126	0.108

3.2. Two-Level Multi-Level Model

Single-level modeling assumes that there is no correlation between individual crashes, which means that the residuals of the model are independent. However, if this assumption is inappropriate for a given set of data, the model may produce biased results, underestimated standard errors, and indicate confidence intervals that are too narrow. Therefore, when there is a correlation between the factors that affect individual crashes and the upper group of the factors, multi-level modeling is suitable because it prevents inaccurate or biased estimates caused by ignoring the hierarchical structure of the crash data.

In this paper, we briefly described the specification of a two-level multi-level model. Yoon et al. (2017) provided a detailed description of the model [1]. We applied a random intercept model that has been used extensively in previous studies. The model has a constant slope, but the intercept of the model depends on the upper group, which represents random effects. Given crash data, N, nested within J groups, the equation for a random intercept model, which is a coupled model that includes an individual-level model (level 1) and a group-level model (level 2), is as follows:

$$Y_{ij} = \gamma_{00} + \sum_{p=1}^{P} \gamma_{p0} X_{pij} + \sum_{q=1}^{Q} \gamma_{0q} Z_{qj} + u_{0j} + e_{ij}, \quad (1)$$

where the outcome, Y_{ij}, for crash observation i in group j with the explanatory (level 1) factors, X_p and γ_{p0}, is the regression coefficient for the individual-level factors. The random parameter, e_{ij}, for individual crashes is assumed to be distributed normally. For the group-level model (level 2), with an arbitrary intercept, γ_{00} is the intercept of level 2; Z_{qj} represents the group-level factors for the jth group; γ_{0q} is the regression coefficient related to the group-level factors; and u_{0j} is the random parameter of level 2. The intercept of the model varies across groups, but the coefficient of the slope is assumed to be fixed.

For the continuous outcome, Y_{ij}, it is important to check the normality of the error term (residual) to obtain unbiased results. Some studies have analyzed the effect of non-normal error distributions (e.g., mixtures of normal distributions, lognormal, and chi-squared distributions), and a few studies have developed a multi-level model that assumes a non-normal error distribution [39–41]. However, the models can be difficult to apply because of the complexity of estimation. Therefore, if the non-normality of errors (residuals) cannot be easily solved by, for example, by transforming the dependent variable, it is worth exploring other approaches, such as non-parametric estimation of random effects.

The variance of outcome (Y_{ij}) is divided into two components. One is the variance of e_{ij}, which represents the within-group variability, and the other is the variance of u_{0j}, which represents the between-group variability. The variances are used to compare the relative impacts on the outcome between the attributes of individuals and groups. In the study, we analyzed whether the severity of injury was influenced to a greater extent by the attributes of individual crashes or by the attributes of a higher group, i.e., the company and the region.

3.3. Cross-Classified Multi-Level Model (CCMM)

Because the CCMM has two upper groups, it considers both group-level factors and random parameters for two different groups. Let us assume that we have crash data, N, cross-classified with J groups and K other groups. Since a crash belongs to two non-nested groups, i.e., J and K, at the same time, the CCMM has group-level explanatory factors, W_q and Z_r, and random parameters, u_{01j} and u_{02k}, for each group. In a cross-classified multi-level model, Equation (1) is modified as follows:

$$Y_{ij} = \theta_0 + \sum_{p=1}^{P} \theta_p X_{pij(jk)} + \sum_{q=1}^{Q} \gamma_{01q} W_{qj} + \sum_{r=1}^{R} \gamma_{02r} Z_{rk} + u_{01j} + u_{02k} + e_{ij}, \quad (2)$$

where θ_p is the regression coefficient for the individual-level (level 1) factors. For the group-level model (level 2), the random intercept, θ_0, is the intercept at level 2, and γ_{01q} and γ_{02r} are the regression coefficients related to the group-level factors.

3.4. Two-Level and Cross-Classified Ordered Multi-Level Models

In the analysis of the injury severity of crashes, the outcome, Y_{ij}, is the ordinal outcome. To capture the ordered nature of the given data in the model, it is necessary to use the cumulative probabilities of the ordinal outcomes. O'Donnell and Connor (1996) provided a detailed mathematical description of ordered logit models [42]. The multi-level model for ordinal outcomes is conceptually identical to the model described above except for the ordinal outcome. If the number of categories for the ordinal outcomes is m, the ordered observed outcomes (Y_{ij}) can be generated from the latent continuous response as follows:

$$Y_{ij} = \left\{ \begin{array}{c} 1\ if\ Y^*_{ij} \leq \delta_1 \\ 2\ if\ \delta_1 < Y^*_{ij} \leq \delta_2 \\ \vdots \\ m\ if\ \delta_{m-1} < Y^*_{ij} \end{array} \right\}, \tag{3}$$

where Y^*_{ij} is the latent continuous response that represents the levels of injury severity for observation i and upper group j, and δ_1, δ_2, and δ_{m-1} are ancillary parameters (known as cutoff points or thresholds). The probabilities of injury outcomes, π_{1ij}, π_{2ij}, and π_{Mij}, are estimated as follows:

$$\pi_{1ij} = \Pr\left(Y_{ij} = 1\right),\ \pi_{2ij} = \Pr\left(Y_{ij} = 2\right),\ \cdots,\ \pi_{Mij} = \Pr\left(Y_{ij} = M\right). \tag{4}$$

The idea of the cumulative probabilities of ordinal outcomes leads to a cumulative logit (π^*_{mij}):

$$\pi^*_{mij} = \Pr\left(Y_{ij} \leq m\right) = \pi_{1ij} + \pi_{2ij} + \cdots + \pi_{mij}\ (m \leq M). \tag{5}$$

With this background, we developed a single-level model and three multi-level models. A single-level ordered logit model (model 1) was developed to compare with the three multi-level models. Assuming that the model has an ordered outcome (Y_i) observed with m categories, the single-level ordered logit model is as follows:

$$\log\left(\frac{\pi^*_{mi}}{1 - \pi^*_{mi}}\right) = \gamma_0 + \sum_{p=1}^{P} \gamma_p X_p + \delta_m. \tag{6}$$

We also developed a two-level multi-level model for company level factors only (model 2), and named it the company-only multi-level model (COMM). Assuming that the model has an ordered observed outcome for a crash, i, nested in a given company, j, COMM is as follows:

$$\log\left(\frac{\pi^*_{mij}}{1 - \pi^*_{mij}}\right) = \gamma_{00} + \sum_{p=1}^{P} \gamma_{p0} X_{pij} + \sum_{q=1}^{Q} \gamma_{0q} Z_{qj} + u_{0j} + \delta_m. \tag{7}$$

Model 3 is a two-level multi-level model for regional level factors only. We called this model the region-only multi-level model (ROMM). Assuming the model has an ordered observed outcome for a crash, i, nested in a given region, k, ROMM is as follows:

$$\log\left(\frac{\pi^*_{mik}}{1 - \pi^*_{mik}}\right) = \gamma_{00} + \sum_{p=1}^{P} \gamma_{p0} X_{pik} + \sum_{r=1}^{R} \gamma_{0r} Z_{rk} + u_{0k} + \delta_m. \tag{8}$$

An advanced cross-classified multi-level model (CCMM) that combines company and regional level factors (model 4) is as follows:

$$\log\left(\frac{\pi^*_{mi(jk)}}{1 - \pi^*_{mi(jk)}}\right) = \theta_0 + \sum_{p=1}^{P} \theta_p X_{pi(jk)} + \sum_{q=1}^{Q} \gamma_{01q} W_{qj} + \sum_{r=1}^{R} \gamma_{02r} Z_{rk} + u_{01j} + u_{02k} + \delta_m. \tag{9}$$

3.5. Intra-Class Correlation Coefficient (ICC)

In order to consider the adequacy of multi-level modeling, the proportions of the variance within and between groups are required. Using the variance within and between groups, we can calculate the intra-class correlation coefficient (*ICC*) to examine the proportion of level specific variance. *ICC* is calculated from unconditional multi-level models (models without any input factors). If the value of the *ICC* is close to zero, the variation between groups is small, and a single-level model may be appropriate. In general, *ICC*s that range from 5% to 25% are considered to be appropriate for the use of multi-level modeling in social science [1]:

$$ICC_1 = \frac{\sigma^2_{u1}}{\sigma^2_{u1} + \sigma^2_{u2} + \sigma^2_e}(\text{intra} - \text{group }1), ICC_2 = \frac{\sigma^2_{u2}}{\sigma^2_{u1} + \sigma^2_{u2} + \sigma^2_e}(\text{intra} - \text{group }2). \tag{10}$$

An intra-cell correlation can also be calculated:

$$ICC_{12} = \frac{\sigma^2_{u1} + \sigma^2_{u2}}{\sigma^2_{u1} + \sigma^2_{u2} + \sigma^2_e}(\text{intra} - \text{cell}). \tag{11}$$

The intra-cell correlation refers to the correlation between the severities of two crashes that involve the same company and occur in the same region. However, the variance parameter, σ^2_e, at level 1 is not available due to the absence of the error term, e_{ij}, in Equations (4)–(6). To overcome this limitation, the models were re-fitted with an explanatory variable that was specified as being normally distributed. As a result, in this analysis, the variance parameter, σ^2_e, is equal to 1 [9].

3.6. Model Evaluation Measures

We used several measures to evaluate the models we developed. First, the Akaiked information criterion (*AIC*) and Bayesian information criterion (*BIC*) were used as follows:

$$AIC = 2k - 2LL(full), \tag{12}$$

$$BIC = kln(n) - 2LL(full), \tag{13}$$

where n is the number of observations, k is the number of parameters, and *LL*(*full*) is the log-likelihood for the full model.

Besides *AIC* and *BIC*, additional performance measures (i.e., precision, recall, F-measure, and G-mean) were used to compare the effectiveness of the models. Based on the classification tables, we calculated the true positive (*TP*), true negative (*TN*), false positive (*FP*), and false negative (*FN*). The precision, recall (or sensitivity), and specificity were calculated as follows:

$$Precision = \frac{TP}{TP + FP}, \tag{14}$$

$$Recall\ (or\ sensitivity) = \frac{TP}{TP + FN}, \tag{15}$$

$$Specificity = \frac{TN}{TN + FP}. \tag{16}$$

Precision is a measure of a model's exactness, and models that have higher precision values are better models. The recall (or sensitivity) measures the effectiveness of the model on the positive/minority case, while specificity measures the effectiveness of the model on the negative/majority case. To balance between the *FP* and *FN*, we also used some combined performance measures, such as the F-measure and G-mean [43,44]:

$$F - measure = \frac{2 \times \text{sensitivity} \times \text{precision}}{\text{sensitivity} \times \text{precision}}, \tag{17}$$

$$G - mean = \sqrt{Sensitivity \times Specificity}. \tag{18}$$

The measures were estimated according to the level of severity, and the weighted average was also estimated based on the observations.

4. Results

We had two main objectives in this study, i.e., (1) to examine the effect of two upper level (company and region) factors on the injury severity of bus–pedestrian crashes, and (2) to identify the factors that influenced the injury severity of bus–pedestrian crashes. Table 4 summarizes for all of the models the within-group and between-group variances and the *ICC*s, which are the proportions of the variance within and between groups. The variances are the results of the unconditional models (models without any input factors), and the *ICC*s were calculated based on Equations (7) and (8). Table 5 shows the results of the evaluation of the model using seven performance measures. Table 6 shows the results, including the fixed and random effects. The level 2 factors include company-level and regional-level factors, and the level 1 factors include the crash-level factors as well as the company and regional factors that were not considered as upper-levels. The models in this study were estimated using HLM 7, i.e., hierarchical linear and nonlinear modeling software [12].

Table 4. Estimation results for within-group and between-group variances and intra-class correlation coefficients.

	COMM	ROMM	CCMM
Unconditional company (level 2) variance	0.202 ***	NA	0.197 ***
Unconditional region (level 2) variance	NA	0.054 ***	0.049 ***
Crash (level 1) variance	1.000 ***	1.000 ***	1.000 ***
Intra-class correlation coefficient (*ICC*)	16.8%	5.1%	19.7%

Note. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

Table 5. Model evaluation results.

		Single-Level Model	COMM	ROMM	CCMM
AIC		37,511.87	37,253.05	37,373.88	37,137.24
BIC		37,677.42	37,404.21	37,546.62	37,281.21
Precision	Minor Injury	0.451	0.451	0.456	0.472
	Major Injury	0.473	0.474	0.473	0.492
	Fatal Injury	0.129	0.119	0.124	0.136
	Weighted average	0.448	0.447	0.450	0.467
Recall (sensitivity)	Minor Injury	0.491	0.473	0.488	0.520
	Major Injury	0.403	0.415	0.417	0.422
	Fatal Injury	0.231	0.227	0.202	0.218
	Weighted average	0.435	0.433	0.439	0.457
Specificity	Minor Injury	0.516	0.533	0.528	0.533
	Major Injury	0.547	0.560	0.546	0.560
	Fatal Injury	0.494	0.513	0.504	0.513
	Weighted average	0.531	0.546	0.536	0.546
F-measure	Minor Injury	0.470	0.462	0.471	0.495
	Major Injury	0.435	0.442	0.443	0.454
	Fatal Injury	0.166	0.156	0.153	0.168
	Weighted average	0.439	0.438	0.443	0.460
G-mean	Minor Injury	0.503	0.502	0.507	0.524
	Major Injury	0.470	0.482	0.477	0.467
	Fatal Injury	0.338	0.341	0.319	0.327
	Weighted average	0.479	0.485	0.484	0.486

Table 6. Model estimation results for bus–pedestrian crashes.

Fixed Effects			Single-Level Model	COMM	ROMM	CCMM
LEVEL2	**Company Related**					
	Intercity bus		NA	0.105 (0.119)	NA	0.100 (0.117)
	Average age of drivers		NA	−0.023 (0.015)	NA	−0.023 (0.014)
	Number of violations		NA	−0.024 (0.027)	NA	−0.022 (0.026)
	Region Related					
	Elderly ratio		NA	NA	0.028 (0.006) ***	0.026 (0.006) ***
	Ratio of transportation infrastructure budget		NA	NA	−0.004 (0.002) **	−0.005 (0.002) **
	Number of doctors		NA	NA	−0.028 (0.009) ***	−0.028 (0.009) ***
LEVEL1	**Company Related**					
	Intercity bus		0.104 (0.060) *	NA	0.101 (0.060) *	NA
	Average age of drivers		−0.017 (0.010) *	NA	−0.017 (0.010) *	NA
	Number of violations		−0.020 (0.009) **	NA	−0.019 (0.009) **	NA
	Region Related					
	Elderly ratio		0.028 (0.005) ***	0.026 (0.006) ***	NA	NA
	Ratio of transportation infrastructure budget		−0.005 (0.002) ***	−0.005 (0.002) ***	NA	NA
	Number of doctors		−0.031 (0.006) ***	−0.030 (0.006) ***	NA	NA
	Crash Related					
	Crash location	Intersection	0.276 (0.046) ***	0.280 (0.046) ***	0.281 (0.046) ***	0.284 (0.047) ***
		Exclusive bus lane	0.593 (0.149) ***	0.598 (0.151) ***	0.614 (0.151) ***	0.609 (0.152) ***
	Road alignment	Right curve	0.506 (0.119) ***	0.488 (0.120) ***	0.502 (0.120) ***	0.485 (0.120) ***
		Left curve	0.029 (0.149)	0.019 (0.150)	0.020 (0.149)	−0.026 (0.151)
	Road type	Rural road	0.231 (0.101) **	0.242 (0.102) **	0.217 (0.103) **	0.224 (0.104) **
		Residential street	−0.366 (0.109) ***	−0.366 (0.110) ***	−0.360 (0.111) ***	−0.358 (0.111) ***
	Road width	Medium road (13–20 m)	0.095 (0.057) *	0.098 (0.058) *	0.092 (0.057)	0.095 (0.058)
		Wide road (>20 m)	0.269 (0.061) ***	0.277 (0.062) ***	0.275 (0.062) ***	0.281 (0.062) ***
	Driverbehavior	Speeding	1.779 (0.339) ***	1.845 (0.340) ***	1.791 (0.340) ***	1.854 (0.340) ***
		Signal violation	0.163 (0.073) **	0.180 (0.074) **	0.158 (0.073) **	0.174 (0.074) **
		Speed 30–60 km/h	0.948 (0.059) ***	0.942 (0.060) ***	0.949 (0.059) ***	0.942 (0.060) ***
		Speed over 60 km/h	2.391 (0.176) ***	2.322 (0.177) ***	2.389 (0.176) ***	2.322 (0.178) ***
	Vehicle	Vehicle age (≥7 years)	0.167 (0.056) ***	0.181 (0.057) ***	0.160 (0.056) ***	0.175 (0.057) ***
	Roadway/Environment	Weekend	−0.125 (0.048) ***	−0.114 (0.048) **	−0.127 (0.048) **	−0.116 (0.048) **
		Wet road	0.230 (0.060) ***	0.232 (0.061) ***	0.238 (0.061) ***	0.239 (0.061) ***
Random Effects						
Company (level 2) variance			NA	0.1785 ***	NA	0.1687 ***
Region (level 2) variance			NA	NA	0.0346 ***	0.0298 ***
Crash (level 1) variance			NA	1.000 ***	1.000 ***	1.000 ***

Note. Standard errors appear in parentheses in the estimates of the parameters.; * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

In order to justify the multi-level approach, we examined the intra-class correlation coefficient (*ICC*) values obtained from different models. The *ICC* values for the COMM and ROMM were 16.8% and 5.1%, respectively (see Table 4). It indicated that the variances in the severity of bus–pedestrian crashes among companies and regions were greater than 5%, which justified the use of the COMM and ROMM. In addition, the use of the CCMM with 19.7% of the *ICC* was justified based on the bus–pedestrian crash data.

As shown in Table 4, the variances of the company and region groups estimated by the CCMM were used to assess the effects of the regional and the company's characteristics on the severity of bus–pedestrian crashes. The unconditional company variance in the CCMM was 0.197, which was significantly larger than the region variance, i.e., 0.049. This implies that the unobserved heterogeneity was greater among companies than among regions, and the effects of the companies' characteristics were dominant in bus–pedestrian crashes. Therefore, an identification of the company-level risk factors is valuable in the severity of bus–pedestrian crashes.

In addition, the Hausman test was performed to determine whether multi-level modeling was appropriate. The Hausman test, which can be applied to a wide variety of possible model misspecifications, determines whether the coefficient estimates from the multi-level model are significantly statistically different from the fixed effects estimates (assumed to be unbiased). If the null hypothesis holds, i.e., that higher-level random effects are not correlated with any crash-level factors, both estimates are consistent and efficient, so we can apply multi-level modeling. In this study, the *p*-values of the test were 0.633 and 0.451 for the company-level and regional-level, respectively, which indicates that we can apply multi-level models to the given data.

Prior to discussing the risk factors, we evaluated the models that were developed in the study, as shown in Table 5. In terms of statistical fit, we found that the CCMM had lower *AIC* and *BIC* values than the other three models (i.e., single-level model, COMM, and ROMM). In terms of classification accuracy, the CCMM also had a lower F-measure and G-mean than the other three models. The results indicated that the CCMM had the best performance in this study.

We identified the risk factors that significantly increased the likelihood of fatal injuries, i.e., three regional-level factors and 13 crash-level factors in the CCMM (Table 6). The discussion of the factors is presented in the next section.

5. Discussion

5.1. Company-Level Factors

Of the company-level factors, the intercity bus, the average age of drivers, and the number of violations were found to be statistically significant in the single-level model and in ROMM, indicating that the factors were potentially important risk factors in increasing or decreasing the severity of bus–pedestrian crashes. In the COMM and CCMM, however, those factors were not statistically significant. The single-level model and ROMM produced underestimated standard errors of those factors. This finding showed that ignoring the effects of upper-level groups, i.e., company effects in the bus–pedestrian crashes, produced type I statistical errors. We confirmed that the CCMM with the ability to take into account the factors of the non-nested groups simultaneously produced unbiased estimates.

5.2. Regional-Level Factors

Among the factors at the regional level, the elderly ratio, the ratio of the transportation infrastructure budget, and the number of doctors turned out to be statistically significant in all models. The elderly ratio factor had a positive coefficient in the CCMM, which means that the higher elderly ratio was associated with an increased severity of bus–pedestrian crashes. In areas with a high percentage of elderly people, elderly people are more susceptible to these crashes, and the potential for a serious crash is greater. The results are consistent with those of previous studies [1,7,45]. In addition, the result may have been influenced by the regional distribution of residents by age group in South Korea. Regions with significant elderly populations are generally rural areas, and these areas lack the budget and qualified personnel for preventing traffic crashes compared to urban areas. This shortage can lead to more severe injuries. Therefore, regions with high elderly populations require more care and effort to prevent bus–pedestrian crashes.

The ratio of the transportation infrastructure budget had a negative coefficient in the CCMM. This means that municipalities with relatively large budgets for their transportation infrastructure may be able to reduce the number of severe crashes. The transportation infrastructure budget can affect facilities that are related to traffic safety and road environments, such as curve radius, drainage, and pavement condition on the roads. For example, in South Korea, those crashes involving children have been reduced significantly by greatly expanding the budget for safety in school zones.

The number of doctors had a negative coefficient in CCMM. This means that an increase in the number of doctors is associated with a reduction in the severity of the injury of bus–pedestrian crashes. The number of doctors represents the level of healthcare available when a crash occurs, such as emergency medical systems and the quality of service at the local hospital [7]. The results are consistent with those of previous studies [7,20].

5.3. Crash-Level Factors

The coefficient directions of the crash-level factors were the same in all models. In addition, the standard errors of the factors in the multi-level models tended to be slightly larger. Fifteen crash-level factors were statistically significant in all models.

In the event of a crash, factors related to intersections and exclusive bus lanes were associated with more severe crashes. In the case of exclusive bus lanes, the excessive speed of buses with a separate travel priority may increase the injury severity of pedestrians in crashes, which was supported by the findings of previous studies [13,26]. However, the effect of intersections has proven to be controversial.

Some studies have pointed out that pedestrian crashes that occurred at intersections were relatively severe [26]. However, Mohamed et al. (2013) found that the injury severity of pedestrians in crashes that occurred at intersections was relatively lower [22]. They also suggested that pedestrian crashes at intersections were affected by various conditions, such as the intersection type, control condition, and the number of pedestrians.

As for the roadway alignment, the right-curve factor was statistically significant with a positive coefficient, while this was not the case for the left-curve factor. This indicates that the injury severity on right-curved roads is higher than on straight roads. In South Korea, the driver's seat is on the left side of the vehicle. This can make it more difficult to see pedestrians approaching from the right side of a road that curves to the right due to the long sight distance and the large size of the bus. The rural road and residential street factors were statistically significant. The coefficient of the rural road factor was positive, which means that crashes on rural roads are associated with an increased injury severity. Since rural roads are located outside the city, vehicles can travel faster, drivers may not expect to encounter pedestrians, and it is difficult to get medical treatment promptly, which can result in serious injury. Unlike rural roads, the risk factors for residential streets have a negative coefficient, indicating that crashes on residential streets are associated with a reduced severity of injuries. In general, the speed of buses on residential streets is low and the drivers can easily see the pedestrians. For a similar reason, wider roads lead to increases in the injury severity when bus–pedestrian crashes occur [7,28].

Regarding the bus driver's behavior, speeding and signal violations were statistically significant and had positive coefficients. These illegal behaviors of bus drivers are more likely to increase the injury severity [25]. Higher speeds just before a crash were also associated with increasing the injury severity in pedestrian crashes. Therefore, traffic calming policy can greatly contribute to reducing the severity of bus–pedestrian crashes.

In the vehicle and roadway environment, older vehicles have been found to be associated with more severe bus–pedestrian crashes [1]. Weekend and dry road factors are also associated with decreasing the injury severity of pedestrians in bus–pedestrian crashes [25,27].

6. Conclusions and Recommendation

Multi-level modeling, which is able to consider the upper groups, is very important in the establishment of policies to prevent bus–pedestrian crashes. This is because most of the government's policies related to bus crashes have been developed and implemented on a company basis rather than an individual basis. In addition, investments in safety facilities for pedestrians and traffic are made by individual municipalities. Thus, it is necessary to be able to judge the impact of group-level factors on bus crashes so that proper policies and investments can be established and implemented. Thus, this study addressed the very important issue of bus–pedestrian crashes.

The objective of this study was to identify the risk factors associated with the injury severity incurred in bus–pedestrian crashes. The bus–pedestrian crashes that are affected by company and regional characteristics have a cross-classified hierarchical structure, which is difficult to address properly using a single-level model or even a multi-level model. Therefore, we used the cross-classified multi-level model (CCMM) to consider simultaneously the unobserved heterogeneities in two distinct groups, i.e., companies and regions.

The results of the study showed that the company effect on the injury severity in bus–pedestrian crashes was 16.8% (ICC = 0.168), and the regional effect on the severity of these injuries was 5.1% (ICC = 0.051). This justified the application of the COMM and ROMM and, in particular, it suggested the need for the CCMM to consider both of the upper groups. Also, we found that ignoring the effects of the upper-level groups, i.e., company effects in bus–pedestrian crashes, produced type I statistical errors. As a result, it showed the advantage of the CCMM in comparison with a single-level model and a two-level model.

In the study, we identified the risk factors that significantly increase the probability of fatal injuries, i.e., three regional-level factors and 13 crash-level factors. The results provided useful insights concerning bus–pedestrian crashes. The injury severity may be high in regions where there is a high percentage of elderly people and a low transportation infrastructure budget. Therefore, local governments that have large populations of elderly people should pay more attention to traffic plans for bus systems to reduce the severity of injuries. Also, local governments should recognize that pedestrian crashes are linked directly to peoples' lives and increase their budget for traffic safety infrastructure. As a matter of course, the federal government should provide funding for local governments that may find it difficult to increase their budgets. The high speed of buses may increase the severity of bus–pedestrian crashes. On roads where buses travel at high speeds, e.g., wide roads or roads with an exclusive bus lane, improvements of the designs of the roads and the installation of safety facilities are required to ensure complete separation of pedestrians and buses. The introduction of a device capable of detecting pedestrians and automatically braking may also reduce the injury severity. Also, it is necessary for the managers of bus companies to encourage their drivers to refrain from speeding.

The factors identified in this study are beneficial, but given the constraints associated with the limited resources, the recommendations should be addressed in the order of their priorities. These priorities can be presented according to the level of various factors. In general, micro-level factors (i.e., crash-level factors) can contribute directly to the occurrence of crashes, and more direct countermeasures should be developed. These micro-level factors (i.e., company-level factors and regional-level factors) provide useful insights in terms of long-term improvements in traffic safety, which lead to many aspects of the creation of policies and decisions concerning safety investments. Therefore, improvements in crash-level factors, such as speeding, can be considered as a priority, while the effect of company-level and regional-level factors (e.g., socioeconomic and demographic factors) should be considered from the long-term perspective.

Future research can enhance the understanding the characteristics of bus–pedestrian crashes. We found that, in South Korea, the effects of a company's characteristics were more dominant in bus–pedestrian crashes than the effects of regional characteristics. However, we did not identify a key factor that is related to the companies. Therefore, efforts are needed to collect additional factors related to bus companies, such as drivers' workloads and safety management practices, and to identify significant factors. In addition, in order to have sustainable transportation systems, an essential effort is to identify the factors related to their sustainability and provide countermeasures for such systems.

It is necessary to analyze the latest data to identify current issues in bus–pedestrian safety. We used three datasets from 2011 to 2015, but various bus safety measures, such as installations of collision avoidance systems and emergency braking systems and the reduction of bus driver's maximum daily working time, have been implemented since 2015. An up-to-date analysis could provide new insights on bus–pedestrian safety. In this study, we used random intercept models of which the intercept only varies according to upper groups. It is easy to interpret the results, but it is difficult to analyze the change of effects of crash level factors based on the company and regional-level factors. A random slope model that has random effects on the coefficients of crash level factors would be an improvement despite the difficulty of interpretation.

Author Contributions: Conceptualization, H.-C.P., S.-Y.K., and D.-K.K.; methodology, H.-C.P., D.-K.K., and B.-J.P.; validation, H.-C.P., S.-Y.K., D.-K.K., and B.-J.P.; formal analysis, H.-C.P. and Y.-J.J.; data curation, H.-C.P. and Y.-J.J.; writing—original draft preparation, H.-C.P. and Y.-J.J.; writing—review and editing, S.-Y.K., D.-K.K., and B.-J.P.; supervision, D.-K.K.; project administration, D.-K.K.; funding acquisition, D.-K.K.

Funding: This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT & Future Planning (2016R1C1B1008492).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Yoon, S.; Kho, S.-Y.; Kim, D.-K. Effect of Regional Characteristics on Injury Severity in Local Bus Crashes. *Transp. Res. Rec. J. Transp. Res. Board* **2017**, *2647*, 1–8. [CrossRef]

2. Haleem, K.; Alluri, P.; Gan, A. Analyzing Pedestrian Crash Injury Severity at Signalized and Non-Signalized Locations. *Accid. Anal. Prev.* **2015**, *81*, 14–23. [CrossRef] [PubMed]
3. Karndacharuk, A.; Wilson, D.J.; Dunn, R. A Review of the Evolution of Shared Space Concepts in Urban Environments. *Transp. Rev.* **2014**, *34*, 190–220. [CrossRef]
4. Elvik, R. The Effect on Accidents of Technical Inspections of Heavy Vehicles in Norway. *Accid. Anal. Prev.* **2002**, *34*, 597–603. [CrossRef]
5. Li, Y.; Itoh, K. Factors Contributing to Driving Safety in the Trucking Industry. *IIE Trans. Occup. Ergon. Hum. Factors* **2013**, *1*, 246–257. [CrossRef]
6. Newnam, S.; Greenslade, J.; Newton, C.; Watson, B. Safety in Occupational Driving: Development of a Driver Behavior Scale for the Workplace Context. *Appl. Psychol.* **2011**, *60*, 576–599. [CrossRef]
7. Kim, M.; Kho, S.-Y.; Kim, D.-K. Hierarchical Ordered Model for Injury Severity of Pedestrian Crashes in South Korea. *J. Saf. Res.* **2017**, *61*, 33–40. [CrossRef]
8. Helai, H.; Chor, C.H.; Haque, M.M. Severity of Driver Injury and Vehicle Damage in Traffic Crashes at Intersections: A Bayesian Hierarchical Analysis. *Accid. Anal. Prev.* **2008**, *40*, 45–54. [CrossRef]
9. Jones, A.P.; Jørgensen, S.H. The Use of Multi-Level Models for the Prediction of Road Accident Outcomes. *Accid. Anal. Prev.* **2003**, *35*, 59–69. [CrossRef]
10. Lenguerrnad, E.; Martin, L.; Laumon, B. Modelling the Hierarchical Structure of Road Crash Data-Application to Severity Analysis. *Accid. Anal. Prev.* **2006**, *38*, 43–53. [CrossRef]
11. Quddus, M. Effects of Geodemographic Profiles of Drivers on Their Injury Severity from Traffic Crashes Using Multi-Level Mixed-Effects Ordered Logit Model. *Transp. Res. Rec.* **2015**, *2514*, 149–157. [CrossRef]
12. Raudenbush, S.W.; Bryk, A.S.; Cheong, Y.F.; Congdon, R.T.; De Toit, M. *Hierarchical Linear and Nonlinear Modeling (HLM7)*; Scientific Software International: Lincolnwood, IL, USA, 2011.
13. Park, H.-C.; Kim, D.-K.; Kho, S.-Y.; Park, P.Y. Cross-Classified Multi-Level Models for Severity of Commercial Motor Vehicle Crashes Considering Heterogeneity Among Companies and Regions. *Accid. Anal. Prev.* **2017**, *106*, 305–314. [CrossRef] [PubMed]
14. Huang, H.; Abdel-Aty, M. Multi-Level Data and Bayesian Analysis in Traffic Safety. *Accid. Anal. Prev.* **2010**, *42*, 1556–1565. [CrossRef] [PubMed]
15. Dupont, E.; Papadimitriou, E.; Martensen, H.; Yannis, G. Multi-Level Analysis in Road Safety Research. *Accid. Anal. Prev.* **2013**, *60*, 402–411. [CrossRef] [PubMed]
16. Montella, A.; Aria, M.; D'Ambrosio, A.; Mauriello, F. Data Mining Techniques for Exploratory Analysis of Pedestrian Crashes. *Transp. Res. Rec.* **2011**, *2237*, 107–116. [CrossRef]
17. Pande, A.; Abdel-Aty, M. Market Basket Analysis of Crash Data from Large Jurisdictions and Its Potential as a Decision Support Tool. *Saf. Sci.* **2009**, *47*, 145–154. [CrossRef]
18. Ghomi, H.; Bagheri, M.; Fu, L.; Miranda-Moreno, L.F. Analyzing Injury Severity Factors at Highway Railway Grade Crossing Accidents Involving Vulnerable Road Users: A Comparative Study. *Traffic Inj. Prev.* **2016**, *17*, 833–841. [CrossRef]
19. Rifaat, S.M.; Tay, R.; De Barros, A. Effect of Street Pattern on the Severity of Crashes Involving Vulnerable Road Users. *Accid. Anal. Prev.* **2011**, *43*, 276–283. [CrossRef]
20. Moudon, A.V.; Lin, L.; Jiao, J.; Hurvitz, P.; Reeves, P. 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. *Accid. Anal. Prev.* **2011**, *43*, 11–24. [CrossRef]
21. Kaplan, S.; Prato, C.G. Risk Factors Associated with Bus Accident Severity in the United States: A Generalized Ordered Logit Model. *J. Saf. Res.* **2012**, *43*, 171–180. [CrossRef]
22. Mohamed, M.G.; Saunier, N.; Miranda-Moreno, L.F.; Ukkusuri, S.V. A Clustering Regression Approach: A Comprehensive Injury Severity Analysis of Pedestrian-Vehicle Crashes in New York, US and Montreal, Canada. *Saf. Sci.* **2013**, *54*, 27–37. [CrossRef]
23. Aziz, H.M.A.; Ukkusuri, S.V.; Hasan, S. Exploring the Determinants of Pedestrian-Vehicle Crash Severity in New York City. *Accid. Anal. Prev.* **2013**, *50*, 1298–1309. [CrossRef] [PubMed]
24. Tefft, B.C. Impact Speed and a Pedestrian's Risk of Severe Injury or Death. *Accid. Anal. Prev.* **2013**, *50*, 871–878. [CrossRef] [PubMed]
25. Prato, C.G.; Kaplan, S. Bus Accident Severity and Passenger Injury: Evidence from Denmark. *Eur. Transp. Res. Rev.* **2013**, *6*, 17–30. [CrossRef]

26. Islam, S.; Jones, S.L. Pedestrian At-Fault Crashes on Rural and Urban Roadways in Alabama. *Accid. Anal. Prev.* **2014**, *72*, 267–276. [CrossRef] [PubMed]
27. Osman, M.; Paleti, R.; Mishra, S.; Golias, M.M. Analysis of Injury Severity of Large Truck Crashes in Work Zones. *Accid. Anal. Prev.* **2016**, *97*, 261–273. [CrossRef] [PubMed]
28. Pour-Rouholamin, M.; Zhou, H. Investigating the Risk Factors Associated with Pedestrian Injury Severity in Illinois. *J. Saf. Res.* **2016**, *57*, 9–17. [CrossRef] [PubMed]
29. Li, D.; Ranjitkar, P.; Zhao, Y.; Yi, H.; Rashidi, S. Analyzing Pedestrian Crash Injury Severity Under Different Weather Conditions. *Traffic Inj. Prev.* **2017**, *18*, 427–430. [CrossRef]
30. Toran Pour, A.; Moridpour, S.; Tay, R.; Rajabifard, A. Modelling Pedestrian Crash Severity at Mid-Blocks. *Transp. A* **2017**, *13*, 273–297. [CrossRef]
31. Dunn, E.C.; Richmond, T.K.; Milliren, C.E.; Subramanian, S. Using Cross-Classified Multi-Level Models to Disentangle School and Neighborhood Effects: An Example Focusing on Smoking Behaviors Among Adolescents in the United States. *Health Place* **2015**, *31*, 224–232. [CrossRef]
32. Muntaner, C.; Li, Y.; Xue, X.; Thompson, T.; O'Campo, P.; Chung, H.; Eaton, W.W. County Level Socioeconomic Position, Work Organization and Depression Disorder: A Repeated Measures Cross-Classified Multi-Level Analysis of Low-Income Nursing Home Workers. *Health Place* **2006**, *12*, 688–700. [CrossRef] [PubMed]
33. Johnson, B.D. Cross-Classified Multi-Level Models: An Application to the Criminal Case Processing of Indicted Terrorists. *J. Quant. Criminol.* **2012**, *28*, 163–189. [CrossRef]
34. Alkahtani, K.F.; Abdel-Aty, M.; Lee, J. A Zonal Level Safety Investigation of Pedestrian Crashes in Riyadh, Saudi Arabia. *Int. J. Sustain. Transp.* **2018**, *13*, 1–13. [CrossRef]
35. Cai, Q.; Abdel-Aty, M.; Lee, J.; Huang, H. Integrating Macro and Micro-Level Safety Analyses: A Bayesian Approach Incorporating Spatial Interaction. *Transp. A* **2019**, *15*, 285–306. [CrossRef]
36. Montella, A.; Marzano, V.; Mauriello, F.; Vitillo, R.; Fasanelli, R.; Pernetti, M.; Galante, F. Development of Macro-Level Safety Performance Functions in the City of Naples. *Sustainability* **2019**, *11*, 1871. [CrossRef]
37. Abdel-Aty, M.; Lee, J.; Siddiqui, C.; Choi, K. Geographical Unit Based Analysis in the Context of Transportation Safety Planning. *Transp. Res. Part A Policy Pract.* **2013**, *49*, 62–75. [CrossRef]
38. Huang, H.L.; Song, B.; Xu, P.P.; Zeng, Q.; Lee, J.; Abdel-Aty, M. Macro and Micro Models for Zonal Crash Prediction with Application in Hot Zones Identification. *J. Transp. Geogr.* **2016**, *54*, 248–256. [CrossRef]
39. Maas, C.J.M.; Hox, J.J. The Influence of Violations of Assumptions on Multilevel Parameter Estimates and Their Standard Errors. *Comput. Stat. Data Anal.* **2004**, *46*, 427–440. [CrossRef]
40. Arpino, B.; Varriale, R. Assessing the Quality of Institutions' Rankings Obtained Through Multilevel Linear Regression Models. *J. Appl. Econ. Sci.* **2010**, *11*, 7–22.
41. Solaro, N.; Ferrari, P.A. Robustness of Parameter Estimation Procedures in Multilevel Models When Random Effects are MEP Distributed. *Stat. Methods Appl.* **2007**, *16*, 51–67. [CrossRef]
42. O'Donnell, C.; Connor, D. Predicting the Severity of Motor Vehicle Accident Injuries Using Models of Ordered Multiple Choice. *Accid. Anal. Prev.* **1996**, *28*, 739–753. [CrossRef]
43. Liu, X.-Y.; Wu, J.; Zhou, Z.-H. Exploratory Undersampling for Class-Imbalance Learning. *IEEE Trans. Syst. Man Cybern. B Cybern.* **2009**, *39*, 539–550. [CrossRef] [PubMed]
44. Chen, C.; Zhang, G.; Tarefder, R.; Ma, J.; Wei, H.; Guan, H. A Multinomial Logit Model-Bayesian Network Hybrid Approach for Driver Injury Severity Analyses in Rear-End Crashes. *Accid. Anal. Prev.* **2015**, *80*, 76–88. [CrossRef] [PubMed]
45. Simončič, M. Road Accidents in Slovenia Involving a Pedestrian, Cyclist or Motorcyclist and a Car. *Accid. Anal. Prev.* **2001**, *33*, 147–156. [CrossRef]

© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).