



Application of Random Effects Nonlinear Model for Analyzing Motorized and Nonmotorized Traffic Safety Performance

Juneyoung Park, Ph.D., Aff.M.ASCE¹; and Mohamed Abdel-Aty, Ph.D., P.E., F.ASCE²

Abstract: In this paper, we investigated the influence of the cross-sectional characteristics of bicycle lanes on motorized and nonmotorized traffic safety. The safety effects of bicycle lanes with different widths were assessed through the estimation of safety performance functions (SPFs) and crash modification factors (CMFs). Four different types of crash prediction models [i.e., generalized linear model (GLM), random effects GLM, generalized nonlinear model (GNM), and random effects GNM] were examined to develop more reliable SPFs. The results indicated that the random effects GNMs provided the most reliable estimates. The goodness of fit was higher for the GNMs than for the GLMs because of the nonlinear relationship between the width of the bicycle lanes and the crash rates. In addition, the random effects models indicated better performance than the GLMs and GNMs. The results indicated that the installation of bicycle lanes is an effective safety measure to reduce four different types of crashes in general. The results of the estimated CMFs using the random effects GNMs indicated that the safety effects of bicycle lanes had nonlinear variations based on different widths. In general, the installation of bicycle lanes with widths of 1.5–1.8 m (5–6 ft) was most effective for reducing motorized crashes, whereas bicycle lanes with widths of 1.8–2.1 m (6–7 ft) were more appropriate for reducing nonmotorized crashes. According to the findings of this study, specific guidance on minimum bicycle lane widths for various roadway characteristics, traffic flows, and roadway types based on empirical evidence regarding traffic safety can be provided. DOI: 10.1061/JTEPBS.0000485. © 2020 American Society of Civil Engineers.

Author keywords: Safety effects; Bicycle lanes; Random effects models; Generalized nonlinear models; Nonmotorized traffic safety.

Introduction

Whereas support for bicycle lanes on urban and suburban roadways has been growing as the levels of bicycling have increased in the United States, the safety issue between drivers and bicyclists has been recently recognized as one of the critical traffic safety problems. According to the National Highway Traffic Safety Administration (NHTSA 2018), approximately 840 bicyclists were killed in 2016 in the United States. Bicyclists generally use roadway shoulders and ride next to vehicles (veh). For shared lanes, the chances of conflict between motorized and nonmotorized traffics are higher because bicyclists can occupy and ride within the travel lane (shared lane) (Park et al. 2015).

For this reason, nearly every major city in the United States has made an effort in recent years to install bicycle lanes to provide bicyclists with a designated lane (path) to reduce the number of conflicts by separating bicyclists from vehicles and pedestrians. Implementation of bicycle lanes is challenging in different traffic conditions, and their safety effectiveness has not been well examined because the bicycle lane is a relatively new feature in the urban roadway cross section. Although safety issues related

to implementing bicycle lanes have been considered important, few studies have investigated the safety effects for different traffic flows (Fowler and Koorey 2006; Dumbaugh and Li 2010; Robartes and Chen 2017), roadway characteristics (Park et al. 2015; Abdel-Aty et al. 2016; Brunson et al. 2017), socioeconomic factors (Park et al. 2015; Osama and Sayed 2017; Lee and Abdel-Aty 2018), installation locations (Sadek et al. 2007; Reynolds et al. 2009; Kondo et al. 2018; Park et al. 2019), and geometric conditions (Lusk et al. 2011; Chen et al. 2012; Nosal and Miranda-Moreno 2012). According to the guidebook for bicycle facilities (AASHTO 2012), the minimum bicycle lane width should be 1.2 m (4 ft), and the recommended width is 1.5 m (5 ft). However, no specific guidance exists on the minimum bicycle lane widths for various roadway characteristics, traffic flows, and roadway types. In addition, the variations in the safety effects of bicycle lanes based on different widths have not been investigated.

According to the Highway Safety Manual (HSM) (AASHTO 2010), a crash modification factor (CMF) can express potential changes in crashes after treatment (or countermeasure) is implemented on a roadway segment or intersection. Observational before and after (B/A) studies and the cross-sectional (CS) method have been used to estimate the CMF (Gross et al. 2010; Srinivasan et al. 2010; Carter et al. 2012; Wu et al. 2015).

To develop CMFs using the CS method, the development of crash prediction models [i.e., safety performance functions (SPFs)] is required. It is commonly known that the generalized linear model (GLM) with negative binomial (NB) distribution can be utilized to develop SPFs. Using the coefficient of a specific variable associated with treatment (or countermeasure), a CMF can be calculated (Carter et al. 2012; Park and Abdel-Aty 2015b). However, the estimated CMFs from GLM cannot account for the nonlinear effects of the treatment because the coefficients in the GLM are assumed to be fixed. In an effort to account for the nonlinear effects of crash

¹Assistant Professor, Dept. of Transportation and Logistics Engineering, Hanyang Univ., 55 Hanyangdaehak-ro, Sangnok, Ansan, Gyeonggi-do 15588, South Korea (corresponding author). Email: juneyoung@hanyang.ac.kr

²Professor and Chair, Dept. of Civil, Environmental and Construction Engineering, Univ. of Central Florida, 12800 Pegasus Dr., Suite 211, Orlando, FL 32816. Email: m.aty@ucf.edu

Note. This manuscript was submitted on March 31, 2019; approved on September 14, 2020; published online on October 31, 2020. Discussion period open until March 31, 2021; separate discussions must be submitted for individual papers. This paper is part of the *Journal of Transportation Engineering, Part A: Systems*, © ASCE, ISSN 2473-2907.

predictors, many previous researchers have used the logarithm of annual average daily traffic (AADT) instead of AADT in their analysis. In contrast, some previous studies indicated nonlinear relationships between crash frequency and roadway characteristics with multiple variations (Park and Abdel-Aty 2015a; Lee et al. 2015). Due to the limitation of GLM in handling the nonlinearity of an explanatory variable, a methodology for applying the generalized nonlinear model (GNM) to analyze crash data has been suggested (Lao et al. 2014; Lee et al. 2015; Park and Abdel-Aty 2016). In the application of GNM, the development of a nonlinearizing link function to reflect the nonlinear effects with multiple variations is required.

The NB regression model assumes that a stochastic component is introduced into the relationship between traffic crashes and covariates. Although the NB regression model has been widely used in the past decades due to its strength for addressing overdispersion in crash prediction models, several important potential issues remain (e.g., spatial effects, heterogeneity) (Hauer et al. 2004; Lord and Mannering 2010). For this reason, random effects models have been used as an alternative approach in the application of traffic crash data (Shankar et al. 1998). Because roadway geometric and traffic parameters are likely to have location-specific effects, the random effects models might be more appropriate to account for the unobserved heterogeneity by introducing random variations (Chin and Quddus 2003; Anastasopoulos and Mannering 2009; Venkataraman et al. 2011; Mitra and Washington 2012; Yu et al. 2013; Bhat et al. 2014; Chen and Tarko 2014; Barua et al. 2015; Mannering et al. 2016).

The objective of this study was to evaluate the safety effects of bicycle lanes based on their different widths by developing SPFs and CMFs for different crash types and severities. Four different statistical modeling approaches (i.e., GLM, random effects GLM, GNM, and random effects GNM) were explored and compared to achieve more reliable estimates. In this paper, the KABCO severity scale could be defined as follows: fatal (K), incapacitating injury (A), nonincapacitating injury (B), possible injury (C), and property damage only (O). The crash types were defined as follows: all crash types with all severity levels (KABCO) as “Motorized,” all crash types with KABC severity levels as “Motorized-Injury,” bicycle-related crashes (including crashes with pedestrians) with all severity levels as “Nonmotorized,” and bicycle-related crashes with KABC severities as “Nonmotorized-injury.”

Methodology

Safety Performance Functions

The GLM with the NB distribution model is most commonly used to develop SPFs because the function can account for overdispersion. SPFs relate crash frequency to traffic and roadway characteristics. Two types of SPFs, which are full SPFs and simple SPFs, have been primarily used in the literature. Full SPFs relate the frequency of crashes to both traffic and roadway characteristics, whereas simple SPFs consider a traffic parameter (e.g., AADT) as an explanatory variable. It should be noted that the HSM provides CMFs calculated based on simple SPFs only. However, simple SPFs are oversimplified functions for reflecting the relationship between crash frequency and roadway characteristics because crash frequency is not only affected by traffic volume (Park et al. 2015). Therefore, in this study, a full SPF was developed to calculate CMFs using the CS method. The roadway length was considered as an offset in the model, as suggested by the HSM (AASHTO 2010). The functional form of SPF

for fitting the NB regression models is provided in Eq. (1) as follows:

$$N_{predicted,i} = \exp[\beta_0 + \ln(L_i) + \beta_1 \ln(AADT_i) + \dots + \beta_k(X_{ki})] \quad (1)$$

where $N_{predicted,i}$ = predicted crash frequency on segment i ; β_k = coefficients for the variable k ; $AADT_i$ = annual average daily traffic of segment i (veh/day); L_i = length of segment i ; and X_{ki} = linear predictor k of segment i .

Cross-Sectional Method

The CS method is a useful approach to estimate CMFs with insufficient crash data from before and after the actual application of a specific treatment. CS studies can be used to estimate CMFs when the date of the treatment installation is unknown, and the data for the period before treatment installation are not available (AASHTO 2010). CMF is calculated using the ratio of the average crash frequency of sites with the feature to the average crash frequency of sites without the feature (Carter et al. 2012). It should be noted that the safety effects estimated by the random effects models could vary by the random location-specific effects term and other covariates. For this reason, marginal effects have been used to provide the safety effects (i.e., CMF) of a specific parameter (Anastasopoulos and Mannering 2009). The marginal effects could be presented as the averages over the roadway segment to indicate the normalized safety impact (Anastasopoulos and Mannering 2009; Park et al. 2014, 2016). Moreover, the delta method was applied to present the variance in the estimated safety effects from the random effects models. This approach took the variance by using a function of a random variable about its mean based on a one-step Taylor approximation and expanding it to vector-valued functions of random vectors on the basis of deriving the asymptotic variance of maximum-likelihood estimators (Oehlert 1992; Rice 1994; Feiveson 1999).

Generalized Nonlinear Model

To account for the nonlinear effects of independent variables, an application of GNM using a nonlinearizing link function to assess the safety effects of treatments was proposed (Lao et al. 2014). The nonlinearizing link function could be described in any functional form, including linear, quadratic, log, and power, among others, for different values of y (Lee et al. 2015). The functional form of the nonlinearizing link function $[U(y)]$ was determined based on the relationship between the logarithm of the crash rate and the variable y . The functional form of GNM is provided in Eq. (2) as follows:

$$N_{predicted,i} = \exp(\beta_0 + \beta_1 \ln(AADT_i) + \beta_k(X_{ki}) + \gamma_l(U(y_{li}))) \quad (2)$$

where γ_l = coefficients for the nonlinear predictor l ; and y_{li} = nonlinear predictor l of segment i .

Since $U(y)$ varies with y , the CMF using GNM can be estimated by Eq. (3) as follows:

$$CMF = \exp\{\gamma_l \times (U(y_{lt}) - U(y_{lb}))\} \quad (3)$$

where y_{lt} = nonlinear predictor l of treated sites; and y_{lb} = nonlinear predictor l of untreated sites (baseline condition).

Random Effects Model

A fixed effects model can be estimated by treating it as a fixed effect linear regression model (Polachek and Yoon 1996). Basically,

the fixed effects model assumes that the individual specific effects are correlated with the independent variables, whereas the random effects model assumes that the individual specific effects are uncorrelated with the independent variables (Torres-Reyna 2007). Moreover, in the random effects model, the unobserved individual-specific heterogeneity is assumed to be completely unrelated to the explanatory-variable vector, which is a rather strong assumption (Mannering et al. 2016).

For this reason, random effects models have been applied to account for the unobserved heterogeneity by introducing random variations because roadway geometric and traffic parameters are likely to have location-specific effects (i.e., group-specific variations) (Chin and Quddus 2003; Anastasopoulos and Mannering 2009; Agüero-Valverde 2013; Yu et al. 2013; Chen and Tarko 2014; Naznin et al. 2016). As stated by Hausman et al. (1984), the unobserved heterogeneity effects can be captured by the random effects modeling approach considering both the fixed and random definitions of the individual effects. Based on this finding, an application to utilize the random effects models in traffic crash analysis was introduced (Shankar et al. 1998). The results of the study indicated that the random effects NB models could significantly improve the explanatory power of crash prediction models.

The random effects modeling approach can be adopted by introducing a random location-specific effects term into the relationship between the expected number of crashes ($\tilde{\lambda}_{it}$) and the covariate, X_{it} , of observation unit i in a given period t as in Eq. (4) as follows (Chin and Quddus 2003):

$$\tilde{\lambda}_{it} = \exp(X_{it}\beta + \delta_i) \quad (4)$$

where δ_i = random location-specific effects; and β = vector of regression of parameter estimates.

In this study, two different spatial geographic units (i.e., Block Groups and Census Tracts) were considered to introduce a random location-specific effects term into the GLMs and GNMs. For motorized crashes, using the models with a random effects term based on Block Groups was found to provide better performance than the models with the Census Tracts random effects term. In contrast, the models with the Census Tracts random effects term were more appropriate for nonmotorized crashes.

Data Preparation

In this study, two datasets were used for urban arterials in Seminole County, Florida: roadway characteristic data from the Roadway Characteristics Inventory (RCI) system and crash data from the Crash Analysis Reporting System (CARS) database for five years (2008–2012). The RCI database provides current and historical roadway characteristics data and reflects features of specific segments for the selected dates. In addition to the five years of crash data, crash records for 2007 were obtained to ensure a sufficient bicycle-related crash frequency. A total of 254 roadway segments with a total length of approximately 82.1 km (51 mi) were identified for the analysis. To identify the geographic entity, the spatial zone unit data were obtained from the US Census Bureau website using the PLANSafe Census Too 1 (Levine et al. 1995; Washington et al. 2010). The distributions of each variable among the target segments are presented in Table 1. From the comparison of crash frequencies between nonmotorized and nonmotorized-injury crashes, approximately 83% of nonmotorized crashes were injury crashes, whereas approximately 60% of all crashes were injury crashes. This result implied that a high risk of injury exists when nonmotorized crashes occurred.

Table 1. Descriptive statistics

Variable	Mean	Standard deviation
Motorized crashes	7.055	8.156
Motorized-injury crashes	4.244	4.890
Nonmotorized crashes	0.272	0.706
Nonmotorized-injury crashes	0.224	0.610
Natural logarithm of AADT (veh/day)	10.206	0.493
Length (km)	0.325	0.348
Number of lanes	4.114	1.547
Median width (m)	7.707	4.721
Median type (2 = median with barrier, 1 = median with no barrier, 0 = no median)	2 = 98 sites, 1 = 130 sites, 0 = 28 sites	
Roadway type (1 = multilane, 0 = two-lane or one-way)	1 = 247 sites, 0 = 9 sites	
Bicycle lane width (m)	1.397	0.436
Bicycle lane (1 = bicycle lane, 0 = no bicycle lane)	1 = 117 sites, 0 = 139 sites	

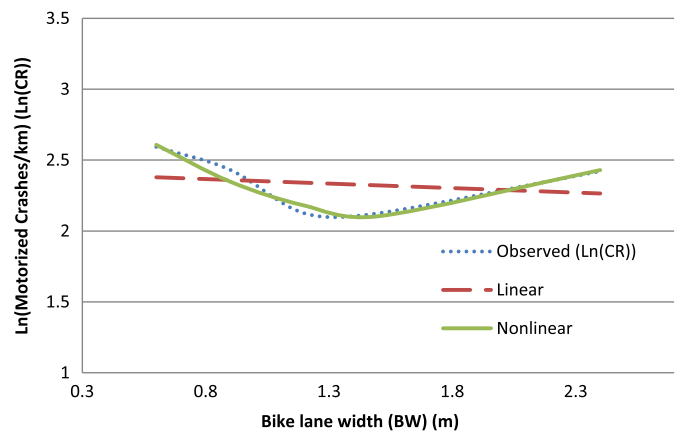


Fig. 1. Nonlinearizing link function for motorized crashes.

Results and Discussion

Nonlinearizing Link Function

The nonlinearizing link functions were developed to reflect the nonlinearity of bicycle lane widths on motorized and nonmotorized crashes, as indicated in Figs. 1 and 2. The relationship between the logarithm of crash rates and bicycle lane widths was plotted to determine the form of the nonlinearizing link function. The crash rate was defined as the number of crashes per mile. Various nonlinear equational forms were considered to associate the suitable form (Park and Abdel-Aty 2015b), and the quadratic nonlinear functional form was best fit for the relationship between crash rates and bicycle lane widths for motorized and nonmotorized crashes.

It should be noted that any correlation impacts and interaction effects between the crash rates and other parameters were also examined, but the nonlinear effects from any of the other parameters were not captured. The developed nonlinearizing link functions can be used as a nonlinear predictor in an analysis to improve model fit (Lao et al. 2014; Lee et al. 2015; Park and Abdel-Aty 2016).

The developed nonlinearizing link functions can be given as follows:

$$U_{BW_A} = 5.4438 - 0.7834 \times BW + 0.0736 \times BW^2 \quad \text{and}$$

$$U_{BW_B} = 5.8422 - 2.0657 \times BW + 0.1611 \times BW^2 \quad (5)$$

where U_{BW_A} = nonlinearizing link function for motorized crashes; and U_{BW_B} = nonlinearizing link function for nonmotorized crashes.

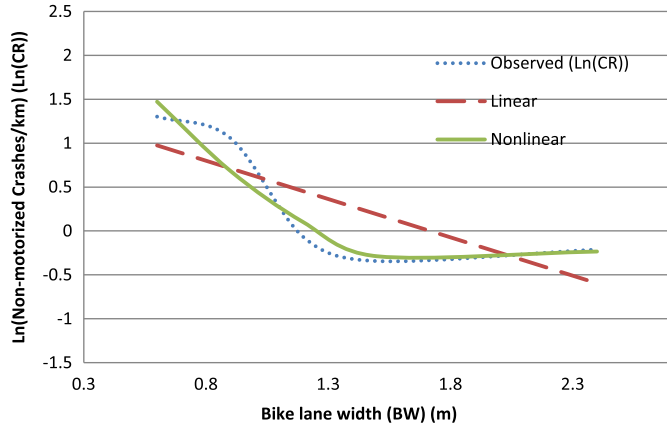


Fig. 2. Nonlinearizing link function for nonmotorized crashes.

Evaluation of SPFs and CMFs

The GLMs, random effects GLMs, GNMs, and random effects GNMs were evaluated for motorized, motorized-injury, nonmotorized, and nonmotorized-injury crashes, as indicated in Tables 2–5. Overall, the estimated parameters were statistically significant at a 90% confidence level. In Tables 4 and 5, the explanatory variable for bicycle lane width for the GLM was included for comparison purposes even though it was statistically insignificant. Moreover, the random effects GLMs for nonmotorized crashes were estimated; however, the level of between-group variability was not sufficient enough to warrant incorporating random effects into the models.

In general, the results indicated that crash frequency increased for the roadway segments with higher AADT values. Moreover, the results of the GLMs (for four different types of crashes) and random effects GLMs (for motorized crashes) indicated that the installation of bicycle lanes with wider widths reduced crash frequency. In contrast, although the evaluated and random effects GNMs also indicated that implementing bicycle lanes was effective for safety in general, the nonlinear effects of bicycle lanes based on different widths were identified. For motorized crashes, the crash frequency was higher for multilane roadways than for two-lane and one-way roadway segments.

The results indicated that the random effects GNMs provided the most reliable estimates. In detail, the goodness of fit was higher for the GNMs than GLMs due to the nonlinear relationship

Table 2. Estimated parameters of different models for motorized crashes

Parameter	GLM		Random effects GLM		GNM		Random effects GNM	
	Coefficient	p-value	Coefficient	p-value	Coefficient	p-value	Coefficient	p-value
Intercept	-7.6889	<0.0001	-9.3166	<0.0001	-9.4070	<0.0001	-10.5271	<0.0001
Log (AADT)	0.9252	<0.0001	1.0795	<0.0001	0.8703	<0.0001	1.0026	<0.0001
Bicycle lane width	-0.0611	0.0291	-0.0673	0.0379	N/A		N/A	
U_{BW_A}		N/A		N/A	0.1766	0.0153	0.1766	0.0287
Roadway type		N/S		N/S	1.3631	0.0092	1.0784	0.0480
Dispersion		1.0341		0.9441		1.0080		0.9196
δ		N/A		0.1373		N/A		0.0915
Log likelihood		1,478.52		1,473.15		1,472.56		1,469.75
AIC		1,486.52		1,483.15		1,482.56		1,481.75
BIC		1,500.67		1,492.07		1,500.25		1,492.46
HQIC		1,492.21		1,486.46		1,489.68		1,485.72

Note: δ = variance of random effects; N/S = not significant; N/A = not available; AIC = Akaike information criterion; BIC = Bayesian information criterion; and HQIC = Hannan–Quinn information criterion.

Table 3. Estimated parameters of different models for motorized-injury crashes

Parameter	GLM		Random effects GLM		GNM		Random effects GNM	
	Coefficient	p-value	Coefficient	p-value	Coefficient	p-value	Coefficient	p-value
Intercept	-7.0284	<0.0001	-7.9761	<0.0001	-8.6912	<0.0001	-9.1969	<0.0001
Log (AADT)	0.8124	<0.0001	0.9020	<0.0001	0.7593	<0.0001	0.8208	<0.0001
Bicycle lane width	-0.0563	0.0531	-0.0588	0.0660	N/A		N/A	
U_{BW_A}		N/A		N/A	0.1414	0.0739	0.1357	0.0857
Roadway type		N/S		N/S	1.4651	0.0145	1.3426	0.0290
Dispersion		0.9904		0.9074		0.9677		0.9162
δ		N/A		0.0751		N/A		0.0457
Log likelihood		1,257.17		1,255.01		1,252.22		1,252.22
AIC		1,265.17		1,265.01		1,262.22		1,262.22
BIC		1,279.32		1,273.93		1,279.91		1,260.27
HQIC		1,270.86		1,268.32		1,269.34		1,256.98

Note: δ = variance of random effects; N/S = not significant; N/A = not available; AIC = Akaike information criterion; BIC = Bayesian information criterion; and HQIC = Hannan–Quinn information criterion.

Table 4. Estimated parameters of different models for nonmotorized crashes

Parameter	GLM		GNM		Random effects GNM	
	Coefficient	p-value	Coefficient	p-value	Coefficient	p-value
Intercept	-20.7384	<0.0001	-20.6409	<0.0001	-21.6598	0.0073
Log (AADT)	1.8705	<0.0001	1.8070	<0.0001	1.8961	<0.0001
Bicycle lane width	-0.1012	0.1225		N/A		N/A
U_{BW_B}		N/A	0.1107	0.0715	0.1017	0.0983
Dispersion		1.8071		1.7409		1.5573
δ		N/A		N/A		0.1465
Log likelihood		301.91		299.31		298.69
AIC		309.91		307.31		308.69
BIC		324.06		321.46		306.73
HQIC		315.60		313.00		303.44

Note: δ = variance of random effects; N/S = not significant; N/A = not available; AIC = Akaike information criterion; BIC = Bayesian information criterion; and HQIC = Hannan–Quinn information criterion.

Table 5. Estimated parameters of different models for nonmotorized-injury crashes

Parameter	GLM		GNM		Random effects GNM	
	Coefficient	p-value	Coefficient	p-value	Coefficient	p-value
Intercept	-17.9120	<0.0001	-18.3610	<0.0001	-19.4590	0.0101
Log (AADT)	1.5857	<0.0001	1.5680	0.0001	1.6586	<0.0001
Bicycle lane width	-0.5237	0.1160		N/A		N/A
U_{BW_B}		N/A	0.1245	0.0527	0.1175	0.0655
Dispersion		1.5525		1.5053		1.2835
δ		N/A		N/A		0.1990
Log likelihood		273.83		272.44		271.49
AIC		281.83		280.44		281.49
BIC		295.98		294.59		279.54
HQIC		287.52		286.13		276.25

Note: δ = variance of random effects; N/S = not significant; N/A = not available; AIC = Akaike information criterion; BIC = Bayesian information criterion; and HQIC = Hannan–Quinn information criterion.

between the bicycle lane width and the crash rate. The random effects models also indicated better performance than the GLMs and GNMs. This result indicated that the inclusion of a location-specific random effects term in the regression models could enhance prediction reliability due to its strength in accounting for the unobserved heterogeneity across target segments. This result might have also been due to the geographic zonal unit being correlated with the socioeconomic and demographic parameters, such as income level, education level, and employment rate. These factors could contribute to crash risk (Huang et al. 2010; Abdel-Aty et al. 2013). Bicycle lane width should be noted as being alternatively considered as a categorical variable but was insignificant.

The CMFs for bicycle lanes with different widths for different crash types were developed using the CS method based on the random effects GNMs, as indicated in Fig. 3. All of the CMF values were statistically significant at a minimum confidence interval of 90% and were mostly significant at a 95% confidence interval. In the CS method, the segments with no bicycle lane were selected as the baseline. The results of the CMFs indicated that installing bicycle lanes is effective and safe for reducing motorized and nonmotorized crashes in general. The results also indicated that bicycle lanes are safer and more effective for reducing nonmotorized crashes than motorized crashes. For motorized crashes, the safety effects of the implementation of bicycle lanes were higher for all types of crashes than for injury crashes, whereas the opposite was detected for nonmotorized crashes. The estimated CMFs for motorized and nonmotorized crashes indicated that the safety effects of bicycle lanes decreased until a certain point [i.e., 1.5 m (5 ft)

bicycle lane width] and then increased clearly after this point. Although the nonlinearity of the developed CMFs for nonmotorized crashes presented similar variations to those of motorized crashes, the crash frequency decreased as the bicycle lane width increased until a width of 1.8 m (6 ft). Then, the crash frequency increased when the bicycle lane width exceeded 1.8 m (6 ft) for nonmotorized crashes. This result may have been due to the fact that drivers tend to regard a bicycle lane as a normal vehicle lane or parking area when the bicycle lane width is similar to the width of the vehicle travel lane, and adequate marking or signs are not correctly used (AASHTO 2012). In addition, drivers may be less cautious when they perceive that the bicycle lane has enough space for bicyclists and are unlikely to have conflicts with bicyclists. Similarly, bicyclists may not be aware of vehicles when they are using a wide bicycle lane (Park et al. 2015). Thus, this result indicated that the estimated CMFs using the GLMs (linear based regression models) could misrepresent the actual safety effects of bicycle lanes with different widths. The differences among the safety effects of bicycle lanes with a width of 1.8 m (6 ft) and greater for nonmotorized crashes were relatively small compared with those of the CMFs for motorized crashes.

Conclusions and Recommendations

The main objective of this study was to evaluate the safety effectiveness of the installation of bicycle lanes with different widths. GNMs were developed to reflect the nonlinear relationship between

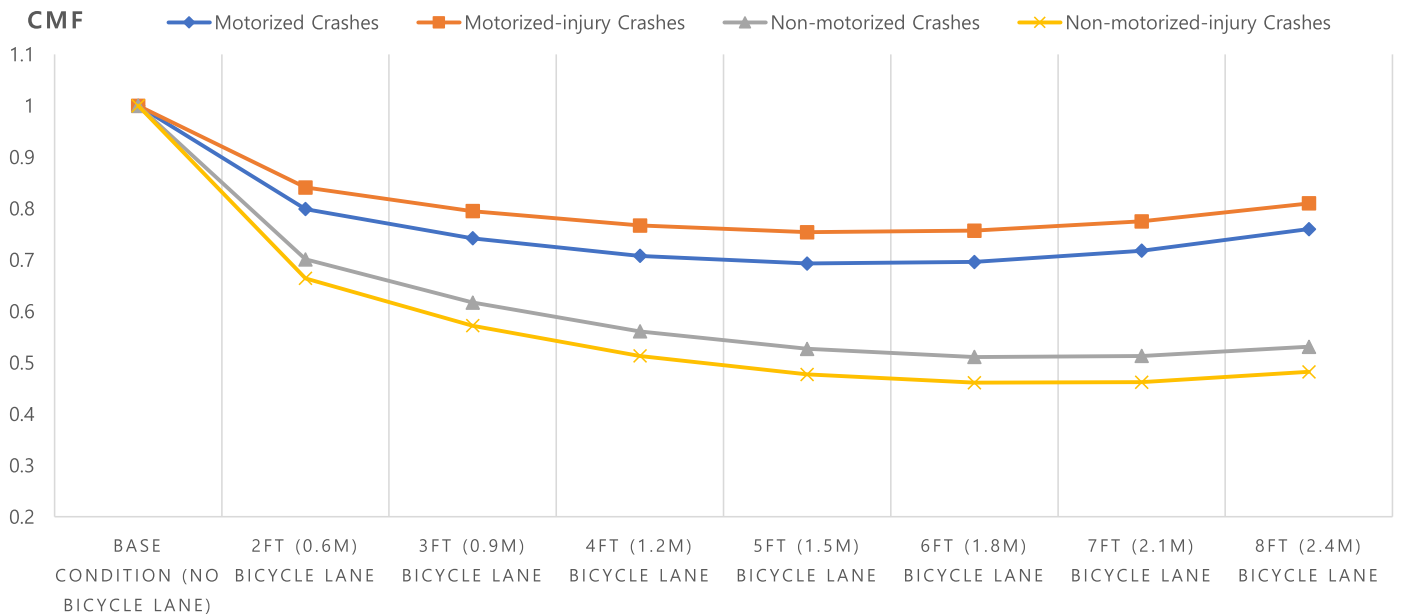


Fig. 3. Developed CMFs for implementing bicycle lanes with different widths.

the crash rates and the bicycle lane widths. For this purpose, the developed nonlinearizing link functions were used in the analysis. The quadratic nonlinear functional form was the best fit for this relationship. The developed nonlinearizing link functions were used in the GNMs to account for the nonlinear effects of changes to bicycle lane widths. The results indicated that the GNMs generally provided a better model fit than the GLMs. Therefore, the conclusion could be reached that including the nonlinearizing link functions in GNMs improved the goodness of fit of the models if the crash rates had a nonlinear relationship with specific parameters. Moreover, the random effects models were utilized to overcome the unobserved heterogeneity issue. Spatial geographic units based on Block Groups and Census Tracts were adopted to introduce a random location-specific effects term into the GLMs and GNMs. The inclusion of site-specific random effects terms in the regression models provided a better model fit than the traditional models. This result implied that the random effects modeling approach could increase the accuracy and reliability of crash prediction models.

In this study, four different types of regression models (i.e., GLMs, random effects GLMs, GNMs, and random effects GNMs) were applied and compared to develop more reliable crash prediction models based on the model performances. The random effects GNMs displayed advantages over the GLMs, random effects GLMs, and GNMs in terms of goodness of fit statistics and the relative performance of estimates.

The results of the CMFs indicated that the installation of bicycle lanes is a safe and effective method for reducing motorized and nonmotorized crashes. Bicycle lanes are safer and more effective at reducing nonmotorized crashes than motorized crashes. The results of the estimated CMFs using the random effects GNMs indicated that the safety effects of bicycle lanes had nonlinear variations based on their different widths. In general, installation of bicycle lanes with 1.5–1.8 m (5–6 ft) widths was the most effective at reducing motorized crashes, whereas bicycle lanes with widths of 1.8–2.1 m (6–7 ft) were more appropriate at reducing nonmotorized crashes. According to the guide book for bicycle facilities (AASHTO 2012), the minimum bicycle lane width should be 1.2 m (4 ft), and the recommended width is 1.5 m (5 ft). However, the

findings from this paper suggest that more specific guidance on minimum bicycle lane widths for various roadway characteristics, traffic flows, and roadway types should be developed using empirical evidence on traffic safety.

Despite the fact that the CS method has been broadly used to estimate safety performance, it should be applied with caution because of several potential limitations, such as correlation, spatial effects, causal effects, and omitted variable bias, among others (Persaud et al. 1999; Hauer 2004; Hauer et al. 2004; Lord and Bonneson 2007; Lord and Mannering 2010; Park and Abdel-Aty 2016). For this reason, CS studies can be applied when (1) the data are not available for the B/A approaches, and (2) checking the effectiveness of a single treatment separately from the effects of other treatments is difficult (Harkey et al. 2008; Bahar 2010), and applying the prescribed B/A studies on the changes in the widths of the roadway cross-sectional features is not feasible (e.g., lane width, median width, and shoulder width) (Carter et al. 2012; Park et al. 2015).

In future studies, the GNMs should be further improved by increasing the sample size and including additional, more detailed roadway characteristics. Also recommended is that the nonlinear relationships between the other treatments and crash rates should be investigated to reflect nonlinear variations in CMFs using GNMs. Whereas the random effects GNMs indicated better model performance due to their strength of simultaneously accounting for both nonlinearity and unobserved heterogeneity issues, the conflict between model complexity and applicability needs to be discussed further.

Although a random, location-specific effects term was introduced into the GLMs and GNMs to consider the different bike travel patterns according to geographic units, actual bicyclist volumes at each site were not available at the time of this study. Therefore, further work is needed to enhance the CMF development by including the volume of bicyclists. Lastly, the safety performance of specific roadway infrastructure under a connected and automated vehicle (CAV) environment needs to be evaluated (Wang et al. 2020).

Data Availability Statement

Some or all of the data, models, or code that support the findings of this study are available from the corresponding author on reasonable request.

Acknowledgments

This study was supported by the project titled “Development of Port Risk Prediction and Intelligent Port Safety Management Technologies,” funded by the Ministry of Oceans and Fisheries, Korea (20190399-08). All of the opinions and results are solely those of the authors.

Notation

The following symbols are used in this paper:

$AADT_i$ = annual average daily traffic of segment i (veh/day);

L_i = length of segment i ;

$N_{predicted,i}$ = predicted crash frequency on segment i ;

U_{BW-A} = nonlinearizing link function for motorized crashes;

U_{BW-B} = nonlinearizing link function for nonmotorized crashes;

X_{ki} = linear predictor k of segment i ;

y_{li} = nonlinear predictor l of segment i ;

y_{lt} = nonlinear predictor l of treated sites;

y_{lb} = nonlinear predictor l of untreated sites (baseline conditions);

β_k = coefficients for variable k ;

β = vector of regression of parameter estimates;

γ_l = coefficient for nonlinear predictor l ; and

δ_i = random location-specific effects.

References

- AASHTO. 2010. *Highway safety manual*. Washington, DC: AASHTO.
- AASHTO. 2012. *Guide for the development of bicycle facilities*. Washington, DC: AASHTO.
- Abdel-Aty, M., J. Lee, C. Siddiqui, and K. Choi. 2013. “Geographical unit based analysis in the context of transportation safety planning.” *Transp. Res. Part A: Policy Pract.* 49 (Mar): 62–75. <https://doi.org/10.1016/j.tra.2013.01.030>.
- Abdel-Aty, M., J. Park, J. H. Wang, and M. Abuzwidah. 2016. *Validation and application of highway safety manual (part D) and developing Florida CMF manual, phase 2*. Tallahassee, FL: Florida DOT.
- Agüero-Valverde, J. 2013. “Full Bayes Poisson gamma, Poisson lognormal, and zero inflated random effects models: Comparing the precision of crash frequency estimates.” *Accid. Anal. Prev.* 50 (Jan): 289–297. <https://doi.org/10.1016/j.aap.2012.04.019>.
- Anastasopoulos, P. C., and F. L. Mannering. 2009. “A note on modeling vehicle accident frequencies with random-parameters count models.” *Accid. Anal. Prev.* 41 (1): 153–159. <https://doi.org/10.1016/j.aap.2008.10.005>.
- Bahar, G. 2010. *Methodology for the development and inclusion of crash modification factors in the first edition of the highway safety manual*. Transportation Research E-Circular (E-C142). Washington, DC: Transportation Research Board.
- Barua, S., K. El-Basyouny, and M. T. Islam. 2015. “Effects of spatial correlation in random parameters collision count-data models.” *Anal. Methods Accid. Res.* 5 (Jan): 28–42. <https://doi.org/10.1016/j.amar.2015.02.001>.
- Bhat, C. R., K. Born, R. Sidharthan, and P. C. Bhat. 2014. “A count data model with endogenous covariates: Formulation and application to roadway crash frequency at intersections.” *Anal. Methods Accid. Res.* 1 (Jan): 53–71. <https://doi.org/10.1016/j.amar.2013.10.001>.
- Brunson, C., A. Getman, S. Hostetter, and R. Viola. 2017. “Left-turn pedestrian and bicycle crash study.” In *Proc., 96th Annual Meeting of the Transportation Research Board*. Washington, DC: Transportation Research Board.
- Carter, D., R. Srinivasan, F. Gross, and F. Council. 2012. *Recommended protocols for developing crash modification factors*. Final Rep. NCHRP 20-7 (314). Washington, DC: Transportation Research Board.
- Chen, E., and A. P. Tarko. 2014. “Modeling safety of highway work zones with random parameters and random effects models.” *Anal. Methods Accid. Res.* 1 (Jan): 86–95. <https://doi.org/10.1016/j.amar.2013.10.003>.
- Chen, L., C. Chen, R. Srinivasan, C. E. McKnight, R. Ewing, and M. Roe. 2012. “Evaluating the safety effects of bicycle lanes in New York City.” *Am. J. Public Health* 102 (6): 1120–1127. <https://doi.org/10.2105/AJPH.2011.300319>.
- Chin, H. C., and M. A. Quddus. 2003. “Applying the random effect negative binomial model to examine traffic accident occurrence at signalized intersections.” *Accid. Anal. Prev.* 35 (2): 253–259. [https://doi.org/10.1016/S0001-4575\(02\)00003-9](https://doi.org/10.1016/S0001-4575(02)00003-9).
- Dumbaugh, E., and W. Li. 2010. “Designing for the safety of pedestrians, cyclists, and motorists in urban environments.” *J. Am. Plann. Assoc.* 77 (1): 69–88. <https://doi.org/10.1080/01944363.2011.536101>.
- Feiveson, A. H. 1999. *Explanation of the delta method*. College Station, TX: StataCorp.
- Fowler, M., and G. Koorey. 2006. “The effects of the pages road cycle lane on cyclist safety and traffic flow operations.” In *Proc., IPENZ Transportation Group Technical Conf.* Wellington, New Zealand: Transportation Group New Zealand.
- Gross, F., B. Persaud, and C. Lyon. 2010. *A guide to developing quality crash modification factors*. Rep. No. FHWA-SA-10-032. Washington, DC: Federal Highway Administration.
- Harkey, D. L., et al. 2008. *Accident modification factors for traffic engineering and ITS improvements*. NCHRP Rep. No. 617. Washington, DC: Transportation Research Board.
- Hauer, E. 2004. “Statistical road safety modeling.” *Transp. Res. Rec.* 1897 (1): 81–87. <https://doi.org/10.3141/1897-11>.
- Hauer, E., F. Council, and Y. Mohammedshah. 2004. “Safety models for urban four-lane undivided road segments.” *Transp. Res. Rec.* 1897 (1): 96–105. <https://doi.org/10.3141/1897-13>.
- Hausman, J. A., B. H. Hall, and Z. Griliches. 1984. *Econometric models for count data with an application to the patents-R&D relationship*. Rep. No. t0017. Cambridge, MA: National Bureau of Economic Research.
- Huang, H., M. Abdel-Aty, and A. Darwiche. 2010. “County-level crash risk analysis in Florida: Bayesian spatial modeling.” *Transp. Res. Rec.* 2148 (1): 27–37. <https://doi.org/10.3141/2148-04>.
- Kondo, M. C., C. Morrison, E. Guerra, E. J. Kaufman, and D. J. Wiebe. 2018. “Where do bike lanes work best? A Bayesian spatial model of bicycle lanes and bicycle crashes.” *Saf. Sci.* 103 (Mar): 225–233. <https://doi.org/10.1016/j.ssci.2017.12.002>.
- Lao, Y., G. Zhang, Y. Wang, and J. Milton. 2014. “Generalized nonlinear models for rear-end crash risk analysis.” *Accid. Anal. Prev.* 62 (Jan): 9–16. <https://doi.org/10.1016/j.aap.2013.09.004>.
- Lee, C., M. Abdel-Aty, J. Park, and J. H. Wang. 2015. “Development of crash modification factors for changing lane width on roadway segments using generalized nonlinear models.” *Accid. Anal. Prev.* 76 (Mar): 83–91. <https://doi.org/10.1016/j.aap.2015.01.007>.
- Lee, J., and M. Abdel-Aty. 2018. “Macro-level analysis of bicycle safety: Focusing on the characteristics of both crash location and residence.” *Int. J. Sustainable Transp.* 12 (8): 553–560. <https://doi.org/10.1080/15568318.2017.1407973>.
- Levine, N., K. E. Kim, and L. H. Nitz. 1995. “Spatial analysis of Honolulu motor vehicle crashes. II: Zonal generators.” *Accid. Anal. Prev.* 27 (5): 675–685. [https://doi.org/10.1016/0001-4575\(95\)00018-U](https://doi.org/10.1016/0001-4575(95)00018-U).
- Lord, D., and J. Bonneson. 2007. “Development of accident modification factors for rural frontage road segments in Texas.” *Transp. Res. Rec.* 2023 (1): 20–27. <https://doi.org/10.3141/2023-03>.
- Lord, D., and F. Mannering. 2010. “The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives.” *Transp. Res. Part A: Policy Pract.* 44 (5): 291–305.

- Lusk, A. C., P. G. Furth, P. Morency, L. F. Miranda-Moreno, W. C. Willett, and J. T. Dennerlein. 2011. "Risk of injury for bicycling on cycle tracks versus in the street." *Injury Prev.* 17 (2): 131–135. <https://doi.org/10.1136/ip.2010.028696>.
- Mannering, F. L., V. Shankar, and C. R. Bhat. 2016. "Unobserved heterogeneity and the statistical analysis of highway accident data." *Anal. Methods Accid. Res.* 11 (Sep): 1–16. <https://doi.org/10.1016/j.amar.2016.04.001>.
- Mitra, S., and S. Washington. 2012. "On the significance of omitted variables in intersection crash modeling." *Accid. Anal. Prev.* 49 (Nov): 439–448. <https://doi.org/10.1016/j.aap.2012.03.014>.
- Naznin, F., G. Currie, D. Logan, and M. Sarvi. 2016. "Exploring causes of tram-involved crashes using a random effects negative binomial model." In *Proc., Transportation Research Board 95th Annual Meeting*. Washington, DC: Transportation Research Board.
- NHTSA (National Highway Traffic Safety Administration). 2018. *Traffic safety facts, 2016 data: Bicyclists and other cyclists*. Washington, DC: USDOT.
- Nosal, T., and L. F. Miranda-Moreno. 2012. "Cycle-tracks, bicycle lanes, and on-street cycling in Montreal, Canada: A preliminary comparison of the cyclist injury risk." In *Proc., Transportation Research Board 91st Annual Meeting*. Washington, DC: Transportation Research Board.
- Oehlert, G. W. 1992. "A note on the delta method." *Am. Statistician* 46 (1): 27–29.
- Osama, A., and T. Sayed. 2017. "Evaluating the impact of socio-economics, land use, built environment and road facility on cyclist safety." In *Proc., 96th Annual Meeting of the Transportation Research Board*. Washington, DC: Transportation Research Board.
- Park, B. J., D. Lord, and C. Lee. 2014. "Finite mixture modeling for vehicle crash data with application to hotspot identification." *Accid. Anal. Prev.* 71 (Oct): 319–326. <https://doi.org/10.1016/j.aap.2014.05.030>.
- Park, B. J., D. Lord, and L. Wu. 2016. "Finite mixture modeling approach for developing crash modification factors in highway safety analysis." *Accid. Anal. Prev.* 97 (Dec): 274–287. <https://doi.org/10.1016/j.aap.2016.10.023>.
- Park, J., and M. Abdel-Aty. 2015a. "Assessing the safety effects of multiple roadside treatments using parametric and nonparametric approaches." *Accid. Anal. Prev.* 83 (Oct): 203–213. <https://doi.org/10.1016/j.aap.2015.07.008>.
- Park, J., and M. Abdel-Aty. 2015b. "Development of adjustment functions to assess combined safety effects of multiple treatments on rural two-lane roadways." *Accid. Anal. Prev.* 75 (Feb): 310–319. <https://doi.org/10.1016/j.aap.2014.12.012>.
- Park, J., and M. Abdel-Aty. 2016. "Evaluation of safety effectiveness of multiple cross sectional features on urban arterials." *Accid. Anal. Prev.* 92 (Jul): 245–255. <https://doi.org/10.1016/j.aap.2016.04.017>.
- Park, J., M. Abdel-Aty, and J. Lee. 2019. "School zone safety modeling in countermeasure evaluation and decision." *Transportmetrica A: Transp. Sci.* 15 (2): 586–601. <https://doi.org/10.1080/23249935.2018.1519646>.
- Park, J., M. Abdel-Aty, J. Lee, and C. Lee. 2015. "Developing crash modification functions to assess safety effects of adding bike lanes for urban arterials with different roadway and socio-economic characteristics." *Accid. Anal. Prev.* 74 (Jan): 179–191. <https://doi.org/10.1016/j.aap.2014.10.024>.
- Persaud, B., C. Lyon, and T. Nguyen. 1999. "Empirical bayes procedure for ranking sites for safety investigation by potential for safety improvement." *Transp. Res. Rec.* 1665 (1): 7–12. <https://doi.org/10.3141/1665-02>.
- Polachek, S. W., and B. J. Yoon. 1996. "Panel estimates of a two-tiered earnings frontier." *J. Appl. Econom.* 11 (2): 169–178. [https://doi.org/10.1002/\(SICI\)1099-1255\(199603\)11:2<169::AID-JAE373>3.0.CO;2-#](https://doi.org/10.1002/(SICI)1099-1255(199603)11:2<169::AID-JAE373>3.0.CO;2-#).
- Reynolds, C. C., M. A. Harris, K. Teschke, P. A. Cripton, and M. Winters. 2009. "The impact of transportation infrastructure on bicycling injuries and crashes: A review of the literature." *Environ. Health* 8 (1): 47. <https://doi.org/10.1186/1476-069X-8-47>.
- Rice, J. 1994. *Mathematical statistics and data analysis*. 2nd ed. Berkeley, CA: Duxbury Press.
- Robartes, E., and T. D. Chen. 2017. "Virginia automobile and bicycle crash safety analysis." In *Proc., 96th Annual Meeting of the Transportation Research Board*. Washington, DC: Transportation Research Board.
- Sadek, A. W., A. S. Dickason, and J. Kaplan. 2007. "Effectiveness of green, high-visibility bike lane and crossing treatment." In *Proc., Transportation Research Board 86th Annual Meeting*. Washington, DC: Transportation Research Board.
- Shankar, V., R. Albin, J. Milton, and F. Mannering. 1998. "Evaluating median crossover likelihoods with clustered accident counts: An empirical inquiry using the random effects negative binomial model." *Transp. Res. Rec.* 1635 (1): 44–48. <https://doi.org/10.3141/1635-06>.
- Srinivasan, R., J. Baek, and F. Council. 2010. "Safety evaluation of transverse rumble strips on approaches to stop-controlled intersections in rural areas." *J. Transp. Saf. Secur.* 2 (3): 261–278. <https://doi.org/10.1080/19439962.2010.508571>.
- Torres-Reyna, O. 2007. *Panel data analysis fixed and random effects using Stata (version 4.2)*. Princeton, NJ: Princeton Univ.
- Venkataraman, N., G. Ulfarsson, V. Shankar, J. Oh, and M. Park. 2011. "Model of relationship between interstate crash occurrence and geometrics: Exploratory insights from random parameter negative binomial approach." *Transp. Res. Rec.* 2236 (1): 41–48. <https://doi.org/10.3141/2236-05>.
- Wang, L., H. Zhong, W. Ma, M. Abdel-Aty, and J. Park. 2020. "How many crashes can connected vehicle and automated vehicle technologies prevent: A meta-analysis." *Accid. Anal. Prev.* 136 (Mar): 105299. <https://doi.org/10.1016/j.aap.2019.105299>.
- Washington, S., I. van Schalkwyk, D. You, K. Shin, and J. P. Samuelson. 2010. *Plansafe: Forecasting the safety impacts of socio-demographic changes and safety countermeasures*. Rep. No. CRP-CD-78. Washington, DC: Transportation Research Board.
- Wu, L., D. Lord, and Y. Zou. 2015. "Validation of crash modification factors derived from cross-sectional studies with regression models." *Transp. Res. Rec.* 2514 (1): 88–96. <https://doi.org/10.3141/2514-10>.
- Yu, R., M. Abdel-Aty, and M. Ahmed. 2013. "Bayesian random effect models incorporating real-time weather and traffic data to investigate mountainous freeway hazardous factors." *Accid. Anal. Prev.* 50 (Jan): 371–376. <https://doi.org/10.1016/j.aap.2012.05.011>.