Accident Analysis and Prevention 97 (2016) 274–287

Contents lists available at ScienceDirect

Accident Analysis and Prevention

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

Finite mixture modeling approach for developing crash modification factors in highway safety analysis

Byung-Jung Park[a], Dominique Lord[b], Lingtao Wu[c,*]

[a] Department of Transportation Engineering, Myongji University, Republic of Korea
[b] Zachry Department of Civil Engineering, Texas A&M University, 3136 TAMU, College Station, TX 77843-3136, United States
[c] Texas A&M Transportation Institute, Texas A&M University System, 3135 TAMU, College Station, TX 77843-3135, United States

ARTICLE INFO

Article history:
Received 3 June 2016
Received in revised form 17 October 2016
Accepted 18 October 2016

Keywords:
Finite mixture model
Negative binomial model
Combined safety effects
Highway safety
Crash modification factor

ABSTRACT

This study aimed to investigate the relative performance of two models (negative binomial (NB) model and two-component finite mixture of negative binomial models (FMNB-2)) in terms of developing crash modification factors (CMFs). Crash data on rural multilane divided highways in California and Texas were modeled with the two models, and crash modification functions (CMFunctions) were derived. The resultant CMFunction estimated from the FMNB-2 model showed several good properties over that from the NB model. First, the safety effect of a covariate was better reflected by the CMFunction developed using the FMNB-2 model, since the model takes into account the differential responsiveness of crash frequency to the covariate. Second, the CMFunction derived from the FMNB-2 model is able to capture nonlinear relationships between covariate and safety. Finally, following the same concept as those for NB models, the combined CMFs of multiple treatments were estimated using the FMNB-2 model. The results indicated that they are not the simple multiplicative of single ones (i.e., their safety effects are not independent under FMNB-2 models). Adjustment Factors (AFs) were then developed. It is revealed that current *Highway Safety Manual*'s method could over- or under-estimate the combined CMFs under particular combination of covariates. Safety analysts are encouraged to consider using the FMNB-2 models for developing CMFs and AFs.

1. Introduction

Highway safety has been a major research topic in transportation studies since highway crashes account for more than 90% of all transportation-related fatalities and cause enormous socio-economic costs. Recently, increased emphasis has been placed on improving the explicit role of highway safety in making decisions on transportation planning, design, and operations. This can be achieved by quantifying the safety effects of geometric design elements for various transportation facilities, and incorporating the safety information in the planning and design stages of the project development process (Bonneson et al., 2007). In this regard, the first edition of *Highway Safety Manual* (*HSM*) uses the concept of crash modification factor (CMF) to evaluate the safety performance for various highway facilities before they are open to traffic (AASHTO, 2010).

A CMF represents the change in safety when a particular geometric design element changes in size with respect to the base (or typical) condition or some treatment is taken at a problematic site. A CMF greater than 1.0 indicates the situation where the design change is associated with more crashes whereas a CMF less than 1.0 represents fewer crashes. CMFs can be developed by various techniques which include the before-and-after study, cross-sectional study, use of expert panels, and regression-based models (Bonneson and Lord, 2005; Li et al., 2010; Shahdah et al., 2014). CMFs are ideally to be developed through before-after studies, in particular with empirical Bayes (EB) analysis (Hauer, 2010). However, it is nearly impossible to evaluate the CMFs for some highway features or treatments using such method in practice, especially when the treatments are costly (e.g., pavement width, horizontal curve, etc.). For these highway features, safety analysts frequently use cross-sectional analysis, practically using regression model methods, for assessing their safety effects. In a cross-sectional analysis, the safety performances of two or several groups of highway segments with different characteristics in terms

* Corresponding author.
E-mail addresses: bjpark@mju.ac.kr (B.-J. Park), d-lord@tamu.edu (D. Lord), wulingtao@gmail.com, l-wu@tti.tamu.edu (L. Wu).

http://dx.doi.org/10.1016/j.aap.2016.10.023
0001-4575/

of the feature of interest are compared. The difference is attributed to that highway features. In regression models, the safety effects (i.e., CMFs) are estimated directly from the coefficients of the crash prediction models or safety performance functions (SPFs). Usually, CMFs developed using regression models are believed to be less reliable than that with before-after studies, mainly because there are some limitations with regression models, e.g., unobserved heterogeneity, confounding variable or omitted variable bias, misspecification in functional form, independence assumption, etc. (Hauer 2013; Jovanis and Gross, 2008; Lord and Mannering, 2010; Mannering et al., 2016; Park and Abdel-Aty, 2016; Wu et al., 2015; Wu and Lord, 2016). Some researchers have criticized the use of regression models for developing CMFs since SPFs cannot capture the cause-effect relationship between variables (Hauer, 2010; Hauer 2015). Even though regression models may still remain one of the most common methods for developing CMFs in the near future due to the limitations and infeasibility of before-after studies (see Lord and Kuo, 2012). As such, it is important to investigate how to improve the robustness and accuracy of CMFs developed from regression models.

Negative binomial (NB) model with additive link functions has been commonly used to develop SPFs in the past decades, and CMFs are then estimated from the SPFs. Numerous studies have used this approach for developing CMFs, including Fitzpatrick et al. (2008), Lord and Bonneson (2007) and Washington et al. (2005). On the other hand, Bonneson et al. (2007) and Gross et al. (2009) have argued that the interaction between design features should be included in the development of CMFs. In line with this effort, Li et al. (2010) tried to incorporate the interactions by using general additive models. Addressing this issue, however, is beyond the scope of this study.

The commonly used NB model explicitly assumes that each covariate is independent, and the model parameters are assumed independent (the terms covariate, variable and treatment will be used interchangeably). In addition, with the traditional NB models, the safety effects of variables are independent, and the CMFs are multipliable, as the *HSM* has documented (referred to as *HSM* method thereafter). Once CMFs are obtained for various highway geometric design elements, they are applied multiplicatively for adjusting crash frequency estimated from a baseline model. The baseline model represents the calibrated statistical model using data that meet specific base conditions, such as 12-ft lane width and 8-ft shoulder width for divided rural multilane highway segments. Therefore, the final predicted number of crashes is computed as follows:

$$\mu_{final} = \mu_{baseline} \times CMF_1 \times \cdots \times CMF_n \times CF \quad (1)$$

Where,

$CMF_1, \cdots, CMF_n$ =crash modification factors;
μ_{final} = final predicted number of crashes per unit of time;
$\mu_{baseline}$ = baseline predicted number of crashes per unit of time; and,
CF = calibration factor to adjust to local conditions.

It is worth mentioning that, however, in practice CMFs may not be completely independent since changes in geometric design characteristics on highways are usually not done separately (e.g., lane and shoulder width may be changed simultaneously) and the combinations of these changes can influence crash risk differently. Although experience in deriving CMFs in this manner indicates that the independence assumption is in general acceptable and the resulting CMFs can yield useful information about the first-order effect of a given variable on safety, the *HSM* has cautioned that the assumption can lead to over- or under-estimation of actual safety impacts of multiple treatments. Recently, efforts have been made to explore the combined safety effects of multiple treatments (Park and Abdel-Aty, 2015a, 2015b; Park et al., 2014b). It was found that the combined safety effects of multiple treatments estimated using the *HSM* method were usually over-estimated.

Despite the important role of CMFs in highway safety analysis, there are currently no documents that address how CMFs could be derived from the finite mixture models and compared with those produced from traditional models, such as the NB models. The finite mixture models, both fixed and varying weight parameter models, have been shown to be useful for explaining the heterogeneity and the nature of the dispersion in crash data (Park and Lord, 2009; Zou et al., 2013; Mannering et al., 2016). More recently, semi-parametric mixture models have been proposed for conducting safety analyses (Shirazi et al., 2016; Heydari et al., 2016). Given the superior performance of the finite mixture model, there is a need to investigate whether this type of model would result in important differences with the development of CMFs. The crash modification function (referred to as CMFuntion hereafter) for the finite mixture models is not as simple as that in the single NB models since the conditional mean takes on the mix of additive and multiplicative terms. Therefore, the main objective of this paper is to compare the relative performance of two models (i.e., two-component finite mixture of NB models (FMNB-2) and the NB model) in terms of the difference in determining CMFs as a result of different model coefficients. More specifically, this paper describes in details the procedure on how to derive a CMFunction from the FMNB-2 model and its characteristics are discussed by comparing it with that from the traditional NB model. Another objective of this paper is to estimate combined safety effects of multiple treatments (i.e., combined CMFs) using FMNB-2 models and compare them with those from NB models, and to further develop adjustment factors (AFs) if the safety effects of multiple treatments are found to be dependent for FMNB-2 models.

2. Derivation of CMFunctions

As mentioned previously, researchers have proposed that regression models or SPFs can be used for developing CMFs. This section presents how CMFunctions are derived from NB and FMNB-2 models, respectively.

2.1. The negative binomial model

In additive models, such as a linear regression with $\hat{\mu}_i = \mathbf{x}_i\hat{\boldsymbol{\beta}}$, the coefficient $\hat{\beta}_j$ for a covariate x_j is readily interpreted as the effect of a one-unit change in x_j on the conditional mean. That is, a unit increase in x_j is associated with a $\hat{\beta}_j$ increase in $\hat{\mu}_i$. In multiplicative models, such as the Poisson or NB regression models, the conditional mean functional form is usually expressed as a log-linear form: $\ln\hat{\mu}_i = \mathbf{x}_i\hat{\boldsymbol{\beta}}$. In such a case, the difference between two conditional means ($\Delta\hat{\mu}_i$) induced by a one-unit change in x_j is no longer constant across sites and depends on the values of the covariates. A more convenient way to examine the effect of a covariate is to take the ratio of the two conditional means, which results in $\exp(\hat{\beta}_j)$. The ratio is now constant across all sites without depending on the values of any covariates.