Despite the various crash modification factors (CMFs) that have been calculated and introduced in the Highway Safety Manual (HSM) (AASHTO, 2010), there are critical limitations that require investigation. First, the HSM provides CMFs for single treatments, but not for multiple treatments to roadway segments. The HSM suggests that CMFs are multiplied to estimate the combined safety effects of single treatments. However, the HSM cautions that the multiplication of CMFs may over- or under-estimate combined effects of multiple treatments. In this dissertation, several methodologies are proposed to estimate more reliable combined safety effects in both observational before-after studies and the cross-sectional method. Averaging two best combining methods is suggested to use to account for the effects of over- or under-estimation. Moreover, it is recommended to develop an adjustment factor and function to apply to estimate more accurate safety performance in assessing safety effects of multiple treatments. The multivariate adaptive regression splines (MARS) modeling is proposed to avoid the over-estimation problem through consideration of interaction impacts between variables in this dissertation.

Second, the variation of CMFs with different roadway characteristics among treated sites over time is ignored because the CMF is a fixed value that represents the overall safety effect of the treatment for all treated sites for specific time periods. Although few studies developed crash modification functions (CMFunctions) to overcome this limitation, there is a lack of prior studies on the variation in the safety effects of treated sites with different multiple roadway characteristics over time. In this study, adopting various multivariate linear and nonlinear modeling techniques is suggested to develop CMFunctions. Multiple linear regression modeling can be utilized to consider different multiple roadway characteristics. To reflect nonlinearity of predictors, a regression model with nonlinearizing link function needs to be developed. The Bayesian approach can also be adopted due to its strength to avoid the problem of over fitting that occurs when the number of observations is limited and the number of variables is large. Moreover, two data mining techniques (i.e. gradient boosting and MARS) are suggested to use 1) to achieve better performance of CMFunctions with consideration of variable importance, and 2) to reflect both nonlinear trend of predictors and interaction impacts between variables at the same time.

Third, the nonlinearity of variables in the cross-sectional method is not discussed in the HSM. Generally, the cross-sectional method is also known as safety performance functions (SPFs) and generalized linear model (GLM) is applied to estimate SPFs. 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. In this dissertation, applications of using generalized nonlinear model (GNM) and MARS in the cross-sectional method are proposed. In GNM, the nonlinear effects of independent variables to crash analysis can be captured by the development of nonlinearizing link function. Moreover, the MARS accommodate nonlinearity of independent variables and interaction effects for complex data structures.
Chung-Ching Wang, Statistics

Approved for distribution by Mohamed Abdel-Aty, Committee Chair, on June 4, 2015.

The public is welcome to attend.