Accident Analysis and Prevention 92 (2016) 245–255

Contents lists available at ScienceDirect

Accident Analysis and Prevention

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

Evaluation of safety effectiveness of multiple cross sectional features on urban arterials

Juneyoung Park*, Mohamed Abdel-Aty

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

ARTICLE INFO

Article history:
Received 12 November 2015
Received in revised form 6 March 2016
Accepted 12 April 2016
Available online 22 April 2016

Keywords:
Safety effectiveness
Crash modification factors
Generalized nonlinear models
Cross-Sectional method
Multiple treatments

ABSTRACT

This research evaluates the safety effectiveness of multiple roadway cross-section elements on urban arterials for different crash types and severity levels. In order to consider the nonlinearity of predictors and obtain more reliable estimates, the generalized nonlinear models (GNMs) were developed using 5-years of crash records and roadway characteristics data for urban roadways in Florida. The generalized linear models (GLMs) were also developed to compare model performance. The cross-sectional method was used to develop crash modification factors (CMFs) for various safety treatments. The results from this paper indicated that increasing lane, bike lane, median, and shoulder widths were safety effective to reduce crash frequency. In particular, the CMFs for changes in median and shoulder widths consistently decreased as their widths increased. On the other hand, the safety effects of increasing lane and bike lane widths showed nonlinear variations. It was found that crash rates decrease as the lane width increases until 12 ft width and it increases as the lane width exceeds 12 ft. The crash rates start to decrease again after 13 ft. It was also found that crash rates decreases as the bike lane width increases until 6 ft width and it increases as the bike lane width exceeds 6 ft. This paper demonstrated that the GNMs clearly captured the nonlinear relationship between crashes and multiple roadway cross-sectional features, which cannot be reflected by the estimated CMFs from the GLMs. Moreover, the GNMs showed better model fitness than GLMs in general. Therefore, in order to estimate more accurate CMFs, the proposed methodology of utilizing the GNMs in the cross-sectional method is recommended over using conventional GLMs when there are nonlinear relationships between the crash rate and roadway characteristics.

1. Introduction

A crash modification factor (CMF) is a factor that can express potential changes in crashes after a treatment (countermeasure) is implemented on a roadway segment or intersection. Among four main parts of the Highway Safety Manual (HSM) (AASHTO, 2010), part D provides a variety of CMFs. The CMF can be estimated by observational before-after studies or the cross-sectional method (Gross et al., 2010; Carter et al., 2012). There are generally five approaches used to perform observational before-after studies; (1) naïve before-after study, (2) before-after study with yoked comparison, (3) before-after study with comparison group (CG), (4) before-after study with the empirical Bayes (EB) approach, and (5) Full Bayes (FB) before-after method (Hauer, 1997; Hauer et al., 2002; Gross et al., 2010). It is known that observational before-after studies with EB and CG methods are the more common approaches among the various before-after studies (Abdel-Aty et al., 2014).

The cross-sectional method also has been widely applied to calculate CMFs (Lord and Bonneson, 2007; Stamatiadis et al., 2009; Li et al., 2011; Park et al., 2014; Lee et al., 2015a; Park and Abdel-Aty, 2015a). However, there are several important potential issues (e.g., correlation, spatial effect, etc.) and biases (e.g., selection bias, omitted variable bias, etc.) with the cross-sectional analysis (Hauer, 2004; Lord and Mannering, 2010; Carter et al., 2012; Hauer, 2013). In particular, according to Persaud et al. (1999) and Hauer et al. (2004), traffic volume can be a significant confounding variable (i.e. confounder) for crash frequency and also be associated with several roadway geometric conditions. A confounder is a significant independent variable that completely or partially accounts for the apparent association between an outcome and a predictor variable (Collett, 2003). In this study, the interaction effects among explanatory variables including traffic volume were investigated to control potential confounding effects.

It is known that the traditional regression model cannot account for the differences among observations (i.e. heterogeneity) (Elvik,

* Corresponding author.
E-mail addresses: jypark@knights.ucf.edu (J. Park), m.aty@ucf.edu (M. Abdel-Aty).

http://dx.doi.org/10.1016/j.aap.2016.04.017

2011). As stated by Lord and Mannering (2010), unobserved heterogeneity is referred to as omitted variable bias when the unobserved characteristics are correlated with a predictor that is included in a model. In order to address this issue, Bonneson and Pratt (2008) suggested one approach using matched pairs (i.e. pairs with and without treatment). Full Bayes and hierarchical Bayes methods also can be applied to account for the unobserved heterogeneity problem (Qin et al., 2005; Park and Lord, 2007; Persaud et al., 2009; El-Basyouny and Sayed, 2010; Yu and Abdel-Aty, 2013; Ahmed et al., 2015).

Hauer and Banfo (1997) discussed that overfitting of prediction models can occur when the model is too complex and developed with too many parameters. Washington et al. (2005) suggested an application of cross-validation process to deal with this problem. In this study, the generalized nonlinear models (GNMs) with nonlinearizing link functions (Lao et al., 2013) were developed and used to estimate CMFs in the cross-sectional analysis process. Although the GNM has several segmented sections for each parameter that has nonlinearity, the changes by the break points were already accounted for in the nonlinearizing link functions. Therefore, in GNM, the nonlinear predictor is still considered as a single parameter and not multiple parameters.

The study by Elvik (2011) discussed that prediction models using data that are aggregated or averaged can lead to biased estimates. Use of disaggregate data (e.g., hourly traffic volume instead of annual average daily traffic (AADT)) can be one way to account for this bias (Lord et al., 2005; Abdel-Aty and Pande, 2009). However, it might cause additional issues such as low mean value, temporal correlation effect, and finding a proper source of disaggregated data.

Moreover, Hauer (2004) indicated that the selection of appropriate functional form is critical to produce more accurate estimates and enhance the model reliability. For this reason, in this study, the nonlinearizing link functions were developed to reflect the nonlinear relationships between crash rates and predictors. Hence, various nonlinear functional forms were compared to identify the best fitted nonlinearizing link function for each segmented section.

Although there are several limitations in the cross-sectional analysis as discussed above, it has been widely applied to calculate CMFs since (1) it is easier to obtain data compared to the before-after approaches, (2) it is difficult to isolate the effect of a single treatment from the effects of the other treatments applied at the same time using the before-after method (Harkey et al., 2008; Bahar, 2010), and (3) it would be practically infeasible to conduct the prescribed before-after study on specific treatments related to the changes of widths of roadway cross section elements (e.g., lane width, median width, shoulder width, etc.) (Carter et al., 2012).

Several alternative approaches such as propensity scoring matching (Rosenbaum and Rubin, 1983; Aul and Davis, 2006; Wood and Porter, 2013; Sasidharan and Donnell, 2013; Wood et al., 2014, 2015) and case-control analysis (Gross and Jovanis, 2007a,b) were proposed to overcome the limitations of cross-sectional regression models. However, according to the Guo and Fraser (2010), it is difficult to apply the propensity scores-potential outcomes framework to treatments with continuous treatment values. Also, since the case-control analysis is more often used to show the relative effects of treatments, the case-control studies cannot be suggested to measure the probability of an event (e.g., crash, severe injury, etc.) in terms of expected frequency (Gross et al., 2010). Furthermore, both approaches cannot account for the nonlinear effects of predictors.

To estimate CMFs using the cross-sectional method, it is required to develop safety performance functions (SPFs) or crash prediction models (CPMs). The generalized linear model (GLM) with negative binomial (NB) distribution has been commonly used to develop SPFs to account for over-dispersion (Shankar et al., 1995; Abdel-Aty and Radwan, 2000). In the cross-sectional method, the coefficient associated with a variable for specific treatment obtained from the SPF is used to estimate the CMF (Lord and Bonneson, 2007; Harkey et al., 2008; Stamatiadis et al., 2009; Carter et al., 2012; Abdel-Aty et al., 2014). Since the GLM is linear-based analysis and is controlled by its linear model specification, it may bias estimates when the explanatory variable shows a nonlinear relationship with response variable. Thus, the CMF developed using the GLM cannot account for nonlinear effects of the treatment since the CMF is a fixed value in the GLM (Lee et al., 2015a).

In order to account for the nonlinear relationship between crashes and roadway characteristics, the use of the logarithm of AADT instead of AADT has been widely applied (Hauer, 1995; Abdel-Aty and Radwan, 2000; Harwood et al., 2000; Wong et al., 2007; Abdel-Aty and Haleem, 2011; Park and Abdel-Aty, 2015b). Moreover, few studies have applied data mining techniques such as generalized additive models (GAM) and multivariate adaptive regression splines (MARS) to reflect the nonlinearity of crash predictors in the development of CMFs (Li et al., 2011; Zhang et al., 2012; Haleem et al., 2013). However, according to Li et al. (2011) and Lee et al. (2015a), the evaluations of GAM and MARS are complex because it include more parameters and the rate of change is assumed to be fixed within a given range of a variable although the rate can vary within that range.

For this reason, an application of using GNM for crash analysis has been recommended (Lao et al., 2013; Park and Abdel-Aty, 2015a; Lee et al., 2015a; Park et al., 2015a). The GNM requires the development of nonlinearizing link function to account for nonlinear effects. Lao et al. (2013) demonstrated that right shoulder width, AADT, grade percentage, and truck percentage have nonlinear effects on rear-end crashes through evaluation of GNMs. It was found that GNMs can better reflect the nonlinear relationships than GLMs. However, the study investigated only the main effects of each variable, but not the effects of interaction between variables.

Similar to this study, Lee et al. (2015a) estimated CMFs for changes of lane width using GNMs. A set of nonlinearizing link functions were developed to reflect the nonlinear effects of lane width and speed limit on crash frequency for all types of roadways. This study showed that the CMFs estimated using the GNMs indicated that both narrower lane (i.e. lane width less than 12 ft) and wider lane (i.e. lane width greater than 12 ft) reduce crash frequency. 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. (2015a) found that nonlinear relationship exist between safety effects of widening urban roadways and time changes (i.e. changes of CMF by different years). The study evaluated crash modification functions (CMFunctions) using a Bayesian regression model including the developed nonlinearizing link function to incorporate the changes in safety effects of the treatment over time. Park and Abdel-Aty (2015a) developed both GNMs and MARS models to reflect nonlinearity of predictors for multiple roadside treatments on rural multilane roadways. It was found that the MARS models showed better model fit than the GNMs due to its strength to capture the interaction impacts among treatments implemented at the same location of the roadway cross-section (i.e. roadside).

A number of studies addressed the safety effects of various roadway cross-sectional elements. In general, it was found that an increase in widths of cross-section elements (i.e. lane width, bike lane width, median width, and shoulder width) reduces crash frequency (Knuiman et al., 1993; Hadi et al., 2000; Hauer, 1997; Karlaftis and Golias, 2002; Lord and Bonneson, 2007; Potts et al., 2007; Harkey et al., 2008; Gross et al., 2009; Stamatiadis et al., 2009; Yanmaz-Tuzel and Ozbay, 2010; Labi, 2011; Haleem et al., 2013; Zeng and Schrock, 2013; Park et al., 2014; Torbic et al., 2014; Park et al., 2015a). On the other hand, some studies showed the nonlinear effects of changes of widths (Xie et al., 2007; Jovanis and