# Finite Mixture Modeling Approach for Developing Crash Modification Factors in Highway Safety Analysis

**Byung-Jung Park**
Department of Transportation Engineering
Myongji University, Korea
Phone: 82-31-330-6499, fax: 82-31-330-2885
E-mail: bjpark@mju.ac.kr

**Dominique Lord**
Zachry Department of Civil Engineering
Texas A&M University, 3136 TAMU
College Station, TX 77843-3136
Phone: 979-458-3949, fax: 979-845-6481
E-mail: d-lord@tamu.edu

**Lingtao Wu***
Texas A&M Transportation Institute
Texas A&M University System, 3135 TAMU
College Station, TX 77843-3135
Phone: 979-845-7214, fax: 979-845-6481
Email: wulingtao@gmail.com

October 17, 2016

* Corresponding author

**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.

Keywords: Finite mixture model, negative binomial model, combined safety effects, highway safety, crash modification factor

# 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 (ASSHTO, 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 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:

$$\mu_{final} = \mu_{baseline} \times CMF_1 \times \cdots \times CMF_n \times CF \tag{1}$$

Where,
$CMF_1, \cdots, CMF_n$ = crash modification factors;
$\mu_{final}$ = 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; Park and Abdel-Aty, 2015b; Park et al. 2014). 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 $\hat{\mu}_i = \mathbf{x}_i\widehat{\boldsymbol{\beta}}$, 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: $\ln \hat{\mu}_i = \mathbf{x}_i\widehat{\boldsymbol{\beta}}$. 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. Hence, the effect of a covariate is interpreted as follows: a one-unit increase in $x_j$ is associated with a factor of $\exp(\hat{\beta}_j)$ increase in $\hat{\mu}_i$ (Long, 1997). In developing the CMF for a covariate $x_j$, however, we are not interested in the safety effect of a covariate $x_j$ by changing a one-unit, but interested in the safety effect of $x_j$ when it changes from its base condition value. In this case, the CMF for $x_j$ can be derived in a continuous functional form with respect to $x_j$. Therefore, the CMFunction for $x_j$ under the NB model is now derived as follows:

$$CMF_{x_j}^{NB} = \frac{\exp(\widehat{\beta}_0+\widehat{\beta}_1 x_1+\cdots+\widehat{\beta}_j x_j+\cdots+\widehat{\beta}_p x_p)}{\exp(\widehat{\beta}_0+\widehat{\beta}_1 x_1+\cdots+\widehat{\beta}_j x_j^{base}+\cdots+\widehat{\beta}_p x_p)} = \exp\left[\hat{\beta}_j(x_j - x_j^{base})\right] \quad (2)$$

Without loss of generality, the subscript $i$ was removed from Equation (2) since the $CMF_{x_j}$ is identical for all sites. This way, the $CMF_{x_j}$ represents the change in the expected crash frequency when the variable $x_j$ changes from its base condition value, and it follows an exponential function with $CMF_{x_j} = 1$ when $x_j = x_j^{base}$. If $\hat{\beta}_j > 0$, the $CMF_{x_j}$ is an strictly increasing function, and if $\hat{\beta}_j < 0$, it is an strictly decreasing function. This relationship is depicted in Figure 1.

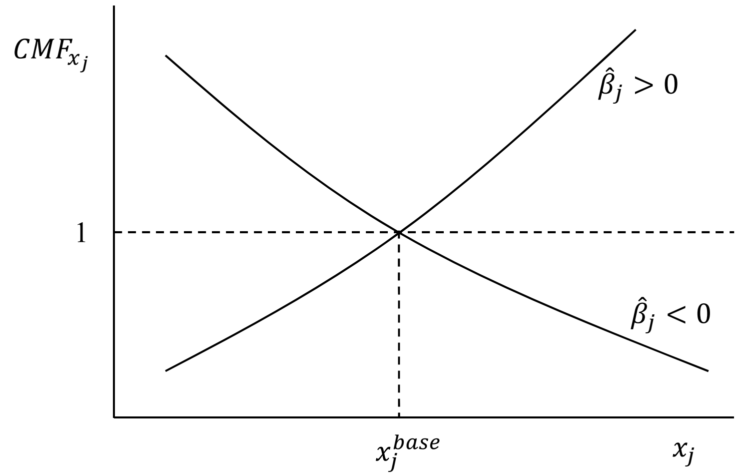



Figure 1. The concept of a CMFunction under NB model

The combined CMFs, $CMF_{Comb.}^{NB}$, for multiple treatments (i.e., $x_1$, $x_2$, …, $x_n$) can be derived through mathematical transformations, as shown in Equation (3). For additional details, please see Appendix A.

$$CMF_{Comb.}^{NB} = CMF_1 \times \cdots \times CMF_n \tag{3}$$

**2.2. The FMNB-2 model**

Finite mixture models assume that the observations of a sample arise from two or more unobserved components with unknown proportions, which allows a great modeling flexibility over traditional single aggregate models. The probability density function, mean and variance of the K-component finite mixture of negative binomial regression models (i.e., FMNB-K) are expressed as follows:

$$p(y_i|\mathbf{x}_i, \mathbf{\Theta}) = \sum_{k=1}^{K} w_k NB(\mu_{i,k},\ \phi_k) = \sum_{k=1}^{K} w_k \left[\frac{\Gamma(y_i+\phi_k)}{\Gamma(y_i+1)\Gamma(\phi_k)}\left(\frac{\mu_{i,k}}{\mu_{i,k}+\phi_k}\right)^{y_i}\left(\frac{\phi_k}{\mu_{i,k}+\phi_k}\right)^{\phi_k}\right] \tag{4}$$

$$\mu_i = E(y_i|\mathbf{x}_i, \mathbf{\Theta}) = \sum_{k=1}^{K} \mu_{i,k}\, w_k \tag{5}$$

$$Var(y_i|\mathbf{x}_i, \mathbf{\Theta}) = E(y_i|\mathbf{x}_i, \mathbf{\Theta}) + \left(\sum_{k=1}^{K} w_k \mu_{i,k}^2 (1 + 1/\ \phi_k) - E(y_i|\mathbf{x}_i, \mathbf{\Theta})^2\right) \tag{6}$$

Where,

$y_i$ = a random variable of $i^{\text{th}}$ observation ($i = 1, 2, \dots, n$);
$w_k$ = weight of component $k$ which sum to 1 ($\sum_{k=1}^{K} w_k = 1$);
$\mu_{i,k} = \exp(\mathbf{x}_i\boldsymbol{\beta}_k)$ is the mean of component $k$;
$\mathbf{x}_i$ = a vector of covariates;
$\boldsymbol{\beta}_k$ and $\phi_k$ are the regression coefficients and the dispersion parameter of the NB distribution for component $k$; and
$\mathbf{\Theta} = \{(\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_K), (\phi_1, \dots, \phi_K), (w_1, \dots, w_K)\}$ is a vector of all unknown parameters.

It can be seen that when $\phi_k = \infty$ in each component the FMNB-K model reduces to the finite mixture of Poisson regression models (FMP-K). The FMNB models, therefore, allow for additional heterogeneity within components not captured by the covariates. The unknown parameters can be estimated via either a maximum likelihood estimation method or a Bayesian method. In order to determine the number of components (K) in the mixture, a series of models with increasing number of components are fitted and then the most plausible model can be selected by various model selection criteria: information-based criteria and Bayes factor via marginal likelihoods. Park and Lord (2009) and Park et al. (2014) have shown a two-component finite mixture of NB regression models (FMNB-2) was quite enough to characterize the uncertainty about the crash occurrence and it provided more opportunities for interpretation of the data which were not available from the standard NB model. For the model structure and parameter estimation method, readers are referred to the aforementioned references.

In the FMNB-2 model, the conditional mean functional form is expressed as $\hat{\mu}_i = \widehat{w}_1 \exp(\mathbf{x}_i \widehat{\boldsymbol{\beta}}_1) + \widehat{w}_2 \exp(\mathbf{x}_i \widehat{\boldsymbol{\beta}}_2)$, where $\widehat{w}_1$ and $\widehat{w}_2$ are the estimated weight parameters which sum to 1. The interpretation of coefficients is not as straightforward as in the NB model since the relationship between the conditional mean and the covariates is a mix of additive and multiplicative forms. The effect of an individual covariate on the conditional mean is determined by two sets of interactions between parameters and covariates. The difficulty arises because the conditional mean ratio varies across all sites, and also depends on the coefficients of the other covariates. Two options can be considered when we want to report a single value for the effect of a one-unit change in $x_j$. One option is first to calculate the ratio of the conditional means for all sites and then to take the average value. Another option is to evaluate the ratio at selected values of the covariates (e.g., sample average).

Analogous to Equation (2), the CMF of $x_j$ in the FMNB-2 model is expressed as follows:

$$CMF_{x_j}^{FMNB-2} = \frac{\widehat{w}_1 \exp\left(\widehat{\beta}_{0,1} + \widehat{\beta}_{j,1} x_{ji} + \sum_{k=1,k\neq j}^{p} \widehat{\beta}_{k,1} x_{ki}\right) + \widehat{w}_2 \exp\left(\widehat{\beta}_{0,2} + \widehat{\beta}_{j,2} x_{ji} + \sum_{k=1,k\neq j}^{p} \widehat{\beta}_{k,2} x_{ki}\right)}{\widehat{w}_1 \exp\left(\widehat{\beta}_{0,1} + \widehat{\beta}_{j,1} x_j^{base} + \sum_{k=1,k\neq j}^{p} \widehat{\beta}_{k,1} x_{ki}\right) + \widehat{w}_2 \exp\left(\widehat{\beta}_{0,2} + \widehat{\beta}_{j,2} x_j^{base} + \sum_{k=1,k\neq j}^{p} \widehat{\beta}_{k,2} x_{ki}\right)} \quad (7)$$

In this case, the $CMF_{x_j}$ differs across sites by depending on the values of covariates. In order to obtain a single continuous function of the $CMF_{x_j}$ with respect to $x_j$ like the one in Figure 1, we need to fix each covariate (except for the interest covariate $x_j$) at a selected value. For this purpose, we used the sample average of each covariate. This is, $x_{ki}$ values for site $i$ is replaced with $\bar{x}_k$ which is $(1/n)\sum_{i=1}^{n} x_{ki}$. This leads the Equation (7) to the following form:

$$CMF_{x_j}^{FMNB-2} = \frac{\widehat{w}_1 \exp\left(\widehat{\beta}_{0,1} + \widehat{\beta}_{j,1} x_j + \sum_{k=1,k\neq j}^{p} \widehat{\beta}_{k,1} \bar{x}_k\right) + \widehat{w}_2 \exp\left(\widehat{\beta}_{0,2} + \widehat{\beta}_{j,2} x_j + \sum_{k=1,k\neq j}^{p} \widehat{\beta}_{k,2} \bar{x}_k\right)}{\widehat{w}_1 \exp\left(\widehat{\beta}_{0,1} + \widehat{\beta}_{j,1} x_j^{base} + \sum_{k=1,k\neq j}^{p} \widehat{\beta}_{k,1} \bar{x}_k\right) + \widehat{w}_2 \exp\left(\widehat{\beta}_{0,2} + \widehat{\beta}_{j,2} x_j^{base} + \sum_{k=1,k\neq j}^{p} \widehat{\beta}_{k,2} \bar{x}_k\right)} \quad (8)$$

Following a similar concept, the combined CMF for multiple treatments can also be derived from FMNB-2 models. Taking two covariates, $x_i$ and $x_j$ ($i \neq j$), as an example, the combined CMF, $CMF_{comb.x_i x_j}^{FMNB-2}$, is shown in Equation (9).

$$CMF_{comb.x_i x_j}^{FMNB-2} = \frac{a+b}{c+d} \quad (9a)$$

$$a = \widehat{w}_1 \exp\left(\hat{\beta}_{0,1} + \hat{\beta}_{i,1} x_i + \hat{\beta}_{j,1} x_j + \sum_{k=1,k\neq i,k\neq j}^{p} \hat{\beta}_{k,1} \bar{x}_k\right) \quad (9b)$$

$$b = \widehat{w}_2 \exp\left(\hat{\beta}_{0,2} + \hat{\beta}_{i,2} x_i + \hat{\beta}_{j,2} x_j + \sum_{k=1,k\neq i,k\neq j}^{p} \hat{\beta}_{k,2} \bar{x}_k\right) \quad (9c)$$

$$c = \widehat{w}_1 \exp\left(\hat{\beta}_{0,1} + \hat{\beta}_{i,1} x_i^{base} + \hat{\beta}_{j,1} x_j^{base} + \sum_{k=1,k\neq i,k\neq j}^{p} \hat{\beta}_{k,1} \bar{x}_k\right) \quad (9d)$$

$$d = \widehat{w}_2 \exp\left(\hat{\beta}_{0,2} + \hat{\beta}_{i,2} x_i^{base} + \hat{\beta}_{j,2} x_j^{base} + \sum_{k=1,k\neq i,k\neq j}^{p} \hat{\beta}_{k,2} \bar{x}_k\right) \quad (9e)$$

Note that under NB models (i.e., Equation 3), the combined CMF for multiple covariates equals to the multiplicative of the single ones. However, this is not the case under FMNB-2 models (i.e., Equation 9), except when the parameters of the two components are identical, which becomes a NB model.

It is important to note that the covariates are considered to be independent within each component of an FMNB-2 model, and their coefficients are also estimated independently. However, this does not mean that their safety effects are independent. For more discussion on the combined safety effects with FMNB-2 models, please see Appendix A.

Finally, the standard error values of the CMFs for both NB and FMNB-2 models can be approximated using delta method (Rice, 2007). Assuming the CMFunction is $CMF = G(\boldsymbol{\beta})$, and the expectation of vector $\boldsymbol{\beta}$ is $\widehat{\boldsymbol{\beta}}$. The variance of CMF at a specific point, $var(CMF)$, is calculated as Equation 10.

$$var(CMF) \approx \frac{\partial G(\widehat{\boldsymbol{\beta}})}{\partial \widehat{\boldsymbol{\beta}}} \times Cov(\boldsymbol{\beta}) \times t(\frac{\partial G(\widehat{\boldsymbol{\beta}})}{\partial \widehat{\boldsymbol{\beta}}}) \quad (10)$$

Where,

$\frac{\partial G(\widehat{\boldsymbol{\beta}})}{\partial \widehat{\boldsymbol{\beta}}}$ = a row vector of partial derivatives of $G(\boldsymbol{\beta})$ at point $\widehat{\boldsymbol{\beta}}$;

$Cov(\boldsymbol{\beta})$ = variance-covariance matrix of $\boldsymbol{\beta}$; and

$t(\bullet)$ = transpose of • in the parentheses.

Particularly, with NB models, the variance of a CMF for feature $x_j$ is shown as Equation 11.

$$var(CMF) \approx \left(\frac{dG(\widehat{\beta_j})}{d\widehat{\beta_j}}\right)^2 \times var(\beta_x) = \left((x_j - x_j^{base}) \times e^{\widehat{\beta_j}(x_j - x_j^{base})}\right)^2 \times var(\beta_x) \quad (11)$$

## 3. DATA AND MODELING RESULTS

This section briefly describes the dataset used to develop CMFunctions and documents the modeling results.

### 3.1. Data description

This study utilized the rural multilane segment crash data on divided highways in California and Texas, which were also analyzed during the National Cooperative Highway Research Program (NCHRP) 17-29 project (Lord et al., 2009). The California data were originally obtained from the Federal Highway Administration (FHWA)'s Highway Safety Information System (HSIS) maintained by the University of North Carolina, and the Texas data were obtained from the Department of Public Safety (DPS) and the Texas Department of Transportation (TxDOT). The dataset contained a total of 2,587 roadway segments with 12-ft lane width only in order to estimate the NB regression models with baseline conditions, and used for developing CMFs for divided rural multilane highways. Table 1 shows the summary statistics of the input data for modeling. It should be noted that one roadway segment may include intersections if there are no roadway geometric changes along the segment. However, intersection and intersection-related crashes were all removed from the analysis.

Table 1. Summary statistics of the dataset

| Variable | Maximum | Minimum | Average | Standard Deviation |
|---|---|---|---|---|
| Average AADT ($F$), (veh/day) | 89,264 | 158 | 13,799 | 11,281 |
| Segment length ($L$), (mile) | 11.21 | 0.1 | 0.82 | 1.05 |
| Median width[a] ($MW$), (feet) | 240 | 1 | 47.07 | 29.41 |
| Right-shoulder width[b] ($RSW$), (feet) | 19 | 0 | 7.68 | 1.98 |
| Injury crashes[c] | 148 | 0 | 3.17 | 6.30 |

NOTE: [a] Median width includes the left shoulder widths; [b] Average right-shoulder width (both sides); [c] Injury crashes include only KAB crashes for five to ten years (K=fatal, A=incapacitating injury, and B=non-incapacitating injury).

### 3.2. Modeling results

This paper builds on the modeling results of earlier work in Park et al. (2014). With the same data, they applied various finite mixture models based on the Bayesian estimation method, and concluded that the regular FMNB-2 model and the constrained FMNB-2 model (termed as a CFMNB-2) were the best models to describe the dataset. The CFMNB-2 model was estimated by constraining the parameters of the median and right-shoulder widths in one component to be zero because their estimates were not much different from zero at a 95% significance level. The component-wise mean functional form was as follows: $\mu_{i,k} = t_i L_i F_i^{\beta_{1,k}} \exp(\beta_{0,k} + \beta_{2,k} MW_i + \beta_{3,k} RSW_i)$, where $t_i$ is the number of years, and $\{\beta_{0,k}, \beta_{1,k}, \beta_{2,k}, \beta_{3,k}\}$ are the parameters to be estimated for component $k$. The modeling results of each model are reproduced in Table 2. According to the result for the CFMNB-2 model, it suggests that the population consists of two distinct sub-populations whose regression parameters and degrees of dispersion are different from each other. With the coefficients estimated in Table 2, the sample averages of the estimated means for Component 1 and Component 2 were computed as $\bar{\mu}_1 = 2.96$ (crashes/year) and $\bar{\mu}_2 = 4.39$ (crashes/year), respectively. This indicates that Component 1 is associated with smaller-mean value observations and Component 2 with higher-mean value observations. The over-dispersion parameter in the NB model ($\hat{\phi} = 3.225$) has been split into two values: i.e., $\hat{\phi}_1 = 6.448$ for Component 1 and $\hat{\phi}_2 = 1.893$ for Component 2. This indicates that the higher-mean value observations (Component 2) are more dispersed than the smaller-mean value observations (Component 1).

Table 2. Modeling results for NB, FMNB-2, and CFMNB-2 models

| Estimated Parameters | NB | FMNB-2 | | CFMNB-2 | |
|---|---|---|---|---|---|
| | | Component 1 | Component 2 | Component 1 | Component 2 |
| $\hat{\beta}_{0,k}$ (Intercept) | -8.5574 (0.2397)* | -8.5272 (0.2862) | -6.8581 (1.2664) | -8.4073 (0.2712) | -6.8646 (1.4437) |
| $\hat{\beta}_{1,k}$ (Average AADT) | 0.9015 (0.0234) | 0.8387 (0.0286) | 0.9078 (0.1151) | 0.8344 (0.0283) | 0.9168 (0.1327) |
| $\hat{\beta}_{2,k}$ (Median Width) | -0.0015 (0.0006) | 0.0013 (0.0008) | -0.0191 (0.0067) | - | -0.0184 (0.0065) |
| $\hat{\beta}_{3,k}$ (Right-Shoulder Width) | -0.0455 (0.0094) | 0.0014 (0.0113) | -0.1509 (0.0412) | - | -0.1643 (0.0447) |
| $\hat{\phi}_k$ (Dispersion parameter) | 3.225 (0.222) | 6.7945 (1.143) | 2.149 (0.803) | 6.448 (1.006) | 1.893 (0.789) |
| $\hat{w}_k$ (Component proportion) | - | 0.857 (0.040) | 0.143 (0.040) | 0.880 (0.033) | 0.120 (0.033) |
| Model Comparison Criteria | | | | | |
| -2LL[a] | 9432.7 | 9300.7 | | 9304.8 | |
| AIC[b] | 9442.7 | 9322.7 | | 9322.8 | |
| Log(ML[c]) | -4752.2 | -4708.3 | | -4691.7 | |

Source: Park et al. (2014)
NOTE: * Values in parenthesis indicate the standard error of each coefficient.
[a] LL=log likelihood; [b] AIC=Akaike information criterion; [c] ML=marginal likelihood.

## 4. COMPARISON OF THE CMFUNCTIONS

This section analyzes the CMFunctions developed from NB and FMNB-2 models. Section 4.1 compares the CMFunction for a single treatment and Section 4.2 discusses the combined safety effects and the development of adjustment factors.

### 4.1.CMFunctions for a single treatment

Based on the CMFunctions provided in Equations (2) and (8), a comparison was carried out between the models with the parameter estimation results in Section 3. For the base conditions of each variable, the values recommended in NCHRP Project 17-29 (Lord et al., 2009) were basically adopted: i.e., 30ft for median width including left shoulder widths and 8ft for right-shoulder width. Since the median width used for modeling also included the left shoulder width for both sides, we can use 30ft as a base condition for the median width. The summary statistics of the variables are shown in Table 3 along with the respective base condition value. Note that the base condition value for the right-shoulder width is very close to the average of the sample data, while the base condition value for the median width is much smaller than the sample average. In the dataset, the median width greater than 120ft accounted for a very small proportion of the observations (2.24%), but it was found that wider median widths greater than 120ft have an unduly influence on crash occurrence. Therefore, after a careful consideration and based on previous work on median width

(Miaou et al., 2005), we limited the range of the median width to 120ft as the maximum width when deriving the CMFunctions for median width.

Table 3. Summary statistics of variables and their base condition values

| Variable | Min. | Max. | Average | Standard Deviation | Base Condition |
|---|---|---|---|---|---|
| Median width (feet) | 1 | 240 | 47.07 | 29.41 | 30 |
| Right-shoulder width (feet) | 0 | 19 | 7.68 | 1.98 | 8 |

The resultant CMFunctions for median width and right-shoulder width are presented in Figure 2, respectively. While the CMFunctions of the NB model are approximately straight lines for both variables, those of the CFMNB-2 model take on a more marked curve-shape. The difference in the shape mainly results from the fact that the NB model takes the average effect of a covariate across all sites, whereas the CFMNB-2 model takes into account the differential responsiveness of crash frequency to the covariate. As already noticed in the parameter estimation results in Section 3, the observations assigned to Component 1 (smaller-mean component) were not influenced by the median width and the right-shoulder width, whereas the observations in Component 2 (higher-mean component) were significantly affected by these variables. This effect is reflected in the shape of CMFunction derived from the CFMNB-2 model.

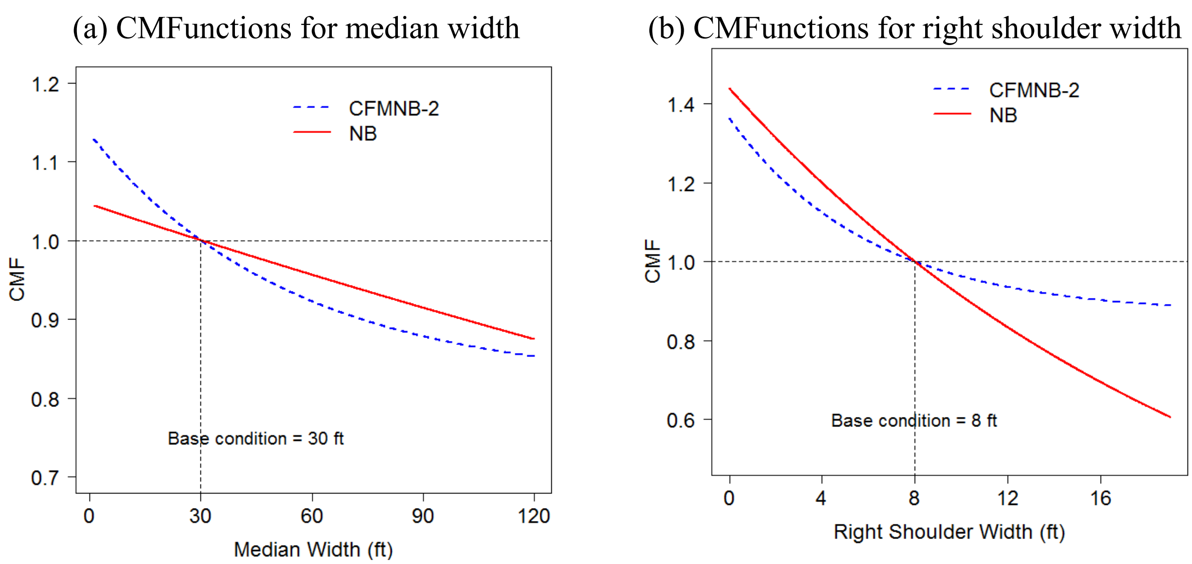



Figure 2. CMFunction comparisons between NB and CFMNB-2 models

Another good property about the shape of CMFunction in the CFMNB-2 model is that the safety effect of a covariate eventually levels off as the covariate increases significantly from the base condition. For example, it can be seen from Figure 2(a) that the safety effect of median width stabilizes after around 110ft. The same tendency is noticed for right-shoulder width after around 16ft. These trends are not observable in the NB model. This is partly supported by a few researchers who have noted that design elements, such as shoulder or lane width could follow a U-shaped relationship with safety (Hauer, 2000; Xie et al., 2007; Li et al., 2008). In a U-shaped relationship, narrow and wide widths experience more crashes. McLean (1996) explained the U-shaped relationship between safety and shoulder width by suggesting that very wide shoulders can often be used as an additional lane, which may lead to an increase in accident rates (Hauer, 2000).

(a) CMFunctions for median width (b) CMFunctions for right shoulder width

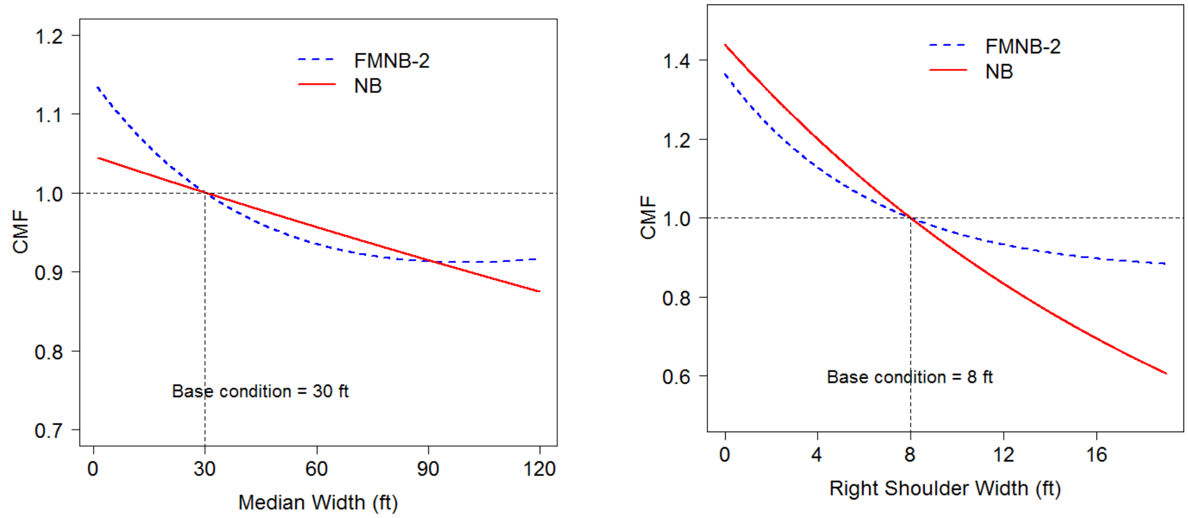



Figure 3. CMFunction comparisons between NB and FMNB-2 models

The CMF curves from the CFMNB-2 model did not exhibit a complete U-shaped relationship within the sample boundaries, but when the coefficients from the FMNB-2 model (see Table 2) were used, the CMFunction for median width revealed a U-shaped relationship, as shown in Figure 3(a). After a median width larger than around 100ft an increase in crashes can be observed. However, note that there are very few observations beyond median width of 100ft. Although this relationship can be still debatable, the bottom line is that the CMFunctions derived from the FMNB-2 model is more flexible and leave much more possibilities about the true effect of a design element on crash occurrence. On the other hand, the CMF curve for right-shoulder width from the FMNB-2 model (Figure 3(b)) did not exhibit a U-shaped curve and remained almost unchanged from the CFMNB-2 model. This is because the absolute of the coefficient of the right-shoulder width in Component 2 was much larger than that in Component 1 (i.e., -0.1509 vs. 0.0014). This small value in Component 1, even with a large weight, exercised little influence on the calculation of the CMF curve within the sample boundaries. Theoretically, if the maximum right shoulder width was increased to a very large value, the U-shape relationship may be observable.

**4.2.Combined CMF of multiple treatments**

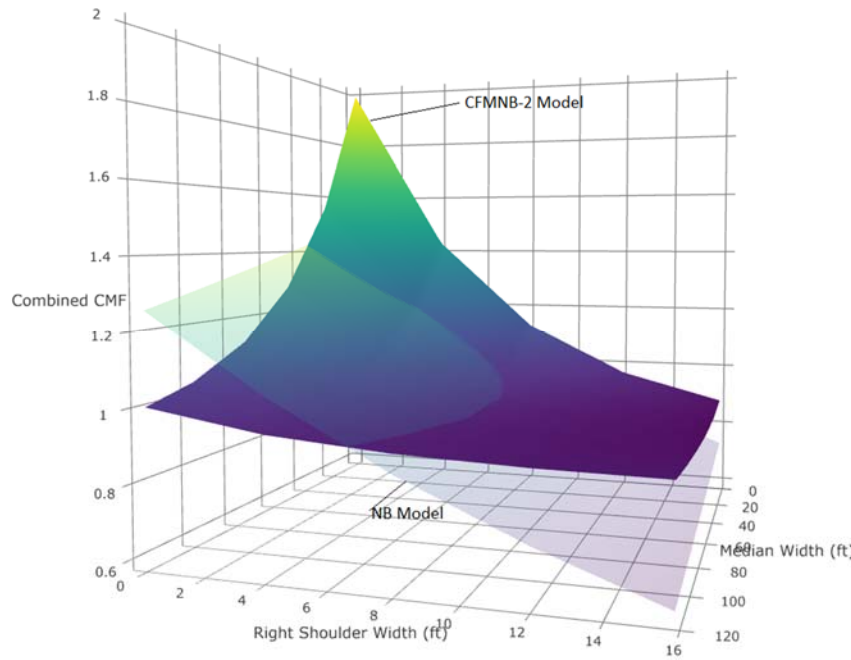
The combined CMFs can be calculated based on Equations (3) and (9). With the dataset in this paper, the combined CMFs for the two covariates (i.e., median width and right shoulder width) were derived from the models. Under the NB model, the combined CMF equal to the multiplicative of the two single ones, since their effects are assumed to be independent. While under CFMNB-2 and FMNB-2 models, the combined CMFunction is a function of the two covariates. The combined CMFunctions are illustrated in Figure 4 in 3-D relationship. The combined CMFs with standard error values for some typical combinations of median and right shoulder widths are presented in Table 4 for explicit comparison. It can be seen that the CMFunction under NB model is a plane over the range of the two covariates (in logarithm form). No matter what the value of one covariate is, one unit change in the other covariate brings the same amount of change in the combined CMF.

However, the combined CMFunctions under both CFMNB-2 and FMNB-2 models are curvature surfaces. When the two covariates are large (i.e., both median and right shoulder widths are wide), the surface is relatively flat. Conversely, when they are small (i.e., narrow median and right shoulder widths), the surface becomes steep. That is to say, when both median and right shoulder widths are narrow, widening one or both of them by some units is more effective than that when both are wide. For example, when the median width is 25ft, widening the right shoulder width from 0 to 4ft could reduce the expected crashes by 22.7% (1-1.19/1.54) with CFMNB-2 model, or 21.9% (1-1.21/1.55) with FMNB-2 model. When the median width is 100ft, doing so could only reduce 8.0% (1-0.92/1.00) and 7.7% (1-0.96/1.04) of expected crashes, respectively, with the two models. This is consistent with the practical engineering experience. In other words, the two covariates are not independent in affecting crash occurrences. Thus the independence assumption with the NB model is invalid.

In Table 4, numbers in parenthesis indicate the standard error value of each combined CMF. Most of the combined CMFs are statistically significant from 1.0 at a 90% level, except for some ranges where the estimated CMFs are approximately 1.0. However, the standard error of CMFs derived from CFMNB-2 and FMNB-2 models are always higher than that derived with NB model. One possible reason is the former two CMFs are generally greater than the latter. Another possible reason is that in CFMNB-2 and FMNB-2 models CMFunctions were developed with nine or seven parameters, respectively, which introduced more uncertainty into the resultant CMFs. By comparison, NB models used two parameters for estimating the combined CMFs.

Another finding worth mentioning is the differences with the combined CMFs between the two mixture models when both covariates are relatively large. The combined CMF derived from the CFMNB-2 model decreases continuously as they become wider. On the other hand, the combined CMF produced from the FMNB-2 model only increases slightly after further widening the median and right shoulder width. In the figures, the former combined CMF plateaus around 0.83, while the latter combined CMF caps at 0.91. This might be caused by the “U-shape” effect of the median width captured within FMNB-2 model, as discussed in the previous section. This needs further analysis in order to investigate the true effects of wide pavements. Nevertheless, both models generally provide a more reasonable combined CMF than the multiplication of two independent CMFs produced from the traditional NB model.

(a) CFMNB-2 Model




(b) FMNB-2 Model

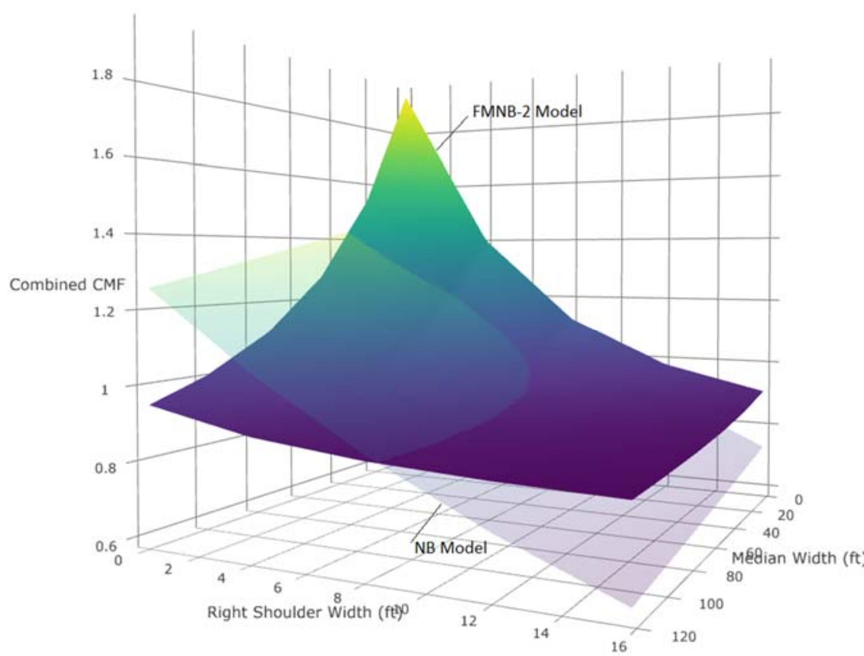



Figure 4. Combined CMF comparisons between NB, CFMNB-2 and FMNB-2 models

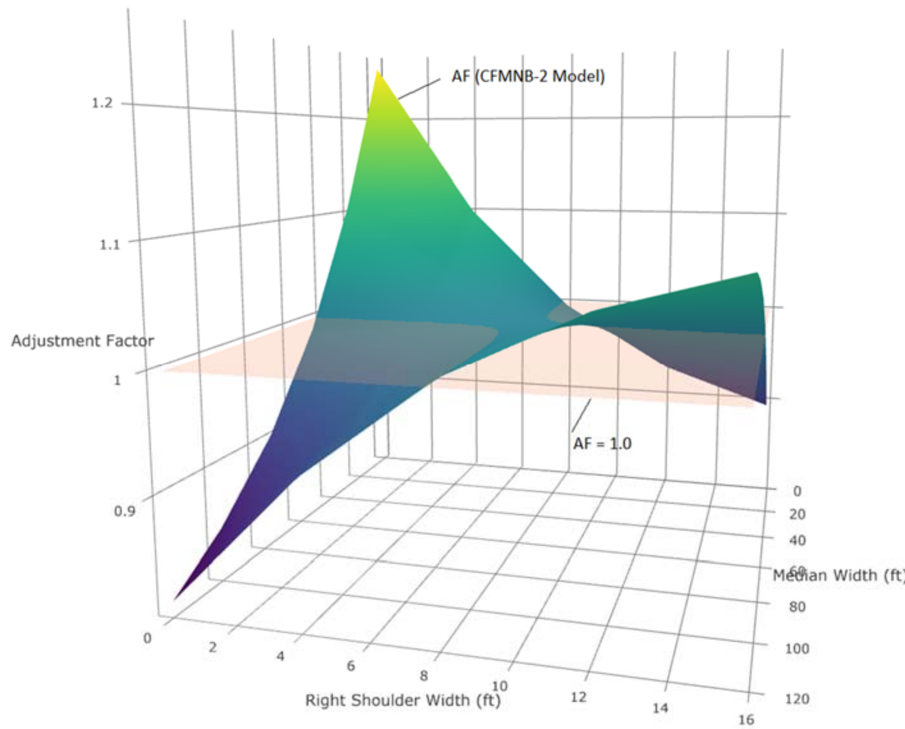
Table 4. Combined CMF for typical median and right shoulder widths

| Model | RSW (ft) | MW (ft) | | | | | |
|---|---|---|---|---|---|---|---|
| | | 1 | 25 | 50 | 75 | 100 | 120 |
| NB | 0 | 1.50 (0.111) | 1.45 (0.109) | 1.40 (0.111) | 1.35 (0.117) | 1.30 (0.125) | 1.26 (0.132) |
| | 4 | 1.25 (0.048) | 1.21 (0.045) | 1.16 (0.050) | 1.12 (0.059) | 1.08 (0.070) | 1.05 (0.079) |
| | 8 | 1.04 (0.019) | 1.01 (0.003) | 0.97 (0.012) | 0.93 (0.027) | 0.90 (0.041) | 0.87 (0.051) |
| | 12 | 0.87 (0.040) | 0.84 (0.032) | 0.81 (0.030) | 0.78 (0.033) | 0.75 (0.039) | 0.73 (0.045) |
| | 16 | 0.73 (0.059) | 0.70 (0.054) | 0.67 (0.050) | 0.65 (0.049) | 0.63 (0.050) | 0.61 (0.051) |
| CFMNB-2 | 0 | 1.93 (0.377) | 1.54 (0.241) | 1.27 (0.183) | 1.11 (0.145) | 1.00 (0.112) | 0.95 (0.090) |
| | 4 | 1.40 (0.121) | 1.19 (0.064) | 1.06 (0.058) | 0.97 (0.053) | 0.92 (0.046) | 0.89 (0.040) |
| | 8 | 1.12 (0.052) | 1.02 (0.006) | 0.95 (0.015) | 0.90 (0.024) | 0.87 (0.027) | 0.86 (0.030) |
| | 12 | 0.98 (0.044) | 0.92 (0.023) | 0.89 (0.022) | 0.87 (0.027) | 0.85 (0.031) | 0.84 (0.033) |
| | 16 | 0.91 (0.041) | 0.88 (0.031) | 0.86 (0.031) | 0.85 (0.033) | 0.84 (0.036) | 0.83 (0.037) |
| FMNB-2 | 0 | 1.98 (0.372) | 1.55 (0.233) | 1.28 (0.166) | 1.13 (0.130) | 1.04 (0.113) | 1.00 (0.109) |
| | 4 | 1.43 (0.123) | 1.21 (0.068) | 1.07 (0.057) | 0.99 (0.058) | 0.96 (0.064) | 0.95 (0.072) |
| | 8 | 1.13 (0.050) | 1.02 (0.006) | 0.95 (0.016) | 0.93 (0.030) | 0.92 (0.043) | 0.92 (0.054) |
| | 12 | 0.96 (0.052) | 0.91 (0.038) | 0.89 (0.038) | 0.89 (0.043) | 0.90 (0.052) | 0.91 (0.061) |
| | 16 | 0.88 (0.075) | 0.86 (0.071) | 0.86 (0.072) | 0.87 (0.075) | 0.89 (0.080) | 0.91 (0.087) |

NOTE: Values in parenthesis indicate the standard error of each combined CMF. Numbers with underline indicate significantly different from 1.0 at a 90% level.

Park and Abdel-Aty (2015a) proposed using adjustment factor (AF) to capture the dependence of simultaneously implemented treatments. An adjustment factor is defined as the ratio between the combined CMF to the multiplicative of single ones. An AF greater than 1.0 means the combined safety effect is smaller than the "sum" of individuals, and vice versa. If an AF equals to 1.0, the multiple treatments are independent of each other. Using the same concept, the adjustment functions were derived from the CFMNB-2 and FMNB-2 models (for the detailed description about deriving AFs, please refer to Appendix B.) Their surfaces are shown in Figure 5. It can be seen that the adjustment function surfaces are similar under the two finite mixture models. Over some areas, the AF is greater than 1.0, and over others it is smaller than 1.0. The AF is about 1.0 around the base condition (i.e., 30ft for median width and 8ft for right shoulder width). It is greater than 1.0 when both covariates are narrow or wide, and the value increases as they become narrower or wider. Over the area where one of the two is narrow and the other is wide, the AF is smaller than 1.0. The further the covariate combination is away from the base condition, the more bias the *HSM* method brings. Some specific AF values as well as standard error for typical median and right shoulder width combinations are shown in Table 5. All of the AFs are significantly different from 1.0 at a 95% level. Based on this result, the *HSM* method might over-estimate the combined CMFs under particular conditions and under-estimate them under some other conditions. This result basically agrees with the previous study conducted by Park and Abdel-Aty (2015a), in which they also reported that the *HSM* method produces biased or erroneous CMFs for combined treatments.

(a) CFMNB-2 Model




(b) FMNB-2 Model

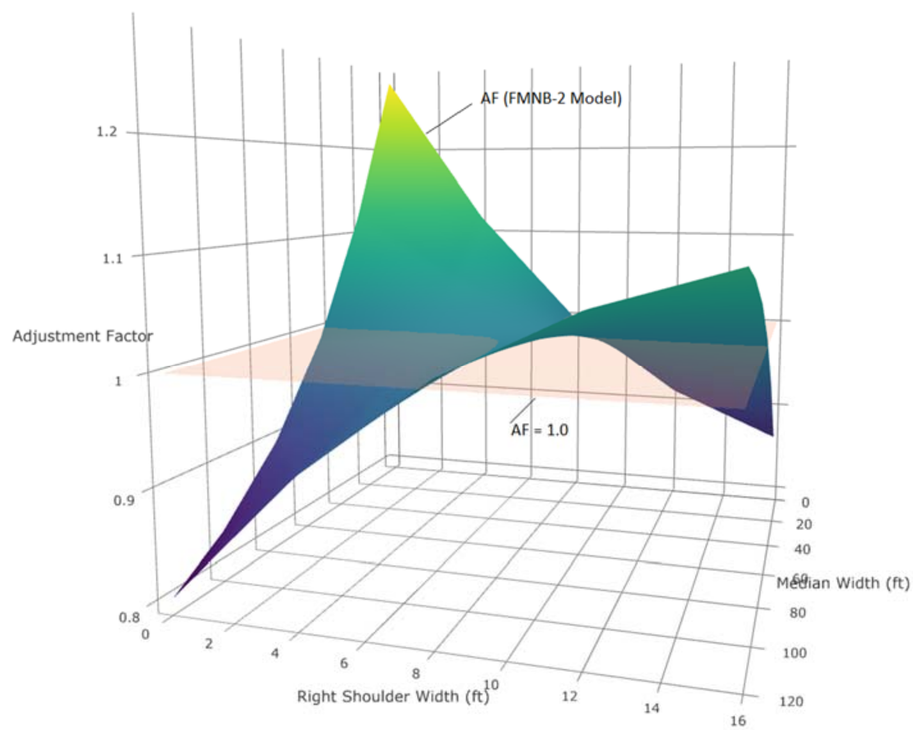



Figure 5. Adjustment factors derived from CFMNB-2 and FMNB-2 models

Table 5. Adjustment factor derived from CFMNB-2 and FMNB-2 models

| Model | RSW (ft) | MW (ft) | | | | | |
|---|---|---|---|---|---|---|---|
| | | 1 | 25 | 50 | 75 | 100 | 120 |
| CFMNB-2 | 0 | 1.26 (0.127) | 1.11 (0.048) | 0.99 (0.004) | 0.91 (0.034) | 0.85 (0.052) | 0.82 (0.062) |
| | 4 | 1.10 (0.046) | 1.04 (0.018) | 0.99 (0.001) | 0.96 (0.012) | 0.94 (0.019) | 0.93 (0.023) |
| | 8 | 1.00 (0.002) | 1.00 (0.000) | 1.00 (0.001) | 1.00 (0.001) | 1.01 (0.002) | 1.01 (0.002) |
| | 12 | 0.93 (0.026) | 0.97 (0.009) | 1.01 (0.002) | 1.03 (0.008) | 1.05 (0.011) | 1.06 (0.014) |
| | 16 | 0.89 (0.037) | 0.96 (0.013) | 1.01 (0.002) | 1.05 (0.010) | 1.08 (0.015) | 1.09 (0.019) |
| FMNB-2 | 0 | 1.28 (0.114) | 1.12 (0.046) | 0.99 (0.004) | 0.90 (0.036) | 0.84 (0.057) | 0.80 (0.069) |
| | 4 | 1.12 (0.044) | 1.05 (0.018) | 1.00 (0.001) | 0.96 (0.014) | 0.93 (0.022) | 0.92 (0.027) |
| | 8 | 0.99 (0.002) | 1.00 (0.000) | 1.00 (0.001) | 1.01 (0.002) | 1.01 (0.002) | 1.01 (0.003) |
| | 12 | 0.91 (0.025) | 0.96 (0.010) | 1.01 (0.002) | 1.04 (0.010) | 1.06 (0.016) | 1.07 (0.019) |
| | 16 | 0.86 (0.036) | 0.94 (0.014) | 1.01 (0.002) | 1.06 (0.014) | 1.09 (0.022) | 1.10 (0.028) |

NOTE: Values in parenthesis indicate the standard error of the adjustment factors. All of the adjustment factors are significantly different from 1.0 at a 95% level.

## 5. SUMMARY AND CONCLUSION

Recently, applications of a finite mixture regression model have gained an interest from researchers because of its considerable potential for addressing the unobserved heterogeneity in highway vehicle crash data. 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 in various highway safety analyses as compared to traditional models. In this respect, this study aimed to investigate the relative performance of the two models (i.e., FMNB-2 model vs. NB model) in terms of developing CMFs. The procedure on how to derive the CMFunctions for both single and multiple covariates from the FMNB-2 model was descried in details and its characteristics were discussed by comparing it with that from the traditional NB model.

The CMFunction for the FMNB-2 model was not as simple as the one derived from the NB model since the conditional mean is a mixture of additive and multiplicative terms. However, the CMFunction derived from the FMNB-2 model showed advantages over that from the NB model since it considers the interactions between parameters and covariates, and hence can better account for the differential responsiveness of crash frequency with respect to a specific covariate. Based on the modeling results in Park et al. (2014), the CMF curves for the median width and right-shoulder width were derived for the constrained FMNB-2 model (i.e., CFMNB-2 model) and they were compared with those from the NB model. The CMF curve shapes produced by the CFMNB-2 model had a better property in that the safety effect of a covariate eventually leveled off as the values of the covariate increase significantly from its base condition value. On the other hand, when the regular FMNB-2 model – which is inferior to the CFMNB-2 model, but superior to the NB model – was used, the CMF curve for median width showed a U-shaped relationship. Further, the combined CMFs of the two covariates were estimated and compared with those estimated using the *HSM* method (i.e., multiplicative of single CMFs developed using the same models). The comparison indicates the two were not independent. Widening the median or right shoulder on narrow highways (recall that the lane width in the dataset is fixed) are more effective than that for wide ones. This result was supported by several previous research studies. The AFs were then estimated, showing that the *HSM* method could over- or under-estimate the combined CMFs under particular combination of covariates. The further the combination is away from the base condition, the higher the bias *HSM* method tends to induce.

Various CMFs for single treatments are available in the current literature (e.g., *HSM*, CMF Clearinghouse). Most of them were assumed to have linear effects or relationships on safety and to be independently affecting crash risk. Recent studies have indicated the linear and independent assumptions may not always be true. Methods for revealing nonlinear relationships have been proposed, and efforts are being made to evaluate the combined safety effects of multiple treatments (Lao et al. 2014; Park and Abdel-Aty 2015a). However, very limited approaches have showed the ability to adequately capture nonlinear effect and interaction impacts between variables simultaneously (e.g., multivariate adaptive regression splines (MARS) model proposed by Park and Abdel-Aty (2015b)). This paper shows that the FMNB-2 models are able to capture the nonlinear effects of single treatments as well as the "non-independent" combined effects of multiple treatments. Another option is provided to safety analysts for developing nonlinear and combined CMFs. To find the true safety effects of design elements and to develop accurate CMFs, it is suggested in the future that the FMNB-2 models be compared with others (e.g., MARS) to assess if the CMFs can be reproduced consistently. Finally, it is worth mentioning that this study used the same concept as that of traditional NB models to estimate combined CMFs. No interaction terms (e.g., median width × right shoulder width) were included in the models. The difference between combined CMFs and the multiplicative of single ones may partially come from nonlinearity of single CMFs. This also needs further analysis.

## APPENDIX A - ANALYTICAL ANALYSIS OF COMBINED SAFETY EFFECTS

*Based on NB Models*

To simplify the analysis, we only considered two covariates, denoted as $x_1$ and $x_2$, and the expected crash mean μ equals

$$\mu = C_{NB} \times \exp(\beta_1 \times x_1 + \beta_2 \times x_2) \tag{A.1}$$

Where,

$\beta_1, \beta_2$ = coefficients for $x_1$ and $x_2$, respectively; and,

$C_{NB}$ = a constant (i.e., scale factor) which does not depend on $x_1$ or $x_2$.

The CMFunctions for $x_1$ and $x_2$ are shown in Equations A.2 and A.3, respectively.

$$CMF_{x1}^{NB} = \frac{C_{NB}\times\exp(\beta_1\times x_1+\beta_2\times x_2)}{C_{NB}\times\exp(\beta_1\times x_1^{base}+\beta_2\times x_2)} = C_{NB,1}\times\exp(\beta_1\times x_1) \quad \text{(A.2)}$$

$$CMF_{x2}^{NB} = \frac{C_{NB}\times\exp(\beta_1\times x_1+\beta_2\times x_2)}{C_{NB}\times\exp(\beta_1\times x_1+\beta_2\times x_2^{base})} = C_{NB,2}\times\exp(\beta_2\times x_2) \quad \text{(A.3)}$$

Where,

$CMF_{x1}^{NB}$ = specific (single) CMF for $x_1$;
$CMF_{x2}^{NB}$ = specific (single) CMF for $x_2$;
$x_1^{base}$ = base condition for $x_1$;
$x_2^{base}$ = base condition for $x_2$; and,
$C_{NB,1}$ and $C_{NB,2}$ = two constants, neither depends on $x_1$ or $x_2$.

And the combined CMF for $x_1$ and $x_2$ is shown in Equation A.4.

$$CMF_{Comb,x1,x2}^{NB} = \frac{C_{NB}\times\exp(\beta_1\times x_1+\beta_2\times x_2)}{C_{NB}\times\exp(\beta_1\times x_1^{base}+\beta_2\times x_2^{base})}$$
$$= C_{NB,Comb}\times\exp(\beta_1\times x_1)\times\exp(\beta_2\times x_2) \quad \text{(A.4)}$$

Where,

$CMF_{Comb,x1,x2}^{NB}$ = combined CMF for $x_1$ and $x_2$; and,
$C_{NB,Comb}$ = a constant that does not depend on $x_1$ or $x_2$.

It is trivial to show that $C_{NB,Comb} = C_{NB,1} \times C_{NB,2}$, and further we have

$$CMF_{Comb,x1,x2}^{NB} = CMF_{x1}^{NB}\times CMF_{x2}^{NB} \quad \text{(A.5)}$$

Equation A.5 implies that under the NB model, the combined CMF of two variables equals to the multiplicative of the two single ones. This is also commonly said the safety effects of the two variables are independent. If we further take the logarithm of both sides of Equation A.4, it leads to Equations A.6

$$\log\left(CMF_{Comb,x1,x2}^{NB}\right) = \log\left(C_{NB,Comb}\right)+\beta_1\times x_1+\beta_2\times x_2 \quad \text{(A.6)}$$

As can be seen, Equation A.6 does not contain any interaction term between $x_1$ and $x_2$ (such as, $x_1\times x_2$, $e^{x_1+x_2}$, $x_1^{x2}$, etc.) Take the partial derivative of Equation A.6 with respect to $x_1$ and $x_2$ respectively, we have Equations A.7 and A.8.

$$\frac{\partial\log\left(CMF_{Comb,x1,x2}^{NB}\right)}{\partial x1} = \beta_1 \quad \text{(A.7)}$$

$$\frac{\partial\log\left(CMF_{Comb,x1,x2}^{NB}\right)}{\partial x2} = \beta_2 \quad \text{(A.8)}$$

Equation A.7, which is partial derivative with respect to $x_1$, is free of $x_2$. Similarly, Equation A.8 is free of $x_1$.

In other words, the statement that the safety effects of two variables are independent is equivalent to that the combined CMF equals to the multiplicative of two single ones. This independence also implies: (1) the logarithm of the combined CMF does not contain any interaction term between the two variables; and (2) the partial derivative of logarithm of the combined CMF with respect to one variable does not depend on the other one.

*Based on FMNB-2 Models*

To simplify the analysis, we only consider two variables, $x_1$ and $x_2$, which is similar with that of NB models. The expected crash mean μ is given as Equation A.9.

$$\mu = w_1 \times \exp(\beta_{0,1} + \beta_{1,1} \times x_1 + \beta_{2,1} \times x_2) + w_2 \times \exp(\beta_{0,2} + \beta_{1,2} \times x_1 + \beta_{2,2} \times x_2) \quad \text{(A.9a)}$$

Or equivalently,

$$\mu = C_{FMNB-2,1} \times \exp(\beta_{1,1} \times x_1 + \beta_{2,1} \times x_2) + C_{FMNB-2,2} \times \exp(\beta_{1,2} \times x_1 + \beta_{2,2} \times x_2) \quad \text{(A.9b)}$$

Where,

$w_1$, $w_2$ = the weight factors of the two components, respectively;

$\beta_{j,k}$ = coefficient for variable $j$ in component $k$, $j$ = 1 or 2, $k$ = 1 or 2; $\beta_{0,1}$ and $\beta_{0,2}$ are intercepts in the two components, respectively; and,

$C_{FMNB-2,1}$ and $C_{FMNB-2,2}$ = two constants that do not depend on $x_1$ or $x_2$.

As has been documented in the manuscript (Section 2.2), the (single) CMF for $x_1$ and $x_2$ are produced below as Equations A.10 and A.11.

$$CMF_{x1}^{FMNB-2} = \frac{w_1 \times \exp(\beta_{1,1} \times x_1 + \beta_{2,1} \times \overline{x_2}) + w_2 \times \exp(\beta_{1,2} \times x_1 + \beta_{2,2} \times \overline{x_2})}{w_1 \times \exp(\beta_{1,1} \times x_1^{base} + \beta_{2,1} \times \overline{x_2}) + w_2 \times \exp(\beta_{1,2} \times x_1^{base} + \beta_{2,2} \times \overline{x_2})}$$
$$= C_{FMNB-2,x1,1} \times \exp(\beta_{1,1} \times x_1) + C_{FMNB-2,x1,2} \times \exp(\beta_{1,2} \times x_1) \quad \text{(A.10)}$$

$$CMF_{x2}^{FMNB-2} = C_{FMNB-2,x2,1} \times \exp(\beta_{2,1} \times x_2) + C_{FMNB-2,x2,2} \times \exp(\beta_{2,2} \times x_2) \quad \text{(A.11)}$$

Where,

$C_{FMNB-2,x1,1}$ and $C_{FMNB-2,x1,2}$ = two constants that do not depend on $x_1$; and,

$C_{FMNB-2,x2,1}$ and $C_{FMNB-2,x2,2}$ = two constants that do not depend on $x_2$.

And the combined CMF for $x_1$ and $x_2$ can be derived as Equation A.12.

$$CMF_{Comb.x1,x2}^{FMNB-2} = \frac{w_1 \times \exp(\beta_{1,1}x_1 + \beta_{2,1}x_2) + w_2 \times \exp(\beta_{1,2}x_1 + \beta_{2,2}x_2)}{w_1 \times \exp(\beta_{1,1}x_1^{base} + \beta_{2,1}x_2^{base}) + w_2 \times \exp(\beta_{1,2}x_1^{base} + \beta_{2,2}x_2^{base})}$$
$$= C_{comb,1} \times \exp(\beta_{1,1}x_1 + \beta_{2,1}x_2) + C_{comb,2} \times \exp(\beta_{1,2}x_1 + \beta_{2,2}x_2) \quad \text{(A.12)}$$

Where,

$C_{comb,1}$ and $C_{comb,2}$ = two constants that do not depend on $x_1$ or $x_2$.

Notice that $CMF_{Comb.x1,x2}^{FMNB-2} = CMF_{x1}^{FMNB-2} \times CMF_{x2}^{FMNB-2}$ does not hold (unless the coefficients in the two components are identical, which becomes an NB model). For example, the right hand side contains two terms of $\exp(\beta_{1,2}x_1 + \beta_{2,2}x_2)$ and $\exp(\beta_{1,1}x_1 + \beta_{2,2}x_2)$, which are not included in the left hand side (i.e., Equation A.11).

The logarithm of $CMF_{Comb.x1,x2}^{FMNB-2}$ is shown in Equation A.13.

$$\log\left(CMF_{Comb.x1,x2}^{FMNB-2}\right) = \log[C_{comb,1}\exp(\beta_{1,1}x_1 + \beta_{2,1}x_2) + C_{comb,2}\exp(\beta_{1,2}x_1 + \beta_{2,2}x_2)] \tag{A.13}$$

It is not simple to show the partial derivative of $\log\left(CMF_{Comb.x1,x2}^{FMNB-2}\right)$ with respect to $x_1$ or $x_2$, but obviously neither is free of $x_2$ or $x_1$.

It can be concluded that, under FMNB-2 models, the safety effects of multiple treatments are not independent, although in each of the two components the variables had been assumed to be independent.

## APPENDIX B - DERIVATIVE OF ADJUSTMENT FACTORS

*Based on NB Models*

As defined in the previous literature (e.g., Park and Abdel-Aty, 2015a), an Adjustment Factor (AF) can be estimated as dividing the combined CMF by the multiplicative of single ones. Using the results of Appendix A, the *AF* based on an NB model is shown in Equation B.1.

$$AF_{NB} = \frac{CMF_{Comb,x1,x2}^{NB}}{CMF_{x1}^{NB} \times CMF_{x2}^{NB}} \tag{B.1}$$

By inserting Equation A.5 into B.1, it can be seen that $AF_{NB} = 1.0$ for all $x_1$ and $x_2$. This is equivalent to the fact that the safety effects of the two variables are independent.

*Based on FMNB-2 Models*

Using the results of Appendix A, the *AF* based on an FMNB-2 model is shown in Equation B.2.

$$AF_{FMNB-2} = \frac{CMF_{Comb.x1,x2}^{FMNB-2}}{CMF_{x1}^{FMNB-2} \times CMF_{x2}^{FMNB-2}} \tag{B.2}$$

Inserting Equations A.10 to A.12 into B.1, $AF_{FMNB-2}$ becomes:

$$AF_{FMNB-2} = \frac{C_1 \times \exp(\beta_{1,1}x_1 + \beta_{2,1}x_2) + C_2 \times \exp(\beta_{1,2}x_1 + \beta_{2,2}x_2)}{[C_3\exp(\beta_{1,1}x_1) + C_4\exp(\beta_{1,2}x_1)] \times [C_5\exp(\beta_{2,1}x_2) + C_6\exp(\beta_{2,2}x_2)]} \tag{B.3}$$

Where, $C_1$, $C_2$, $C_3$, …, $C_6$, are constants, none of them depend on $x_1$ or $x_2$. Although it is not easy to simplify the expression of $AF_{FMNB-2}$, it is clearly a function of both $x_1$ and $x_2$. That is to say, the AF estimated using FMNB-2 models is not a constant, and it varies depending on specific conditions (i.e., different $x_1$'s and/or $x_2$'s).

## REFERENCES

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

Bonneson, J.A. and Lord, D. (2005) Role and Application of Accident Modification Factors in the Highway Design Process. Report No. FHWA/TX-05/0-4703-2, Texas Transportation Institute, College Station, Texas.

Bonneson, J.A., Lord, D., Zimmerman, K., Fitzpatrick, K., and Pratt, M. (2007) Development of Tools for Evaluating the Safety Implications of Highway Design Decisions. *Report No. FHWA/TX-07/0-4703-4*, Texas Transportation Institute, College Station, Texas.

Fitzpatrick, K., Lord, D., and Park, B.-J. (2008) Accident modification factors for medians on freeways and multilane highways. *Transportation Research Record* 2083, pp. 62-71.

Früwirth-Schnatter, S., 2006. *Finite Mixture and Markov Switching Models*. Springer Series in Statistics, Springer, New York

Gross, F., Jovanis, P.P., and Eccles, K.A. (2009) Safety effectiveness of lane and shoulder width combinations on rural, two-lane, undivided roads. *Transportation Research Record* 2103, pp. 42-49.

Hauer, E. (2000) Shoulder width, shoulder paving and safety. Unpublished manuscript prepared for the Federal Highway Administration, Toronto, ON. Retrieved April 2009 from http://ca.geocities.com/hauer@rogers.com/Pubs/Shoulderwidth.pdf.

Hauer, E. (2010) Cause, Effect and Regression in Road Safety: A Case Study. *Accident Analysis & Prevention,* Vol. 42, No. 4, pp. 1128-1135.

Hauer, E. (2013) Even perfect regression may not tell the effect of interventions. Transportation Research Board 92nd Annual Meeting. Washington, D.C.

Heydari, S., Fu, L., Lord, D., and Mallick, B.K. (2016) Multilevel Dirichlet process mixture analysis of railway grade crossing crash data. *Analytic Methods in Accident Research*, Vol. 9, pp. 27-43.

Hauer, E. (2015) *The Art of Regression Modeling in Road Safety*. Springer, USA.

Lao, Y., Zhang, G., Wang, Y., and Milton, J. (2014) Generalized nonlinear models for rear-end crash risk analysis. *Accident Analysis & Prevention* 62, pp. 9-16.

Li, X., Lord, D., and Zhang, Y. (2010) Development of accident modification factors for rural frontage road segments in Texas using results from generalized additive models. *ASCE Journal of Transportation Engineering* 42(4), pp. 74-83.

Li, X., Lord, D., Zhang, Y., and Xie, Y. (2008) Predicting motor vehicle crashes using support vector machine models. *Accident Analysis & Prevention* 40(4), pp. 1611-1618.

Long, J.S. (1997) *Regression Models for Categorical and Limited Dependent Variables*. SAGE Publications, Thousand Oaks, California.

Lord, D. and Bonneson, J.A. (2007) Development of accident modification factors for rural frontage road segments in Texas. *Transportation Research Record* 2023, pp. 20-27.

Lord, D., Geedipally, S.R., Persaud, B.N., Washington, S.P., van Schalkwyk, I.,Ivan, J.N., Lyon, C., and Jonsson, T. (2009) Methodology for Estimating the Safety Performance of Multilane Rural Highways. *NCHRP Web-Only Document 126*, National Cooperation Highway Research Program, Washington, D.C.

Lord, D. and Mannering, F. (2010) The Statistical Analysis of Crash-Frequency Data: A Review and Assessment of Methodological Alternatives. *Transportation Research Part A*, Vol. 44, No. 5, pp. 291-305.

Lord, D., and Kuo, P-F. (2012) Examining the Effects of Site Selection Criteria for Evaluating the Effectiveness of Traffic Safety Improvement Countermeasures. *Accident Analysis & Prevention*, Vol. 47, pp. 52-63.

Mannering, F., Shankar, V., and Bhat, C. (2016) Unobserved heterogeneity and the statistical analysis of highway accident data. *Analytic Methods in Accident* Research 11, pp. 1-16.

McLean, J. (1996) Review of Accidents and Rural Cross Section Elements Including Roadsides. *ARR 297*, ARR, ARRB, Victoria, Australia.

Miaou, S.-P., Bligh, R.P., Lord, D. (2005) Developing Median Barrier Installation Guidelines: A Benefit/Cost Analysis using Texas Data. Transportation Research Record 1904, pp. 3-19.

Park, B.-J. and Lord, D. (2009) Application of finite mixture models for vehicle crash data analysis. *Accident Analysis & Prevention* 41(4), pp. 683-691.

Park, B.-J., Lord, D., and Lee, C. (2014) Finite mixture modeling for vehicle crash data with application to hotspot identification. *Accident Analysis & Prevention* 71, pp. 319–326.

Park, J., Abdel-Aty, M., and Lee, C. (2014) Exploration and comparison of crash modification factors for multiple treatments on rural multilane roadways. *Accident Analysis & Prevention* 70, pp. 167–177.

Park, J., and Abdel-Aty, M. (2015a) Development of adjustment functions to assess combined safety effects of multiple treatments on rural two-lane roadways. *Accident Analysis & Prevention* 75, pp. 310–319.

Park, J., and Abdel-Aty, M. (2015b) Assessing the safety effects of multiple roadside treatments using parametric and nonparametric approaches. *Accident Analysis & Prevention* 83, pp. 203–213.

Park, J., and Abdel-Aty, M. (2016) Evaluation of safety effectiveness of multiple cross sectional features on urban arterials. *Accident Analysis & Prevention* 94, pp. 245–255.

Rice, J. (2007) *Mathematical statistics and data analysis. The 3rd Edition.* Duxbury advanced series, USA.

Shahdah, U., Saccomanno, F., and Persaud, B. (2014) Integrated traffic conflict model for estimating crash modification factors. *Accident Analysis & Prevention* 71, pp. 228-235.

Shirazi, M., Lord, D., Dhavala, S.S., and Geedipally, S.R. (2016) A Semiparametric Negative Binomial Generalized Linear Model for Modeling Over Dispersed Count Data with a Heavy Tail: Characteristics and Applications to Crash Data. *Accident Analysis & Prevention*, Vol. 91, pp. 10-18.

Washington, S.P., Persaud, B.N., Lyon, C., Oh, J. (2005) Validation of Accident Models for Intersections. *Report No. FHWA-RD-03-037*, Federal Highway Administration, Washington, D.C.

Wu, L., Lord, D., and Zou, Y. (2015) Validation of Crash Modification Factors Derived from Cross-Sectional Studies with Regression Models. *Transportation Research Record* 2514, pp. 88-96.

Wu, L., and Lord, D. (2016) Investigating the Influence of Dependence between Variables on Crash Modification Factors Developed Using Regression Models. *Transportation Research Board 95th Annual Meeting*, TRB, Washington D.C.

Xie, Y., Lord, D., and Zhang, Y. (2007) Predicting motor vehicle collisions using Bayesian neural networks: an empirical analysis. *Accident Analysis & Prevention* 39 (5), pp. 922-933.

Zou, Y., Zhang, Y., and Lord, D. (2013) Application of finite mixture of negative binomial regression models with varying weight parameters for vehicle crash data analysis. *Accident Analysis & Prevention* 50, pp. 1042-1051.