공학석사학위논문

Development of Crash Prediction Model Considering Characteristics of Ramp Types

램프 유형별 특성을 고려한 사고예측모형 개발

2015년 2월

서울대학교 대학원

건설환경공학부

이 태 헌

Development of Crash Prediction Model Considering Characteristics of Ramp Types

램프 유형별 특성을 고려한 사고예측모형 개발

지도교수 고 승 영

이 논문을 공학석사 학위논문으로 제출함

2014년 10월

서울대학교 대학원

건설환경공학부

이 태 헌

이 태 헌 의 논문을 공학석사 학위논문으로 인준함

2014년 12월

위 원 장 이 정 원 (인)

부위원장 고 승 영 (인)

위 원 김 동 규 (인)

Abstract

Lee, Taehun
Department of Civil and Environmental Engineering
The Graduate School
Seoul National University

Freeway ramps have a relatively higher number of crashes per unit traffic and distance than freeway mainlines. Moreover, freeway ramps also have different crash characteristics and geometric designs than those of the mainline. As such, a crash prediction model specific to freeway ramps should be developed. The trumpet interchange, where more than 70% of ramp crashes occur, has six ramp types, based on their ramp configurations (loop/semi-direct/direct) and ramp functions (on/off ramp). Of these six types, each ramp type has a different crash rate. Therefore, the crash prediction model to be developed must consider the different ramp-type characteristics.

Most previous studies have used a generalized linear model when developing a crash prediction model, and this type of model cannot take into account different ramp-type characteristics. This is because the generalized linear model assumes that every piece of data observed is independent of any other. The purpose of this study is to develop a

crash prediction model based on a multilevel model to address the limitations of the conventional model. A multilevel model can consider intra-class correlations derived from several individuals included in a group. Thus, this model has the advantage of being able to take into account different ramp-type characteristics.

In this study, the ramp crash data including 1,155 crashes were obtained from 2007 to 2010 for that occurred on three Korean freeway lines (Kyungbu, Yeongdong, and Seohaean lines) that had the highest number of crashes among all Korean freeway lines. Then a multilevel model is developed that considers different ramp-type characteristics and a generalized linear model that does not consider these characteristics. The annual average daily traffic (AADT), ramp length, and ramp-type dummy variables are used as the independent variables in the models. The multilevel model consists of a 1-level model for individual ramps and a 2-level model for other ramp types. The generalized linear model has the same structure as the 1-level model in the multilevel model.

The parameter estimation results in the multilevel model include the different crash risk for each ramp type and the impacts of different ramp lengths on the number of crashes, according to type. In contrast, the generalized linear model estimates the same parameters for all ramp types and does not reflect the different type characteristics. To validate the models, quantitative indicators including root mean square error

(RMSE), median absolute deviation (MAD), and cumulative residuals (CURE) plot are adopted. Both the RMSE and MAD values revealed the prediction accuracy of the multilevel model to be superior to that of the generalized linear model. Also, the CURE plot implies that the goodness-of-fit of the multilevel model is relatively higher than that of the generalized linear model.

The multilevel model proposed in this study can improve upon the limitations of the generalized linear model. Moreover, this model would make possible the accurate investigation of crash hot-spots, due to its higher crash prediction reliability. Consequently, this model can take a role as a quantitative basis for implementation of safety improvements on each type of trumpet interchange ramp.

· Keywords : Crash Prediction Model, Multilevel Model, Regression Analysis, Ramp Types

· Student Number : 2013-20936

List of Contents

List of Tables

List of Figures

Ⅰ. INTRODUCTION

1.1. Background

Crashes occurring on freeways can cause severe injury to drivers due to the high speed of the vehicles and the potential for secondary crashes in the vehicles following. Moreover, freeway traffic flow can be interrupted from the disturbances to vehicle progress. For this reason, a number of safety improvement projects have been undertaken, such as modifying roadway geometry and installing safety facilities. However, the efficient use of the limited resources available for freeway safety improvement projects requires that clear guidelines be established for the selection of hot-spots and the prioritization of projects.

In general, conventional crash prediction studies have developed a safety performance function, which is a regression model for predicting the number of crashes with consideration to the factors affecting crash occurrences. The number of crashes predicted by this safety performance function is then used to prioritize the safety improvement projects as the basis for the potential for safety improvement (PSI). In addition, it can set the long-term direction for road safety improvements on freeways through the investigation of environmental factors that affect the occurrences of crashes.

The majority of these studies have developed safety performance

functions that are focused on the freeway mainline. These studies have analyzed the impact of crash factors on accident occurrences and developed safety performance functions to predict the number of crashes. Although a freeway consists of various sections, such as the ramp for transfers and the tollgate for the collection of tolls, research focused on sections other than the mainline have been relatively rare. According to statistical analysis results, ramps, as the most representative non-mainline section, are more dangerous than the mainline. The average number of crashes per year·million traffic·km that occur on ramps is 1.82, but the average for the mainline is only 0.05. Furthermore, ramps have different characteristics from the mainline with respect to roadway geometry and crash occurrence.

In particular, the proportion of ramp crashes caused by speeding, driver drowsiness, and negligence is significantly different from that of mainline crashes. The proportion of crashes by speeding on ramps is 50.5% that is 1.5 times larger than the 27.7% proportion on the mainline. In contrast, the proportions of crashes due to driver drowsiness and negligence account for 8.0% and 6.3%, respectively, which is relatively smaller than the 23.0% and 16.3% proportions on the mainline, respectively. Likewise, ramps have different characteristics than the mainline with respect to their geometric design and crash type. As such, the development of crash prediction models specifically for ramps should be developed.

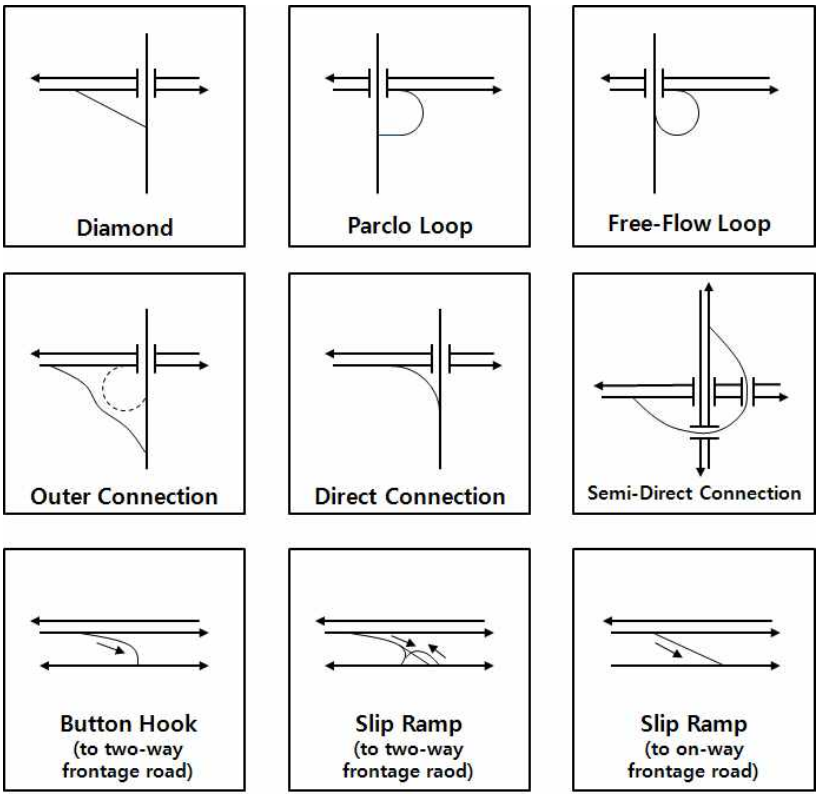
[Table 1.1] Crash Types on Ramp and Mainline

Crash Type	Mainline		Ramp	
	Number of Crashes	Rate(%)	Number of Crashes	Rate(%)
Total	12,750	100.0	1,706	100.0
Speeding	3,529	27.7	861	50.5
Drowsy Driving	2,932	23.0	137	8.0
Negligence	2,079	16.3	107	6.3
Overhandling	2,661	20.9	366	21.5
Others	1,549	12.1	235	13.8

* Figures are based on the data for 2007~2010, Kyeongbu, Yeongdong, and Seohaean line

Freeway ramps connect two freeways or connect a freeway with a national highway. A junction is the section connecting two freeways, and an interchange connects a freeway and a national highway. In general, there are more than two ramps on each interchange or junction.

Ramps can be classified into several types, according to their configurations and on/off forms. A technical report for the U.S. Federal Highway Administration (FHWA) by Bauer and Harwood (1998) describes nine representative ramp configurations. As shown in Figure 1.1, these configurations also include on/off forms, so they can be further classified into 18 ramp types.

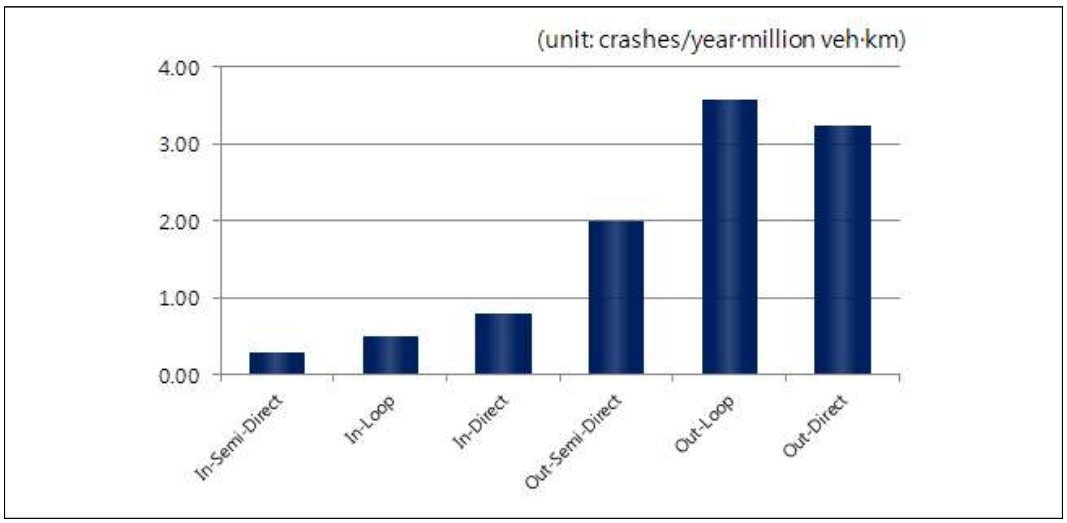


[Figure 1.1] Representative Ramp Configurations[1)]

The trumpet interchange is the most representative interchange type, accounting for 78 percent of all interchanges in the Korean freeway system. It consists of a free-flow loop, direct connections, and semi-direct connections, as shown in the Figure 1.1[1)]. In other words, the trumpet interchange has 6 ramp types, including on-loop, off-loop, on-semi-direct, off-semi-direct, on-direct, and off-direct.

1) Bauer K. M. and Harwood D. W.. 1998. Statistical Models of Accidents on Interchange Ramps and Speed-Change Lanes

Different trumpet interchange ramp types have different road geometries and traffic flow characteristics and different risk levels for crash occurrence. Figure 1.2 shows a comparison of the annual average number of crashes per unit traffic and distance for each ramp type, revealing significant differences among the ramp types. A comparison of the number of crashes occurring on the on/off ramps shows that more crashes have occurred on off ramps. This is because the average speed of vehicles is relatively higher on the off ramps than on the on ramps. With respect to ramp configuration, semi-direct ramps have the longest average length and the least number of crashes. These figures were obtained from the trumpet interchange data of the Kyeongbu, Yeongdong, and Seohaean freeway mainlines from 2007 to 2010.



[Figure 1.2] Annual Average Number of Crashes on Unit Traffic and Distance

Differences in the crash occurrences by ramp type indicate that there is heterogeneity by ramp type. Huang and Abdel-Aty (2010) reported that the conventional crash prediction model that is based on the generalized linear model cannot take into account the multilevel structure of the safety data because it assumes that every observation is independent. As a result, this reduces the reliabilities of the parameter estimations and statistical inferences. The reliability of the ramp crash prediction model will also be decreased unless the heterogeneity of ramp types is considered.

For these reasons, the development of a ramp crash prediction model that considers the heterogeneity in ramp types is necessary in order to improve model accuracy and statistical inference reliability. Therefore, the purpose of this study is to develop a crash prediction model that considers the heterogeneity of trumpet interchange ramp types, which are the most representative interchange type in Korean freeways, to remove the limitations of the conventional generalized linear model.

1.2. Composition

In this study, the literature review to compare the crash characteristics of each ramp type, data analysis, model estimation, and comparison of the multilevel model and the generalized linear model are implemented.

In the 2nd chapter, a review of conventional crash prediction model studies is conducted. Then the literature reviews on the generalized linear model and the multilevel model are implemented. The 3rd chapter outlines the study methodology, including the research flow, the multilevel structure of the ramp crash data, the multilevel model concept, and the basic form of the proposed model. In the 4th chapter, the temporal and spatial scope of this study, the results of our basic statistical analysis and the model variables are discussed. The 5th chapter presents the design of a crash prediction model based on the multilevel model and the generalized linear model. Then the superiority of the multilevel model is investigated by comparing the two models in the 6th chapter. Finally, the conclusion of this study and a future feasible study are suggested in the 7th chapter.

Ⅱ. LITERATURE REVIEW

2.1. Crash Prediction Model

Hauer (1995) defined the safety performance function as a regression model for predicting crash frequencies that takes into account the characteristics of the roadway, including the annual average daily traffic (AADT), the length of the section, and the geometry of the roadway.

The safety performance function can be used to calculate the PSI at certain sites by predicting crash frequency. PSI infers the potential number of crashes that can be reduced by safety improvements, and this value can be used as the quantitative basis for prioritizing target sites for safety improvements.

Safety performance functions can be classified into the simple safety performance function and the inclusive safety performance function. The form for each representative model is shown in Eq. 2.1 and 2.2, respectively. The simple safety performance function has a simple model form and is easy to apply, so the majority of prior studies have used it. Harwood et al. (2010) defined the simple safety performance function as a crash prediction model, for a roadway section in basic conditions, and it can estimates the relationships between crash frequency and the several major factors including AADT and length of segment.

· **Simple Safety Performance Function**

$$Y = \exp(\alpha + \beta_{AADT} \cdot AADT + \beta_L \cdot Length) \cdot CMF_1 \cdot \cdots \cdot CMF_i \qquad \text{[Eq. 2.1]}$$

· **Inclusive Safety Performance Function**

$$Y = \exp(\alpha + \beta_1 \cdot AADT + \beta_2 \cdot var_2 + \beta_3 \cdot var_3 + \cdots) \qquad \text{[Eq. 2.2]}$$

where, Y : expected number of crashes

α : constant

β_i : regression coefficients of parameter

var_i : independent variables

CMF_i : crash modification factors

The simple safety performance function can be combined with crash modification factors to improve the accuracy of the model. Using these crash modification factors, it can be estimated that the effect of road geometry on the frequency of crashes and assess the need for safety improvements.

In contrast, the inclusive safety performance function can take into account various roadway characteristics that affect crash occurrence as model variables. Moreover, this function has the advantage that it can

estimate the marginal effect of factors on crash frequencies as a coefficient of the independent variables.

Most previous research has involved the development of simple safety performance functions using a few independent variables. According to the American Association of State Highways (2010), the advantage of these models is that they do not require additional variables to take into account. Their main disadvantage is their low reliability.

Alluri and Ogle (2012) developed a safety performance function for local areas in an effort to remove this limitation, and the reliability of the model was increased, based on estimates using local data. Also, some studies have considered various independent variables in developing the safety performance function. Lu et al. (2013) reported that these models could be improved with the addition of traffic flow data and geometric design characteristics.

2.2. Prior Crash Prediction Model Studies

2.2.1. Generalized Linear Model

Previous research has mostly focused on the development of safety performance functions using the generalized linear model and traffic crash data. The initial studies mainly targeted freeway mainlines. However, following recognition of the need to develop safety performance functions targeting sub-facilities, studies then began to include ramps and tollgates, due to their high crash risks.

Bauer and Harwood (1997) suggested the development of a crash prediction model to investigate the relationship between crash frequency and the geometric design of roadways or traffic volume. According to their analysis, the AADT of ramps is the most reliable independent variable. The on/off type and geometric designs of ramps are also reliable variables that affect the number of crashes.

Park and Ryu (2002) developed a multivariate regression model to investigate the relationship between crash frequency and interchange types for trumpet interchanges. According to their results, different factors affect the crash frequency for each interchange type, such as the on/off ramp, the mainline radius of curvature, and the mainline vertical curve.

Lord and Bonneson (2005) suggested crash prediction models for each ramp configuration, including the diagonal, non-free-flow loop,

free-flow loop, and outer connection for a local area (urban and suburban). They revealed that crash factors have different effects according to ramp configurations and on/off types. According to the results, the different impacts of crash factors need to be considered in this study. But they developed the linear regression model with the coefficient of traffic volume except for other independent variables based on the assumption that crash frequency on the ramp is affected by only the traffic volume. For this reason, it is different from this study with respect to the model form that consists of multilevel structure.

Yoon (2007) suggested crash prediction models for different geometric designs and traffic flow conditions on freeway trumpet interchanges. These models were developed only for the off ramps in each ramp configuration (loop, semi-direct, and direct), and considered several independent variables including the AADT, the radius of curvature, and the slope of the mainline. The scope of Yoon's study was limited to closed sections of freeways, based on the acquisition of traffic volume data. For this reason, the lack of samples resulted in low reliability of the parameters defined in the model.

Park (2006) developed a crash prediction model for the connected section of the freeway interchange, using various independent variables, including the on/off ramp, the radius of curvature, the traffic volume, and the rate of heavy vehicles. According to the results of this study,

the most significant factor is the traffic volume of the ramp. The rates of heavy vehicles, radius of curvature, and on/off ramp also affect crash frequency. However, the radius of curvature is not a suitable independent variable in the model because its method of measurement in each study is different and therefore not standardized. Moreover, Bauer and Harwood (1997) reported that the radius of curvature was not a reliable independent variable. Thus, it would not be appropriate to take this factor into account in the present study.

Abdel-Aty et al. (2007) suggested a 2-level nested logit model for estimating the impact of traffic flow on on/off ramp crash occurrences. The results of their analysis show that higher speeds on the upper mainline stream, lower speeds on the mainline downstream, and lower ramp traffic volume causes a higher number of crashes on on ramps and adversely affect crashes occurring on off ramps. These results could explain why more crashes occur on off ramps with low traffic volume, as drivers may lose control due to their high speeds of approach. As such, these results bear consideration in the design of the multilevel model structure for each ramp type in the present study.

Having reviewed previous research efforts on the development of crash prediction models for freeway ramps, it was founded that in the majority of studies, AADT was the primary crash factor affecting the frequency of ramp crashes. While the length of the ramp, length of the accelerate/decelerate lane, radius of curvature, and slope of the mainline

were also used to develop the models, their statistical reliability was not significant. These results agree with those of Lord and Bonneson (2005) who found that crash factors have different impacts on crash occurrence for different ramp configurations and on/off ramps. For this reason, the development of a model based on the assumption that every piece of observed data is independent, and with no consideration of the ramp type, can result in low model reliability.

In order to address this problem, several studies have developed crash prediction models for different interchanges and trumpet interchange ramp configurations of a relatively standardized form, and the reliabilities of models were thereby improved. In these cases, however, the models have the limitation of not being able to compare the parameters of a certain independent variable in two different models because the model parameters were estimated by a maximum likelihood estimator. For this reason, this study developed a crash prediction model that considers the characteristics of each ramp type and takes into account the multilevel structure of the ramp crash data, using a multilevel model to improve on the limitations of conventional models.

2.2.2. Multilevel Model

Several studies have used the multilevel model to consider intra-class correlations for the multilevel structure of freeway crash data. But the majority of these used the multilevel model to model the

relationship between the crash factors and the crash frequency or crash severity occurring on the freeway mainline only.

Shankar et al. (1998) compared the random effect negative binomial (RENB) regression model with the negative binomial (NB) regression model to predict median crossover crash frequencies on the mainline. Unless the NB model considers the relationship of the exogenous factors to the spatial and temporal effects, the RENB model is more reliable than the NB model. So, the RENB model can be used to take into account the spatial and temporal effects indirectly when data for temporal- and spatial-specific variables cannot be obtained.

Jones and Jorgensen (2003) used a multilevel model to consider the impact of crash factors on the severity of crashes at the time of occurrence. With respect to fatal injuries, the multilevel model design includes a 1-level model for crash casualties, a 2-level model for crashes, and a 3-level model for municipalities. The analysis results reveal that the proportions of variation due to accident level heterogeneities and the municipality level of the total unexplained variations accounted for 16% and 1%, respectively. According to cumulative residuals (CURE) plots of the generalized linear and multilevel models, the multilevel model could help to reduce unavoidable randomness by smoothing the residuals.

Huang and Abdel-Aty (2010) reported that generalized linear models have the limitations of not being able to consider the multilevel

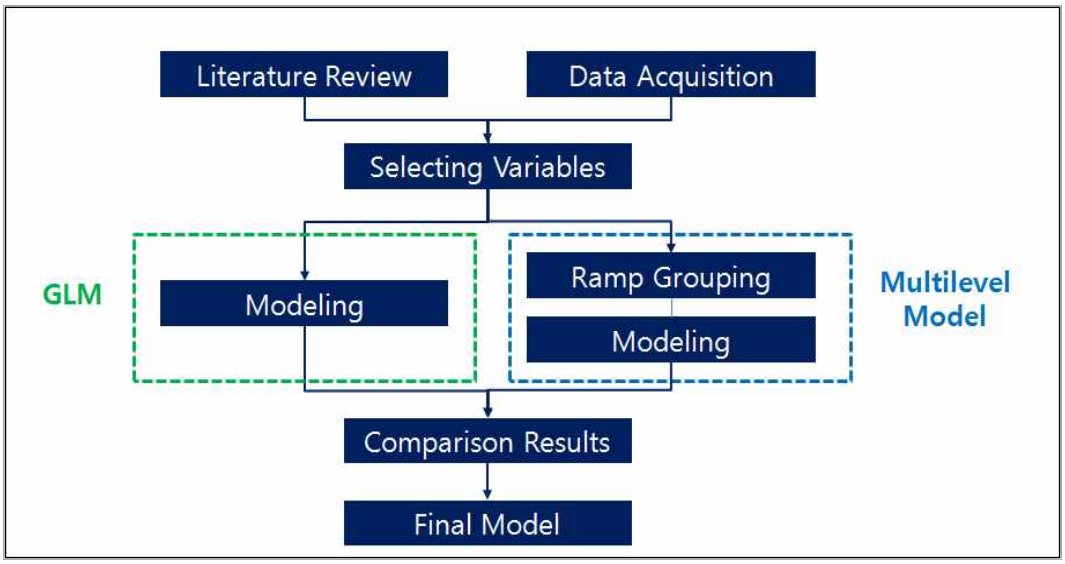
structure of data. This is because they assume that every individual observation is independent. Also, they suggest the 5 × ST (spatio-temporal)-hierarchy as a conceptual framework for safety data, which could be a fundamental aspect of crash prediction when using a multilevel model. This hierarchy consists of a geographic regional level, a traffic site level, a traffic crash level, a driver-vehicle unit level, and an occupant level. Using this hierarchy, a multilevel model can be developed that considers the traffic crash level and traffic site level.

In summary, most prior studies for predicting crash frequencies and severities using multilevel models developed crash prediction models targeted to the mainline. They found the multilevel model to be superior to the generalized linear model when not taking into account the variables that account for variations from spatial and temporal effects. This study differs from prior studies in its focus on ramp section and the different characteristics of each ramp type. In the present study, the crash characteristics of different ramp types are compared with consideration of the findings of previous studies. Then a crash prediction model considering the multilevel structure is developed.

Ⅲ. Methodology

3.1. Research Flow

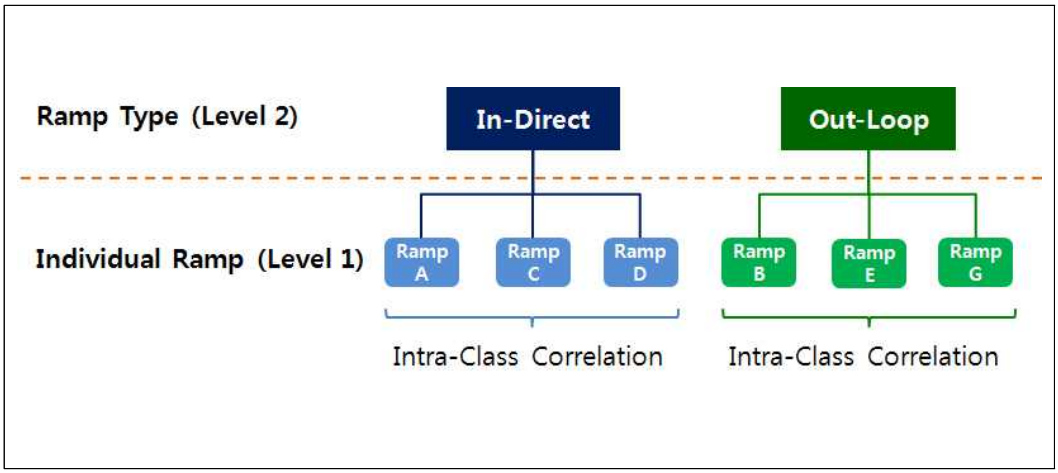
At first, the factors affecting the crash frequency are investigated and the model variables are selected with consideration of the literature review. Then, a basic statistical analysis of data is conducted. After that, the multilevel model considering the different characteristics of each ramp type and the generalized linear model without consideration are estimated. Finally, the results of estimated parameters are compared to determine the superior model. Figure 3.1 illustrates the research flow of this study.



[Figure 3.1] Research Flow

3.2. Multilevel Structure

The crash data from freeway ramp sections have a multilevel structure when more than two individual ramps belong to the same ramp type. In this case, several individual ramps included in the same group share an intra-class correlation. As shown in Figure 3.2, the ramp crash data from trumpet interchanges have a multilevel structure according to ramp types.

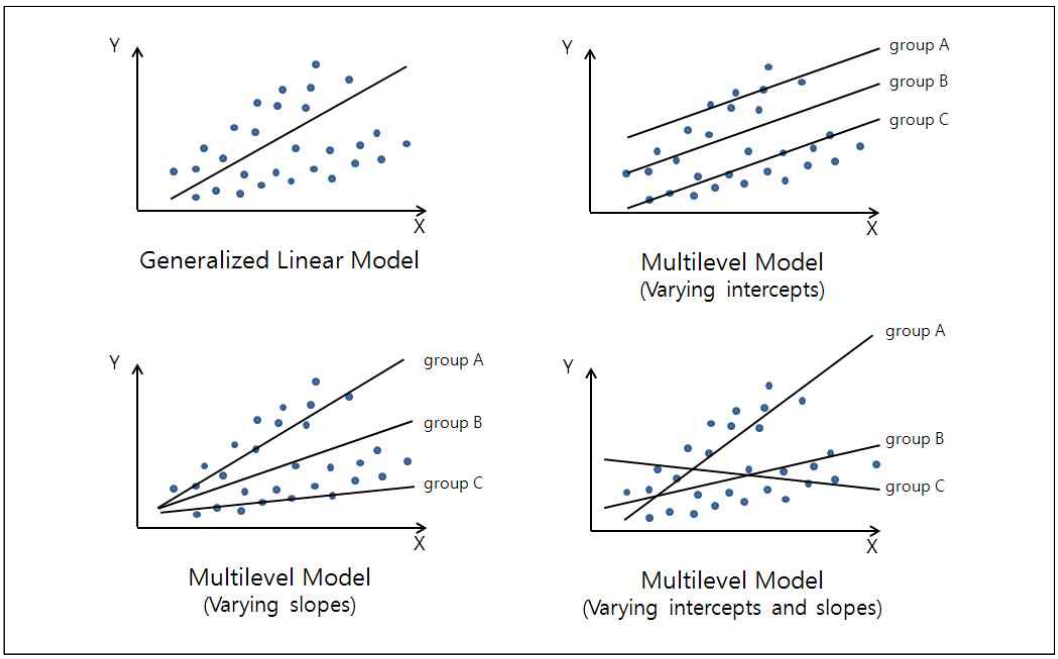


[Figure 3.2] Multilevel Structure of Ramp Crash Data

3.3. Multilevel Model

3.3.1. Definition and Concept

Gelman and Hill (2007) defined the multilevel model as a regression model with the parameters estimated as probability models, rather than fixed coefficients. So the multilevel model is different form the generalized linear model in that it can consider multilevel data structures. As shown in Figure 3.3, the generalized linear model estimates a fixed parameter for all samples, but the multilevel model estimates different parameters according to their upper level grouping.



[Figure 3.3] Concepts of Multilevel Model

3.3.2. Model Form

Multilevel model consists of more than two levels. In this study, the multilevel model consists of a 1-level model and a 2-level model. In general, the 1-level model predicts the dependent variables and the 2-level model estimates the parameters of the 1-level model as probability variables according to their group. The combined model can be calculated by incorporating the 2-level model into the 1-level model.

· Form of Multilevel Model

Level	Model Form
1-level Model	$Y = \beta_0 + \beta_1 \cdot X_1 + \epsilon$
2-level Model	$\beta_0 = \beta_{00} + \beta_{01} \cdot X_{01} + \epsilon_{00}$
	$\beta_1 = \beta_{10} + \beta_{11} \cdot X_{11} + \epsilon_{10}$

[Eq. 3.1]

where,

Y : dependent variable

X_{ij} : independent variable of individual i in group j

β_{ij} : coefficient of individual i in group j

ϵ_{ij} : error term of individual i in group j

Ⅳ. Data

4.1. Scope of Study

4.1.1. Temporal Scope

The temporal scope of this study is 4 years, from 2007 to 2010, a time period chosen in this study due to consistency considerations with respect to the ramp crash data. As the crash data are recorded, they are classified into A-D levels according to the severity of the crash. However, no D class data were recorded prior to 2007. Therefore, this study limited the research period to years after 2007 for data consistency.

4.1.2. Spatial Scope

This study limited the spatial scope of this study to the ramps of trumpet interchanges on 3 freeways (Kyungbu, Yeongdong, and Seohaean lines). These freeways have the highest annual average number of crashes of all Korean freeways. During the study period, the number of ramp crashes that occurred on these 3 lines was 1,155, accounting for 37.5% of the total Korean freeway crashes. The trumpet interchange is the most representative interchange type, accounting for more than 78% of the total number of Korean freeway interchanges. Accordingly, this study limited the range of research area to ramp crashes that occurred on trumpet interchanges.

[Figure 4.1] Spatial Scope of Study

4.2. Basic Statistical Analysis

4.2.1. Composition of Analysis Data

For the analysis, 15,438 crashes data obtained from the three freeway lines. The number of mainline crashes, 11,681, accounted for 75.7% of the total number of crashes. The number of ramp crashes, 1,632, accounted for 10.6% of the crashes-the 2nd highest proportion after that of mainline crashes.

Among the ramp crashes, 1,155 crashes occurred on trumpet interchanges, accounting for 70.8% of the total number of ramp crashes. Considering that trumpet interchange crashes accounted for more than 70% of total number of ramp crashes, the trumpet interchange is the most representative type of ramp on Korean freeways. This is why the prior studies focused on the ramp crashes of trumpet interchanges. Likewise, this present study also focuses on trumpet interchange, but the different estimation method is used for the model development.

[Table 4.1] Crashes on Each Site of Expressway

Site	Number of Crashes	Rate(%)
Total	15,438	100.0
Mainline	11,681	75.7
Ramp	1,632	10.6
Tollgate	1,257	8.1
Others	868	5.6

This study conducted analysis on data from the 1,155 crashes that occurred on 1,070 ramps from 2007 to 2010. Tables 4.2 and 4.3 show the number of ramps and the crashes for each ramp type, respectively. Comparing the number of crashes with regard to ramp configurations, the loop ramp had 434 crashes, the highest number of crashes, and off ramps had a much higher number of crashes than in ramps.

[Table 4.2] Number of Ramps of Each Type

Number of Ramps	SUM	Loop	Semi-Direct	Direct
SUM	1,070	267	268	535
In	535	115	152	268
Out	535	152	116	267

[Table 4.3] Number of Crashes of Each Ramp Type

Number of Crashes	SUM	Loop	Semi-Direct	Direct
SUM	1,155	434	322	399
In	173	56	52	65
Out	982	378	270	334

The reasons for crashes in the trumpet interchange ramp and the mainline are compared in order to identify the different crash factors for ramps. The result of analysis showed that speeding accounted for

51.6% of the total ramp crashes, which is significantly higher than that for mainlines, where speeding accounts for just 19.0% of the crash total. Over-handling accounted for 21.2% of ramp crashes, which is also relatively higher than that of mainline where over-handling accounted for the 14.9% of the total. The higher proportion of crashes due to speeding and over-handling is related to the number of off ramp crashes, which is much higher than the on ramp total. Considering the driving characteristics of vehicles travelling on the off ramp, the long driving time on the mainline with no steep curves could result in crashes on the ramp section where there is a sudden requirement for cautious handling and deceleration. As such, the crash prediction model presented in this study has the potential for use in improving ramp safety for each ramp type.

[Table 4.4] Crash Reasons for Ramp and Mainline

Crash Reason	Trumpet IC Ramp		Mainline	
	Crashes	Rate(%)	Crashes	Rate(%)
Total	1,155	100.0	11,681	100.0
Speeding	596	51.6	2,224	19.0
Overhandling	245	21.2	1,735	14.9
Drowsy Driving	92	8.0	1,815	15.5
Negligence	79	6.8	1,371	11.8
Others	143	12.4	4,536	38.8

To investigate the vehicle types that are especially vulnerable to ramp crashes, this study compared crash vehicle types for ramps and mainlines. As shown in Table 4.5, compact car accounts for the highest proportion of crashes for both ramps and mainlines. The truck/trailer/special vehicle is the 2nd highest vehicle type likely to crash. According to these results, there is no significant difference in vehicle type between ramp and mainline crashes.

[Table 4.5] Vehicle Types of Crash for Ramp and Mainline

Vehicle Type	Trumpet IC Ramp		Mainline	
	Crashes	Rate(%)	Crashes	Rate(%)
Total	1,155	100.0	11,681	100.0
Compact Car	785	68.0	7,632	65.3
Van	75	6.5	988	8.5
Truck, Trailer, and Special Vehicle	290	25.1	2,948	25.2
Others	2	0.2	26	0.2
Blank	3	0.2	87	0.8

4.2.2. Crash Characteristics of Each Ramp Type

Prior to developing the crash prediction model, it was necessary to investigate the different crash characteristics of each ramp type through basic statistical analysis. The number of crashes per unit distance,

crashes per unit vehicle, and crashes per unit distance and unit vehicle are compared to investigate the difference of crash rates for different ramp configurations of trumpet interchange. The results show that off ramps had a much higher number of crashes than on ramps. The number of crashes per unit distance was highest on loop ramps. The direct on ramp and the loop off ramp had the highest number of crashes per unit vehicle. The number of crashes per unit distance and vehicle was also highest on direct on ramps and loop off ramps. These comparison results show that the basic crash risk is very different for each ramp type. So ramp types must be classified into 6 groups, with combinations of on/off ramps and loop/semi-direct/direct ramps.

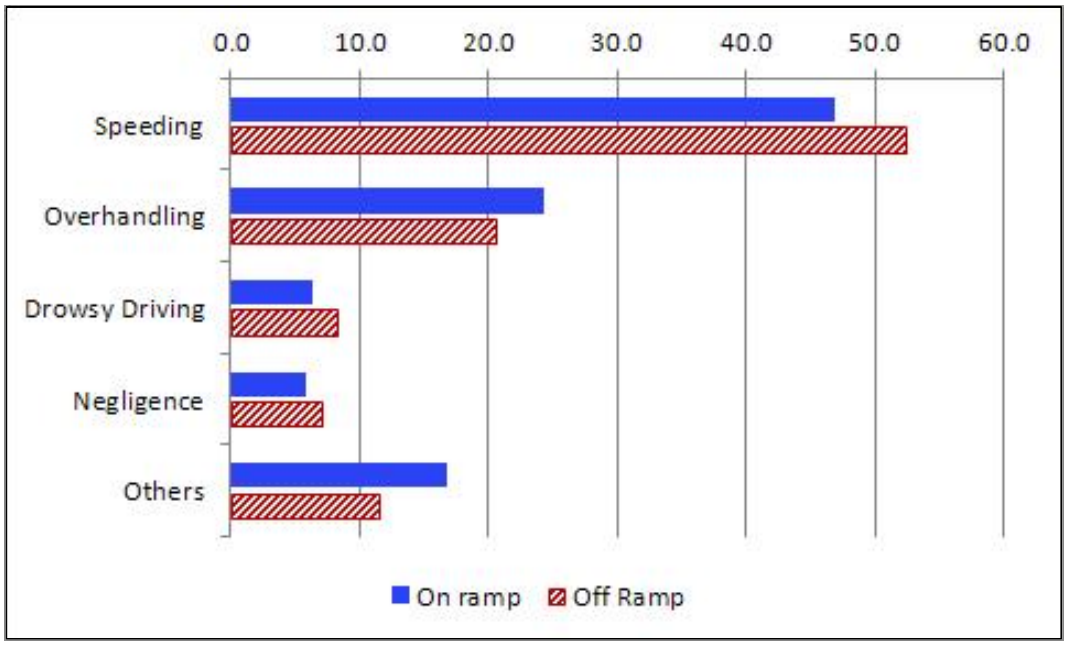
[Table 4.6] Annual Crashes per unit distance and vehicles for Each Ramp Type

On/Off	Configuration	Crashes/km	Crashes /million veh	Crashes /km·million veh
Total		0.87	2.1	1.82
On	**Semi-Direct**	0.19	0.50	0.28
	Loop	0.34	0.71	0.50
	Direct	0.23	0.77	0.79
Off	**Semi-Direct**	1.36	3.19	1.99
	Loop	2.32	3.91	3.56
	Direct	1.10	3.44	3.23

After that, the reasons for the crashes for the different ramp types are compared. Firstly, considering that the much higher number of crashes on off ramps than on the on ramps, the crash reasons are compared. Results show that speeding accounted for 52.4% of the off ramp crashes, while it accounted for 46.8% of on ramp crashes. These results are due to the fact that drivers travelling on off ramps had been driving over long distances at high speed, and they could not control their vehicles when they encountered the off ramp. Consequently, the vehicle driving characteristics on off ramps led to different crash rates for on/off ramps.

[Table 4.7] Crash Reasons for On/Off Ramps

Crash Reason	On Ramp		Off Ramp		Total	
SUM	173	100.0	982	100.0	1155	100.0
Speeding	81	46.8	515	52.4	596	51.6
Overhandling	42	24.3	203	20.7	245	21.2
Drowsy Driving	11	6.4	81	8.2	92	8.0
Negligence	10	5.8	69	7.0	79	6.8
Others	29	16.8	114.0	11.6	143.0	12.4

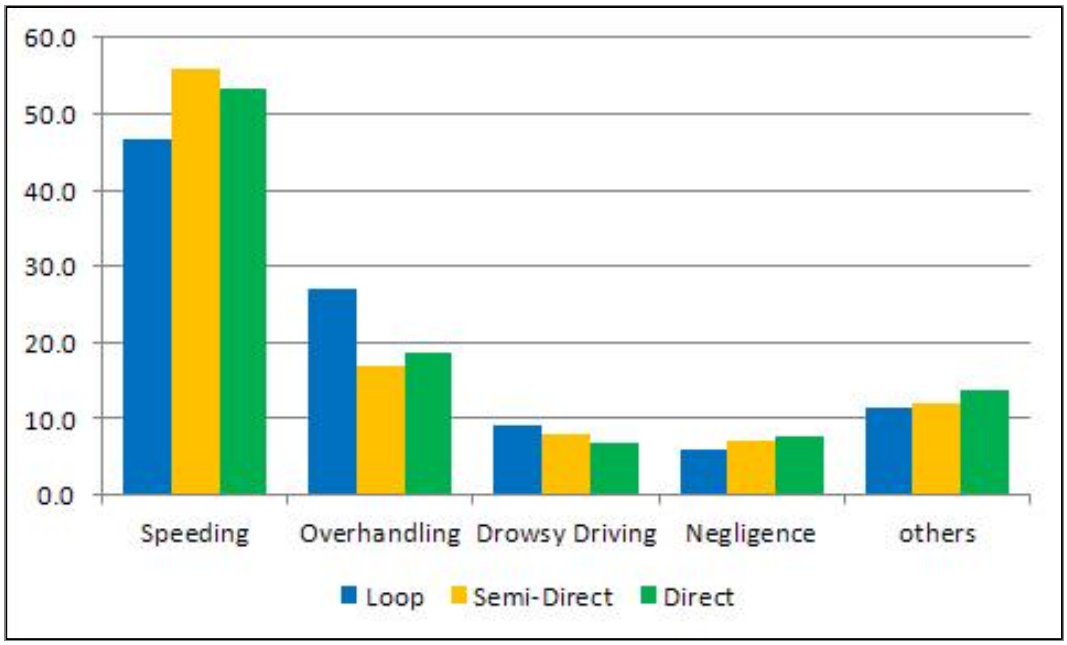


[Figure 4.2] Crash Reasons for On/Off Ramps

Table 4.8 shows the reasons for crashes in different ramp configurations. The proportion of crashes due to over-handling is the only different result among the three configurations. For loop ramps, the speeding rate is relatively lower and the over-handling rate is relatively higher than for the other ramps. For semi-direct ramps, the speeding rate is higher than on other ramps and the over-handling rate is lower. These results indicate that the reasons for crashes differ for different ramp configurations.

[Table 4.8] Crash Reasons for Loop/Semi-Direct/Direct Ramp

Crash Reason	Loop		Semi-Direct		Direct		Total	
SUM	434	100.0	322	100.0	399	100.0	1,155	100.0
Speeding	203	46.8	180	56.0	213	53.3	596	51.6
Overhandling	117	27.0	54	16.7	74	18.6	245	21.2
Drowsy Driving	39	9.0	26	8.1	27	6.8	92	8.0
Negligence	26	6.0	23	7.1	30	7.5	79	6.8
Others	49	11.3	39	12.1	55	13.8	143	12.4

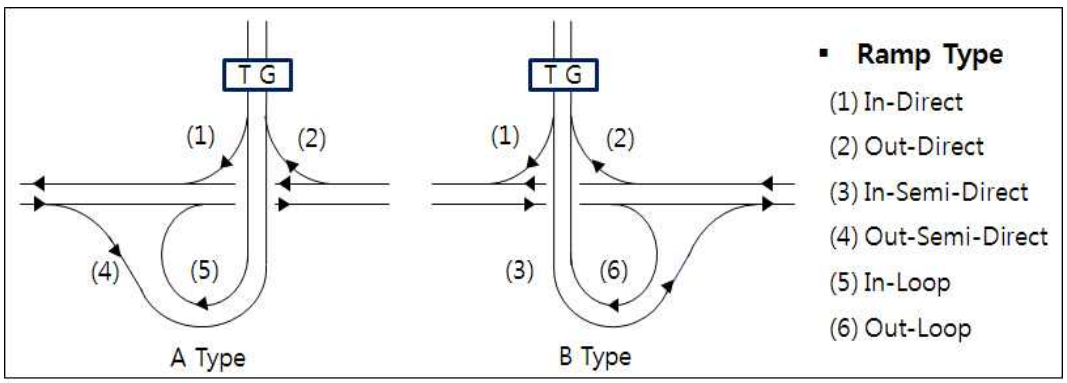


[Figure 4.3] Crash Reasons for Each Ramp Configuration

4.3. Data Measurement

4.3.1. Ramp Types

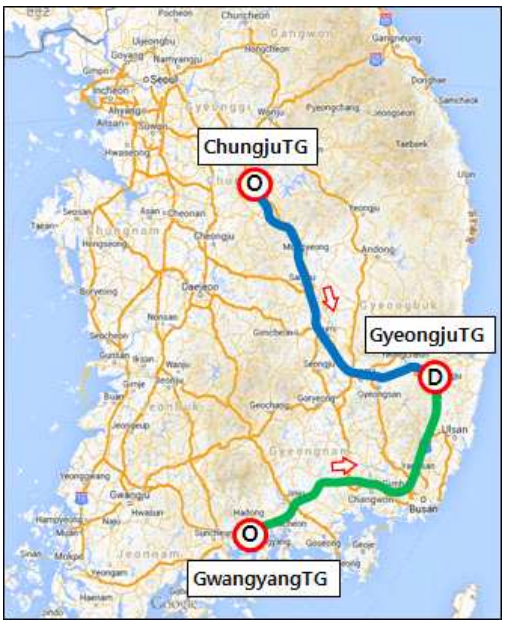
From a basic statistical analysis, the number of crashes per unit distance and vehicle for different ramp types are compared. The results of this comparison show that the differences in the crash occurrence rates should be considered in the model,. Therefore, the 6 ramp types are suggested in for this study. These types consist of combinations of the on/off ramps with the loop/semi-direct/direct ramps. As shown in Figure 4.4, the trumpet interchange has two form types depending on whether the loop ramp is an on or off ramp. The 6 ramp types classifications are the on-loop, on-semi-direct, on-direct, off-loop, off-semi-direct, and off-direct. These types constitute the 2-level group in the multilevel model.



[Figure 4.4] Two Types of Trumpet Interchange

4.3.2. Traffic Volume of Ramp

To develop a ramp crash prediction model, AADT data, the most representative exposure variable, is essential. However, there is no mechanism for obtaining the traffic volume data of individual ramps. This is why previous studies conducted research only on the closed sections of freeways. However, it is so limited area to obtain the enough samples to estimate the parameters of the model. As such, due to it having been estimated with a limited number of samples, this model has low transferability to other areas. Therefore, the traffic volume of individual ramp is calculated by using the origin-destination traffic data between tollgates.



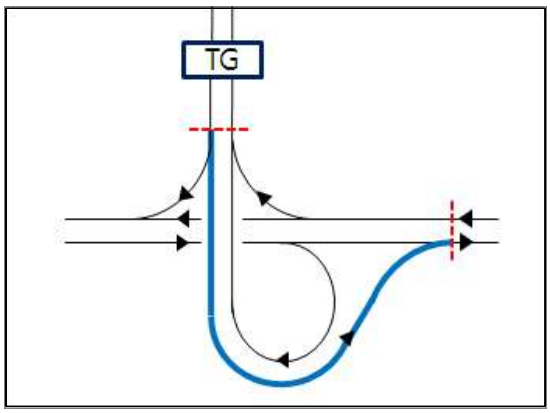
[Figure 4.5] Selection of Shortest Route between O-D

The traffic volume is calculated by considering the direction of the mainline and the relative location of the tollgate, under the assumption that the origin-destination traffic would have chosen the shortest route. For instance, as shown in Figure 4.5, the traffic moving from the Chungju tollgate to the Kyungju tollgate would opt to travel the blue route because it is the shortest. Likewise, the traffic moving from the Kwangyang tollgate to the Kyungju tollgate would have chosen the green route. Consequently, the traffic on the blue route that merged from the Chungju tollgate to the mainline necessarily travelled on the on-direct ramp toward Busan. And the traffic diverging from the mainline to the Kyungju tollgate travelled on the off-direct ramp toward Seoul.

[Figure 4.6] Selection of Ramp for Each Route

4.3.3. Ramp Length

In most previous studies, the AADT and segment length are used as the independent variables in crash prediction models. The length of the segment is the most representative exposure variable and the number of crashes increases as the length is extended. But there is no conviction that crashes that occur on ramps also increase in number as the length of the ramp is extended. Therefore, the lengths of individual ramps are measured to take this variable into account as an independent model variable.



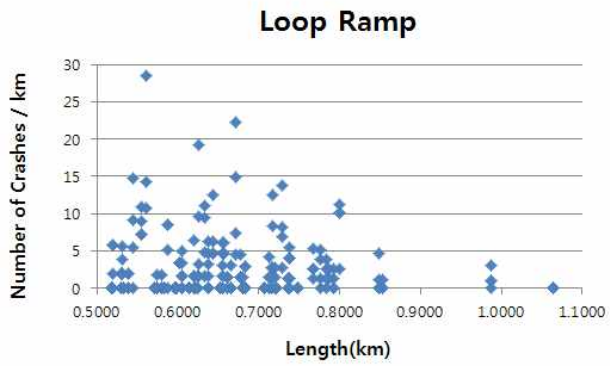
[Figure 4.7] Measurement of Ramp Length

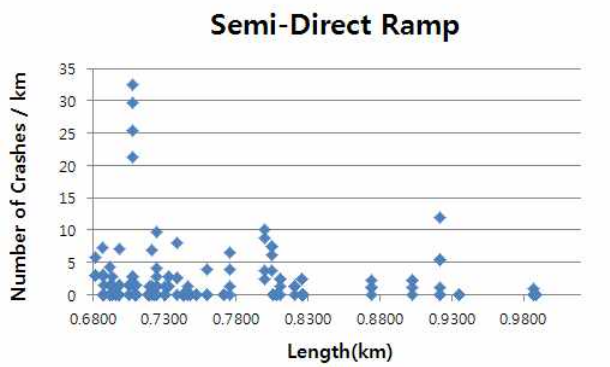
As shown in Figure 4.7, the length of ramp measured in this study as the section between the point where it divides from the mainline and the point where it is joined to other ramps. In general, the accelerate lane or decelerate lane are located near the point where the ramp is

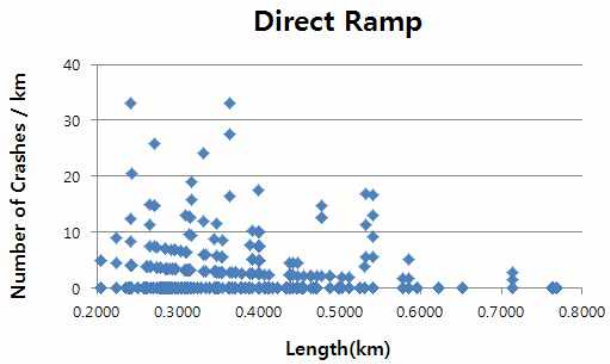
connected to the mainline. Crashes also occurred in the accelerate/decelerate lanes, but these crashes occurred in areas affected only by the traffic volume or the length of the accelerate/decelerate lanes. For these reasons, the accelerate/decelerate lanes are excluded from the study area and the length of ramp.

The length of a ramp is simply an exposure variable in the crash prediction model with respect to the mainline, but it is related to the geometric design and the radius of curvature in ramp models. Park (2006) used ramp length as the independent variable, but this parameter had an adverse impact with respect to the ramp configurations. So it could not be assured that the number of ramp crashes increases according to the length of ramp.

When compare the number of crashes per unit distance for each ramp configuration, the number of crashes tends to decrease as the length of the ramp increases, as shown in Figure 4.8. Each ramp configuration has a different average ramp length, however, so the relationship between the ramp length and the number of crashes cannot be determined without classifying the ramp configurations. The multilevel model proposed in this study classifies these configurations in order to properly investigate the impact of ramp length on the number of crashes. Then, the parameter estimation results for the multilevel model can be compared with those of the generalized linear model.







[Figure 4.8] Crashes per Unit Distance for Each Ramp Configuration

Ⅴ. Modeling

5.1. Model Specification

Based on the literature review and a basic statistical analysis, the multilevel model is designed as a 1-level model for individual ramps and a 2-level model for different ramp types. The generalized linear model was designed only for individual ramps. At the individual ramp level, the dependent variable is the number of annual crashes on a ramp (crashes/year) and the independent variables are the AADT (veh/day) and the ramp length (m). In the level for different ramp types, the independent variables consist of dummy variables that include on/off ramps and loop/semi-direct/direct ramps. Table 5.1 shows the variables for these two models.

[Table 5.1] Variables of Multilevel Model and GLM

Model	Dependent Variables	Independent Variables
Multilevel Model	Annual Crashes (crashes/year)	· Level-1 ln(AADT), ln(Length) · Level-2 On/Off Ramp (dummy variables), Loop/Semi-Direct/Direct (dummy variables)
Generalized Linear Model	Annual Crashes (crashes/year)	ln(AADT), ln(Length)

5.2. Multilevel Model

The multilevel model consists of a 1-level model for individual ramps and a 2-level model for the different ramp types. In the 1-level model, the log-transformed AADT and the ramp length are used as the independent variables.

The reason why the log-transformed variables are used in the model is that they can offset the impact of outliers and improve the model's goodness-of-fit. In the 2-level model, the intercept and slopes of the 1-level model (β_0, β_{AADT}, and β_{Length}) are estimated as the model parameters for the different ramp types.

To consider the different basic risk of a crash for each ramp type, as shown in the Eq. 5.1, β_0 is estimated with using the ramp dummy variables (off, semi, and loop) for the six types. Also, β_{Length} is estimated with using the dummy variables (semi and loop) to consider the different impacts of ramp length on the crash. But, β_{Length} is not estimated in the 2-level model based on the assumption that the impacts of the AADT are not different for each ramp types.

Level-1 $Y = \exp(\beta_0 + \beta_{AADT} \cdot \ln(AADT) + \beta_{Length} \cdot \ln(Length))$

Level-2

$$\beta_0 = \beta_{00} + \beta_{0,Off} \cdot Off + \beta_{0,Semi} \cdot Semi + \beta_{0,Loop} \cdot Loop$$

$$\beta_{AADT} = \beta_{AADT,0}$$

$$\beta_{Length} = \beta_{Length,0} + \beta_{Length,Semi} + \beta_{Length,Loop}$$

(Eq. 5.1)

where,

β_0 : constant (intercept)

β_i : coefficient of variable i

$\beta_{i,j}$: coefficient of individual variable i in group j

AADT : annual average daily traffic

Length : length of ramp

5.3. Generalized Linear Model

The generalized linear model is developed to be for comparison with the multilevel model. As in the 1-level model of the multilevel model, the log-transformed AADT and length of ramp are used as the independent variables in the model. As such, the generalized model cannot reflect the different impacts of each ramp type on the crash frequency. Therefore, the need for a multilevel model to improve accuracy and reliability can be confirmed by comparing these two models.

The Poisson distribution has been used in many previous studies, and the Poisson model assumes the average and variation. However, the crash count data has a larger variation than the average, and this is called an over-dispersion. So a negative binomial distribution is more appropriate for the crash prediction model. The dependent variable is assumed to follows the negative binomial distribution in this study. Eq. 5.2 shows the generalized linear model.

GLM $$Y = \exp(\beta_0 + \beta_{AADT} \cdot \ln(AADT) + \beta_{Length} \cdot \ln(Length))$$

where, β_0 : constant (intercept)

β_i : coefficient of variable i

AADT : Annual Average Daily Traffic

Length : Length of Ramp

(Eq. 5.2)

Ⅵ. Results

6.1. Results of Model Estimation

In this study, the multilevel model and the generalized linear model are developed as the crash prediction models and the model parameter estimation results show that all the variable coefficients are statistically reliable within a 95% significance level in the generalized linear model. However, some of the parameters are not reliable within a 90% significance level in the multilevel model. This is because of the different number of samples used. The generalized linear model has more samples because it does not consider ramp types. If a larger number of samples were obtained, the statistical reliability of these multilevel model parameters would be improved.

In the multilevel model, β_{00} has a negative coefficient, and $\beta_{0,off}$, $\beta_{0,semi}$, and $\beta_{0,Loop}$ have positive coefficients. This means that the off ramps have a higher number of crashes on average, and the semi and loop ramps also have a higher number of crashes than the direct ramp. The negative coefficient of the intercept is due to the exponential form of the model. The AADT parameter has positive coefficients in both models and their values are similar.

The ramp length parameter has three coefficients in the 2-level part of the multilevel model. The $\beta_{Length,0}$ has a positive coefficient and

the $\beta_{Length,\,Semi}$ and $\beta_{Length,\,Loop}$ have negative coefficients. This means that the number of crashes increase as the ramp length increases in direct ramps. But the number of crashes increases as the ramp length decreases in semi-direct and loop ramps. These results differ from those of the generalized linear model that show that the number of crashes increases only when ramp lengths increase. In this study, the negative effect of ramp length on the number of crashes was investigated by plotting the number of crashes per unit distance for each ramp configuration. According to the study by Park (2006), the direct ramp has a positive coefficient whereas semi-direct and loop ramps have negative coefficients in the individual crash prediction models. Considering these results, the generalized linear model is revealed that it cannot estimate the parameters for different ramp-type characteristics.

[Table 6.1] Model Estimation Results for Multilevel Model and GLM

Multilevel Model			Generalized Linear Model		
Parameters	Coefficient	Standard Error	Parameters	Coefficient	Standard Error
β_{0}			β_{0}	-9.89562**	1.341359
β_{00}	- 8.454836**	2.478182			
$\beta_{0,Off}$	1.706563**	0.188966			
$\beta_{0,Semi}$	6.193064	6.252065			
$\beta_{0,Loop}$	5.438950	5.751697			
β_{AADT}	0.761881**	0.078386	β_{AADT}	0.723789**	0.080477
β_{Length}			β_{Length}	0.629366**	0.188912
$\beta_{Length,0}$	0.105235	0.408223			
$\beta_{Length,Semi}$	- 0.876152	0.965750			
$\beta_{Length,Loop}$	- 0.732609	0.903740			

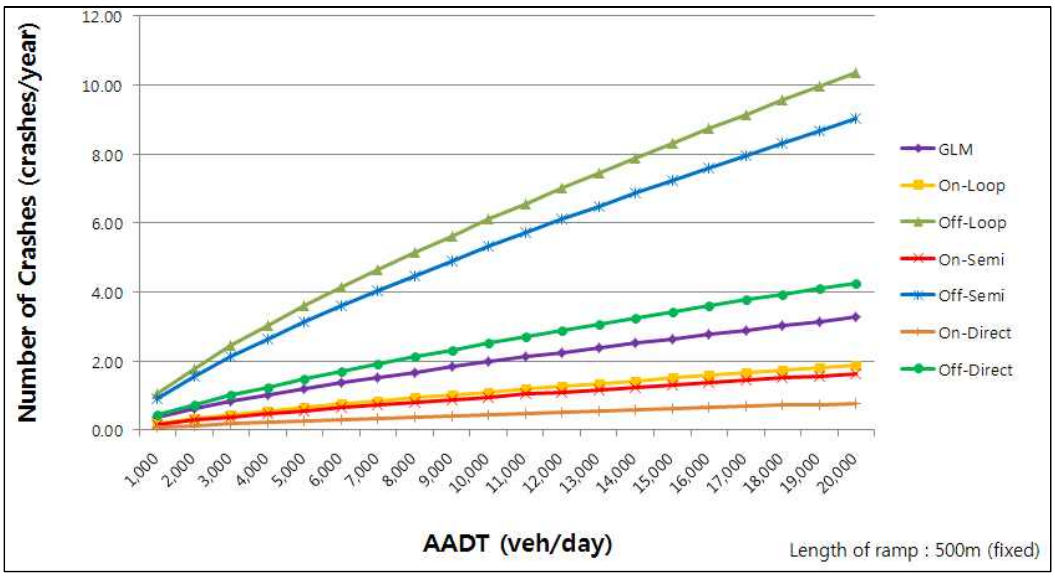
* $|P| < 0.1$, ** $|P| < 0.05$

6.2. Model Comparison

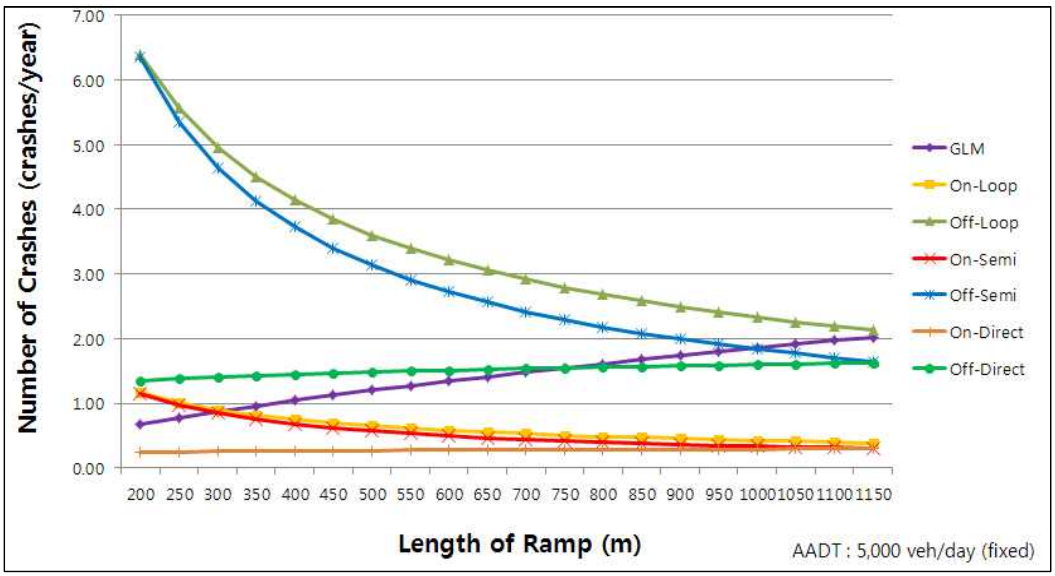
To more easily compare the prediction results of the two models, Figures 6.1 and 6.2 show plots of the predicted number of crashes for the AADT and for ramp length, respectively. However, in the actual data, ramps of the same length and type could have different AADT values. So the model validation would be implemented for the appropriate comparisons.

Figure 6.1 shows the number of crashes for the AADT for ramps with a fixed length of 500 m, which is similar to the average length of all freeway ramps. Comparing the predicted number of crashes in the multilevel model, all off ramps have higher numbers of crashes than on ramps. The different prediction results of the multilevel model can be seen in the graph. In contrast, the generalized linear model prediction of the number of crashes is roughly the average of the multilevel model results. This is because the generalized linear model cannot make distinctions between different ramp-type characteristics.

Figure 6.2 shows the predicted number of crashes for the two models with respect to ramp length under a fixed AADT of 5,000 veh/day, which is the average value for all freeway ramps. Similar to the results of Figure 6.1, off ramps have the higher number of crashes than other ramp types. While the number of crashes for direct ramps increases as the ramp length increases, the number of crashes for semi-direct and loop ramps decrease. However, the generalized linear model could not make these distinctions between ramp types.



[Figure 6.1] Predicted Number of Crashes for AADT



[Figure 6.2] Predicted Number of Crashes for Length of Ramp

6.3. Model Validation

This study investigated the accuracies of the two crash prediction models by comparing their goodness-of-fits. The quantitative indicators including the RMSE and MAD and the cumulative residuals (CURE) plot are adopted which is based on the residuals between the real number and the predicted number of crashes.

Washington et al. (2010) reported that the RMSE and MAD may be used to compare prediction accuracies in transportation modeling. Therefore, this study uses these as quantitative indicators.

The RMSE and MAD can be obtained using the residuals between the real number of crashes and the predicted number. Thus, a lower value would indicate a superior model for both indicators. The RMSE and MAD can be calculated using Eq. 6.1.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(X_i - F_i)^2}{n}} \qquad MAD = \frac{\sum_{i=1}^{n}|X_i - F_i|}{n}$$

(Eq. 6.1)

where,
X_i : ith observed number of crashes
F_i : ith predicted number of crashes
n : number of observations from 1 to i

Table 6.2 shows the results calculated for RMSE and MAD. The RMSE and MAD of the multilevel model are 2.10 and 1.03, respectively. Those of the generalized linear model are 2.37 and 1.27, respectively. Consequently, the multilevel model is confirmed to be more capable than the generalized linear model with respect to prediction accuracy.

[Table 6.2] RMSE and MAD of Multilevel model and GLM

Model	**Multilevel Model**	**Generalized Linear Model**
Root Mean Square Error	2.10	2.37
Mean Absolute Deviation	1.03	1.27

The CURE plot is also adopted in this study to compare the goodness-of-fit of the two models. Hauer (2004) suggested the CURE plot as a method for confirming the fit of a prediction model. In the CURE plot, cumulative residuals oscillating around zero, within $\pm 2\sigma^*(n)$, indicates that a model has excellent goodness-of-fit. Cumulative residuals that have only a positive or negative value indicate that a model has poor goodness-of-fit. If there is a range in which the cumulative residuals increase or decrease continuously, the model may be improved by transforming its functional form. If the cumulative residuals ascend or descend perpendicularly, the model may be affected by data outliers. The cumulative residuals and the upper and lower limits can be calculated using Eq. 6.2 and 6.3, respectively.

$$\hat{\sigma}(n) = \sum_{i=1}^{n}(X_i - F_i)$$

where,

$\hat{\sigma}(n)$: cumulative residuals (CURE)

X_i : ith observed number of crashes

F_i : ith predicted number of crashes

n : number of observations

(Eq. 6.2)

$$2\sigma^*(n) = 2\hat{\sigma}(n)\sqrt{1 - \frac{\hat{\sigma}^2(n)}{\hat{\sigma}^2(N)}}$$

$$\hat{\sigma}^2(n) = \sum_{i=1}^{n}(X_i - F_i)^2$$

(Eq. 6.3)

where,

$\pm 2\sigma^*(n)$: upper/lower limits of CURE plot

n : number of observations

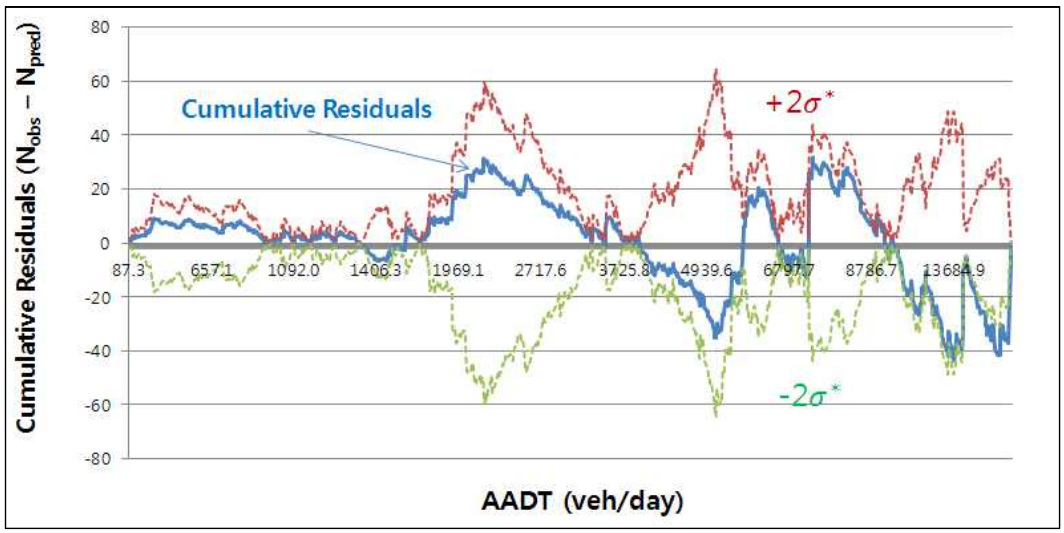
N : number of total observations

Bissonette and Cramer (2008) used the CURE plot to investigate the accuracy of crash prediction model for all values of AADT. According to this study, if the cumulative residuals fall extremely outside the standard deviation limits for a certain AADT range, the accuracy of the model decreases significantly for that AADT range.

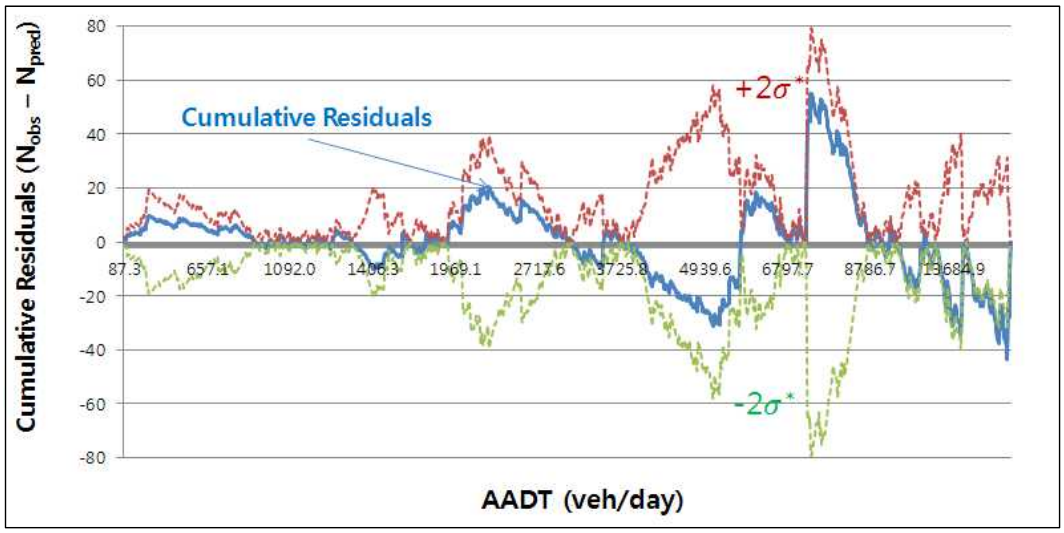
In this study, the cumulative residuals for the AADT are plotted on the graph. The CURE plots for the multilevel and generalized linear models are shown in Figures 6.3 and 6.4.

The results show that both of the models are within the range of $\pm 2\sigma^*(n)$, except for ranges of more than 15,000 veh/day. Moreover, both models oscillate around zero, so this means that both model forms are appropriate.

However, the perpendicular increase in width of the cumulative residuals around 7,000 veh/day in the generalized linear model is higher than that in the multilevel model. This means that the reduction of accuracy due to the impact of outliers is bigger in the generalized linear model than in the multilevel model. Consequently, the generalized linear model, which cannot consider the ramp-type characteristics, is affected more significantly by outliers. A comparison of the CURE plots of the two models shows that both of the models are marginally acceptable for the entire range of AADT.



[Figure 6.3] Cumulative Residuals versus AADT (Multilevel Model)



[Figure 6.4] Cumulative Residuals versus AADT (GLM)

Ⅶ. Conclusions

In this study, the two crash prediction models are developed based on the multilevel model and the generalized linear model. According to the parameter estimation results of the multilevel model, each ramp type has a different basic number of crashes and the impact of ramp length on the number of crashes is different for different ramp types. In contrast, the generalized linear model estimates the same parameters for all ramp types so it cannot reflect the different type characteristics. In other words, the generalized linear model does not consider different ramp-type characteristics.

To compare the goodness-of-fit of the two models, the RMSE and MAD are adopted as quantitative indicators, and the CURE plot based on the cumulative residuals is adopted. The RMSE and MAD reveal that the multilevel model is superior to the generalized linear model with respect to prediction accuracy. Also, the results of the CURE plot imply that the generalized linear model is more affected by data outliers than is the multilevel model.

The crash prediction model proposed in this study is able to take into account the different characteristics of ramp types in trumpet interchanges. However, ramps of the same type may have different geometric designs and traffic flow characteristics. As such, improved reliability of the proposed crash prediction model may be achieved by

considering these factors in a future study. In addition, the investigation for different impacts on crash frequency of accelerate/decelerate lanes and safety facilities for each ramp type on the crash frequency is suggested towards the development of safety improvement strategies.

References

Abdel-Aty, M., Pande, A., Lee, C., Gayah, V., Santos, C. D. Crash Risk Assessment Using Intelligent Transportation Systems Data and Real-Time Intervention Strategies to Improve Safety on Freeways. *Journal of Intelligent Transportation Systems*, Vol.11, No.3, pp.107-120, 2007.

American Association of State Highway and Transportation Officials. Highway Safety Manual 1st Edition. AASHTO, Washington, D.C., 2010.

Alluri, P. and Ogle, J. Effects of State-Specific SPFs, AADT Estimations, and Overdispersion Parameters on Crash Predictions Using SafetyAnalyst. *Proceeding from Transportation Research Board 91st Annual Meeting*, Washington, D.C., 2012.

Bauer, K. M. and Harwood, D. W. Statistical Models of Accidents on Interchange Ramps and Speed-Change Lanes. Publication FHWA-RD-97-106. FHWA, U.S. Department of Transportation, 1998.

Bissonette, J. A. and Cramer, P. C. Evaluation of the Use and Effectiveness of Wildlife Crossings. *Transportation Research Board,* NCHRP report. Vol.615, 2008.

Gelman, A. and Hill, J. Data Analysis Using Regression and Multilevel/Hierarchical Models. *Cambridge University Press*. 2007.

Harwood, D. W., Torbic D. J., Richard K. R. and Meyer M. M. SafetyAnalyst: Software Tools for Safety Management of Specific Highway Sites. Publication FHWA-HRT-10-063. FHWA, U.S. Department of Transportation, 2010.

Hauer, E. On Exposure and Accident Rate. *Traffic Engineering & Control*. Vol.36, No.3, pp.134-138, 1995.

Hauer, E. Statistical Road Safety Modeling. *Transportation Research Record: Journal of the Transportation Research Board*, Vol.1897, pp.81-87, 2004.

Huang, H. and Abdel-Aty, M. Multilevel Data and Bayesian Analysis in Traffic Safety. *Accident Analysis and Prevention*, Vol.42, No.6, pp.1556-1565, 2010.

Jones, A. P. and Jorgensen S. H. The Use of Multilevel Models for the Prediction of Road Accident Outcomes. *Accident Analysis and Prevention*, Vol.35, No.1, pp.59-69, 2003.

Korea Institute of Civil Engineering and Building Technology. Interchange Design Manual. 2005.

Lord, D. and Bonneson, J. A. Calibration of Predictive Models for Estimating Safety of Ramp Design Configurations. *Transportation Research Record: Journal of the Transportation Research Board*, Vol.1908, No.1, pp.88-95, 2005.

Miaou, S. P. The Relationship Between Truck Accidents and Geometric Design of Road Sections: Poisson versus Negative Binomial

Regressions. *Accident Analysis and Prevention*, No.26, pp.471-482, 1994.

Park, B. H. and Ryu, S. O. Relations Between Types of Trumpet Interchanges and Traffic Accident Rate. *Journal of the Korean Regional Development Association*, Vol.14, No.1, pp.107-124, 2002.

Park, H. S. Development of Accident Prediction Models on Freeway Interchange Ramps. M.Eng. Dissertation, Yonsei University, 2006.

Shankar, V. N., Albin, R.B., Milton, J.C., and Mannering, F.L. Evaluating Median Crossover Likelihoods with Clustered Accident Counts: An Empirical Inquiry Using the Random Effects Negative Binomial Model. *Transportation Research Record: Journal of the Transportation Research Board*, Vol.1635, No.1, pp.44-48, 1998.

Washington, S. P., Karlaftis, M. G., and Mannering, F. L. Statistical and Econometric Methods for Transportation Data Analysis. *CRC press*, 2010.

Yoon, B. J. Development of Traffic Accident Forecasting Models for Trumpet IC Ramp Types of Freeway. *Journal of the Korean Society of Civil Engineers*, D:27, No.1D, pp.81-87, 2007.

국문 초록

이 태 헌
서울대학교 대학원
건설환경공학부

고속도로 램프 구간은 본선 구간에 비해 단위 교통량 및 단위 거리 당 사고건수가 높은 사고다발지점일 뿐만 아니라, 사고 발생 특성과 기하구조 측면에서 차이가 존재하여 별도의 사고예측모형 개발이 필요하다. 국내 고속도로 램프 사고 중 70% 이상을 차지하는 트럼펫 IC의 경우, 램프의 형태(루프/준직결/직결)와 진입/진출에 따라 총 6가지 유형의 램프로 구성된다. 그러나 램프 유형에 따라 사고 발생률이 상이하여, 유형별 특성을 고려한 사고예측모형의 개발이 요구된다.

대다수의 기존 연구들에서 이용된 전통적 사고예측모형 구축 방법론인 일반화 선형모형은 모든 관측 자료가 독립적이라는 가정을 하기 때문에 램프 유형별 특성 차이를 모형에 고려하지 못한다. 따라서 본 연구에서는 기존 방법론이 갖는 한계점을 개선하기 위해 다층모형을 기반으로 한 사고예측모형을 구축하는 것을 목적으로 한다. 다층 모형은 다수의 객체가 하나의 그룹에 속하여 그룹내 상관성(intra-class correlation)을 갖는다는 점을 반영하여, 사고예측모형에서 램프 유형별 특성을 고려할 수 있다는 장점을 갖는다.

본 연구에서는 국내 고속도로 중 사고가 가장 많이 발생하는 상위 3개 노선(경부선, 영동선, 서해안선)을 대상으로 2007년부터 2010년까지 4년간 수집

된 1,155건의 램프 사고 자료를 분석에 이용하였다. 램프 유형별 특성을 고려할 수 있는 다층 모형과 고려할 수 없는 일반화 선형모형을 기반으로 두 사고예측모형들을 개발하였다. 모형에서는 램프 AADT, 램프 길이, 램프 유형을 독립변수로 이용하였다. 다층모형의 경우 1수준 모형은 개별 램프를 대상으로, 2수준 모형은 램프 유형을 대상으로 설계되었다. 일반화 선형모형은 램프 유형별 차이를 고려하지 못하므로 다층 모형의 1수준 모형과 동일한 구조로 설계되었다.

다층 모형 추정 결과, 램프 유형별로 기본 사고발생률에 차이가 존재하며, 램프 길이의 변화가 사고 발생에 미치는 영향력도 램프 유형별로 차이가 있는 것으로 나타났다. 반면 일반화 선형모형은 모든 램프들에 대해 동일한 파라미터들을 추정하여, 다층모형과 같이 램프 유형별 차이를 고려하지 못하는 것으로 나타났다. 두 모형의 적합도 비교를 위해, 본 연구에서는 정량적 비교지표인 RMSE, MAD와 CURE plot을 활용하였다. RMSE와 MAD 두 지표 모두에서 다층 모형의 정확도가 우수한 것으로 나타났으며, CURE plot 비교 결과 역시 다층 모형의 적합도가 상대적으로 우수한 것으로 나타났다.

본 연구에서 제시된 다층 모형은 기존의 일반화 선형모형이 가진 한계점을 개선한다는 측면에서 의의가 있다. 보다 신뢰도 높은 사고건수의 예측을 통해 정확한 사고다발지점의 선정이 가능하며, 나아가 램프 유형별 안전성 개선사업의 정량적 근거로 활용될 수 있을 것으로 예상된다.

· 주요어 : 사고예측모형, 다층 모형, 회귀 분석, 램프 유형
· 학 번 : 2013-20936