Analysis of Design Elements for Urban Highway Safety

Lee, SooBeom Ph.D.
Associate Professor, Korea
Dept. of Transportation Engr.
University of Seoul
mendota@uos.ac.kr

ABSTRACT

Traffic safety depends on humans, vehicles, and highway conditions. These factors influence traffic safety separately or in combination. It is then necessary to clarify the relationship between the traffic accidents and the geometric conditions and road appurtenances to prevent accidents and to provide a comfortable and safe driving environment. This study developed a traffic accident frequency prediction model based on the geometric and highway appurtenance condition(s) on urban highways. The data were collected on various type(s) of highway in Seoul and the Kyunggi area. The concept of homogeneous highway segments was adopted in the classification of the highways to analyze the accident characteristics. A homogeneous highway segment is defined by the homogeneity of the influencing factors such as median, lane width, number of lanes, and horizontal/vertical alignment In the first stage, the urban highway was classified into three types: four-lane or less, six-lane, and eight-lane or over highways. The model was developed based on the Poisson regression theory. The common influencing factors were traffic volume, length of segment, and the number of mid-segment entrances/exits. The vertical and horizontal alignments were the additional factors on the four-lane or less highways. The influencing factors on six-lane highways were the illegal on-street parking, the lighting conditions, and the lane width. The land use type and the lane width influence traffic safety on eight-lane highways. The developed model in this study is helpful for the design of safe highways. Additionally, it will contribute to identifying the potentially hazardous locations for highways already in service and contribute to the treatment of safety improvements.

INTRODUCTION

The physical and mental pain of traffic accident victims and their families and the amount of property damage has became an important part of the national agenda in many countries, so it is necessary to continue to maximize our efforts to reduce the number of traffic accidents. It is essential to identify and analyze the accident inducing factors and come up with accident reduction policies.

This study attempts to develop an accident frequency prediction model. The major role of the developed model is to judge the safety level of highway segments by predicting the accident potential. To improve the safety of the study area, the relationship between the explaining variables and the accident frequency was analyzed.

The developed model is for urban highways due to the fact that the accident and highway characteristics of urban and rural areas are different in nature. The highways are categorized into three groups, i.e., 4-lane and under, 6-lane, and 8-lane and over highway groups. The Poisson regression model, Negative Binomial regression model, and Zero Inflated Poisson regression model were used to develop the model

LITERATURE REVIEW

The existing urban street accident frequency prediction models are summarized in Table 1 below.

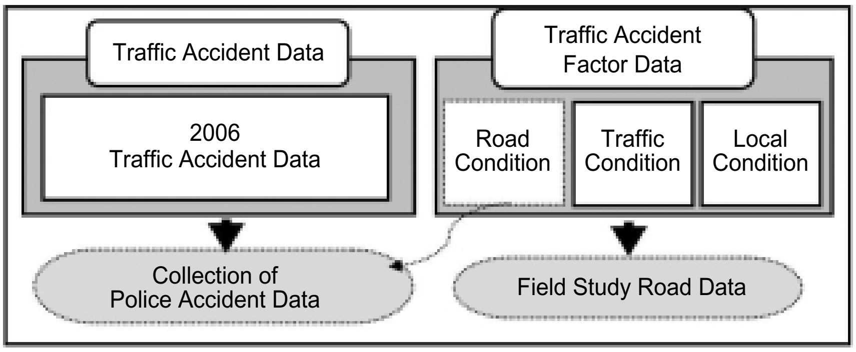
<Table 1> Urban Street Accident Frequency Prediction Model

Hadi Model	$C_{u,2} = 0.25ADT^{0.913}(1000L)^{0.9930}e^{B_{u,2}}$ with $B_{u,2} = -11.415-0.0489(W_1+W_{p8})-0.0201V+0.056N_i-0.0342W_{u8}$ where $C_{u,2}$ = frequency of injury-causing crashes on undivided two-lane urban streets, crashes/yr W_1 = lane width, ft; W_{p8} = paved shoulder width, ft; V = speed limit, mph; N_i = number of intersections; and W_{u8} = unpaved shoulder width, ft
	$C_{u,4} = 0.25ADT^{0.8317}(1000L)^{0.8831}e^{B_{u,4}}$ $B_{u,4} = -9.584-0.1037W_1-0.0150V+0.0395N_i-0.3318I_{curb}$ where $C_{u,4}$ = frequency of injury-causing crashes on undivided four-lane urban streets, crashes/yr; and I_{curb} = presence of curb (1 if curb present, 0 otherwise)
Hauer Model	$C = C_{off\text{-}road} + C_{on\text{-}road}$ where $C_{off\text{-}road}$ = frequency of severe off-road crashes on undivided streets, crashes/yr; $C_{on\text{-}road}$ = frequency of severe on-road crashes on undivided streets, crashes/yr
Hauer Model	$C_{off\text{-}road} = 0.36\left(\frac{ADT}{10{,}000}\right)^{0.815} Le^{B+C}$ with, $B = -0.069\frac{ADT}{10{,}000} + 0.148I_{curb} + (1.0-I_{curb})\ln(0.76+0.083SWC)-0.184I_{twltl}$ and, $C = 0.230(1-I_{park})+\frac{320.9}{R} + 0.252I_{V\leq 30} + 0.381I_{V\leq 45}$ where SWC = shoulder width category (1 if 0 to 1 ft, 2 if 2 to 3 ft, 3 if 4 to 6 ft, 4 if 7 to 9 ft, 5 if 10 to 11 ft, 6 if over 11 ft); R = curve radius, ft 1n(X) = natural log of x I_{twltl} = presence of two-way left-turn lane (1 if lane is present, 0 otherwise); I_{park} = presence of on-street parking (1 if parking is present, 0 otherwise); $I_{V\leq 30}$ = low speed limit (1 if speed limit is 30 mph or less, 0 otherwise); and $I_{V\leq 45}$ = high speed limit (1 if speed limit is 45 mph or more, 0 otherwise)
Hauer Model	$C_{om\text{-}road} = 0.90L\left[\left(\frac{ADT}{10{,}000}\right)^{1.830} Ae^{B} + 0.0205D_{d,b}\right]$, with, $B = -0.000\frac{ADT}{10{,}000} + 0.1148I_{curb} + (1.0-I_{curb})\ln(0.96+0.040SWC)-0.229I_{twltl}$ and, $A = \frac{1}{150}(2e^{-0.059P_t} + 0.017P_t)\left(e^{-2298/R} + \frac{343.8}{R}\right)(V^{2.066}e^{-0.0689V})$ where $D_{d,b}$ = density of business or commercial driveways (two-way total), driveways/mi; P_t = percent trucks represented in ADT, %;

Bonneson and McCoy Model	$C^*_R = ADT^{0.910}(5280L)^{0.852} e^{B_R} (1.0-0.01PDO)$, with $B_R = -15.162 - 0.296I_{b/0} - 0.596(1.0-I_{b/0})+0.00478(D_d + D_{u/a})I_{b/0} + 0.0255PDO$ where C^*_R = frequency of severe, mid-signal crashes on urban streets with raised-curb medians, crashes/yr; $I_{b/0}$ = business or office land use (1 if business or office, 0 otherwise); D_d = driveway density (two-way total), driveways/mi; $D_{u/a}$ = unsignalized public street approach density (two-way total), approaches/mi; PDO = property-damage-only crashes as a percentage of total crashes (=68percent)
Bonneson and McCoy Model	$C^*_T = ADT^{0.910}(5280L)^{0.852} e^{B_T} (1.0-0.01PDO)$, with $B_T = -15.162 - 0.018I_{b/o} - 0.093(1.0-I_{b/o})+0.00478(D_d + D_{u/a})I_{b/o} + 0.0255PDO$ where C^*_T = frequency of severe, mid-signal crashes on urban streets with a TWLTL, crashes/yr. $C^*_{U,b/o} = ADT^{0.910}(5280L)^{0.852} e^{B_{U,b/o}} (1.0-0.01PDO)$, with $B_{U,b/o} = -15.162+0.570I_{park}+0.00478(D_d + D_{u/a}) + 0.0255PDO$ where $C^*_{U,b/o}$ = frequency of severe, mid-signal crashes on undivided urban streets serving a business or office land use, crashes/yr
	$C^*_T = ADT^{0.910}(5280L)^{0.852} e^{B_T} (1.0-0.01PDO)$, with $B_{U,r/i} = -25.666+0.570I_{park}+0.0255PDO$ where $C^*_{U,r/i}$ = frequency of severe, mid-signal crashes on undivided urban streets serving residential or industrial land use, crashes/yr.

DATA DESCRIPTION

Traffic accident data can be classified into two groups. One is the historical traffic accident data, and the other is the traffic accident inducing factors data. The vehicle-to-vehicle historical accident data are extracted from the police accident database. Data for the accident inducing factors are classified into three groups - road conditions, traffic conditions, and the local environments. They are from field studies and existing reports.



<Figure 1> Collection of Data

Field studies were conducted on 13 segments of urban highway which showed typical urban highway characteristics; this included highways in the metro Seoul area. The highway segments analyzed measured 77 km.

<Table 2> Location of Highway Segment Data Collection

Research Section (Total Segment Length)	
Gyeonggi Sungnamsi	Sujeongro (3km), Beonyungro (4km)
Gyeonggi Goyangsi	Ilsanro (4km), Kangseokro (3km)
Gyeonggi Gwangmyeongsi	Oriro (5km)
Gyeonggi Gurisi	Kyungchunro (3km), Mt.Ahcha Street (4km)
Seoul Dongdaemun	Jeonnongro (3km), Jegiro (3km), Hongneungro (2km)
Seoul Eunpyeong	Tongilro (5km)
Seoul Seocho	Unjuro (5km), Nambu Loop (33km)

Twenty-five variables were selected and the database was constructed through field studies. The dependent variable was the traffic accident frequency, and the independent variables were geometric highway elements including traffic volume.

<Table 3> Geometric & Traffic Condition DB for the Survey Result

Variable	Description of Variable	Variable	Description of Variable
Frequency	the number of traffic accidents	Speed Control	the number of speed control systems/ devices
AADT	annual average daily traffic volume [veh/day]	Terrain	terrain [level=1, rolling=2, mountainous=3]
Length	highway segment length [m]	Delineation	the presence of delineation systems [Yes=1, No=2]
H.C	the presence of horizontal curves [m]	Md_Type	the median type [None=0, Concrete=1, Guardrail=2, greenbelt=3, Others=4]
V.C	the presence of vertical curves [%]	Md_Width	the median width [m]
Radius	horizontal curvature [m]	Speed limit	posted speed [kph]
G	vertical grades [%]	Crosswalk	the number of crosswalks
Driveway	the number of driveways	Bus Stop	the number of bus stops
Lighting	the number of lights	On Street Parking	presence and type of on-street parking [Absence=0, Presence=1]
Terrain	terrain type [level, rolling, mountainous]	Ex_Buslane	exclusive bus lanes (Roadside)
T_width	traveled lane average Width [m]	Num_Bus	number of bus service routes
Sh_Width	the shoulder width [m]	Ill_Parking	presence/absence of illegal parking [Absence=0, Presence=1]
Sh_Type	the shoulder type [Non=0, Pavement=1, Non Pavement=2, Others=3]	Land-Use	area-use around highway segments [residential=1, commercial=2, farmland=3, industrial=4, residential-commercial=5, farmland-industrial=6, residential-farmland=7, industrial-farmland=8, others=9]
Lane	the number of lanes		

Conversion: 1km/h = 0.621mi/h ; 1m = 3.28ft

Homogeneous Highway Segments

Identifying homogeneous highway segments was necessary in order to analyze the parts of the continuous highway. To divide the continuous highway into the homogeneous highway segments, the Korean highway design regulations and existing reports were consulted.

1) Regulations governing standards of road structure and facilities

Regulations state that the standardized sections are the highway segments in which the same design standards should be applied based on the highway location, local environments and expected traffic volume.

> Article 9 (Design Segment)
>
> The highway segment in which the same design standards should be applied represents the segment between the major intersection (including interchange) and the major highway facilities
>
> The design speed differencial with the adjacent highway segments should be less than 20km/h

<Table 4> Highway Segment Length

Type	Design Standard Length	Exception Case
Freeway, Rural Arterial	30~20km	5km
Rural Highway	15~10km	2km
Urban Highway	Gap between major intersections	

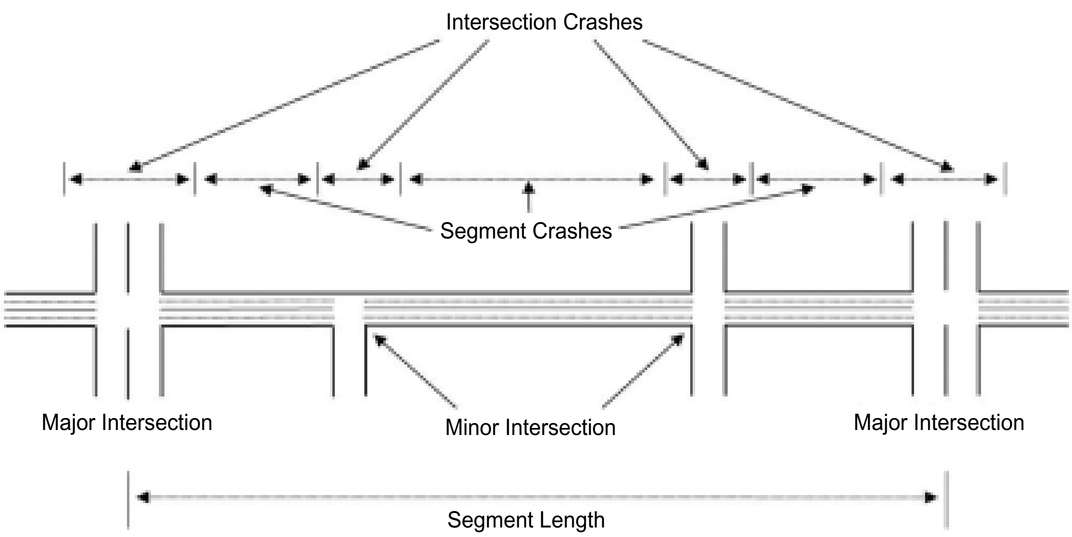
The regulations recommend designing the highway consistently as much as possible to avoid drivers' confusion due to the design change in very short or unusual segments of the highway.

2) Homogeneous Highway Segments in the Highway Safety Manual

According to the Highway Safety Manual (HSM), a homogeneous highway segment is defined as a highway segment which is divided by major intersections, or a change of the horizontal alignment and the geometric elements, or a major change of land use. A highway is considered to be homogeneous if it is not divided into segments and if there is no change in ADT, lane width, shoulder width, presence of median, median width, side slope, and major intersection.

In the HSM, traffic accidents in the highway segment include only those accidents which are not related to the intersections, and the intersection accidents include only those accidents which occurred within 76 m of the center of the intersection.

<Figure 2> Homogeneous Highway Segment in the HSM



3) Homogeneous Highway Segment in the IHSDM

According to the Design Consistency Module in the IHSDM, the homogeneity of the segment is judged by the following 4 elements in which statistically significant factors are related to the number of traffic accidents:

- expected speed drivers would reduce compared to the speed in the upstream straight or curved segments
- average horizontal curvature of all segments
- average vertical curvature of highway segment
- average horizontal curvature of highway segment

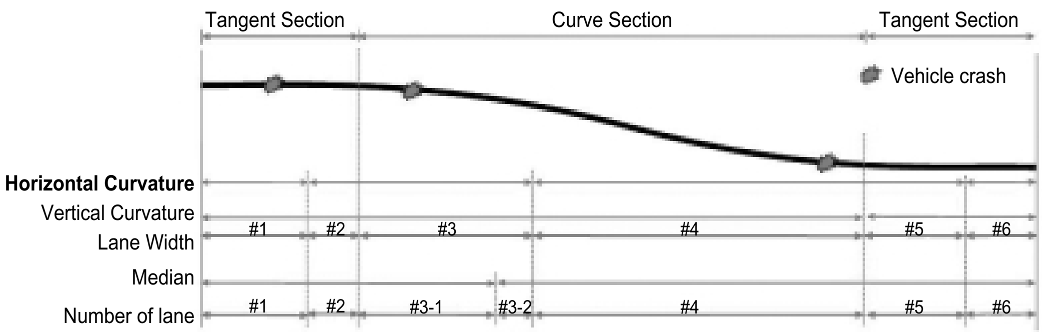
<Table 5> Literature Review for Standard of Homogenous Highway Segment

Type		Factor of Homogenous Highway Segment
HSM	Two Lane	ADT, Lane Width, Shoulder Width (Type), Driveways/mile, Ratio of Risk
	Multi-Lane	ADT, Lane Width, Presence of Median (Width), Shoulder Width, Side Slope
IHSDM	MINNESOTA	• rural two-lane, two-way, paved road • 17 feet < surface width ≤ 24 feet • left and right shoulder width differing by 2 feet of less • average of left and right shoulder width ≤ 12 feet • segment length > 0.1 mile • segment present if both time periods with characteristics unchanged • 5-year average daily traffic (ADT) > 5 vehicles • 5-year average daily commercial traffic (ADT) > 5 vehicles
	WASHINGTON	• rural two-lane, two-way, paved road • 17 feet < surface width ≤ 24 feet • left and right shoulder width differing by 2 feet of less • average of left and right shoulder width ≤ 12 feet • segment length > 0.1 mile • segment present if both time periods with characteristics unchanged • 2-year average daily traffic (ADT) > 5 vehicles • no vertical curves of zero length with grade changes of 1% or more • no horizontal curves of zero length with angular changes of 1° or more

4) Homogeneous highway segment in this study

In this study, the highway segment is considered to be homogeneous as far as none of the following design elements - horizontal curvature, vertical curvature, and lane width - is changing.

< Figure 3> Shows the Homogenous Highway Segment in the study



MODEL

The background information and model formulations concerning the three target models are briefly described below.

Poisson Regression Model

Poisson distribution is usually appropriate to describe the distribution of count data, and Poisson regression is widely used in the prediction of accident likelihood or frequency. In the study, let Y_i be the random variable that represents the accident frequency in the segment i, and y_i is a realization of Y_i. The mean of Y_i, denoted by $\wedge_i$, is also a random variable. For the case of $\wedge_i$. λ_i Y_i is a Poisson distribution with parameter λ_i. That is, the probability of Y_i is shown as follows:

$$P(Y_i = y_i \mid \wedge_i) = \frac{\lambda_i^{y_i} e^{-\lambda_i}}{y_i!} \qquad (1)$$

The expected value of Y_i is $E(Y_i) = \lambda_i$, and the variance of Y_i is $V(Y_i) = \lambda_i$. In the present study, λ_i is always a positive value since it represents the average number of accidents of a specific highway segment *i*. In general, λ_i is defined as an exponential function of the explanatory variables at highway segment *i* as follows:

$$\lambda_i = e^{\sum_j \beta_j x_{ij}} \tag{2}$$

where β_j = The model parameter of explanatory variable x_j

Moreover, the method of maximum likelihood estimation is the most widely used approach to calibrate the parameters in the log-likelihood function, and rho-square (ρ^2) is the statistic of model performance.

NB Regression Model

When the expected value of Y_i is not equal to the variance of Y_i, Poisson regression is not a good choice, and a negative binomial becomes a popular alternative. For the case of $\wedge_i$. $\lambda_i + \varepsilon_i$, and let $\wedge_i$ be subjected to a gamma distribution with shape parameter μ_i and scale parameter κ/μ_i; then, Y_i is described with a negative binomial distribution, with an expected value $E(Y_i) = \mu_i$, and a variance $V(Y_i) = \mu_i(1+\alpha\mu)$, where $\alpha = 1/\kappa$. The gamma distribution of $\wedge_i$ is written as follows:

$$P(\mu_i = \lambda_i) = \frac{(\kappa/\mu_i)\lambda_i^{\kappa-1}e^{-(\kappa/\mu_i)\lambda_i}}{\Gamma(\kappa)} \tag{3}$$

In general, μ_i is always a positive value, and it is defined as an exponential function of the explanatory variables at highway segment *i* as follows:

$$\mu_i = e^{\sum_j \beta_j x_{ij}} \tag{4}$$

where β_j is the model parameter of explanatory variable x_j.

Further, the method of maximum likelihood estimation is the most widely used approach to calibrate the parameters in the log-likelihood function, and rho-square (ρ^2) is the statistic of model performance.

When comparing the expected value and variance of Y_i in Poisson regression to those in the negative binomial regression model, it is evident that they are essentially the same when α is equal to zero. However, Poisson regression is obviously not acceptable, when the calibrated α is significantly different from zero. Negative binomial regression is more appropriate for the case of over-dispersion, i.e. the variance of the random variable Y_i is greater than its expected value. In addition, the examination of the observed distribution of accident risk levels in terms of accident likelihood (frequency) and accident impact (severity) may be helpful in model selection.

Zero-Inflated Poisson (ZIP) Model

As stated above, Poisson and NB models are frequently applied to address count data problems. However, when the analyzed data consist of a significant number of zero events, the data is usually found to be over-dispersed and it may cause excess variation in the Poisson models, i.e., extra-Poisson variation. In dealing with the count data problem with extra zeros, we can apply compound probability models, e.g., zero-inflated count data models. The zero inflated model combines a binary variable c_i with a standard count variable y_i* (with support over the nonnegative integers) such that the observed count y_i:

$$y_i = \begin{cases} 0, \text{if } c_i = 1 \\ y_i^*, \text{ if } c_i = 0 \end{cases} \tag{5}$$

If the probability that is $c_i = 1$ is denoted by p_i, the probability function of y_i can be written compactly as:

$$f(y_i) = p_i d_i + (1-p_i)g(y_i),\ y_i = 0, 1, 2, ... \tag{6}$$

where $d_i = 1-\min\{y_i, 1\}$ and $g(y_i)$ is a regular count data probability function such as the Poisson or the negative binomial probability function.

Oral Presentations

In discussing the Zero-Inflated Poisson (ZIP) model, Mullahy(1986) first proposed the ZIP model with constant p_i. Lambert (1992) extended it by specifying a logit model form in order to capture the influence of covariates on the probability of extra zeros:

$$P_i = \frac{\exp(G_i\gamma)}{1+\exp(G_i\gamma)} \tag{7}$$

Accordingly, the probability density function of a ZIP model can be expressed as follows:

$$Y_i = 0,\ \text{with probability : } P(Y=0) = p_i + (1-p_i)\, e^{-\lambda_i}$$
$$Y_i = y_i^*,\ \text{with probability : } P(Y=y_i) = (1-p_i)\frac{\lambda_i^{y_i} e^{-\lambda_i}}{y_i!} \tag{8}$$

where λ_i as similarly defined in equation (2), is an exponential function of the explanatory variables at Highway Segment i, and the covariates that formulate the mean of accident frequency (λ_i) in a Poisson regression model could be the same as or different from those of explaining the probability of extra zeros (p_i) in a logistic model.

The mean and variance of a ZIP model is expressed as follows:

$$E(y_i) = (1-p_i)\lambda_i \tag{9}$$

$$Var(y_i) = \lambda_i(1-p_i)(1+\lambda_i p_i) \tag{10}$$

Note that when p_i is equal to zero, then the mean of a ZIP model is same as that of a Poisson model, and the ZIP model is essentially the same as a Poisson model. It can be further verified by the variance to mean ratio shown in Equation 11 that a ZIP model is suitable to capture over-dispersed data in view of the fact that its variance is generally greater than its mean value. The ratio $(\frac{p}{1-p})$ in Equation 11 plays a similar role as the dispersion factor α in a NB model and it is employed to capture the over-dispersion characteristic of the analyzed data.

$$\frac{V(y)}{E(y)} = 1+(1-p)\lambda = 1+(\frac{p}{1-p})\,E(y) \tag{11}$$

MODEL DEVELOPMENT

Model Form

The mathematical form used for any accident prediction model must yield logical result, meaning that it must not lead to the prediction of a number of accidents and must ensure a prediction of zero accident frequency for zero values of the exposure variables.

$$E(\wedge) = a_0 \times L^{a_1} \times V^{a_2} \times \exp \sum_{j=1}^{m} b_j x_j$$

where $E(\wedge)$ = Predicted Accident Frequency L=Segment Length; V=Segment AADT; x_j = Any of m variables additional to L and V; and a_0, a_1, a_2, and b_j =Model Parameters

Validation of Model

1) ρ^2 (Rho-Squared)

Maximum likelihood ratio ρ^2 is used to test the fitness of the model developed in this study. ρ^2 is the number between 0 and 1, and the model shows the higher fitness as ρ^2 is close to 1.

In the relationship between the traffic accidents and the highway geometrics and facilities, the value of ρ^2 between 0.2 and 0.4, however, represents rather excellent fitness (McFadden, 1978)

$$\rho^2 = 1 - \frac{L(\beta)}{L(0)} \quad (0 \le \rho^2 \le 1)$$

where, $L(\beta)$: Log likelihood function

$L(0)$: Restricted log likelihood

2) MPB (Mean Prediction Bias)

MPB is used to judge the skewness of the results of the model due to the dependent variables. As long as the value is small, the value predicted by the model tends to be correct.

$$MAB = \frac{\sum_{j=1}^{n}(Y_j - \hat{Y}_j)}{n} \quad (12)$$

where, Y_j : Actually Dependent Variable

$\hat{Y}_j$: Predicted Dependent Variable for Model

3) MAB (Mean Absolute Bias)

MAB indicates how much error is contained on average in the predicted value of the model. MAB is different from MPB in which the values are not canceled out by the negative and positive numbers. As long as the value is close to 0, the predicted value is close to the observed value.

$$MPB = \frac{\sum_{j=1}^{n}\left| Y_j - \hat{Y}_j \right|}{n} \quad (13)$$

where, Y_j : Actually Dependent Variable

$\hat{Y}_j$: Predicted Dependent Variable for Model

Selection of the Optimum Traffic Accident Frequency Prediction Model

To select the optimum model, a Vuong statistic is used. If $F_1(y_i / x_i)$ and $F_2(y_i / x_i)$ are opposing predicted probability distributions, the Vuong statistic is computed as follows:

$$m_i = \log\left[\frac{F_1(y_i/x_i)}{F_2(y_i/x_i)}\right] \quad (14)$$

$$V = \frac{\sqrt{n}\left[\frac{1}{n}\sum_{i=1}^{n} m_i\right]}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(m_i - \bar{m})^2}} \quad (15)$$

where, V= Vuong Statistics, $\bar{m}$ =Mean, n=Number of Sample

$F_1(y_i/x_i)$ represents the data better if the Vuong statistic is larger than 1.96 with 95% confidence, while $F_2(y_i/x_i)$ represents the data better if the Vuong statistic is smaller than -1.96 with 95% confidence.

<Table 6> Decision rule for model selection under ZINB-NB comparisons using the Vuong statistic and overdispersion parameter criteria

		NB Overdispersion parameter a (t-statistic)	
		<2	>2
Vuong Statistic for ZINB-NB comparison	< |1.96|	ZIP or Poisson	NB
	> 1.96	ZIP	ZINB

Quotation: Vhon milton and Fred Mannering(1998), "Modeling Accident Frequencies as Zero-Altered Probability Processes: An Empirical Inquiry" , TRB

Data Summary of Statistics

Table 7 shows the summary statistics of the data used in this study. The homogeneous highway segment which experiences the most accidents is the one with 13 accidents. Also, it is noted that the length of the homogeneous segment is longer and the number of access roads is higher as the number of lanes gets higher.

<Table 7> Data summary of statistics of characteristics of highway segment

	Four or Less (84)				Six (74)				Eight or More (88)			
	Min.	Max.	Mean	S.D	Min.	Max.	Mean	S.D	Min.	Max.	Mean	S.D
Frequency	0	10	1.6	2.0	0	8	2.41	2.1	0	13	4.1	3.0
FrThe Number of Driveways	0	14	2.1	2.5	0	10	2.9	1.8	0	12	3.8	2.1
Highway Segment Length [m]	26	890	126.3	146.9	31	956	140.1	126.2	39.5	449	142.9	76.00
Horizontal Curve [m]	0	870	122.7	214.8	0	1800	160.0	310.7	0	857	88.4	233.6
Vertical Curve [%]	0	4.4	0.9	1.2	0	4	1.1	1.2	0	4.5	1.0	1.2
Traveled Lane Width [m]	2.4	4.46	3.3	0.4	2.53	5.25	3.7	0.6	2.58	4.1	3.3	0.2

Development Accident Frequency Models Result

A statistical package called LIMDEP 8.0 was used to analyze the data. To decide which model fits the data better, the Poisson regression model, the Negative Binomial regression model, and the Zero Inflated Poisson regression model were tested.

The Vuong statistic is computed as less than -1.96 due to the fact that there are few 0 values in the data set. Also, the overdispersion parameter statistic is less than 1.96, so the Poisson regression model is better than the Negative Binomial regression model in this study.

The variables of AADT, segment length, and the number of driveways are included in the three different highway groups - 4-lane and under, 6-lane, and 8-lane and over. In addition to the three common variables, the presence of vertical curvature and the horizontal curvature are the variables identified as significant for the 4-lane and under highway group; illegal parking, lighting, and lane width are significant for the 6-lane highway group, and land use (commercial) and lane width are significant for the 8-lane and over highway group.

In the 4-lane and under highway group, as the horizontal curvature gets smaller, more traffic accidents happen. Also, more accidents happen with greater vertical curvature in the homogeneous segment. For the 6-lane highway group, more accidents happen with more illegal street parking. Illegal street parking is likely to increase the accident potential because of sudden lane changes by the drivers to avoid the illegally parked vehicles. For the 8-lane and above highway group, more accidents occurs when the land use pattern next to the highway is commercial.

Multicollinearity testing was done to ensure the independence of the independent variables. Multicollinearity doesn't occur when the tolerance limit is close to 1 and the VIF is under 10. As the tests show, multicollinearity didn't occur for all the models considered.

When the value of ρ^2 is between 0.2 and 0.4, the fit is excellent. In this study, however, there is a case where ρ^2 is less than 0.2, Even though the value is not statistically significant, the model is still applicable because it can be logically explained

<Table 8> Developed Model Result

Variable name		Four or Less	Six	Eight or More
Constant	Parameter	-7.345	-5.769	-4.514
	T-statistic	-37.658	-10.425	-4.908
	P-value	0.000	0.000	0.000
EXPO (LengthXADTX365X10^{-6})	Parameter	Fixed parameter	Fixed parameter	Fixed parameter
	T-statistic			
	P-value			
Density of Driveways (km/Num)	Parameter	0.017	0.027	0.019
	T-statistic	2.581	6.591	5.988
	P-value	0.010	0.000	0.000
Lane Width (m)	Parameter	-	-0.450	-1.106
	T-statistic	-	-2.853	-4.050
	P-value	-	0.043	0.001
Presence of Vertical Curvature (%)	Parameter	0.524	-	-
	T-statistic	2.779	-	-
	P-value	0.006	-	-
Horizontal Curvature (1/R)	Parameter	0.108	-	-
	T-statistic	2.114	-	-
	P-value	0.345	-	-
Lighting Density (km/num)	Parameter	-	-0.009	
	T-statistic	-	-2.456	
	P-value	-	0.141	
Land-use (Commercial)	Parameter	-	-	0.358
	T-statistic	-	-	2.958
	P-value	-	-	0.031
Illegal Parking	Parameter	-	0.568	-
	T-statistic	-	2.892	-
	P-value	-	0.004	-
Log Likelihood Function		-110.542	-125.203	-187.007
Restricted Log Likelihood		-133.913	-157.910	-229.543
α(t-statistic)		0.000	0.001	0.003
$\rho 2$		0.175	0.207	0.185
MPB		0.081	-0.073	0.185
MAD		0.979	1.190	1.833
Vuong Statistic		-3.542	-2.846	-2.984

<Table 9> Multicollinearity Testing

Variable Name	Four or Less		Six		Eight or More	
	Tolerance Limit	VIF	Tolerance Limit	VIF	Tolerance Limit	VIF
EXPO (LengthXADTX365X10^{-6})	0.97	1.03	0.92	1.09	0.91	1.08
Density of Driveways (km/Num)	0.84	1.18	0.79	1.26	0.84	1.18
Lane Width (m)	-	-	0.83	1.20	0.92	1.09
Presence of Vertical Curvature (%)	0.92	1.09	-	-	-	-
Horizontal Curvature (1/R)	0.92	1.09	-	-	-	-
Lighting Density (km/num)	-	-	0.81	1.24	-	-
Land-use (Commercial)	-	-	-	-	0.92	1.09
Illegal Parking	-	-	0.81	1.23	-	-

CONCLUSIONS

This study developed a traffic accident frequency prediction model to analyze the accident characteristics of urban highways. A Poisson regression model and a Negative Binomial regression model were applied to the data set which includes the historical accident data, highway geometrics, facilities, and environments, and the land use data for the homogeneous highway segments.

The developed model shows that there are many vehicle-to-vehicle conflict-avoiding accidents, which is the common characteristic of an urban highway. Geometric elements such as presence of vertical curvature and horizontal curvature are rarely related to the accidents except in the 4-lane and under highway group.

There exists a limitation in this study due to the fact that human factors are not considered in the development of the prediction model. Accordingly, driver characteristics and driving behavior due to the change of the traffic environment are not included in the developed model in this study.

It is necessary to consider policy measures to reduce the vehicle-to-vehicle conflicts in the design of urban highways. Also, policy measures are necessary to deal with illegal street parking, which is rather common in the urban environment including Korea.

References

American Association of State Highway and Transportation Officials "A Policy on Geometric Design of Highways and Streets", 2001

Transportation Research Board, "Highway Capacity Manual", 2000

NCHRP 330. "Effective Utilization of Street Width on Urban Arterials.", Federal Highway Administration, 1990

Texas Transportation Institute, "Roadway Safety Design Workbook", Federal Highway Administration, 2005

Miaou, S. P. and H. Lum. "Modeling Vehicle Accidents and Highway Geometric Design Relationships", Accident Analysis and Prevention, Vol. 25, No. 6, pp. 689-709, 1993.

Shankar, V., J. Milton, and F. Mannering. "Modeling Accident Frequencies as Zero-altered Probability Processes: An Empirical Inquiry", Accident Analysis and Prevention, Vol. 29, No. 6, pp. 829-837, 1997.

Vhon milton, Fred Mannering, "Modeling Accident Frequencies as Zero-Altered Probability Processes: An Empirical Inquiry", TRB, 1998

Sawalha, Z., and T. Sayed, "Evaluating Safety of Urban Arterial Roadways, Journal of Transportation Engineering", American Society of Civil Engineers. pp. 151-158, 2001

Greibe, Poul, "Accident Prediction Models for Urban Roads", Accident Analysis & Prevention 839, pp 1-13, 2002

Shou-Ren Hu, Chi-Kang Lee, "Analysis of Accident Risk at Railroad Grade Crossing", TRB 87th Annual Meeting, 2008

Donald Erdman, Laura Jackson, Arthur Sinko, "Zero-Inflated Poisson and Zero-Inflated Negative Binomial Models Using the COUNTREG Procedure", SAS Global Forum 2008

Lambert, D., "Zero-inflated Poisson Regression with an Application to Defects in Manufacturing." Technometrics, No. 34, pp. 1-14, 1992

Winkelmann, R. "Econometric Analysis of Count Data". Fourth edition, 2003

Mullahy, J., "Specification and Testing in Some Modified Count Data Models", Journal of Econometrics, No. 33, pp. 341-365. 1986

El-Basyouny, K. and T. Sayed, "Comparison of Two-binomial Regression Techniques

in Developing Accident Prediction Models." Transportation Research Record: Journal of the Transportation Research Board, No. 1950, pp. 9-16, 2006