FACTORS THAT INFLUENCE THE LEVEL OF ACCIDENT SEVERITY IN VEHICLE CRASHES: A CASE STUDY OF ACCIDENTS ON KOREAN EXPRESSWAYS

PARK, Sunghee, Seoul National University, Seoul, Korea, sunny83@snu.ac.kr

PARK, Shin Hyoung, Seoul National University, Seoul, Korea, shpark76@snu.ac.kr

KIM, Dong-Kyu, Seoul National University, Seoul, Korea, kimdk95@snu.ac.kr

CHON, Kyung-Soo, Seoul National University, Seoul, Korea, chonks@snu.ac.kr

ABSTRACT

For the years from 2004 through 2008, 1,381 people were killed, and 5,925 people were injured due to vehicle crashes on expressways in Korea. While accidents on expressways make up only 1.7% of total vehicle crashes on all roads in Korea, the ratio of fatalities to crashes on expressways (113 people per 1,000 crashes) is more than three times higher than the ratio for all roads in Korea. This indicates that the severity of expressway crashes is relatively higher than that of crashes that occur on other types of roads. The goal of this study is to identify the most influential factors that determine the severity of accidents on Korean expressways. In this study, the factors that influence the level of accident severity were investigated using crash data from Korean expressways. These data are categorized into three levels of accident severity(level A/B/C), and an ordered probit model was used to determine the ordinal nature of the severity categories. Also, statistical tests were performed on the parameters based on robust standard errors to draw unbiased interpretations from the estimated parameters. Some of the factors that are expected to increase the level of accident severity on expressways include dozing off, speeding, tire failures, pedestrian violations, two-car accidents, cars hitting pedestrians, more than four cars involved in an accident, stopping or parking on the shoulder of the road, work-zone areas, and left curves that have a radius of more than 500 m. The results of this study will be helpful to transportation planners in understanding which risk factors contribute more to the severity of accidents on the expressways, contribute to better predictions of policy implications, and allow the recommendation and implementation of optimal countermeasures.

Keywords: vehicle crash, accident severity, influential factor, ordered probit model, robust standard error

INTRODUCTION

Road safety has been a national concern in Korea because vehicular accidents have been one of the leading causes of death for many years. To improve the state of public health, it is essential to reduce the number of traffic accidents and the fatality rate. Identifying the major factors that increase the level of accident severity and analyzing the way those factors affect the accidents are fundamental tasks that are needed to enhance road safety.

According to national statistics for the years from 2004 to 2008, 1,076,155 crashes were reported in Korea. In these crashes, 31,302 people were killed and 1,704,317 people were injured. This amounted to an average of slightly more than 17 fatalities every day for five years, three fatalities per 100 accidents, and an average of slightly more than 129 fatalities per million people of the total population of Korea for the last five years. In addition to those who died, an average of almost 934 people was injured every day due to traffic accidents. On average, 158 people were injured per 100 accidents, and an average of slightly more than 7,041 people were injured per million people in the country. Another way of looking at this is that, during the five-year period, approximately 0.7% of the entire Korean population was injured in vehicle crashes.

Crashes that occurred on Korea's expressways had a relatively higher fatality rate than did crashes that occurred on other types of roads. There were 13,704 accidents on expressways for the five-year study period, and 1,381 people died in those accidents. While only 1.7% of all accidents occurred on expressways, the fatalities that occurred on expressways amounted to 4.41% of the total fatalities, which is more than three times the percentage of the accidents that occurred on expressways. This indicates that traffic accidents on expressways are relatively more severe than traffic accidents that occur on other types of roads. Based on these findings, the goals of this paper are to identify the factors that influence the level of accident severity and to determine how these factors affect the severity of accidents.

Data collected by the Korea Expressway Corporation from the 13,704 accidents that occurred from 2004 through 2008 were used in this study. The Corporation classified the level of accident severity into three categories. To identify the factors that influence the level of accident severity by using these data, an ordered probit model was used, and the relative effects for each factor were calculated to identify how the factors affect the level of accident severity.

The remainder of this paper is organized as follows. In section 2, a review of the pertinent literature is provided. In section 3, the characteristics of the data used in this paper are described in detail. In section 4, the methodologies used, the ordered probit model, robust standard error, and relative effects are discussed. In section 5, the results achieved by applying the methodologies to the data are presented. In section 6, the importance of this research, its implications, and suggestions for future research are addressed.

LITERATURE REVIEW

Since evaluation of influential factors is essential for traffic safety, research to analyze the factors that influence the severity level of accidents and the associated deaths and injuries have long been a major topic in discussions of road safety. Previous research that analyzed the influential factors is presented below.

Jang et al. (2010) analyzed influential factors on level of injury in pedestrian accidents. Since the level of injury, which is the dependent variable, has both categorical and ordinal characteristics, an ordered probit model was used. The results of this study suggested that injuries were more severe at nighttime, on weekends, in rainy weather, when alcohol was involved, when large vehicles were involved, and when vehicles was proceeding straight. Since this research dealt mainly with pedestrian accidents, drivers were not exposed to the crashes, and, hence, the drivers' characteristics did not influence the severity of injuries significantly.

Singleton et al. (2004), using the available transport accident data, connected data about severely damaged vehicles with the data that related to severely injured people. By using these combined data, they were able to analyze the risk factors that increase the severity of the injuries sustained by the passengers in vehicles. The results of this study suggested that the significant risk factors were older drivers, female drivers, not wearing seat belts, ejection from the vehicle, the influence of alcohol, vehicle rollovers, vehicle fires, head-on collisions, collisions with fixed objects, whether the roadway was a federal or state roadway, and speed limits over 85 kph. Among these risk factors, ejection from the vehicle was the most influential factor because it was 6.5 times more likely to result in a high severity outcome for the accident. Not wearing a seat belt and fire were the next in line, increasing the probability of a severe outcome from an accident by 3.4% and 2.9%, respectively.

Abdel-Aty (2003) conducted research on the analysis of the severity of drivers' injuries at three locations: roadway sections, signalized intersections, and toll plazas. In this research, three methodical models, i.e., an ordered probit model, a multi-nomial logit model, and a nested logit model, were applied, and their performances were compared. The ordered probit model was chosen as the model of this research because it provided comparable accuracy and a simple application procedure. For all locations, older drivers, female drivers, not wearing a seat belt, speed ratio which is defined as the ratio of the estimated running speed at the time of a crash to the posted speed limit at the location of the crash, driver struck at their side, and passenger car were the influential factors that commonly increased the severity of drivers' injuries. It should be noted that, although alcohol use was not selected as one of the influential factors, alcohol use combined with not wearing a seat belt or E-Pass(electronic toll collection) use results in an increase in the probability of severe injury.

Kockelman and Kweon (2002) also analyzed risk factors associated with driver injury severity based on three different types of models, i.e., models of all crashes, two-vehicle crashes, and single-vehicle crashes. Since injury variables are ordinal, they used the ordered probit regression. In the model of all crashes, older drivers, older vehicles, and alcohol use were determined to be risk factors. Among all crash types, those involving roll-over of the vehicle were determined to be the most injurious. In the model of two-vehicle crashes, female drivers, older drivers, and nighttime driving were selected as risk factors. In the model of single-vehicle crashes, drivers' age close to 50 was determined to be the least injurious. Among vehicle types, pickups and SUVs worked differently, depending on who the injured person was. Both decreased the severity of injury to their drivers, but they increased the severity of injury to people in the other vehicle.

Renski et al. (1998) used the ordered probit model to investigate the effects of policy variables on injury severity. Highway segments for which the speed limits were increased by 10 mph resulted in a higher probability of increased severity than those that were increased by only 5 mph. However, there were no significant changes in injury severity for the highway segments for which the speed limits were increased from 65 to 70 mph.

Klop (1998) also used the ordered probit model to examine the impacts of physical and environmental factors on the severity of injury to bicyclists. Results from separate models estimated for urban and rural locations showed that straight grades, curved grades, darkness, and fog were significant factors that increase injury severity.

O'Donnell and Connor (1996) studied influential factors on the severity of traffic crash injuries. By using both of the ordered probit model and the ordered logit model, they found that increases in the age of the victim and vehicle speed led to slight increases in the probabilities of serious injury and death, and other factors, such as seating position, blood alcohol level, vehicle type, vehicle make, and type of collision were also significant.

The objective, method, and risk factors associated with the previous research reviewed above are summarized in Table 1.

Table 1 - Summary of previous research of influential factors on the severity of transportation accidents

Author	Objective	Method	Risk Factors
Jang et al. (2010)	influential factors on level of pedestrian injury	ordered probit model, marginal effect	· nighttime · weekend · rainy weather · alcohol use · large vehicle · proceeding straight
Singleton et al. (2004)	risk factors on injury severity	ordinal logistic regression with stepwise selection	· older driver · female driver · not wearing seat belt · ejection from vehicle · alcohol use · rollover

			· fire · head-on collision · collision with fixed object · federal/state road way · speed
Abdel-aty (2003)	influential factors on level of driver injury	ordered probit model, multi-nomial logit model, nested logit model	· female driver · older driver · not wearing seat belt · speed ratio · driver struck at her/his side · passenger car
Kockelman and Kweon (2002)	risk factors on driver injury severity	ordered probit model	· older driver · older vehicle · alcohol · rollover · female driver · nighttime
Renski et al. (1998)	the effects of policy variables on injury severity	ordered probit model	· speed limits
Klop (1998)	the impacts of physical and environmental factors on the severity of injury to bicyclists	ordered probit model	· straight grades · curved grades · darkness · fog
O'Donnell and Connor (1996)	influential factors on the severity of traffic crash injuries	ordered probit model/logit model	· age of victims · vehicle speed · seating position · blood alcohol level · vehicle type · vehicle make · type of collision

DATA DESCRIPTION

To analyze influential factors on level of accident severity in vehicle crashes, data from 13,704 accidents that occurred on Korean expressways over the period from 2004 through

2008 were used. These data were provided by the Korea Expressway Corporation, which operates all expressways in South Korea, except for expressways built with private capital. The accident data contain one dependent variable and 51 independent variables. The dependent variable is level of accident severity, which is categorized into three levels (A/B/C; Table 2). The accident severity levels are classified according to the number of casualties, the amount of facility damage, the number of vehicles involved in the crash, and the extent of traffic suspension as a result of the accident, as indicated in Table 2. The Korean Expressway Corporation established the three classification levels at the data collection stage.

Table 2 - Standards for classification of levels of accident severity

	Level A	Level B	Level C
Number of casualties	· More than 3 fatalities · More than 10 casualties · More than 20 injuries	· More than 1 fatality · More than 3 casualties · More than 5 injuries	· No fatalities · less than 2 casualties · less than 4 injuries
Amount of facility damage	· More than 10 million won (about USD 8,600)	· More than 2.5 million won (about USD 2,200)	· More than 300,000 won (about USD 260)
Number of related vehicles	· More than 10 vehicles · More than 5 vehicles with fatal accident	· More than 5 vehicles · More than 3 vehicles with injury accident	· More than 3 vehicles
Extent of traffic suspension	· Full suspension for both directions · More than 1 hour suspension of one direction for 4-lane expressway · More than 3 hours suspension of 1-lane for 4-lane expressway · More than 2 hours suspension of one direction for 2-lane expressway	· More than 30 minutes suspension of one direction for 4-lane expressway · More than 1 hour suspension of 1-lane for 4-lane expressway · More than 1 hour suspension of one direction for 2-lane expressway	· Suspension of one direction for 4-lane expressway · More than 30 minutes suspension of 1-lane for 4-lane expressway

Notation 1) 1 USD = 1,150 won

Dependent variables are classified into four groups of characteristics, i.e., driver characteristics, crash characteristics, environmental characteristics, and geometry characteristics. The detailed information of each variable is listed in Table 3.

Table 3(a) - Dependent variable and descriptive statistics

Variable Categories	Variables	Description	Number of crashes	Percentage

Dependent variable	Level of accident severity	A B C	121 1,925 11,658	0.88% 14.05% 85.07%

Table 3(b) - Driver characteristics variables and descriptive statistics

Variable Categories	Variables	Description	Number of crashes	Percentage
Driver characteristics	DAGE	Younger than 30 Older than 60 Between 20 and 60 Unknown	2,412 507 10,284 501	17.60% 3.70% 75.04% 3.66%
	DFAULT	Driver at fault Otherwise	11,600 2,104	84.65% 15.35%
	DSEX	Male Female Unknown	12,004 1,618 82	87.59% 11.81% 0.60%
	DUI	Driver has been drinking Otherwise	278 13,426	2.03% 97.97%

Table 3(c) - Environment characteristics variables and descriptive statistics

Variable Categories	Variables	Description	Number of crashes	Percentage
Environment characteristics	TIME	6:00 A.M. - Noon Noon – 6:00 P.M. 6:00 P.M. – Midnight Midnight – 6:00 A.M.	3,948 4,062 2,563 3,131	28.81% 29.64% 18.70% 22.85%
	ETIME (Emergency Service Arrival Time)	~10 min 11 ~ 20 min 21 ~ 30 min More than 30 min	7,463 4,385 1,648 208	54.46% 32.00% 12.03% 1.52%
	WEEKEND	Weekdays Weekends	9,348 4,356	68.21% 31.79%
	WEATHER	Clear Cloudy Rainy Misty Otherwise	7,865 2,556 2,758 106 419	57.39% 18.65% 20.13% 0.77% 3.06%
	RCONDITION (Road Condition)	Dry Wet Snowy Icy Otherwise	9,929 3,476 111 30 158	72.45% 25.36% 0.81% 0.22% 1.15%

Table 3(d) - Crash characteristics variables and descriptive statistics

Variable Categories	Variables	Description	Number of crashes	Percentage
Crash characteristics	PCF (Primary Crash Factor)	Dozing Off	3,230	23.57%
		Speeding	2,891	21.10%
		Visual Neglect	1,805	13.17%
		Excessive Steering	2,171	15.84%
		Tire Damage	1,038	7.57%
		Pedestrian Violation	79	0.58%
		Others	2,490	18.17%
	COLLISION PARTNER	Car vs. Car	2,210	16.13%
		Car vs. People	237	1.73%
		Car vs. Facility	10,754	78.47%
		Others	503	3.67%
	PARTIES (Number of Related Cars)	1	9,550	69.69%
		2	3,011	21.97%
		3	678	4.95%
		More than 4	465	3.39%
	VEHTYPE	Passenger car	6,988	50.99%
		Pickup	1,245	9.08%
		Truck	5,246	38.28%
		others	225	1.64%

Table 3(e) - Geometry characteristics variables and descriptive statistics

Variable Categories	Variables	Description	Number of crashes	Percentage
Geometry characteristics	RSHOULDER	Stopping or Parking on Shoulder	348	2.54%
		Otherwise	13,356	97.46%
	PCONDITION (Pavement Condition)	Good	13,615	99.35%
		Otherwise	89	0.65%
	RWORK	Work-zone	465	3.39%
		Otherwise	13,239	96.61%
	RCURVE (Radius of Horizontal Curve)	Right Curve Less Than 500 m	649	4.74%
		Right Curve More Than 500 m	1,756	12.81%
		Left Curve Less Than 500 m	488	3.56%
		Left Curve More Than 500 m	1,933	14.11%
		Straight	8,878	64.78%
	VGRADE (Vertical Grade)	Downhill	3,804	27.76%
		Uphill	2,910	21.23%
		Flat	6,990	51.01%

METHODOLOGY

Since the dependent variables that represent the level of accident severity are categorized into three levels in this study, it is inherently ordered and discrete. For this reason, an ordered probit model and an ordered logit model, which commonly are used to analyze ordinal and categorical data, were used. The only difference between the ordered logit model and the probit model is the assumption that is used for the distribution of error term. In the ordered probit model, it is assumed that the error term follows a standard normal distribution, while the error term of the ordered logit model follows a logistic distribution (e.g., Borooah, 2001). The logistic distribution resembles the normal distribution in shape, but it has heavier tails. As Greene (2000) pointed out, “It is difficult to justify the choice of one distribution over the other on theoretical grounds and, in most applications, it seems not to make much difference.” Thus, there are no noticeable differences between the two models.

It was assumed that the error term follows a standard normal distribution. So, the ordered probit model was chosen as the regression model for this study.

Ordered probit model

In this study, the STATA 10.0 statistical software package was used to estimate the ordered probit model and to compute logarithmic likelihood values. The model was formulated as shown below:

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

where I_p^* is an unobserved and unknown variable that measures the level of accident severity of the p^{th} accident; β is a vector of unknown parameters; X_p is a vector of measurable and observed variables that describe the driver, environment, crash, and geometry characteristics of the p^{th} accident; ε_p is an error term of the p^{th} accident that follows a standard normal distribution whose mean and variance are set to 0 and 1, respectively.

Even though I_p^* cannot be observed directly from any accidents, a discrete level of accident, I_p , can be observed directly from a given dataset:

$$I_P = \begin{cases} 1 \text{ if } -\infty < I_p^* \le \psi_1 & \text{accident level C} \\ 2 \text{ if } \psi_1 < I_p^* \le \psi_2 & \text{accident level B} \\ 3 \text{ if } \psi_2 < I_p^* \le \infty & \text{accident level A} \end{cases},$$

where the threshold values ψ_1 and ψ_2 ($\psi_1 , \psi_2 \ge 0$ and $\psi_1 \le \psi_2$) 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,

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

where $\Pr(I_p = i)$ represents the odds that the p^{th} pedestrian experiences i level of accident severity (i = 1, 2, 3); $\Phi(\,)$ is a standard normal cumulative distribution function of the error term.

The maximum likelihood estimation (MLE) was used to obtain estimators of parameters in the model. Since this log-likelihood function is a function of $\psi_1, \psi_2, \beta_0, \beta_1, \dots\dots, \beta_n$, it can be maximized subject to $\psi_1 \leq \psi_2$ (For more information, see Mckelvey and Zavoina (1975) and Jang et al., (2010).).

Robust standard errors

Since the formulation of the ordered model is constructed as a linear function, numerous assumptions, which linear regressions are required to fulfill, were considered in this study. The assumptions are on error term, and it requires zero mean, homoscedasticity, nonautocorrelation, uncorrelatedness of regressor and normality of error term. When these assumptions are not satisfied, remedial actions should be taken, and, in some cases, alternative modeling approaches should be used (See, e.g., Washington et al., 2003.).

The ordered probit model also makes some assumptions about the error term; therefore, it is necessary to test to determine whether the result will be the same in the condition for which the assumptions are not satisfied. To validate the estimates of the ordered probit model, robust standard errors are used. Robust standard errors are calculated based on the assumption that the requirements are not satisfied. If the results of usual standard errors and robust standard errors are similar, the parameter estimates of the ordered probit model can be validated.

Measure of fit

The extent of the fit of the estimated ordered probit model was assessed using adjusted likelihood ratio index statistics. In general, since the likelihood ratio index (ρ^2) can increase when additional variables are applied to the model, the adjusted likelihood ratio index ($\bar{\rho}^2$) is used to consider changes in the number of degrees of freedom.

In addition, the likelihood ratio test was conducted to complement the adjusted likelihood ratio index. This test was useful to assess two competing models (Washington et al., 2003). In this study, several tests were performed to compare the statistical significance between a variety of null models and alternative models, and the most appropriate model was selected.

Marginal effects

In the ordered probit model, since all variables are binary variables, which means the value of each variable is set to 0 or 1, the marginal effects are instantaneous rates of change calculated for a variable while holding all other variables constant (Jang et al., 2010). Therefore, the magnitude of the marginal effect relies on the values of the other variables and their coefficients. Given an independent variable, X_p with positive parameter, β , a 1-unit change of X_p will make the probability distribution move toward the right and produce an increased change in the probability of severe accident to the magnitude of marginal effect. The negative sign of β will cause reverse results. Consequently, the marginal effects are very useful for estimating the effectiveness to the accident severity of each variable.

RESULTS

Model selection

To select a model that has the most significant goodness-of-fit, the log of the likelihoods of alternative models that are the combinations of the variables' characteristics groups was calculated. By using log likelihoods, the adjusted likelihood ratio index ($\bar{\rho}^2$) was computed. The adjusted likelihood ratio index was, also, applied to the calculation of the likelihood ratio test at a 5% significance level for verification of the selected model. By comparing the likelihood ratio test statistics (Λ) and chi-square statistics for each pair of models, it can be confirmed whether the difference was statistically significant or not. After conducting the above procedures for each pair of models, the best performing model was selected.

The results of log likelihood and the adjusted likelihood ratio index analysis are summarized in Table 4. Also, the results of likelihood ratio test between alternative models are presented in Table 5.

For the log likelihood analysis, a total of 10 models were organized by possible combinations of each variable characteristics group. As presented in Table 4, compared with the model 0 with no independent variable, the log likelihoods of other models decreased. The result of comparing the log likelihoods of models with one characteristic (models 1, 2, 3, and 4) suggests that model 3, which includes only crash characteristics variables, shows the highest value.

In two characteristics analyses, the model with driver characteristics was combined with other characteristics. The result of comparing the log likelihoods of models with two characteristics (models 5, 6, and 7) suggests that model 6, which includes driver and crash characteristics variables, shows the highest value. The adjusted likelihood ratio index of model 6 was higher than that of model 3, thus model 6 outperforms model 3 in goodness of fit. But, the likelihood ratio statistics between models 3 and 6 (test 5) are not higher than the chi-square statistics. So, the result means that the difference between those two models is not statistically significant.

For the analysis of models with three characteristics, the driver and crash characteristics are combined with one of the other characteristics. The result of comparing the log likelihoods of models with three characteristics (models 8 and 9) suggests that model 9, which includes driver, crash, and geometry characteristics variables, shows the highest value. Compared with models 3 and 6, the adjusted likelihood ratio index of model 9 is higher. And the likelihood ratio test statistics between models 3 and 9 (test 7) are higher than the chi-square statistics. The test between models 6 and 9 (test 8) shows the same result. Thus, this result means that model 9 outperforms model 3 and model 6, and it is statistically significant.

Last, the result of comparing models 9 and model 10, which includes all four characteristics groups, suggests that the log likelihood of model 10 is higher than model 9 and the likelihood ratio test statistics between them (test 9) are higher than chi-square statistics as well. This result means that model 10 outperforms model 9 in goodness of fit and this result is statistically significant.

Since model 10 shows the highest log likelihood and the tests with other alternative models verify its significance, model 10 was selected as the best performing model in this study.

Table 4 - Log likelihood of alternative models

Classification	Selected Characteristics				Number of Independent variables	Log likelihood	Adjusted Likelihood Ratio Index ($\bar{\rho}^2$)
	Driver	Environment	Crash	Geometry			
Model 0	X	X	X	X	0	-6235.6461	-
Model 1	O	X	X	X	7	-6178.1678	0.0081
Model 2	X	O	X	X	15	-6160.4050	0.0097
Model 3	X	X	O	X	15	-5412.5036	0.1296
Model 4	X	X	X	O	9	-6045.7941	0.0290
Model 5	O	O	X	X	22	-6097.1206	0.0187
Model 6	O	X	O	X	22	-5409.6027	0.1289
Model 7	O	X	X	O	16	-5999.4818	0.0353
Model 8	O	O	O	X	37	-5382.2507	0.1309
Model 9	O	X	O	O	31	-5361.5988	0.1352
Model 10	O	O	O	O	46	-5333.9185	0.1372

Table 5 - Likelihood ratio test between alternative models

Test	Null model	Alternative model	Likelihood ratio test statistics	Chi-squared	Significance
Test 1	Model 1	Model 2	35.526	15.507	Significant
Test 2	Model 1	Model 3	1,531.328	15.507	Significant
Test 3	Model 1	Model 4	264.747	5.992	Significant

Test 4	Model 3	Model 5	-1,369.234	14.067	Not significant
Test 5	Model 3	Model 6	5.802	14.067	Not significant
Test 6	Model 3	Model 7	-1,173.956	3.842	Not significant
Test 7	Model 3	Model 9	101.810	26.296	Significant
Test 8	Model 6	Model 9	96.008	16.919	Significant
Test 9	Model 9	Model 10	151.368	36.415	Significant

Model estimates

Since the ordered probit model makes some assumptions about the error term, robust standard errors are used to validate the estimates of the ordered probit. The p-value is calculated by using robust standard errors to draw valid interpretation for each variable, within the 10% significance level.

The results of estimates of the ordered probit model and robust standard errors are provided in Table 6.

Table 6 – Ordered probit model estimates

Variable Categories		Variables	Coefficient	Standard Error	Robust standard error	p-value
Driver	DAGE	Younger than 30	-0.030	0.082	0.082	0.719
		Older than 60	0.032	0.102	0.103	0.752
		20 - 60	0.048	0.075	0.076	0.527
	DFAULT	Driver at fault	0.085	0.074	0.082	0.303
	DSEX	Unknown	-0.024	0.173	0.168	0.886
		Male	-0.058	0.051	0.051	0.258
	DUI	Alcohol use	-0.231	0.124	0.129	0.074
Environ ment	TIME	6:00 A.M. - Noon	-0.098	0.039	0.040	0.014
		Noon – 6:00 P.M.	-0.137	0.040	0.040	0.001
		6:00 P.M. –Midnight	-0.098	0.044	0.044	0.024
	ETIME	~10 mins	-0.153	0.043	0.043	0.000
		11 - 20 min	-0.060	0.045	0.046	0.185
		30 - min	0.056	0.110	0.114	0.626
	WEEKEND	WEEKEND	0.016	0.031	0.031	0.611
	WEATHER	Clear	-0.078	0.106	0.104	0.454
		Cloudy	-0.052	0.104	0.102	0.609
		Rainy	-0.223	0.089	0.089	0.012
		Misty	0.028	0.173	0.178	0.875
	RCONDITON	Dry	0.209	0.140	0.143	0.144
		Wet	0.173	0.146	0.153	0.259

		Snowy	0.143	0.214	0.212	0.500
		Icy	0.149	0.321	0.357	0.675
Crash	PCF	Dozing Off	0.185	0.053	0.054	0.001
		Speeding	0.139	0.059	0.060	0.021
		Visual Neglect	-0.072	0.056	0.056	0.200
		Excessive Steering	-0.157	0.062	0.063	0.013
		Tire Damage	0.131	0.078	0.086	0.128
		Pedestrian Violation	0.352	0.164	0.139	0.011
	COLLISION PARTNER	Car vs. Car	0.190	0.080	0.087	0.029
		Car vs. People	0.965	0.115	0.111	0.000
		Car vs. Facility	-0.584	0.074	0.082	0.000
	PARTIES	1	-0.149	0.064	0.065	0.022
		2	-0.057	0.059	0.059	0.328
		More than 4	0.217	0.079	0.081	0.007
	VEHTYPE	Passenger car	-0.484	0.095	0.103	0.000
		Pick up	-0.055	0.101	0.109	0.616
		Truck	-0.185	0.094	0.102	0.072
Geometry	RSHOULDER	Stopping or Parking on Shoulder	0.164	0.072	0.073	0.025
	PCONDTION	Good	-0.238	0.159	0.152	0.119
	RWORK	Work-zone involvement	0.542	0.061	0.062	0.000
	HCURVE	Right Curve Less Than 500 m	-0.023	0.109	0.106	0.829
		Right Curve More Than 500 m	0.066	0.091	0.089	0.459
		Left Curve More Than 500 m	0.179	0.090	0.087	0.040
		Straight	0.121	0.084	0.082	0.138
	VGRADE	Downhill	-0.001	0.040	0.040	0.986
		Flat	0.022	0.037	0.037	0.553

As shown in Table 6, the standard error (SE) of the ordered probit model and robust standard error (RSE) are similar. So, this result means that the estimates of the ordered probit are validated, and the result is statistically unbiased. Through the analysis, among 51 variables, only 20 variables were identified as significant.

The result of analyzing p-value suggests that, first, except for alcohol use, driver characteristics show a statistically significant influence on the level of accident severity. Second, among environmental characteristics, time, emergency service arrival time (ETIME), and rainy weather have significant influences. Third, most of the crash characteristics factors show significant influence. They are dozing off, speeding, excessive steering, and pedestrian violation among the primary crash factors (PCF), collision partners, the number of related vehicles (PARTIES), and the type of vehicle. Last, stopping or parking on the shoulder, work-zone involvement, and radius of left curve more than 500 m among horizontal curve factors (HCURVE) were analyzed as significantly influential factors. Among the characteristics groups, the crash characteristics group had the most significant factors, as shown previously in Table 4.

In Table 7, variables satisfying the 10% significance level, which can be defined as the influential factors on level of accident severity and the thresholds of the ordered probit model, are presented.

Table 7 – Ordered probit model estimates at 10% significance level

Variable Categories		Variables	Coefficient	Standard Error	Robust standard error	p-value
Driver	DUI	Alcohol use	-0.231	0.124	0.129	0.074
Envirionment	TIME	6:00 A.M. - Noon	-0.098	0.039	0.040	0.014
		Noon - 6:00 P.M.	-0.137	0.040	0.040	0.001
		6:00 P.M. –Midnight	-0.098	0.044	0.044	0.024
	ETIME	Within 10 min	-0.153	0.043	0.043	0.000
	WEATHER	Rainy	-0.223	0.089	0.089	0.012
Crash	PCF	Dozing Off	0.185	0.053	0.054	0.001
		Speeding	0.139	0.059	0.060	0.021
		Excessive Steering	-0.157	0.062	0.063	0.013
		Pedestrian Violation	0.352	0.164	0.139	0.011
	COLLISION PARTNER	Car vs. car	0.190	0.080	0.087	0.029
		Car vs. people	0.965	0.115	0.111	0.000
		Car vs. facility	-0.584	0.074	0.082	0.000
	PARTIES	1	-0.149	0.064	0.065	0.022
		More than 4	0.217	0.079	0.081	0.007
	VEHTYPE	Passenger car	-0.484	0.095	0.103	0.000
		Truck	-0.185	0.094	0.102	0.072
Geometry	RSHOULDER	Parking or stopping on shoulder	0.164	0.072	0.073	0.025
	RWORK	Work zone involvement	0.542	0.061	0.062	0.000
	HCURVE	Left curve more than 500 m	0.179	0.090	0.087	0.040
Severity			coefficient		Standard error	
ψ_1 (between level B and C)			0.2742011		0.3093528	
ψ_2 (between level A and B)			1.854624		0.3110536	

Marginal effects

The marginal effects of each influential factor are presented in table 8.

Table 8 – Marginal effects

Variable Categories		Variables	level C	level B	level A
Driver	DUI	Alcohol use	0.040304	-0.03877	-0.00154
Environment	TIME	6:00 A.M. - Noon	0.019079	-0.01825	-0.00083
		Noon – 6:00 P.M.	0.026497	-0.02536	-0.00114
		6:00 P.M. - Midnight	0.018967	-0.01816	-0.00081
	ETIME	within 10 min	0.030844	-0.02944	-0.0014
	WEATHER	Rainy	0.041144	-0.03946	-0.00168
Crash	PCF	Dozing Off	-0.03914	0.037231	0.00191
		Speeding	-0.0292	0.027803	0.0014
		Excessive Steering	0.029407	-0.02819	-0.00122
		Pedestrian Violation	-0.08505	0.079892	0.00516
	COLLISION PARTNER	car vs. car	-0.04099	0.038942	0.002051
		car vs. people	-0.29357	0.261005	0.03257
		car vs. facility	0.140066	-0.13125	-0.00882
	PARTIES	1	0.030767	-0.02932	-0.00145
		more than 4	-0.04872	0.046126	0.00259
	VEHTYPE	Passenger car	0.097624	-0.09295	-0.00467
		Truck	0.036005	-0.03443	-0.00157
Geometry	RSHOULDER	parking or stopping on shoulder	-0.03574	0.033921	0.001821
	RWORK	work-zone involvement	-0.14115	0.131096	0.010053
	HCURVE	left curve more than 500 m	-0.03852	0.0366	0.001925

It should be noted that this paper does not focus on the frequency of accidents but on the level of accident severity. The marginal effects are to interpret the influences of variables on the possibility of traffic accidents at a certain level of accident severity. For example, if a marginal effect of specific factor at a certain accident severity level has a positive sign, it means the corresponding factor increases the risk of an accident at such level.

On the contrary, if a marginal effect of a specific variable at a certain level of accident severity has a negative sign, the variable decreases the risk of an accident at the corresponding level. How the influential factors influence the possibility of the occurrence of a traffic accident can be suggested by interpreting the marginal effects. And the results are described below.

Driver Characteristics

Among driver characteristics, only alcohol use was identified as an influential factor. Unlike the results of previous studies, this factor was analyzed since it decreases the risk of a severe accident. This result would have been shown because of the characteristic of the road. Since expressways in Korea are not the usual roads to use often in daily life, the number of drivers who have been drinking would be relatively small compared to other types of roads. In spite of our estimates though, this factor should be handled carefully.

Factors such as driver age, gender, and driver at fault were not identified as influential factors in this study.

Environment Characteristics

Time of day, emergency service arrival time, and rainy weather were identified as influential factors. The time of day, from 6:00 A.M. to midnight, is influential in that the risk of a severe accident is decreased. Through this result, it can be suggested that the time between midnight and 6:00 A.M is influential in increasing the risk of a severe accident. Among the time of day variables, the risk of severe accident was the lowest during the period of time from noon to 6:00 P.M.

An accident to which emergency service arrives within 10 minutes has a decreased risk of being a severe accident when compared with an accident to which emergency service arrived after 10 minutes. Since the number of casualties is one of the levels of the accident classification standards, if the emergency service arrives faster, the possibility of surviving increases and it will end up being a relatively minor accident. Among weather factors, only rainy weather was identified as an influential factor. When an accident occurred in rainy weather, the risk of the accident being severe tends to be lower.

Crash Characteristics

Crash characteristics show the most significant influence on accident severity. Among the influential factors in table 8, dozing off, speeding, pedestrian violation, car versus car collision, car versus people collision, and more than four related vehicles were analyzed as the factors that increase the risk of a severe accident. On the other hand, excessive steering, car versus facility collision, one related vehicle, passenger car and truck were the factors that decreased the risk of a severe accident. More specifically, compared with passenger car involved crashes, truck involved crashes are approximately three times more prone to have the risk of a severe accident.

Geometry Characteristics

The influential factors were analyzed as parking or stopping on the shoulder, work-zone involvement, and a radius of a left curve of more than 500 m. They all tend to increase the risk of a severe accident and decrease the risk of a minor accident.

CONCLUSIONS

Identifying factors that increase or decrease the risk of accident severity is one of the fundamental tasks to enhance road safety. So, this study analyzed the vehicle crashes that occurred on Korean expressways from 2004 through 2008 to identify influential factors on level of accident severity and how these factors affect accident severity.

The independent variables were classified into four characteristics groups, which are driver, environment, crash, and geometry characteristics. And level of accident severity was used as the dependent variable. Since level of accident severity was based on categorized data, the ordered probit model was applied to identify the influential factors. Using robust standard errors, this result was confirmed to be valid and significant. Also, to determine how the factors affect the level of accident severity, marginal effects were adopted.

As a result of the analysis, influential factors that increase the risk of severe accident were identified as follows: (1) dozing off, speeding and pedestrian violation among primary crash factors, car vs. car collision and car vs. people collision among collision partners, and more than four related vehicles in crash characteristics; (2) parking or stopping on shoulder, work-zone involvement, and the radius of left curve more than 500 m in geometry characteristics.
Meanwhile, influential factors that decrease the risk of a severe accident were identified as follows: (1) time between 6:00 A.M. and midnight, emergency service arrival time within 10 minutes, and rainy weather among environmental characteristics; (2) excessive steering, car vs. facility collision, one related vehicle, and passenger car and truck among crash characteristics;

Most driver characteristics were not influential on the level of accident severity. And all of the influential factors in environmental characteristics tend to decrease the risk of a severe accident, while all influential factors in geometry characteristics tend to increase the risk of a severe accident. Crash characteristics were analyzed as the most influential factors.

By identifying the influential factors on accident severity and how those factors affect accident severity, some political advice can be proposed. Since the crash characteristics factors are physical and practical conditions of the accidents, they are the most applicable for improving road safety. Therefore, the enforcement of regulations against dozing off, speeding, and pedestrian violation on expressways can reduce the risk of severe accidents. And comparing with collision between car and car or between car and people, the collision between car and facility decreases the risk of severe accidents, thus, the installation of protection facilities, such as median strips and guard rails, can be considered as some of the countermeasures.

Among geometry characteristics, since the vehicles parking or stopping on the shoulder affect road safety, this peril should be regulated. When an accident occurs in a work zone, the risk of a severe accident also increases considerably. So, the time of work on expressways should be carefully determined and positioning sufficient warning signs as vehicles approach the work-zone can be suggested as effective measures for decreasing accident severity.

For further research, an investigation of the extent to which alcohol use causes accidents on Korean expressways should be conducted. Since the effect of alcohol use suggested in this study is not the same as the results of previous research, additional research is needed to clarify the effect of alcohol on accident severity.

ACKNOWLEDGEMENTS

This research was supported by the Korea Science and Engineering Foundation (KOSEF) grant funded by the Korea government (MEST) (No. 2009-0075811).

REFERENCES

Abdel-Aty, M. (2003). Analysis of driver injury severity levels at multiple locations using ordered probit models. Journal of Safety Research, 34, 597-603.

Borooah V. K. (2001). Logit and probit: Ordered and multinomial models. Sage Publications, London.

Greene, W. H. (2003). Econometric analysis. 5th edition. Prentice Hall, New Jersey.

Jang, K. T., S. H. Park, S. B. Chung and K. H. Song (2010). Influential factors on level of injury in Pedestrian Crashes: Applications of ordered probit model with robust standard errors. 89th Annual Meeting of the Transportation Research Board, Washington, D.C.

Klop, J. (1998). Factors Influencing Bicycle Crash Severity on Two-Lane Undivided Roadways in North Carolina. 78th Annual Meeting of the Transportation Research Board, Washington, D.C.

Kockelman K. M. and Y. J. Kweon (2002). Driver injury severity: An application of ordered probit models, Accident Analysis & Prevention, Vol. 34, Issue 3, 313-321.

McKelvey, R. D. and W. Zavoina (1975). A statistical model for the analysis of ordinal level dependent variables, Journal of Mathematical Sociology, 4, 103-120.

O'Donnell, C. and Connor, D. (1996) Predicting the Severity of Motor Vehicle Accident Injuries Using Models of Ordered Multiple Choices. Accident Analysis and Prevention, 28(6), 739– 753.

Renski, H., A. Khattak and F. Council (1998). Impact of Speed Limit Increases on Crash Injury Severity: Analysis of Single-Vehicle Crashes on North Carolina Interstate

Highways. 78th Annual Meeting of the Transportation Research Board, Washington, D.C.

Singleton, M., H. Qin and J. Luan (2004). Factors associated with higher level of injury severity in occupants of motor vehicles that were severely damaged in traffic crashes in Kentucky, 2000-2001. Traffic Injury Prevention, 5, 144-150.

Washington, S. P., M. G. Karlaftis and F. L. Mannering (2003). Statistical and econometric methods for transportation data analysis. Chapman & Hall/CRC, Boca Raton, Florida