DEVELOPMENT OF A SAFETY PERFORMANCE FUNCTION FOR KOREAN EXPRESSWAYS

KWAK, Hochan, Seoul National University, Seoul, Korea, colony01@snu.ac.kr

KIM, Dong-Kyu, Seoul National University, Seoul, Korea, klimdk95@snu.ac.kr

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

KHO, Seung-Young, Seoul National University, Seoul, Korea, sykho@snu.ac.kr

ABSTRACT

Traffic accidents have long been a major issue, especially in developed countries. Accident prediction models, also known as safety performance functions, have recently received substantial attention for their use in estimating the expected number of accidents on road sections in order to determine priority sites for safety improvements. The purpose of this paper is to develop a safety performance function for Korean expressways. Goodness-of-fit tests are performed to specify the model structure and to select variables included in the model and, based on the test results, the safety performance function for Korean expressways is formulated as a negative binomial model. The model developed in this paper shows that traffic accidents on Korean expressways may occur most often on sections of road where traffic flows can increase discontinuously, such as those with a small number of lanes near on-ramp interchanges or tollgates and on sections of mainstream expressways with a higher percentage of truck traffic. Cumulative scaled residuals are plotted to evaluate the fitness of the proposed model. The results of this paper can be used as a tool in the safety evaluation of expressways in Korea in order to choose areas for implementation of accident countermeasures.

Keywords: safety performance function, expressways, priority site, negative binomial model, traffic accident, cumulative scaled residual

INTRODUCTION

Safety performance functions (SPFs), which are also referred as accident prediction models, are models that estimate the expected number of traffic accidents per unit of time using independent variables such as traffic flow rates and geometric design features (Zhong et al., 2009). SPFs play an important role in road construction and improvement projects. For road construction, SPFs provide guidelines for determining horizontal and vertical alignments and design speeds of each road section. For road improvement, SPFs can help prioritize sites for road safety improvements when resources are limited.

The budget for road construction and improvement projects in Korea decreased about 15% in 2010, from 9 trillion Won (about USD7.81) in 2003 to 7.8 trillion Won (about USD6.77) in 2010. Therefore, an objective and rational method is urgently needed to determine prioritize the many pending projects. Currently, the order of priority for road safety improvement projects in Korea is simply determined by the number of traffic accidents. This method is limited in that the roads with the higher number of traffic accidents tend to get first priority irrespective of their causes, even though not the road itself but the human factors, such as drinking, speeding, and other violations, may be to blame for most of the accidents. SPFs can provide transportation engineers and planners with the most frequent causes of accidents on a section of road and their influence on the number of accidents on that road in order to provide a more reasonable rationale for prioritization.

Although many studies have been conducted on developing SPFs for a variety of roads, research on SPFs for expressways has not been the focus. This is mainly due to the difficulty of identifying the relationships between geometric features and traffic accidents on expressways. Road sections on expressways usually have little outstanding drawbacks to their geometric features because of their strict design standards. Furthermore, the frequency of traffic accidents is not often enough large for the statistical analysis required by SPFs.

Recently, several studies on SPF for Korean expressways have been conducted (Kang and Lee, 2002; Kang et al., 2002; Park et al., 2007; Yoon, 2007; Han et al. 2008). These studies estimated their models with the assumption that the frequency of traffic accidents is usually distributed in the Poisson or negative binomial distribution. They estimated the parameters in their models using least square estimation (LSE) or maximum likelihood estimation (MLE). Exposure variables and geometric variables such as horizontal curvature and super elevation were used only as independent variables.

One of the most important studies on SPFs for expressways was performed by Zhong et al. (2009), who considered environmental variables (inbound or outbound direction, city or rural area, and interchange area) and traffic variables (percent of truck traffic, speed difference between cars and trucks, standard deviation of truck speed, and standard deviation of car speed), along with the more common exposure (e.g., AADT and the length of sections) and geometric variables. Comparing the output criteria of four models (negative binomial, Poisson, zero-inflated Poisson (ZIP), and zero-inflated negative binomial (ZINB)), Zhong et al. (2009) selected the negative binomial model for the final model.

The purpose of this paper is to develop the SPFs for Korean expressways with geometric and environmental variables and to examine the influential factors on traffic accidents using the method adopted by Zhong et al. (2009). Using 1,459 crashes from 2006 to 2008, we performed goodness-of-fit tests to determine the structure of the model. We selected the variables in the model using the backward method, and used cumulative scaled residuals (CURE) to validate the selected model.

The paper is organized as follows. In section 2, we develop SPFs by determining the model structure and candidate independent variables. Section 3 presents the results of the parameter estimation for several types of regression models. The model proposed in this paper is validated using CURE in section 4. In section 5, we summarize and discuss the results and suggest avenues for future research.

DEVELOPMENT OF SPFS

Data Description and Analysis Unit

Among Korean Expressways, 1,459 accident data of the Gyeongbu Expressway were drawn from Korea Expressway Corporation from 2006 to 2008. The Gyeongbu Expressway has the longest extension, 416.05km, in Korea and extends northwest to southeast through the center of the country from Seoul, the capital of Korea, to Busan, the second biggest city in Korea. The accident data on interchange and within service area are excluded in order to focus on the accidents on main line. Each accident record contains information as follow: occupants' demographics, alignment and geometric element, environmental condition, and accident information such as accident location, time, level, type, cause and severity. Figure 1 shows the Gyeongbu Expressway in a map of South Korea.

Figure 1 – Gyeongbu Expressway

In order to develop safety performance functions, a road section has to be set up as an analysis unit. Generally two methods can be used for this task: a fixed-length method that divides a road into sections of the same length, and a variable-length method that divides a road into road sections based on the attributes of independent variables (Zhong, 2009). In this study, the variable-length method is applied in order to use the homogeneity of independent variables. According to the changes of AADT and horizontal curvature, which are usually known as the most significant variables in the frequency of traffic accidents, the Gyeongbu Expressway is divided into 524 sections on the northbound side and 522 sections on the southbound side. The average lengths of the sections are 0.799km for the northbound side and 0.802km for the southbound side.

Modeling

This study uses the generalized linear regression method, which has been adopted in many recent studies on SPFs. This method assumes that: (1) accidents that occur on a certain road are independent and stochastic events and (2) accidents follow a specific probability distribution. In other words, traffic accidents that occur on a certain road can be assumed to fit statistically into a specific probability distribution. Poisson and negative binomial (NB) distributions have been widely accepted for the development of crash-prediction models (Miou et al., 1993; Maher et al., 1996; Ivan et al., 1997; Caliendo et al., 2007; Zhong et al., 2009). However, we must also be able to account for zero events, since traffic accidents are regarded as rare. Therefore, ZIP and ZINB are also used in order to deal with the "excess" zeroes that are frequently observed in accident count data (Lord et al., 2005).

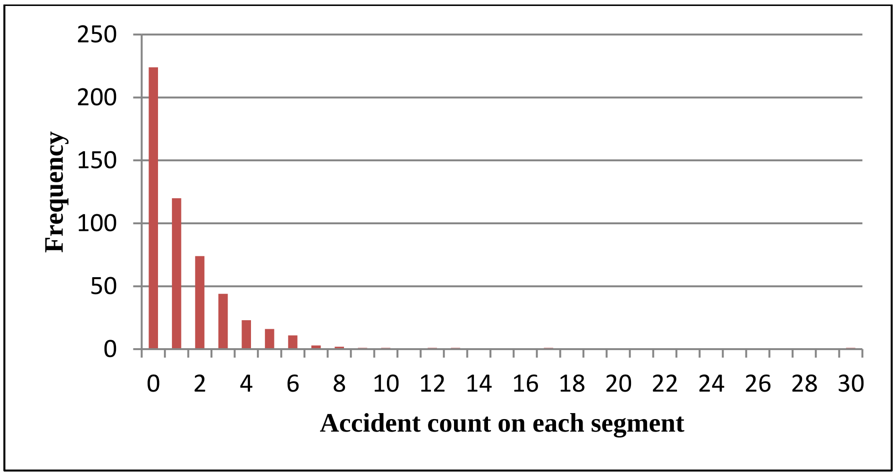

Figure 2 – Distribution of accident count data

As shown in figure 2, the traffic accident count data has a distribution that is similar to a negative binomial distribution. The result of the Kolmogorov-Smirnov test also shows that the count of accidents follows NB distribution under 95% confidence levels. NB distribution, Poisson, ZIP, and ZINB will be used in establishing SPFs as well. The equation for the model is:

$$\gamma_i = \mathrm{EXPO} \cdot \mathrm{Exp}\left(\beta_0 + \sum_{j=1}^{n} \beta_j x_{ij}\right)$$

where
γ_i is the predicted annual accident frequency at road segment i
EXPO is the exposure variable (unit=millions of vehicle-kilometers)

β_i is the model parameter
x_{ij} are explanatory variables at road section i.

Selection of Variables

In order to determine the variables that are significant in predicting to expressway accidents, we used the variables selected in existing studies (Kang et al., 2002; Shankar et al., 1995; Zhong et al., 2009) as well as environmental variables.

Exposure variable

Since the length of the road section and the traffic volume (AADT) are the most closely related to the frequency of traffic accidents, an exposure variable is introduced in the form suggested by Zhong et al. (2009):

$$\text{EXPO} = \frac{\text{AADT} \times 365 \times (\text{number of years in time period}) \times \text{L}}{10^6}$$

where
L is the length of the road section

Alignment variables

Four variables are used to represent alignment attributes: radius of horizontal curve (Hor_radius), grade of vertical curve (Ver_grade), uphill road (Ver_up), and downhill road (Ver_down). Radius of horizontal curve doesn't discriminate between left curve and right curve, and grade of vertical curve has absolute values because uphill and downhill are dummy variables that specify the sign of the slopes.

Environmental variables

In addition to the exposure variable and the alignment variables, we examine environmental variables such as interchange (IC), number of lanes, toll gate (TG), bridge, tunnel, speed camera and service area. Among them, the influence of intersection, which is considered in many studies, is analyzed separately for on-ramp and off-ramp.

In Zhong et al. (2009), since the patterns of crash occurrences are remarkably different depending on the direction (inbound or outbound of downtown) and the area (urban or rural), these factors are considered as environmental variables. However, for the analysis of the Korean expressway, we don't take direction and area into account because there is no basis on which to assume that there are different patterns for direction and area.

Traffic variable

Compared to other types of vehicles, trucks generally move at lower speeds on roadways, and they have a higher impact in a collision with vehicles or roadway facilities. Therefore, the percentage of truck traffic has a considerable effect on crash occurrence. While information on the speed difference between cars and trucks would be useful in examining the effect of trucks in truck-involved collisions, we could not find or even estimate the difference in speed from the current crash data.

The 14 independent variables, consisting of 1 exposure variable, 4 alignment variables, 8 environmental variables and 1 traffic variable, are listed in table 1.

Table 1 – Independent Variables

	parameter	explanation	Illumination
Exposure	EXPO	Exposure	Unit :10^6year-veh-km
Alignment	Hor_radius	Radius of horizontal Curve	Unit : m
	Ver_grade	Grade of vertical Curve	Unit : %
	Ver_up	Uphill road	0 : no, 1 : yes
	Ver_down	Downhill road	0 : no, 1 : yes
Environment	IC_out	On-ramp	0 : no, 1 : yes
	IC_in	Off-ramp	0 : no, 1 : yes
	Lane	Number of lane	
	TG	Toll gate	0 : no, 1 : yes
	Bridge	Bridge	0 : no, 1 : yes
	Tunnel	Tunnel	0 : no, 1 : yes
	Camera	Speed camera	0 : no, 1 : yes
	Service_area	Service area	0 : no, 1 : yes
Traffic	Truck%	Percent of truck	

MODEL ESTIMATION

In this study, model estimation is conducted using the backward method, which rejects the least relevant variables of all the independent variables each time the estimation is made. The estimation is complete when the remaining variables are judged significant according to the p-value of each variable. Table 2 shows the values of parameters estimated from NB, Poisson, ZINB and ZIP.

Table 2 – Results of Estimated Regression Models

parameter	NB	Poisson	ZINB	ZIP
Hor_radius	-	-	-	-
Ver_angle	-	-	-	-
Ver_up	-	-	-	-
Ver_down	-	-	-	-
IC_out	-	-	-	-

IC_in	0.2854692	0.2978242	0.2854692	0.3203059
Lane	-0.234709	-0.2387801	-0.234709	-0.2270864
TG	0.7286191	0.4622066	0.7286196	0.4637691
Bridge	-	-0.0652569	-	-0.0892199
Tunnel	-	-	-	
Camera	-	-	-	
Service_area	-	-	-	
Truck%	2.070565	1.98223	2.070565	2.173431
Constant	-4.212821	-4.105277	-4.212821	-4.014005

The logarithm likelihood ratio test, Akaike's Information Criterion (AIC), Bayesian Information Criterion (BIC), the Vuong statistic and chibar2 are applied to evaluate the goodness-of-fit of these models. AIC and BIC are the most commonly used model selection criteria. AIC can be justified as Bayesian using a "sawy" prior on models that is a function of sample size and the number of model parameters. Furthermore, BIC can be derived as a non-Bayesian result (Burnham et al., 2004). Smaller values of these imply better models. From these, AIC is indicated as below (Akaike, 1977).

$$\mathrm{AIC} = (-2)\ln\mathrm{L} + 2\mathrm{n}$$

where
L is the maximum likelihood
n is the number of parameters

Vuong statistics are for testing the null hypothesis that the competing models are equally close to the true data against the alternative hypothesis that one model is closer than the other model (Vuong, 1989). In this study, Vuong statistics are used for relative comparison between zero-inflated models (ZINP, ZIP) and common models (NP, Poisson). The Criterion results for the 4 models are compared in table 3.

Table 3 – Criterion Results of 4 Regression Models

Models	Log likelihood	AIC	BIC	Vuong	Chibar2
NB	-2512.4776	5036.955	5073.263		63.11
Poisson	-2537.8438	5087.688	5123.996		
ZINB	-2512.478	5038.955	5081.314	-0.06	
ZIP	-2524.272	5062.544	5104.903	2.21	

The Chibar2 rejects the null hypothesis that there is no overdispersion at 95% confidence interval, which means that NB can improve goodness-of-fit better than Poisson does. Comparing ZIP to Poisson through the Vuong test, we found that ZIP is better than Poisson. On the other hand, whether NB is better than ZINB or vice versa cannot be determined because the result of the Vuong test for ZINB is -0.06, which is located between -1.96 and +1.96. However, the result that both the AIC and the BIC of NB are smaller than those of ZINB suggests that NB is better than ZINB; consequently, it is confirmed that the NB is the best model in terms of data fit among the 4 models. Based on this outcome, the best SPF is developed as:

$$\gamma_i = \text{EXPO} \cdot \text{Exp}(-4.212821 + 0.2854692 \cdot \text{IC}_{\text{in}} - 0.234709 \cdot \text{Lane} \\ +0.7286191 \cdot \text{TG} + 2.070565 \cdot \text{Truck\%})$$

In the safety performance function, the coefficients of each variable explain that: (1) as for interchanges, on-ramp tends to increase the probability of crashes, but off-ramp has no noticeable effect in the analyses; (2) as the number of lanes increases, the crash frequency diminishes because the number of lanes have a negative sign as its coefficient; and (3) if there is a toll gate on a road section or if the percentage of trucks in AADT rises, the crash count increases on the section. What's interesting is that the radius of the horizontal curve and vertical alignment were not selected as significant variable for SPF. That is, the elements of alignment have little or no influence on expressway crashes. This result from the highest design criteria applied to expressways.

To compare the influence of the selected variables, elasticity analysis for each independent variable was done. Table 4 lists the result of elasticity analysis.

Table 4 – Result of Elasticity Analysis

Variable	IC_in	Lane	TG	Truck%
Elasticity	0.248339	-0.733629	0.517425	0.737535

As a result, this analysis indicates that the most influential variable is truck% and the least influential one is IC_in.

An example directly supports this result. Location number 103 on northbound side of the expressway is a straight, level line, that is, the radius of the horizontal curve is 0 and the vertical grade is also close to 0 (-0.36%). AADT is 17,566 vehicles per day. The conditions, such as the low traffic volume and high-quality alignment, mean that crashes rarely. However, a lot of crashes (9 crashes) occurred at this section because the percentage of trucks involved in AADT on this 2-lane road section is very high (44.23%). Another section, number 519, is a curve-shaped downhill section whose vertical grade is more than 3% (-4.4%), and the radius of the horizontal curve is 8026.7m. AADT is 80,797 vehicles per day, and there is an off-ramp located in this section. It would be reasonable to infer that the crash count will be high on this section, but since truck% is low (9.16%) and an on-ramp is not in this 4-lane section, the section saw only 2 crashes. This finding confirms that influential variables in crash occurrence are not so much the geometry of the roadway as the truck traffic.

MODEL VALIDATION

CURE (cumulative scaled residuals) has been widely used to check the model fit. A CURE plot that fluctuates around 0 is considered a good result because the residual is the basic element by which to judge fit. A vertical jump in the CURE plot indicates an outlier; a good plot should run within a reasonable range (Hauer, 2004). The CURE equation is:

$$\mathrm{CURE} = \sum_{i:x_{i \leq l}} \frac{y_i - \hat{y}_i}{\sqrt{\hat{y}_i + K(\hat{y}_i)^2}}$$

where
CURE is the value of Cumulative scaled Residuals
K is the overdispersion parameter
y_i is the observed accident count for road section i
$\hat{y}_i$ is the estimated accident count for road section i
l is the range of x_i.

CURE is calculated and plotted for each variable adopted in SPF. Figure 3 shows a CURE plot for exposure (EXPO), which varies between -11.9 and 17.1. Since the threshold values, which are approximately derived from the square root of the number of data points used for this study, are ±38.2, the plot resides in the relevant range. Since the entire plot varies around 0, we can conclude that the estimated model fit is acceptable.

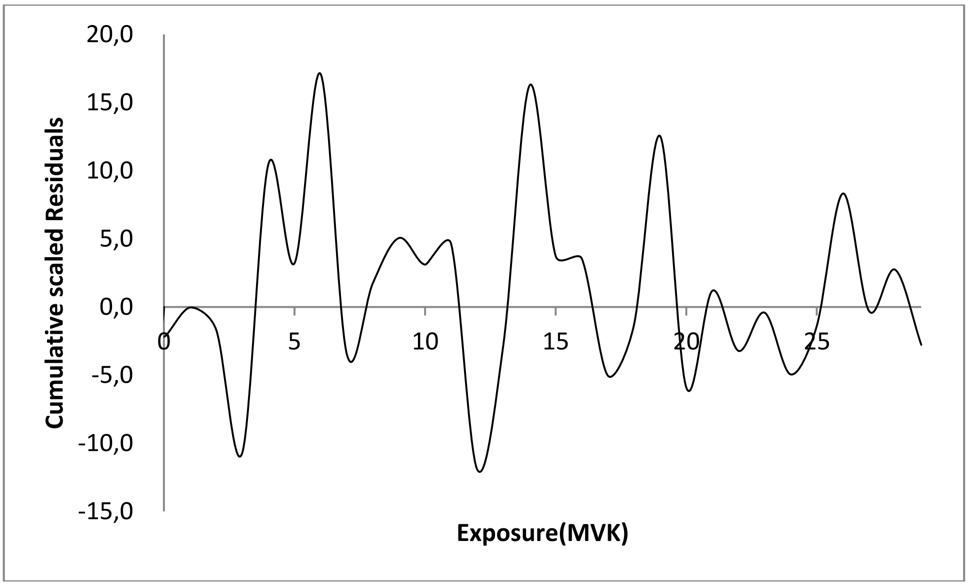

Figure 3 – Cumulative scaled Residual versus EXPO

CONCLUSIONS AND RECOMMENDATIONS

The purpose of this paper is to develop SPF for a Korean expressway. Fourteen independent variables consisting of 4 alignment variables, 8 environmental variables, 1 exposure variable and 1 traffic variable were examined, and 5 variables—EXPO, IC_in, lane, TG, and Truck%—were selected for the final model. Goodness-of-fit tests were performed on the criterion outputs of 4 models, and the NB model was selected to develop the SPF for Korean expressways.

The analysis showed, first, that the alignment variables of expressways do not have much of an effect on the number of traffic accidents. This result can be explained by the high-quality design standards for expressways and the latest steady safety improvement projects in Korea.

Second, the analysis showed that facilities in the mainstream of expressways, such as interchanges and tollgates, significantly affect the number of traffic accidents and that the possibility of traffic accidents increases at the on-ramps, while it does not at off-ramps. These results suggest that traffic accidents are related to the discontinuous increase of traffic flows and unstable traffic conditions that are due to acceleration/deceleration and lane-changing.

Third, the percent of truck traffic has the largest effect on the number of traffic accidents. Because trucks tend to move more slowly than other types of vehicles, an increase in the percent of trucks can cause greater speed variance in the mainstream of traffic, making it more unstable. Crashes may also be affected by the sight obstacles that trucks present and lane-changing that is due to trucks.

Based on these results, some suggestions for enhancing safety performance on Korean expressways follow. First, the road sections where traffic flows can increase discontinuously, such as on-ramps and tollgates, should be designed strictly and minutely. The installment of real-time monitoring systems using traffic information systems and accident management systems could be useful near these sections. In the case of tollgates, especially, traffic flow collisions (by weaving and occurrence of speed difference) between the vehicles using high-pass system which automatically collects tolls from electronic devices and common toll system which is directly collected by person may have occurred because high-pass system was partly installed on inner lanes of tollgates. Therefore high-pass system is necessary to be installed on entire lanes as soon as possible to reduce the vehicle crashes by collisions. Separating trucks from other vehicles can also reduce the number of traffic accidents and, in the long term, the installment of exclusive truck lanes may be suggested to accompany road-widening projects.

In order to generalize the observations and conclusions in this study, additional analyses on other expressways is required. Accurate and detailed data related to geometric features can make the model and the results in this study more useful and meaningful.

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. A. and A. E. Radwan (2000). Modeling traffic accident occurrence and involvement. Accident Analysis and Prevention, 32(5), 633-642.

Akaike, H. (1978). A Bayesian Analysis of the Minimum AIC Procedure. Annals of the Institute of Statistical Mathematics, 30(A), 9-14.

Burnham, K. P. and D. R. Anderson (2004). Multimodel Inference: Understanding AIC and BIC in Model Selection. Sociological Methods Research, 33, 261-304.

Caliendo, C., M. Guida and A. Parisi (2007). A crash-prediction model for multilane roads. Accident Analysis & Prevention, 39(4), 657-670.

Hadi, M. A., J. Aruldhas, L. Chow and J. A. Wattleworth (1995). Estimating safety effects of cross-section design for various highway types using Negative Binomial regression. Transportation Research Record: Journal of the Transportation Research Board, 1500, 169-177.

Hamilton, L. C. (2006). Statistics with STATA. Brooks/Cole, USA.

Han, S., K. Kim and S. Oh (2008). What goes problematic in the Existing Accident Prediction Models and How to Make it Better. Journal of Korean Society of Road Engineers, 10(1), 19-29.

Hauer, E (2004). Statistical Road Safety Modeling. Transportation Research Record: Journal of the Transportation Research Board, 1987, 81-87.

Ivan, J. N. and P. J. O'Mara (1997). Prediction of Traffic Accident Rates Using Poisson Regression. Presented at 76th Annual Meeting of the Transportation Research Board, Washington D. C.

Kang, J. and S. Lee (2002). Traffic Accident Prediction Model by freeway Geometric Types. Journal of Korean Society of Transportation, 20(4), 163-175.

Kang, M., C. Doh and B. Son (2002). Fitting Distribution of Accident Frequency of Freeway Horizontal Curve Sections & Development of Negative Binomial Regression Models. Journal of Korean Society of Transportation, 20(7), 197-204.

Lord, D., S. P. Washington and J. N. Ivan (2005). Poisson, Poisson-gamma and zero-inflated regression models of motor vehicle crashes: balancing statistical fit and theory. Accident Analysis & Prevention, 37(1), 35-46.

Maher, M. and I. Summersgill (1996). A Comprehensive Methodology for the Fitting of Predictive Accident Models. Accident Analysis and Prevention, 8(3), 281-289.

Miaou, S. and H. Lum (1993). Modeling Vehicle Accidents and Highway Geometric Design Relationships. Accident Analysis and Prevention, 25(6), 689-709.

Pardillo Mayora, J., R. Bojorquez Manzo and A. Camarero Orive (2006). Refinement of Accident Prediction Models for Spanish National Network. Transportation Research Record: Journal of the Transportation Research Board, 1950, 65-72.

Park, H., B. Son and H, Kim (2007). Development of Accident Prediction Models for Freeway Interchange Ramps. Journal of Korean Society of Transportation, 25(3), 123-135.

Shankar, V., F. Mannering and W. Barfield (1995). Effect of Roadway Geometrics and Environmental Factors on Rural Freeway Accident Frequencies. Accident Analysis and Prevention, 27(3), 371-389.

Vuong, Q. H. (1989). Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses. Econometrica, 57(2), 307-333.

Yoon, B. (2007). Development of Traffic Accident Forecasting Models for Trumpet IC Ramp Types of Freeway. Journal of Korean Society of Civil Engineers, 27(1D), 81-87.

Zhong, L. D., X. D. Sun, Y. L. He, X. M. Zhong and Y. S. Chen (2009). Safety Performance Function for Freeway in China. Presented at 88th Annual Meeting of the Transportation Research Board, Washington D. C.