Accident Analysis and Prevention 83 (2015) 203–213

Contents lists available at ScienceDirect

Accident Analysis and Prevention

journal homepage: www.elsevier.com/locate/aap

Assessing the safety effects of multiple roadside treatments using parametric and nonparametric approaches

Juneyoung Park*, Mohamed Abdel-Aty[1]

Department of Civil, Environmental and Construction Engineering, University of Central Florida, Orlando, Florida 32816-2450, United States

ARTICLE INFO

Article history:
Received 10 March 2015
Received in revised form 3 July 2015
Accepted 6 July 2015
Available online 25 August 2015

Keywords:
Safety effectiveness
Crash modification factors and functions
Multiple treatments
Cross-sectional method
Generalized nonlinear models
Multivariate adaptive regression splines

ABSTRACT

This study evaluates the safety effectiveness of multiple roadside elements on roadway segments by estimating crash modification factors (CMFs) using the cross-sectional method. To consider the nonlinearity in crash predictors, the study develops generalized nonlinear models (GNMs) and multivariate adaptive regression splines (MARS) models. The MARS is one of the promising data mining techniques due to its ability to consider the interaction impact of more than one variables and nonlinearity of predictors simultaneously. The CMFs were developed for four roadside elements (driveway density, poles density, distance to poles, and distance to trees) and combined safety effects of multiple treatments were interpreted by the interaction terms from the MARS models. Five years of crash data from 2008 to 2012 were collected for rural undivided four-lane roadways in Florida for different crash types and severity levels. The results show that the safety effects decrease as density of driveways and roadside poles increase. The estimated CMFs also indicate that increasing distance to roadside poles and trees reduces crashes. The study demonstrates that the GNMs show slightly better model fitness than negative binomial (NB) models. Moreover, the MARS models outperformed NB and GNM models due to its strength to reflect the nonlinearity of crash predictors and interaction impacts among variables under different ranges. Therefore, it can be recommended that the CMFs are estimated using MARS when there are nonlinear relationships between crash rate and roadway characteristics, and interaction impacts among multiple treatments.

1. Introduction

Crash modification factors (CMFs) are multiplicative factors that express the expected changes of crash frequency as a result of a specific treatment (or countermeasure) implemented on roadways. Among four main parts of the Highway Safety Manual (HSM) (AASHTO, 2010), part D provides a variety of CMFs for different roadway facilities such as rural two-lane, rural multilane roadways, and urban arterials. CMFs in part D have been developed using high-quality observational before-after studies that account for the regression to the mean threat. Observational before-after studies are well known methods for evaluating safety effectiveness and calculating CMFs of specific roadway treatments (Gross et al., 2010). Moreover, the cross-sectional method has been commonly applied to derive CMFs due to the ease with which data can be obtained compared to the before-after approaches. According to the HSM, the cross-sectional method is used when (1) the date of the treatment installation is unknown, (2) the data for the period before treatment installation are not available, and (3) the effects of other factors on crash frequency must be controlled for creating a crash modification function (CMFunction) (Abdel-Aty et al., 2014; Lee et al., 2015a).

Although the current HSM provides various CMFs for single treatments, there are no CMFs for multiple treatments to roadway segments and intersections. Due to the lack of sufficient CMFs for multiple treatments, the HSM provides combining method (i.e. multiplication of single treatments) to assess the combined safety effect. However, it is cautioned in the HSM that the combined safety effect of multiple CMFs may be over or under estimated. In particular, since the roadside elements are usually simultaneously applied to roadways and implemented at the same location, interaction effects among multiple roadside features need to be considered to overcome the issue of over- or under- estimation. In general, most previous studies have estimated single treatment effect with no attention for multiple treatments since it is hard to consider the safety effect of single treatment from other multiple treatments implemented at the same time using the observational

* Corresponding author. Tel.: +1 407 823 0300; fax: +1 407 823 3315.
E-mail addresses: jypark@knights.ucf.edu (J. Park), m.aty@ucf.edu (M. Abdel-Aty).
[1] Tel.: +1 407 823 5657; fax: +1 407 823 3315.

http://dx.doi.org/10.1016/j.aap.2015.07.008
0001-4575/

before-after studies (Harkey et al., 2008; Stamatiadis et al., 2011). According to Bonneson et al. (2007),Gross et al. (2009), Li et al. (2011), Park et al. (2014), and Park et al. (2015b), the CMFs need to be developed with consideration of simultaneous impact of more than one roadway characteristic to account for the combined safety effects of multiple treatments.

In order to assess safety effects of multiple roadway characteristics, CMFs have been evaluated using the cross-sectional method (Lord and Bonneson, 2007; Stamatiadis et al., 2009; Li et al., 2011; Carter et al., 2012; Park et al., 2014; Abdel-Aty et al., 2014; Park et al., 2015a; Lee et al., 2015a). To estimate the CMF using the cross-sectional method, development of safety performance functions (SPFs) or crash prediction models (CPMs) is required. Due to its strength of accounting for over-dispersion, generalized linear model (GLM) with negative binomial (NB) distribution has been widely used to develop SPFs. The CMFs can be calculated from the coefficient of the variable associated with specific treatment. However, the estimated CMFs from GLM cannot account for the nonlinear effect of the treatment since the coefficients in the GLM are assumed to be fixed. As one of the efforts to account for the nonlinear effects of crash predictors, many previous researchers have used the logarithm of Annual Average Daily Traffic (AADT) instead of AADT in the analysis (Abdel-Aty and Radwan, 2000; Harwood et al., 2000; Wong et al., 2007; Abdel-Aty and Haleem, 2011; Park et al., 2014; Wang and Abdel-Aty, 2014). Moreover, some previous studies found a nonlinear relationship between crash frequency and roadway characteristics (e.g. lane width and shoulder width) (Xie et al., 2007; Li et al., 2008; Li et al., 2011; Lee et al., 2015a).

Therefore, researchers have tried to apply different techniques to account for the nonlinearity of variables on crash frequency. For instance, an application of using generalized nonlinear model (GNM) was proposed by Lao et al. (2013). In GNMs, the nonlinear effects of independent variables to crash analysis can be captured by the development of nonlinearizing link function. The study found that GNM performs better than GLM since it can reflect nonlinear effects of AADT, shoulder width, grade, and truck percentage on rear-end crashes. Similar to this study, Lee et al. (2015a) estimated CMFs for changes of lane width using GNMs. The study developed nonlinearizing link functions to reflect the nonlinear effects of lane width and speed limit on crash frequency. The CMFs estimated using the GNMs reflect that narrower lanes reduce crashes for the lane width less than 12 ft whereas wider lanes reduce crashes for lane widths greater than 12 ft. It was concluded that the CMFs estimated using GNMs clearly reflect variations in crashes with lane width, which cannot be captured by the CMFs estimated using GLMs. Park et al. (2015b) found that the nonlinear relationship between safety effects of widening urban roadways and time changes. The study developed crash modification functions (CMFunctions) using a Bayesian regression model including the estimated nonlinearizing link function to incorporate the changes in safety effects of the treatment over time. It was found that including the nonlinearizing link functions in developing CMFunctions shows more reliable estimates, if the variation of CMFs with specific parameters has a nonlinear relationship. Moreover, data mining techniques have been applied in the evaluation of safety impacts of roadway features to consider nonlinear effects. Li et al. (2011) utilized the generalized additive model (GAM) to estimate the safety effects of combinations of lane and shoulder width on rural frontage roads in Texas. Similarly, Zhang et al. (2012) applied the GAM to determine the nonlinear relationships between crash frequency and exposure for different segment types. However, most studies investigated only the main effect of each variable, but not the effects of interaction between variables. In addition, the applicability of random parameters modelling approaches has been discussed and tested in order to account for the variations of the effects of variables (or heterogeneity) across observations (Eluru et al., 2008; Anastasopoulos and Mannering, 2009; Venkataraman et al., 2013; Xu and Huang, 2015). However, although the variation of the effects of variables is not fixed and the approach can account for heterogeneity among different sites, interaction impacts between variables were not considered in most studies.

In order to account for both nonlinear effects and interaction impacts between variables, another data mining technique, the multivariate adaptive regression splines (MARS), have been used in safety evaluation studies. According to Briand et al. (2004), the MARS accommodate nonlinearity of independent variables and interaction effects for complex data structure. Unlike other data mining and machine learning techniques, the MARS is a non-black-box model and making it advantageous in the analysis of traffic safety (Haleem et al., 2013). Harb et al. (2010) applied MARS to assess safety effects of toll-lane processing time. Haleem et al. (2010) used MARS to analyze rear-end crashes at un-signalized intersections in Florida. Both studies found that the MARS can be superior to the traditional models and have high model performance. Haleem et al. (2013) also applied MARS to develop CMFs for changes of median width and inside and outside shoulder widths on urban freeway interchange influence areas for total and injury crashes. The study shows that MARS models outperformed the NB models based on their prediction performance and goodness-of-fit statistics. However, the uniform truncated basis functions were used for both total and injury crashes although the rate of changes can vary within the range for different crash types or severity levels. A number of studies addressed the safety effects of roadside features on roadway crashes. The roadside countermeasures have been known as one of the most important treatments for roadway safety to reduce injury crashes (Elvik et al., 2009). The study summarized the aggregate effects of roadside features on injury crash reduction. Other studies have assessed the safety effects of particular roadside elements such as rumble strips, shoulder widths, guardrails, barriers, poles, bridges, signs, ditches and side slopes (Turner, 1984; Good et al., 1987; Gattis et al., 1993; Hadi et al., 1995; Zegeer and Council, 1995; Viner, 1995; Kennedy, 1997; Reid et al., 1997; Bateman et al., 1998; Ray, 1999; Griffith, 1999; Lee and Mannering, 2002; Carrasco et al., 2004; Patel et al., 2007; Jovanis and Gross, 2008; Harkey et al., 2008; Torbic et al., 2009; Wu et al., 2014; Park et al., 2014; Park and Abdel-Aty, 2015 Wu et al., 2014; Park et al., 2014; Park and Abdel-Aty, 2015). As stated by Park et al. (2014), although it is important to examine the interaction impact of multiple treatments implemented on the same location such as roadside, there is a lack of studies that have dealt with this issue.

Thus, the objectives of this study are (1) to analyse the safety effects of multiple roadways and roadside elements using NB, GNM, and MARS, and (2) to develop the CMFs using cross-sectional method for single and multiple treatments for different crash types and severity levels. The remainder of this study is organized as follows. The second section describes data collection and preparation. The third section describes methodologies. The fourth section presents and discusses the results. The final section draws conclusions. In this paper, we refer to 'All crash types (KABCO severities)' as Total crashes, 'All crash types (KABC severities)' as Injury crashes, 'All crash types (KAB severities)' as Severe crashes, and 'Run-off roadways crashes (KABCO severities)' as ROR crashes for different crash severity levels. Crash severities were categorized according to the KABCO scale as follows: fatal (K), incapacitating injury (A), non-incapacitating injury (B), possible injury (C) and property damage only (O).

2. Data preparation

In this study, the road geometry data for roadway segments were identified for 5 years (2008–2012) and crash records were collected for 5 years (2008–2012) from multiple sources maintained by the Florida Department of Transportation (FDOT). These include the Roadway Characteristic Inventory (RCI) and Crash Analysis Reporting System (CARS) database. The CARS contains crash data for Florida State from 2003. The RCI database provides current and historical roadway characteristics data and reflects features of specific segments for the selected dates. For the application of cross-sectional method, it is recommended in the HSM that crash prediction models are developed using the crash data for both treated and untreated sites for the same time period—typically 3–5 years (AASHTO, 2010). Moreover, the cross-sectional method requires much more samples than the observational before-after study (e.g. 100 ~ 1000 sites) (Carter et al., 2012).

Although the RCI database provide more than 200 roadway characteristics for a specific roadway segment in a given date, it does not have information of more detailed roadside features such as number of utility poles, number of signs, number of isolated trees or groups, number of driveways, distance to poles, distance to signs, distance to trees, etc. Therefore, extensive effort by the research team was needed to use Google Earth and Street-view applications to identify these roadside elements. The Google Earth and Street-view applications have recently started to provide historical images and surrounding views from 2007 to recent. In this study, each roadway segment has uniform geometric characteristics for five years except AADT. Also, AADT in 2010 was used as an average AADT for the period 2008–2012. A total of 222 rural undivided four-lane roadway segments with 81.758 miles in length were identified as target sites. A segment is represented by roadway identification numbers and beginning and end mile points. Segments do not necessarily have equal length. However, very short segments (<0.1 mi) were excluded from the analysis because crash rates (=crash frequency per mile) may be exceptionally high on these segments even for a small number of crashes. This high crash rate for a small number of crashes for short segment might cause a misinterpretation in developing non-linearizing link function (Lee et al., 2015a). It is better noting that the data for roadway pavement condition of each site was also collected from RCI due to its significant effects on crash frequency and severity alongside the traffic and roadway characteristics (Burns, 1981; Shankar et al., 1995; Al-Masaeid, 1997; Tighe et al., 2000; Anastasopoulos et al., 2008; Milton et al., 2008; Mayora and Piña, 2009; Pande and Abdel-Aty, 2009; Anastasopoulos and Mannering, 2011; Ahmed et al., 2011; Anastasopoulos et al., 2012a, b; Buddhavarapu et al., 2013; Fernandes and Neves, 2013; Li et al., 2013; Lee et al., 2015b). However, since (1) there are no much changes of pavement condition for target sites, (2) the RCI data for roadway pavement condition has some missing values, and (3) it was difficult to verify and collect manually through Google Earth images for missing information of pavement condition, it was not used in the analysis. Distributions of each variable among these treated segments are summarized in Table 1.

3. Methodology

3.1. Cross-sectional method

The cross-sectional method is also known as SPFs or crash prediction models. The GLM with NB distribution model is most commonly used to develop a SPF since the function can account for over-dispersion. The SPF relates the crash frequency to traffic and roadway characteristics. The functional form of SPF for fitting the NB regression models is shown in Eq. (1). The roadway length was considered as an offset in the model as suggested by the HSM. It is worth to mention that the NB models with a variable for roadway length were also developed. However, considering roadway length as an offset showed better model fitness than inclusion of roadway length as a variable in the model.

$$N_{\text{predicted},i} = \exp(L_i + \beta_0 + \beta_1 \ln(\text{AADT}_i) + \cdots + \beta_k(X_{ki})) \quad (1)$$

where $N_{\text{predicted},\ i}$ = predicted crash frequency on segment i, L_i = roadway length of segment i, β_k = coefficients for the variables k, AADT_i = annual Average Daily Traffic of segment i (veh/day), X_{ki} = linear predictor k of segment i.

The cross-sectional method is a useful approach to estimate CMFs if there are insufficient crash data before and after a specific treatment that is actually applied. According to the HSM, the cross-sectional 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. As stated by Carter et al. (2012), the CMF is calculated by taking the ratio of the average crash frequency of sites with the feature to the average crash frequency of sites without the feature. Thus, the CMFs can be estimated from the coefficient of the variable associated with the treatment as the exponent of the coefficient when the form of the model is log-linear (Lord and Bonneson, 2007) as shown in Eq. (2).

$$\text{CMF} = \exp\{\beta_k \times (x_{kt} - x_{kb})\} \quad (2)$$

where x_{kt} = linear predictor k of treated sites, x_{kb} = linear predictor k of untreated sites (baseline condition).

Table 1
Descriptive statistics of target segments.

Variable	Mean	S.D.	Min.	Max.
Crash frequency				
Number of total crashes	3.027	5.856	0	37
Number of injury crashes	1.270	2.342	0	19
Number of severe crashes	0.635	1.413	0	15
Number of ROR crashes	0.257	1.134	0	15
Variables related to traffic and basic roadway geometric characteristics				
AADT (veh/day)	14654.604	8650.731	1500	34500
Length (mile)	0.368	0.427	0.1	3.0
Lane Width (ft)	11.243	0.956	9.5	15
Maximum speed limit (mph)	34.82	4.8	25	55
Horizontal curve	One or more curved sections in the segment = 28sites. No curve = 194 sites			
Variables related to roadside characteristics				
Shoulder width (ft)	3.45	2.235	1.5	10
Driveway density (per mile)	28.306	14.993	0	76.749
Density of poles (per mile)	52.910	21.793	2.333	113.208
Average distance to Poles (ft)	3.752	2.378	0.5	19.5
Density of trees (per mile)	31.765	20.267	0	125.0
Average distance to trees (ft)	12.265	7.245	0	58.0

The standard error (SE) of the CMF can be calculated by Eq. (3) as follows (Harkey et al., 2008):

$$SE = \frac{\exp(\beta_k + SE_{\beta_k}) - \exp(\beta_k - SE_{\beta_k})}{2} \tag{3}$$

where SE = standard error of the CMF,
$SE\beta_k$ = standard error of the coefficient β_k.

If a geometric characteristic is expressed in a binary variable (e.g. treatment (=1) or no treatment (=0)), the CMF will be $\exp(\beta_k)$ or the odds ratio of the linear predictor k (x_k). However, it is worth to note that the GLM represents the effect of each predictor x on crash frequency as a single coefficient for all values of x – i.e. β (Lee et al., 2015a).

3.2. Generalized nonlinear models

To account for nonlinear effects of independent variables, Lao et al. (2013) proposed an application of GNM using a nonlinearizing link function to assess safety effects of treatments. The nonlinearizing link function can be described in any functional form including linear, quadratic, log, power, etc. for different values of y (Lee et al., 2015a). The functional form of nonlinearizing link function ($U(y)$) is determined based on the relationship between the logarithm of crash rate and the variable y (Lao et al., 2013). The functional form of GNM is shown in Eq. (4) as follow:

$$N_{\text{predicted},i} = \exp(L_i + \beta_0 + \beta_1 \ln(\text{AADT}_i) + \beta_k(X_{ki}) + \gamma_l(U(y_{li}))) \tag{4}$$

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

Since $U(y)$ varies with y, the CMF using GNM can be estimated by Eq. (5) as follow (Lee et al., 2015a):

$$\text{CMF} = \exp\{\gamma_l \times (U(y_{lt}) - U(y_{lb}))\} \tag{5}$$

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

3.3. Multivariate adaptive regression splines

According to Friedman (1991), the MARS analysis can be used to model complex relationships using a series of basis functions. Abraham et al. (2001) described that MARS as a multivariate piecewise regression technique and the splines can be representing the space of predictors broken into number of regions. Piecewise regression, also known as segmented regression, is a useful method when the independent variables, clustered into different groups, exhibit different relationships between the variables in these groups (Snedecor and Cochran, 1980). The independent variable is partitioned into intervals and a separate line segment is fit to each interval. The MARS divides the space of predictors into multiple knots (i.e. the boundary between regions) and then fits a spline functions between these knots (Friedman, 1991). The MARS model is defined as shown in Eq. (6) (Put et al., 2004). It is worth to note that log form of MARS model was fitted to develop CMFs in this study.

$$\hat{y} = \exp(b_0 + \textstyle\sum_{m=1}^{M} b_m B_m(x)) \tag{6}$$

where $\hat{y}$ = predicted response variable, b_0 = coefficient of the constant basis function, b_m = coefficient of the m_{th} basis function, M = number of non-constant basis functions, $B_m(x)$ = m_{th} basis function.

There are three main steps to fit a MARS model (Put et al., 2004; Haleem et al., 2013). The first step is a constructive phase, in which basis functions are introduced in several regions of the predictors using a forward stepwise selection procedure. The predictor and the knot location that contribute significantly to the model are searched and selected in an iterative way in this step. Also, the introduction of an interaction is checked so as to improve the model at the each iteration. The second step (pruning phase) performs backward deletion procedure to eliminate the least contributed basis functions. Generalized cross-validation (GCV) criterion is generally used in this pruning step to find best model. The GCV criterion can be estimated by Eq. (7). The last step, which is selection phase, selects the optimum MARS model from a group of recommended models based on the fitting results of each (Haleem et al., 2013).

$$\text{GCV}(M) = \frac{1\sum_{i=1}^{n}(y_i - \hat{y})^2}{n(1 - C(M)/n)^2}$$

$$C(M) = M + dM \tag{7}$$

where y_i = response for observation i, n = number of observations, $C(M)$ = complexity penalty function, d = defined cost for each basis function optimization.

4. Results and discussion

4.1. Developed nonlinearizing link functions

The nonlinearizing link functions were developed to reflect the nonlinearity of AADT and driveway density on crashes as shown in Figs. 1 and 2. The relationships between the logarithm of crash rates (ln(CR)) and AADT and driveway density were plotted to determine the form of nonlinearizing link function (Lee et al., 2015a). It is worth noting that interaction effects between the crash rates and other explanatory variables were also investigated, but it did not capture the nonlinear effects clearly from any other parameters. Moreover, AADT and driveway density were alternatively treated as categorical variables instead of continuous variables. Although, goodness-of-fit was improved with the categorical variables instead of a continuous variable, some categories were not statistically significant at a 95% confidence level. Thus, we were unable to detect statistically significant effects of changes in AADT and driveway density on the crash rate. A linear regression line was also fitted to the observed data but it does not clearly reflect the nonlinearity of each predictor. The developed nonlinearizing link function can be used as a nonlinear predictor in analysis to improve the model fit (Lao et al., 2013; Lee et al., 2015a).

The nonlinearizing link functions for AADT are summarized as shown in Eq. (8) as follows:

$$U_{\text{AADT}} \begin{cases} = 1.79 + 1.880(\text{Ln.AADT} - 8) & \text{Ln.AADT} \le 8 \\ = 1.79 - 1.108(\text{Ln.AADT} - 8)^2 & 8 \le \text{Ln.AADT} \le 8.5 \\ = 2.3 + 1.560(\text{Ln.AADT} - 9) & 8.5 < \text{Ln.AADT} \le 9 \\ = 2.3 + 0.482(\text{Ln.AADT} - 9)^2 & 9 \le \text{Ln.AADT} \end{cases} \tag{8}$$

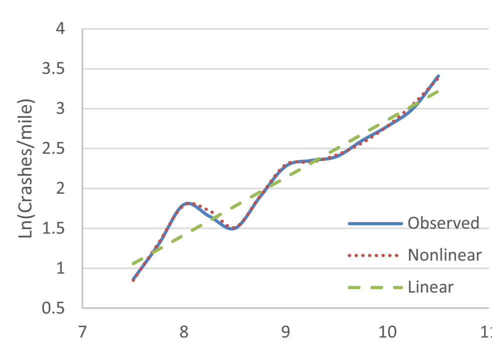
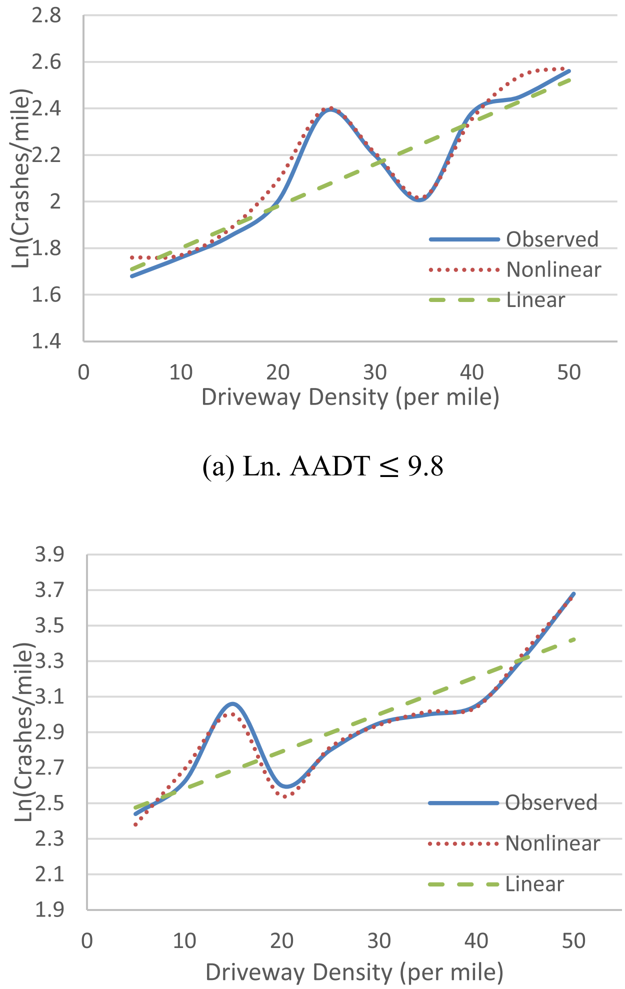


Fig. 1. Development of nonlinearizing link function for AADT.



(a) Ln. AADT ≤ 9.8

(b) Ln. AADT > 9.8

Fig. 2. Development of nonlinearizing link function for driveway density with different AADT levels.

According to the HSM, the safety effectiveness of changes of driveway density is function of driveway density with AADT changes. In this study, it was found that the correlation between driveway density and AADT is relatively high as more driveways tend to be increasing traffic volumes. This correlation can be captured by comparing the relationship between crash rate and driveway density under different AADT levels.

Due to the limitation of sample size, the nonlinearizing link functions for driveway density were developed under two ranges of AADT as shown in Eq. (9).

4.2. Generalized linear and nonlinear models

The GNMs with U_{AADT} and both U_{AADT} and $U_{DrivewayAADT}$ for total, injury, severe, and ROR crashes were developed using the nonlinearizing link functions as shown in Table 2. In order to compare model performance, the GLMs were also developed. In general, the estimated parameters were statistically significant at a 90% confidence level. Although the GNMs generally provided slightly better model fits (i.e. smaller AIC value) than the GLMs, the difference was not significant. This may be because there are interaction impacts among roadside features under different ranges of variables and these were not captured by the GNMs even though the nonlinearizing link functions are reflecting the nonlinearity effects of specific predictors.

Overall, the results of both GLMs and GNMs show that (1) increase of distance to poles, (2) increase of distance to trees, (3) decrease of driveway density, and (4) decrease of poles density reduce crash frequency. The safety effects of driveway density and poles density were selected for all different crash types whereas distance to poles was significant for total, injury, and ROR crashes. Moreover, the distance to trees was significant for total crashes only.

It was found that the GNMs with U_{AADT} only show better model fitness than the GNMs with both U_{AADT} and $U_{DrivewayAADT}$ for total, injury, and severe crashes whereas an opposite result was found for ROR crashes. However, there are no significant differences between the GNMs with U_{AADT} only and both U_{AADT} and $U_{DrivewayAADT}$. This indicates that the effects of inclusion of nonlinearizing link functions in the developing crash prediction models can vary based on different crash types.

4.3. Development of multivariate adaptive regression splines models

In this study, the ADAPTIVEREG procedure in the SAS program (SAS Institute Inc., 2012) was used to fit a MARS model. In the ADAPTIVEREG procedure, it is able to adjust maximum order of interactions using the MAXORDER option. It was found that there are no big difference between selecting the default condition (2-way maximum interactions) and increasing maximum number of interactions (e.g. 3-way or 4-way) in the analysis. Although increasing model complexity by adding more interactions might help improve predictive power for highly structured data, the applicability of model might be decreased. Thus, 2-way maximum order of interactions was used consistently for the different crash severities in this study. Moreover, the basis functions were constructed for each severity level since the rate of changes can vary within the range for different severities. It is worth to note that due to the low crash frequency, the MARS model for ROR crashes was not significant.

Table 3 presents the developed MARS models with NB distribution for total, injury, and severe crashes. In general, the estimated parameters of basis functions were statistically significant at a 90%

(a) Ln. AADT ≤ 9.8

$$U_{DrivewayAADT}\begin{cases} = 2.4 + 0.072(\text{Drwy.Den} - 25) + 0.002(\text{Drwy.Den} - 25)^2 & \text{Drwy.Den} \le 25 \\ = 2.4 - 0.038(\text{Drwy.Den} - 25) & 25 \le \text{Drwy.Den} < 35 \\ = 2.019 + 0.082(\text{Drwy.Den} - 35) - 0.003(\text{Drwy.Den} - 35)^2 & 35 \le \text{Drwy.Den} \end{cases}$$

(b) Ln.AADT > 9.8

$$U_{DrivewayAADT}\begin{cases} = 3.0 + 0.062(\text{Drwy.Den} - 15) & \text{Drwy.Den} \le 15 \\ = 3.0 - 0.092(\text{Drwy.Den} - 15) & 15 \le \text{Drwy.Den} < 20 \\ = 3.04 - 0.001(\text{Drwy.}\mathit{Den} - 40)^2 & 20 \le \text{Drwy.Den} \le 40 \\ = 3.04 + 0.063(\text{Drwy.Den} - 40) & 40 \le \text{Drwy.Den} \end{cases} \quad (9)$$

Table 2
Estimated parameters of GLMs and GNMs for different crash types.

(a) NB (GLM)

	Total crashes			KABC crashes			KAB crashes			ROR crashes		
Parameter	Coefficient	SE	*p*-value	Coefficient	SE	*p*-value	Coefficient	SE	*p*-value	Coefficient	SE	*p*-value
Constant	−10.2411	1.6393	<0.0001	−9.2788	1.5748	<0.0001	−10.7040	1.7656	<0.0001	−17.0584	3.6675	<0.0001
Ln(AADT)	1.0127	0.1668	0.0032	0.8047	0.1650	<0.0001	0.8210	0.1896	<0.0001	1.4405	0.3880	0.0002
Driveway density × Ln(AADT)	0.0024	0.0008	<0.0001	0.0021	0.0008	0.0071	0.0018	0.008	0.0199	0.0023	0.0013	0.0655
Poles density	0.0194	0.0054	0.0003	0.0174	0.0052	0.0008	0.0211	0.0057	0.0002	0.0194	0.0092	0.0355
Distance to poles	−0.1471	0.0590	0.0127	−0.1107	0.0595	0.0628	–	–	–	−0.2496	0.1313	0.0572
Distance to trees	−0.0288	0.0157	0.0672	–	–	–	–	–	–	–	–	–
Curve	1.0264	0.3168	0.0012	1.0185	0.3121	0.0011	1.1556	0.3067	0.0002	1.0397	0.5070	0.0403
Dispersion	1.5000			1.1288			0.7727			1.4532		
Log likelihood	−407.2575			−296.9135			−207.9855			−101.1665		
AIC	830.5149			607.8269			427.9711			216.3331		

(b) GNM with U_{AADT} only

	Total crashes			KABC crashes			KAB crashes			ROR crashes		
Parameter	Coefficient	SE	*p*-value	Coefficient	SE	*p*-value	Coefficient	SE	*p*-value	Coefficient	SE	*p*-value
Constant	−4.2188	0.7411	<0.0001	−4.5657	0.6657	<0.0001	−5.7501	0.6686	<0.0001	−8.0212	1.3603	<0.0001
U_{AADT}	1.4852	0.2443	<0.0001	1.2146	0.2374	<0.0001	1.1948	0.2642	<0.0001	1.9146	0.5050	0.0001
Driveway density × Ln(AADT)	0.0024	0.0008	0.0032	0.0020	0.0008	0.0083	0.0018	0.0008	0.0248	0.0023	0.0013	0.0719
Poles density	0.0178	0.0054	0.0009	0.0160	0.0052	0.0019	0.0197	0.0057	0.0005	0.0179	0.0094	0.0565
Distance to poles	−0.1349	0.0582	0.0205	−0.1029	0.0587	0.0794	–	–	–	−0.2309	0.1304	0.0767
Distance to trees	−0.0306	0.0156	0.0501	–	–	–	–	–	–	–	–	–
Curve	1.0453	0.3160	0.0009	1.0324	0.3091	0.0008	1.1725	0.3037	0.0001	1.0071	0.5057	0.0464
Dispersion	1.4781			1.0862			0.736			1.4795		
Log likelihood	−406.3469			−295.2479			−206.8915			−101.2897		
AIC	828.6938			604.4958			425.7829			216.5794		

(c) GNM with U_{AADT} and $U_{Driveway,\ AADT}$

	Total crashes			KABC crashes			KAB crashes			ROR crashes		
Parameter	Coefficient	SE	*p*-value	Coefficient	SE	*p*-value	Coefficient	SE	*p*-value	Coefficient	SE	*p*-value
Constant	−5.7366	1.0149	<0.0001	−5.8520	0.8890	<0.0001	−6.7111	0.8483	<0.0001	−9.5796	1.5788	<0.0001
U_{AADT}	1.5417	0.2460	<0.0001	1.2424	0.2354	<0.0001	1.2367	0.2615	<0.0001	1.9385	0.4936	<0.0001
$U_{Driveway,\ AADT}$	0.7761	0.3038	0.0106	0.6992	0.2749	0.0110	0.5269	0.2740	0.0545	0.8427	0.4461	0.0589
Poles density	0.0187	0.0054	0.0006	0.0161	0.0051	0.0017	0.0201	0.0057	0.0004	0.0177	0.0093	0.0575
Distance to poles	−0.1371	0.0589	0.0199	−0.1035	0.0588	0.0784	–	–	–	−0.2282	0.1294	0.0779
Distance to trees	−0.0266	0.0157	0.0895	–	–	–	–	–	–	–	–	–
Curve	1.0287	0.3173	0.0012	1.0178	0.3063	0.0009	1.1510	0.2641	0.0001	0.9931	0.498	0.0461
Dispersion	1.5030			1.0765			0.7430			1.4138		
Log likelihood	−407.3205			−295.3472			−207.5420			−101.1121		
AIC	830.6410			604.6945			427.0841			216.2243		

Table 3
Developed MARS models.

(a) MARS model for total crashes

Basis function	Basis function information	Coefficient	Standard error	*p*-value
BF0	Constant	−2.4285	0.5010	<0.0001
BF1	MAX (Poles density−41.852, 0)	0.0333	0.0095	0.0004
BF2	MAX (41.852−Poles density, 0)	−0.0859	0.0256	0.0008
BF3	MAX (Ln. AADT−8.501, 0)	2.5740	0.3938	<0.0001
BF4	MAX (8.501−Ln. AADT, 0)	−3.8338	1.0863	0.0004
BF5	MAX (Distance to trees−9.365, 0)	0.1424	0.0472	0.0025
BF6	MAX (9.365−Distance to trees, 0)	0.3297	0.1063	0.0019
BF7	MAX (Driveways density−25.237, 0)	–	–	–
BF8	MAX (25.237−Driveways density, 0)	−0.0753	0.0170	<0.0001
BF9	Curve (1 if exists; 0 otherwise)	–	–	–
BF10	BF6 × MAX (Driveways density−51.565, 0)	0.0680	0.0159	<0.0001
BF11	BF3 × MAX (Distance to trees−9.365, 0)	−0.1432	0.0413	0.0005
BF12	BF3 × MAX (9.365−Distance to trees, 0)	−0.2129	0.0823	0.0096
BF13	MAX (Poles density−76.233, 0)	−0.0555	0.0211	0.0084
BF14	BF3 × MAX (Distance to poles−4.0, 0)	−0.2105	0.0835	0.0117
BF15	BF3 × MAX (4.0−Distance to poles, 0)	−0.3563	0.2036	0.0802
BF16	BF9 × MAX (9.269−Ln. AADT, 0)	2.4186	0.6188	<0.0001
BF17	MAX (4.0−Distance to poles, 0)	0.4248	0.2519	0.0917
BF18	BF7 × MAX (Ln. AADT−9.815, 0)	−0.2014	0.0445	<0.0001
BF19	BF7 × MAX (16.892−Poles density, 0)	0.0514	0.0176	0.0034
BF20	BF17 × MAX (49.505−Poles density, 0)	0.0266	0.0121	0.0276

Dispersion = 0.8361
Log likelihood = −377.4936
AIC = 794.9871

(b) MARS model for injury crashes

Basis function	Basis function information	Coefficient	Standard error	*p*-value
BF0	Constant	0.7131	0.3206	0.0261
BF1	MAX (Ln. AADT−8.501, 0)	–	–	–
BF2	MAX (8.501−Ln. AADT, 0)	−2.0676	0.5329	0.0001
BF3	MAX (Poles density−93.75, 0)	–	–	–
BF4	MAX (93.75−Poles density, 0)	–	–	–
BF5	BF3 MAX (Driveways density−56.497, 0)	0.9660	0.2270	<0.0001
BF6	BF3 MAX (56.497−Driveways density, 0)	0.0038	0.0017	0.0221
BF7	Curve (1 if exists; 0 otherwise)	0.5760	0.2409	0.0168
BF8	MAX (Driveways density−25.281, 0)	0.0929	0.0233	<0.0001
BF9	MAX (25.281−Driveways density, 0)	−0.0506	0.0173	0.0034
BF10	BF8 MAX (Ln. AADT−8.882, 0)	−0.0545	0.0196	0.0053
BF11	BF8 MAX (8.882−Ln. AADT, 0)	−0.2300	0.0854	0.0071
BF12	BF8 MAX (Distance to poles−3.5, 0)	−0.0368	0.0104	0.0004
BF13	BF8 MAX (3.5−Distance to poles, 0)	−0.0370	0.0118	0.0018
BF14	BF7 MAX (8.854−Ln. AADT, 0)	4.6606	1.1947	<0.0001
BF15	BF2 MAX (Distance to trees−7.5, 0)	0.1085	0.0366	0.0030
BF16	BF2 MAX (7.5−Distance to trees, 0)	0.7279	0.1473	<0.0001
BF17	MAX (Distance to trees−5, 0)	–	–	–
BF18	MAX (5−Distance to trees, 0)	–	–	–
BF19	BF1 MAX (42.357−Driveways density, 0)	0.1606	0.0377	<0.0001
BF20	BF4 MAX (Ln. AADT−9.148, 0)	−0.0164	0.0085	0.0534
BF21	BF4 MAX (9.148−Ln. AADT, 0)	−0.0416	0.0170	0.0144
BF22	BF17 MAX (Poles density−76.233, 0)	−0.0114	0.0039	0.0037
BF23	BF17 MAX (76.233−Poles density, 0)	−0.0012	0.0005	0.0193
BF24	BF18 MAX (Poles density−93.75, 0)	−0.0911	0.0297	0.0022
BF25	BF18 MAX (93.75−Poles density, 0)	−0.0145	0.0042	0.0006

–
Dispersion = 0.2905
Log likelihood = −261.6967
AIC = 567.3934

(c) MARS model for severe crashes

Basis function	Basis function information	Coefficient	Standard error	*p*-value
BF0	Constant	−0.3702	0.2019	0.0668
BF1	MAX (Ln. AADT−9.976, 0)	3.6685	1.5189	0.0157
BF2	MAX (9.976−Ln. AADT, 0)	−2.6215	0.4549	<0.0001
BF3	MAX (Poles density−93.645, 0)	–	–	–
BF4	MAX (93. 645−Poles density, 0)	–	–	–
BF5	BF3 MAX (Driveways density−51.565, 0)	0.2382	0.0509	<0.0001
BF6	Curve (1 if exists; 0 otherwise)	1.1727	0.2471	<0.0001
BF7	BF2 MAX (Driveways density−19.841, 0)	0.0559	0.0146	0.0001
BF8	BF2 MAX (Distance to trees−6, 0)	0.1212	0.0332	0.0003

Table 3 (*Continued*)

(c) MARS model for severe crashes

Basis function	Basis function information	Coefficient	Standard error	*p*-value
BF9	BF2 MAX (6−Distance to trees, 0)	0.7754	0.2193	0.0004
BF10	BF4 MAX (Distance to trees−6, 0)	−0.0015	0.0004	0.0007
BF11	BF4 MAX (6−Distance to trees, 0)	−0.0080	0.0034	0.0177
BF12	BF1 MAX (18.018−Driveways density, 0)	0.5323	0.2246	0.0178
BF13	BF1 MAX (Poles density−50, 0)	−0.1052	0.0529	0.0467
BF14	BF1 MAX (50−Poles density, 0)	−0.5476	0.2467	0.0264

Dispersion = 0.1903
Log likelihood = −191.6311
AIC = 411.2623

confidence level. The basis functions are constructed by using truncated power functions based on knot values (Kuhfeld and Cai, 2013). The knots are automatically chosen in the ADAPTIVEREG procedure. In the MARS model for total crashes, the first basis function, BF0, is the intercept. The second basis function, BF1, is Poles Density—41.852 when Poles Density is greater than 41.852 and is 0 for otherwise (where the knot value is 41.852). Other basis functions are constructed in a similar manner by using different knot values. The results show that the MARS models generally provide better model fits than the GLMs and GNMs. This may be because the MARS can account for both nonlinear effects and interaction impacts between variables.

4.4. Estimation of crash modification factors

Table 4 presents a summary of the CMFunctions to estimate the safety effects (i.e. CMFs) of different roadside features for different severities. As stated previously, in the cross-sectional method, the CMF is estimated using the coefficient of the variable associated with a specific roadway characteristic in the exponential functional form (i.e. CMFunction). Since there were no big differences between GLMs (i.e. traditional NB models) and GNMs, the GLMs were compared with MARS models in Table 4. The results show that various interaction impacts among variables under different ranges based on knot values were found from MARS whereas one interaction impact between AADT and driveway density was found in the NB models. This indicates that the MARS can capture the interacting effects among multiple roadside elements based on different ranges of variables. It was found that for injury crashes, the basis functions related to distance to trees were selected in the MARS whereas it was not significant in the NB model. Similarly, for severe crashes, the basis functions for distance to trees found to be significant in the MARS whereas it was not selected in the NB models.

According to the HSM, the CMFs are multiplied to assess the combined safety effects of single treatments when the CMFs are estimated for same crash types (e.g. total crashes, night time crashes, bike related crashes, ROR crashes, etc.) and severity levels (e.g. injury, fatal, PDO, etc.). However, the HSM cautions that the multiplication of the CMFs may over- or under-estimate combined effects of multiple treatments. For instance, Park and Abdel-Aty (2015) found that the combined safety effects over-estimated the real safety effects of multiple treatments (shoulder rumble strips and widening shoulder width) by 4 to 10 percent when using the HSM procedure (multiply single CMFs to estimate combined safety effectiveness). This over-estimation may be because the two treatments are implemented on the same location (i.e. roadside) of roads. To overcome this limitation, interaction impacts among treatments need to be considered when they are implemented on the same location (e.g. roadside, mainline, median, etc.) of roadways. For this purpose, the MARS models can be recommended to assess the safety effects of multiple treatments due to its strength of accounting for the interaction impacts among variables. Table 5 presents an example of estimation and comparison of CMFs for single and multiple treatments from the GLM and MARS model for total crashes. Since the results from MARS model vary based on different original roadway characteristics (base conditions) whereas the GLM does not account for it, one sample base condition was set in the analysis. In Table 5, the base conditions of sample roadway are as follow: (1) AADT is 15,000 veh/day and no changes, (2) driveway density is 25 per mile, (3) poles density is 55 per mile, (4) distance from roadway to poles is 1 ft, and (5) distance from roadway to trees is 10 ft.

The results show that the single treatments and combinations are safety effective in reducing crashes by both GLM and MARS models. It was found that the CMFs of decreasing poles density and increasing distance to poles are similar whereas there are significant differences between the CMFs of decreasing driveway density for GLM and MARS. Similarly, there are 0.08 differences between the CMFs for increasing distance to trees for GLM and MARS. It can be noted that the standard errors of CMFs from GLM are relatively lower than the MARS since only one parameter from GLM is used to estimate the CMFs whereas multiple parameters including interaction terms are used in the MARS. According to the HSM, a standard error of 0.1 or less indicates that the CMF value is sufficiently accurate, precise, and stable. It also suggests that other related CMFs with standard errors of 0.2 to 0.3 may also be included to account for the effects of the same treatment on other facilities, other crash types or other severities. For example, the CMF of increasing distance to poles by 1 ft for total crashes is 0.788 with 0.073 standard error when the base conditions are as follow: (1) AADT is 15,000 veh/day and no changes, (2) driveway density is 60 per mile, (3) poles density is 30 per mile, (4) distance from roadway to poles is 4.5 ft, and (5) distance from roadway to trees is 7 ft. However, in Table 5, the CMF for increasing distance to poles by 1 ft is 0.894 with standard error of 0.192 for the given base conditions.

The combined safety effects over-estimated the real safety effects of multiple treatments by 8 to 10 percent when using the HSM procedure (multiply single CMFs to estimate combined safety effectiveness) compared to the results of estimation of CMFs from MARS. This result is consistent with Park and Abdel-Aty (2015). Since there is an interaction between driveway density and distance to trees when distance to trees is less than 9.365 ft and the distance to trees in the sample base condition is 10 ft, there was no difference between the combined CMF by HSM procedure and the real safety effect for the combination of decreasing driveway density and increasing distance to trees.

Therefore, it can be recommended that the MARS is used to assess the safety effects of multiple treatments to account for the interaction impacts among treatments, especially when they are

Table 4
Summary of CMFunctions for different crash types.

(a) Total crashes

Treatment	GLM		MARS	
	CMFunctions	Interaction term	CMFunctions	Interaction term
Driveway density (DD)	$\exp\{0.0024 \times (DD - Base_{DD}) \times Ln(AADT)\}$	AADT × DD	$\exp\{(\beta_8 \times BF8 + \beta_{10} \times BF10 + \beta_{18} \times BF18 + \beta_{19} \times BF19)$—Base condition}	DT × DD AADT × DD PD × DD
Poles density (PD)	$\exp\{0.0194 \times (PD - Base_{PD})\}$	–	$\exp\{(\beta_1 \times BF1 + \beta_2 \times BF2 + \beta_{13} \times BF13 + \beta_{19} \times BF19 + \beta_{20} \times BF20)$—Base condition}	PD × DD DP × PD
Distance to poles (DP)	$\exp\{-0.1471 \times (DP - Base_{DP})\}$	–	$\exp\{(\beta_{14} \times BF14 + \beta_{15} \times BF15 + \beta_{17} \times BF17 + \beta_{20} \times BF20)$—Base condition}	DP × AADT DP × PD
Distance to trees (DT)	$\exp\{-0.0288 \times (DT - Base_{DT})\}$	–	$\exp\{(\beta_5 \times BF5 + \beta_6 \times BF6 + \beta_{10} \times BF10 + \beta_{11} \times BF11 + \beta_{12} \times BF12)$—Base condition}	DT × DD AADT × DT

Note: Basis functions (BF_i) with estimated coefficient (β_i) are from Table 3(a).

(b) Injury crashes

Treatment	GLM		MARS	
	CMFunctions	Interaction term	CMFunctions	Interaction term
Driveway density (DD)	$\exp\{0.0021 \times (DD - Base_{DD}) \times Ln(AADT)\}$	AADT × DD	$\exp\{(\beta_5 \times BF5 + \beta_6 \times BF6 + \beta_8 \times BF8 + \beta_9 \times BF9 + \beta_{10} \times BF10 + \beta_{11} \times BF11 + \beta_{12} \times BF12 + \beta_{13} \times BF13 + \beta_{19} \times BF19$—Base condition}	AADT × DD DP × DD PD × DD
Poles density (PD)	$\exp\{0.0174 \times (PD - Base_{PD})\}$	–	$\exp\{(\beta_3 \times BF3 + \beta_4 \times BF4 + \beta_5 \times BF5 + \beta_6 \times BF6 + \beta_{20} \times BF20 + \beta_{21} \times BF21 + \beta_{22} \times BF22 + \beta_{23} \times BF23 + \beta_{24} \times BF24 + \beta_{25} \times BF25$—Base condition}	PD × DD AADT × PD PD × DT
Distance to poles (DP)	$\exp\{0.1107 \times (DP - Base_{DP})\}$	–	$\exp\{(\beta_{12} \times BF12 + \beta_{13} \times BF13$—Base condition	DP × DD
Distance to trees (DT)	–	–	$\exp\{(\beta_{15} \times BF15 + \beta_{16} \times BF16 + \beta_{17} \times BF17 + \beta_{18} \times BF18 + \beta_{22} \times BF22 + \beta_{23} \times BF23 + \beta_{24} \times BF24 + \beta_{25} \times BF25$—Base condition}	AADT × DT PD × DT

Note: Basis functions (BF_i) with estimated coefficient (β_i) are from Table 3(b).

(c) Severe crashes

Treatment	GLM		MARS	
	CMFunctions	Interaction term	CMFunctions	Interaction term
Driveway density (DD)	$\exp\{0.0018 \times (DD - Base_{DD}) \times Ln(AADT)\}$	AADT × DD	$\exp\{(\beta_5 \times BF5 + \beta_7 \times BF7 + \beta_{12} \times BF12$—Base condition}	AADT × DD PD × DD
Poles density (PD)	$\exp\{0.0211 \times (PD - Base_{PD})\}$	–	$\exp\{(\beta_3 \times BF3 + \beta_4 \times BF4 + \beta_5 \times BF5 + \beta_{10} \times BF10 + \beta_{11} \times BF11 + \beta_{13} \times BF13 + \beta_{14} \times BF14$—Base condition}	PD × DD PD × DT PD × AADT
Distance to poles (DP)	–	–	–	–
Distance to trees (DT)	–	–	$\exp\{(\beta_8 \times BF8 + \beta_9 \times BF9 + \beta_{10} \times BF10 + \beta_{11} \times BF11$—Base condition}	AADT × DT PD × DT

Note: Basis functions (BF_i) with estimated coefficient (β_i) are from Table 3(c).

(d) ROR crashes

Treatment	GLM		MARS	
	CMFunctions	Interaction term	CMFunctions	Interaction term
Driveway density (DD)	$\exp\{0.0023 \times (DD - Base_{DD}) \times Ln(AADT)\}$	AADT × DD	–	–
Poles density (PD)	$\exp\{0.0194 \times (PD - Base_{PD})\}$	–	–	–
Distance to poles (DP)	$\exp\{-0.2496 \times (DP - Base_{DP})$	–	–	–
Distance to trees (DT)	–	–	–	–

Table 5
Example of estimation of CMFs for a sample base condition.

Base condition AADT: 15,000/Driveway density: 25/Poles density: 55/Distance to poles: 1/Distance to trees: 10			
After treated condition AADT: 15,000/Driveway density: 20/Poles density: 50/Distance to poles: 2/Distance to trees: 11			
Treatments	GLM (NB)	MARS	
CMFs (S.E) by cross-sectional method			
Decreasing driveway density (DD)	**0.891** (0.001)	**0.686** (0.058)	
Decreasing poles density (PD)	**0.908** (0.005)	**0.847** (0.040)	
Increasing distance to poles (DP)	**0.863** (0.051)	0.894 (0.192)	
Increasing distance to trees (DT)	**0.972** (0.015)	0.896 (0.072)	
	Using HSM combining method (multiplication)		CMFs by cross-sectional method
DD + PD	0.891 × 0.908 = 0.809	0.686 × 0.847 = 0.615	**0.675** (0.120)
DD + DP	0.891 × 0.863 = 0.769	0.686 × 0.894 = 0.613	0.668 (0.260)
DD + DT	0.891 × 0.972 = 0.866	0.686 × 0.896 = 0.581	**0.581** (0.022)
DD + PD + DT	0.891 × 0.908 × 0.972 = 0.786	0.686 × 0.847 × 0.896 = 0.520	**0.571** (0.075)
DD + PD + DP + DT	0.891 × 0.908 × 0.863 × 0.972 = 0.678	0.686 × 0.847 × 0.894 × 0.896 = 0.465	**0.556** (0.197)

Bold: Significant at a 90% confidence level.

implemented on the same location of roadway. However, the traditional NB models can also be used to estimate overall safety effects of treatments with relatively lower standard error.

5. Conclusions and recommendations

There are very few studies on the combined effects of multiple treatments although safety effects of multiple treatments have recently appeared as an important issue of validation of the HSM procedures. Therefore, this study analyses the safety effects of multiple roadside features using the cross-sectional method through development and comparison of GLM, GNM, and MARS models for different crash types and severity levels. In order to reflect the nonlinear effects of predictors, the nonlinearizing link functions were developed and used in the GNM. Also, the MARS models were evaluated to account for both nonlinearity of independent variables and interaction effects for complex data structure.

For the GNMs, the nonlinearizing link functions were developed based on the relationships between the logarithm of crash rates and AADT and driveway density. Although the GNMs generally provided slightly better model fits than the GLMs, the difference was not significant. This may be because the interaction impacts among variables under different ranges were not reflected by the GNMs.

In order to account for both nonlinear effects and interaction impacts between variables, the MARS models were developed for different severity levels in this study. It was found that MARS models generally provide better model fitness than the GLMs and GNMs. However, the MARS model for ROR crashes was not significant due to the low crash frequency. It is worth to note that various interaction impacts among variables under different ranges based on knot values were found from MARS whereas one interaction impact between AADT and driveway density was found in the GLMs and GNMs. The results showed that for injury and severe crashes, the basis functions related to distance to trees were selected in the MARS whereas it was not significant in the GLMs and GNMs.

The results showed that the combined safety effects overestimated the real safety effects of multiple treatments by 8 to 10 percent when using the HSM combining method compared to the estimated CMFs from MARS. This may be because roadside elements are implemented on the same location of roadway and they have interaction effects with each other. Thus, it can be recommended that the MARS is used to assess the safety effects of multiple treatments to account for the interaction impacts among treatments, especially when they are implemented on the same location of roadway.

However, there are some limitations in this study. Due to lack of data, more general relationship between crash rates and variables could not be found in the GNMs. Similarly, the interaction impacts under different ranges of variables were not observed in the MARS for ROR crashes due to the low crash frequency. Although the MARS models showed better model fits and can reflect the nonlinearity and interaction effects, there is a need to optimize the issue between complexity for increasing model accuracy and applicability for the ease of general implementation of model. As with many safety studies including the HSM, due to the lack of detailed data within aggregated time period, the potential within-period variation in explanatory variables may be ignored and result in the loss of potentially important explanatory information. This can introduce error in model estimation as a result of unobserved heterogeneity (Lord and Mannering, 2010). In addition, although we tried to obtain high quality roadway parameters as much as possible, some important variables might not have been included. This may be resulted in biased parameter estimates that can produce erroneous inferences and crash predictions (Washington et al., 2010).

Acknowledgements

The authors wish to thank the Florida Department of Transportation (FDOT) for providing the data that were used in this study, and for funding this research. The authors acknowledge the contribution of Erika Meeker to this study. All opinions and results are solely those of the authors.

References

Abdel-Aty, M., Radwan, E., 2000. Modeling traffic accident occurrence and involvement. Accid. Anal. Prev. 32 (5), 633–642.
Abdel-Aty, M., Haleem, K., 2011. Analyzing angle crashes at unsignalized intersection using machine learning techniques. Accid. Anal. Prev. 43, 461–470.
Abdel-Aty, M., Lee, C., Park, J., Wang, J., Abuzwidah, M., Al-Arifi, S., 2014. Validation and Application of Highway Safety Manual (Part D) in Florida. Florida Department of Transportation.
Abraham, A., Steinberg, D., Philip, N., 2001. Rainfall forecasting using soft computing models and multivariate adaptive regression splines. IEEE SMC Transactions: Special Issue on Fusion of Soft Computing and Hard Computing in Industrial Applications .
Ahmed, M., Huang, H., Abdel-Aty, M., Guevara, B., 2011. Exploring a Bayesian hierarchical approach for developing safety performance functions for a mountainous freeway. Accid. Anal. Prev. 43 (4), 1581–1589.
Al-Masaeid, H.R., 1997. Impact of pavement condition on rural road accidents. Can. J. Civ. Eng. 24 (4), 523–531.
American Association of State Highway Transportation Officials (AASHTO), 2010. Highway Safety Manual, 1st ed. AASHTO, Washington, D.C.
Anastasopoulos, P., Tarko, A., Mannering, F., 2008. Tobit analysis of vehicle accident rates on interstate highways. Accid. Anal. Prev. 40 (2), 768–775.

Anastasopoulos, P.C., Mannering, F.L., 2009. A note on modeling vehicle accident frequencies with random-parameters count models. Accid. Anal. Prev. 41, 153–159.
Anastasopoulos, P., Mannering, F., 2011. An empirical assessment of fixed and random parameter logit models using crash- and non-crash-specific injury data. Accid. Anal. Prev. 43 (3), 1140–1147.
Anastasopoulos, P., Mannering, F., Shankar, V., Haddock, J., 2012a. A study of factors affecting highway accident rates using the random-parameters tobit model. Accid. Anal. Prev. 45, 628–633.
Anastasopoulos, P., Shankar, V., Haddock, J., Mannering, F., 2012b. A multivariate tobit analysis of highway accident-injury-severity rates. Accid. Anal. Prev. 45, 110–119.
Bateman, M., Howard, I., Johnson, A., Walton, J., 1998. Model of the performance of a roadway safety fence and its use for design. Transp. Res. Rec. 1647, 122–129.
Bonneson, J., Lord, D., Zimmerman, K., Fitzpatrick, K., Pratt, M., 2007. Development of Tools for Evaluating the Safety Implications of Highway Design Decisions. Report No. FHWA/TX-07/0-4703-P4. Texas Transportation Institute, College Station, TX.
Briand, L., Freimut, B., Vollei, F., 2004. Using multiple adaptive regression splines to support decision making in code inspections. J. Syst. Software 73, 205–217.
Buddhavarapu, P., Banerjee, A., Prozzi, J.A., 2013. Influence of pavement condition on horizontal curve safety. Accid. Anal. Prev. 52, 9–18.
Burns, J.C., 1981. Roughness and roadway safety. Transp. Res. Rec. 836, 8–14.
Carrasco, O., McFadden, J., Chandhok, P., Patel, R., 2004. Evaluation of the effectiveness of shoulder rumble strips on rural multilane divided highways in minnesota. Presented at 83rd Annual Meeting of the Transportation Research Board, Washington, D.C..
Carter, D., Srinivasan, R., Gross, F., Council, F., 2012. Recommended Protocols for Developing Crash Modification Factors. NCHRP 20-7(314) Final Report.
Eluru, N., Bhat, C.R., Hensher, D.A., 2008. A mixed generalized ordered response model for examining pedestrian and bicyclist injury severity level in traffic crashes. Accid. Anal. Prev. 40 (3), 1033–1054.
Elvik, R., Hoye, A., Vaa, T., Sorensen, M., 2009. The handbook of road safety measures, 2nd ed. Emerald Group Publishing Limited.
Fernandes, A., Neves, J., 2013. An approach to accidents modeling based on compounds road environments. Accid. Anal. Prev. 53, 39–45.
Friedman, J., 1991. Multivariate adaptive regression splines. Ann. Stat. 19, 1–141.
Gattis, J., Varghese, J., Toothaker, L., 1993. Analysis of guardrail-end accidents in Oklahoma. Transp. Res. Rec. 1419, 52–62.
Good, M., Fox, J., Joubert, P., 1987. An in-depth study of accidents involving collisions with utility poles. Accid. Anal. Prev. 19 (5), 397–413.
Griffith, M.S., 1999. Safety Evaluation of Rolled-in Continuous Shoulder Rumble Strips Installed on Freeways. Transportation Research Record, 1665. Transportation Research Board, National Research Council, Washington, D.C, pp. 28–34.
Gross, F., Jovanis, P.P., Eccles, K., 2009. Safety Effectiveness of Lane and Shoulder Width Combinations on Rural, Two-lane, Undivided Roads. Transportation Research Record, 2103. Transportation Research Board of the National Academies, Washington, D.C, pp. 42–49.
Gross, F., Persaud, B., Lyon, C., 2010. A Guide to Developing Quality Crash Modification Factors. Publication FHWA-SA-10-032. FHWA. U.S. Department of Transportation.
Hadi, M.A., Aruldhas, J., Chow, L., Wattleworth, J., 1995. Estimating Safety Effects of Cross-Section Design for Various Highway Types Using Negative Binomial Regression. Transportation Research Record 1500, pp. 169–177.
Haleem, K., Abdel-Aty, M., Santos, J., 2010. Multiple Applications of the Multivariate Adaptive Regression Splines in Predicting Rear-end Crashes at Unsignalized Intersections. Transportation Research Record, 2165. Transportation Research Board of the National Academies, Washington, DC, pp. 33–41.
Haleem, K., Gan, A., Lu, J., 2013. Using multivariate adaptive regression splines (MARS) to develop crash modification factors for urban freeway interchange influence areas. Accid. Anal. Prev. 55, 12–21.
Harb, R., Radwan, E., Su, X., 2010. Empirical analysis of toll-lane processing times using proportional odds augmented MARS. ASCE J. Transp. Eng. 136 (11), 1039–1048.
Harkey, D.L., Srinivasan, R., Baek, J., Council, F.M., Eccles, K., Lefler, N., Gross, F., Persaud, B., Lyon, C., Hauer, E., Bonneson, J.A., 2008. Accident Modification Factors for Traffic Engineering and ITS Improvements. NCHRP Report 617. Transportation Research Board, Washington, D.C.
Harwood, D.W., Council, F.M., Hauer, E., Hughes, W.E., Vogt, A., 2000. Prediction of the Expected Safety Performance of Rural Two-Lane Highways. Publication No. FHWA-RD-99-207. Federal Highway Administration, McLean, Virginia.
Jovanis, P., Gross, F., 2008. Estimation of Safety Effectiveness of Changes in Shoulder Width with Case Control and Cohort Methods. Transportation Research Record, 2019. Transportation Research Board, National Research Council, Washington, D.C, pp. 237–245.
Kennedy, J., 1997. Effect of light poles on vehicle impacts with roadside barriers. Transp. Res. Rec. 1599, 32–39.
Kuhfeld, W.F., Cai, W., 2013. Introducing the new ADAPTIVEREG procedure for adaptive regression. SAS Global Forum: Statistics and Data Analysis Paper 457-2013.
Lao, Y., Zhang, G., Wang, Y., Milton, J., 2013. Generalized nonlinear models for rear-end crash risk analysis. Accid. Anal. Prev. 62, 9–16.
Lee, C., Abdel-Aty, M., Park, J., Wang, J., 2015a. Development of crash modification factors for changing lane width on roadway segments using generalized nonlinear models. Accid. Anal. Prev. 75, 83–91.
Lee, J., Mannering, F., 2002. Impact of roadside features on the frequency and severity of run-off-roadway accidents: an empirical analysis. Accid. Anal. Prev. 34, 149–161.
Lee, J., Nam, B., Abdel-Aty, M., 2015b. Effects of pavement surface conditions on traffic crash severity. J. Transp. Eng. doi:http://dx.doi.org/10.1061/(ASCE)TE.1943-5436.0000785 4015020.
Li, X., Lord, D., Zhang, Y., Xie, Y., 2008. Predicting motor vehicle crashes using support vector machine models. Accid. Anal. Prev. 40 (4), 1611–1618.
Li, X., Lord, D., Zhang, Y., 2011. Development of accident modification factors for rural frontage road segments in texas using generalized additive models. J. Transp. Eng. 137, 74–83.
Li, Y., Liu, C., Ding, L., 2013. Impact of pavement conditions on crash severity. Accid. Anal. Prev. 59, 399–406.
Lord, D., Bonneson, J.A., 2007. Development of accident modification factors for rural frontage road segments in Texas. Transp. Res. Rec. J. Transp. Res. Board 20–27.
Lord, D., Mannering, F.L., 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.
Mayora, J., Piña, R., 2009. An assessment of the skid resistance effect on traffic safety under wet-pavement conditions. Accid. Anal. Prev. 41 (4), 881–886.
Milton, J., Shankar, V., Mannering, F., 2008. Highway accident severities and the mixed logit model: an exploratory empirical analysis. Accid. Anal. Prev. 40 (1), 260–266.
Pande, A., Abdel-Aty, M., 2009. A novel approach for analyzing severe crash patterns on multilane highways. Accid. Anal. Prev. 41 (5), 985–994.
Park, J., Abdel-Aty, M., Lee, C., 2014. Exploration and comparison of crash modification factors for multiple treatments on rural multilane roadways. Accid. Anal. Prev. 70, 167–177.
Park, J., Abdel-Aty, M., 2015. Development of adjustment functions to assess combined safety effects of multiple treatments on rural two-lane roadways. Accid. Anal. Prev. 75, 310–319.
Park, J., Abdel-Aty, M., Lee, J., Lee, C., 2015a. 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, 179–191.
Park, J., Abdel-Aty, M., Wang, J., Lee, C., 2015b. Assessment of safety effects for widening urban roadways in developing crash modification functions using nonlinearizing link functions. Accid. Anal. Prev. in-press.
Patel, R.B., Council, F.M., Griffith, M.S., 2007. Estimating the safety benefits of shoulder rumble strips on two lane rural highways in minnesota: an empirical bayes observational before–after study. Presented at 86th Annual Meeting of the Transportation Research Board, Washington, D.C..
Put, R., Xu, Q., Massart, D., Heyden, Y., 2004. Multivariate adaptive regression splines (MARS) in chromatographic quantitative structure–retention relationship studies. J. Chromatogr. A 1055, 11–19.
Ray, M., 1999. Impact conditions in side-impact collisions with fixed roadside objects. Accid. Anal. Prev. 31 (1), 21–30.
Reid, J., Sicking, D., Faller, R., Pfeifer, B., 1997. Development of a new guardrail system. Transp. Res. Rec. 1599, 72–80.
Shankar, V., Mannering, F., Barfield, W., 1995. Effect of roadway geometrics and environmental factors on rural accident frequencies. Accid. Anal. Prev. 27 (3), 371–389.
Snedecor, G.W., Cochran, W.G., 1980. Statistical Methods, 7th ed. Iowa State University Press.
Stamatiadis, N., Pigman, J., Sacksteder, J., Ruff, W., Lord, D., 2009. Impact of Shoulder Width and Median Width on Safety. NCHRP Report 633. Transportation Research Board, Washington, D.C.
Stamatiadis, N., Lord, D., Pigman, J., Sacksteder, J., Ruff, W., 2011. Safety impacts of design element trade-offs for multilane rural highways. ASCE J. Transp. Eng. 137 (5), 333–340.
Tighe, Susan, Li, Ningyuan, Falls, Lynne Cowe, Haas, Ralph, 2000. Incorporating road safety into pavement management. Transp. Res. Rec. 1699, 1–10.
SAS Institute Inc, 2012. The ADAPTIVEREG Procedure. SAS Institute Inc., Cary, North Carolina, USA.
Turner, D., 1984. Prediction of bridge accident rates. J. Transp. Eng. 110 (1), 45–54.
Venkataraman, N., Ulfarsson, G.F., Shankar, V.N., 2013. Random parameter models of interstate crash frequencies by severity, number of vehicles involved, collision and location type. Accid. Anal. Prev. 59, 309–318.
Viner, J., 1995. Rollovers on sideslopes and ditches. Accid. Anal. Prev. 27 (4), 483–491.
Wang, J., Abdel-Aty, M., 2014. Comparison of safety evaluation approaches for intersection signalization in Florida. Transportation Research Board 93rd Annual Meeting, Washington, D.C..
Washington, S.P., Karlaftis, M.G., Mannering, F.L., 2010. Statistical and Econometric Methods for Transportation Data Analysis, 2nd ed. Chapman Hall/CRC, Boca Raton, FL.
Wong, S.C., Sze, N.N., Li, Y.C., 2007. Contributory factors to traffic crashes at signalized intersections in Hong Kong. Accid. Anal. Prev. 39, 1107–1113.
Wu, K.F., Donnell, E.T., Aguero-Valverde, J., 2014. Relating crash frequency and severity: evaluating the effectiveness of shoulder rumble strips on reducing fatal and major injury crashes. Accid. Anal. Prev. 67, 86–95.
Xie, Y., Lord, D., Zhang, Y., 2007. Predicting motor vehicle collisions using Bayesian neural network models: an empirical analysis. Accid. Anal. Prev. 39 (5), 922–933.
Xu, P., Huang, H., 2015. Modeling crash spatial heterogeneity: random parameter versus geographically weighting. Accid. Anal. Prev. 75, 16–25.
Zegeer, C., Council, F., 1995. Safety relationships associated with cross-sectional roadway elements. Transp. Res. Rec. 1512, 29–36.
Zhang, Y., Xie, Y., Li, L., 2012. Crash frequency analysis of different types of urban roadway segments using generalized additive model. J. Safety Res. 43, 107–114.