Exploration and development of crash modification factors and functions for single and multiple treatments

Juneyoung Park
University of Central Florida

EXPLORATION AND DEVELOPMENT OF CRASH MODIFICATION FACTORS AND FUNCTIONS FOR SINGLE AND MULTIPLE TREATMENTS

by

JUNEYOUNG PARK

B.S. Hanyang University, Korea, 2009
M.S. Hanyang University, Korea, 2011

A dissertation submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in the Department of Civil, Environmental and Construction Engineering
in the College of Engineering and Computer Science
at the University of Central Florida
Orlando, Florida

Summer Term
2015

Major Professor: Mohamed Abdel-Aty

ABSTRACT

Traffic safety is a major concern for the public, and it is an important component of the roadway management strategy. In order to improve highway safety, extensive efforts have been made by researchers, transportation engineers, Federal, State, and local government officials. With these consistent efforts, both fatality and injury rates from road traffic crashes in the United States have been steadily declining over the last six years (2006~2011). However, according to the National Highway Traffic Safety Administration (NHTSA, 2013), 33,561 people died in motor vehicle traffic crashes in the United States in 2012, compared to 32,479 in 2011, and it is the first increase in fatalities since 2005. Moreover, in 2012, an estimated 2.36 million people were injured in motor vehicle traffic crashes, compared to 2.22 million in 2011.

Due to the demand of highway safety improvements through systematic analysis of specific roadway cross-section elements and treatments, the Highway Safety Manual (HSM) (AASHTO, 2010) was developed by the Transportation Research Board (TRB) to introduce a science-based technical approach for safety analysis. One of the main parts in the HSM, Part D, contains crash modification factors (CMFs) for various treatments on roadway segments and at intersections. A CMF is a factor that can estimate potential changes in crash frequency as a result of implementing a specific treatment (or countermeasure). 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 the most common methods for evaluating safety effectiveness and calculating CMFs of specific roadway treatments. Moreover, cross-sectional method has commonly been used to derive CMFs since it is easier to collect the data compared to before-after methods.

Although various CMFs have been calculated and introduced in the HSM, still there are critical limitations that are required to be investigated. First, the HSM provides various CMFs for single treatments, but not CMFs 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 the 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 adjustment factor and function (i.e. weighting 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. Recently, few studies developed crash modification functions (CMFunctions) to overcome this limitation. However, although previous studies assessed the effect of a specific single variable such as AADT on the CMFs, 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 GNMs, 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.

In this dissertation, the CMFs and CMFunctions are estimated for various single and combination of treatments for different roadway types (e.g. rural two-lane, rural multi-lane roadways, urban arterials, freeways, etc.) as below:

- Treatments for mainline of roadway:
    - adding a thru lane, conversion of 4-lane undivided roadways to 3-lane with two-way left turn lane (TWLTL)

- Treatments for roadway shoulder:
  - installing shoulder rumble strips, widening shoulder width, adding bike lanes, changing bike lane width, installing roadside barriers
- Treatments related to roadside features:
  - decrease density of driveways, decrease density of roadside poles, increase distance to roadside poles, increase distance to trees

Expected contributions of this study are to 1) suggest approaches to estimate more reliable safety effects of multiple treatments, 2) propose methodologies to develop CMFunctions to assess the variation of CMFs with different characteristics among treated sites, and 3) recommend applications of using GNM and MARS to simultaneously consider the interaction impact of more than one variables and nonlinearity of predictors.

Finally, potential relevant applications beyond the scope of this research but worth investigation in the future are discussed in this dissertation.

ACKNOWLEDGMENT

The author would like to thank his advisor, Dr. Mohamed Abdel-Aty, for his invaluable guidance, advice and support and encouragement toward successful completion of his doctoral course. The author wishes to acknowledge the support of his committee members, Dr. Essam Radwan, Dr. Naveen Eluru, Dr. Chung-Ching Wang, and Dr. Jaeyoung Lee.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS/ABBREVIATIONS

AADT	Annual Average Daily Traffic
AASHTO	American Association of State Highway & Transportation Officials
ADT	Average Daily Traffic
AIC	Akaike Information Criterion
AMF	Accident Modification Factor
BF	Basis Function
BG	Block Group
BIC	Bayesian Information Criterion
CARS	Crash Analysis Reporting System
CDC	Centers for Disease Control and Prevention
CG	Comparison Group
CPM	Crash Prediction Model
CMF	Crash Modification Factor
CMFunction	Crash Modification Function
CRF	Crash Reduction Factor
CS	Cross-sectional
CT	Census Tract
DIC	Deviance Information Criteria
DOT	Department of Transportation
EB	Empirical Bayes
EACF	Expected Average Crash Frequency
EEACF	Excess Expected Average Crash Frequency
FARS	Fatality Analysis Reporting System

FB	Full Bayes
FDOT	Florida Department of Transportation
FHWA	Federal Highway Administration
FI	Fatal and Injury
GAM	Generalized Additive Model
GCV	Generalized Cross-validation
GIS	Geographic Information System
GLM	Generalized Linear Model
GNM	Generalized Nonlinear Model
HCM	Highway Capacity Manual
HSM	Highway Safety Manual
IS	Influential Segment
LOS	Level of Service
MARS	Multivariate Adaptive Regression Splines
MCMC	Markov Chain Monte Carlo
MVM	Million Vehicle Miles
NB	Negative Binomial
NCHRP	National Cooperative Highway Research Program
NHTSA	National Highway Traffic Safety Administration
OP	Observed Prediction
PDO	Property Damage Only
RCI	Roadway Inventory Characteristics
ROR	Run-off Roadway
RTM	Regression-to-the-mean
SE	Standard Error

SPF	Safety Performance Function
SVROR	Single Vehicle Run-off Roadway
TRB	Transportation Research Board
TWLTL	Two-way Left-turn Lane
VMT	Vehicle-Miles-Traveled

CHAPTER 1: INTRODUCTION

1.1 Overview

Traffic safety is a major concern for the public, and it is an important component of roadway management strategy. In order to improve highway safety, extensive efforts have been made by researchers, transportation engineers, Federal, State, and local government officials. With these consistent efforts, both fatality and injury rates from road traffic crashes in the United States have been steadily declining over the last six years (2006-2011). However, according to the National Highway Traffic Safety Administration (NHTSA, 2013), 33,561 people died in motor vehicle traffic crashes in the United States in 2012, compared to 32,479 in 2011, and it is the first increase in fatalities since 2005. Moreover, in 2012, an estimated 2.36 million people were injured in motor vehicle traffic crashes, compared to 2.22 million in 2011.

Due to the demand of highway safety improvements through systematic analysis of specific roadway cross-section elements and treatments, the Highway Safety Manual (HSM) (AASHTO, 2010) was developed by the Transportation Research Board (TRB) to introduce a science-based technical approach for safety analysis. The HSM presents analytical methods to determine and quantify the safety effectiveness of treatments or improvements on roadways. In particular, part D of the HSM presents a variety of crash modification factors (CMFs) for safety treatments on roadway segments and at intersections. A CMF is a multiplicative factor that can estimate the expected changes in crash frequencies as a result of improvements with specific treatments. The CMFs have been estimated using observational before-after studies that account for the regression-to-the-mean bias. Moreover, cross-sectional method has been commonly used to derive CMFs since it is easier to collect the data compared to before-after methods. The cross-sectional method is also known as safety performance functions (SPFs) or crash prediction models (CPMs). Part C in the HSM provides various SPFs and detailed procedures for their application. Although various CMFs have been calculated and introduced in the HSM, still there are critical limitations that are required to be investigated.

The HSM provides various CMFs for single treatments, but not CMFs 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 the CMFs may over- or under-estimate combined effects of multiple treatments.

Moreover, 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. To overcome this limitation, crash modification functions (CMFunctions) have been utilized to determine the relationship between the safety effects and roadway characteristics. However, although previous studies assessed the effect of a specific single variable such as AADT on the CMFs, there is a lack of prior studies on the variation in the safety effects of treated sites with different multiple roadway characteristics over time.

Lastly, 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 order to account for the nonlinear effects of predictors, generalized nonlinear models (GNM) can be utilized.

In this dissertation, 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).

1.2 Research Objectives

The dissertation focuses on exploration and development of CMFs and CMFunctions for multiple treatments. The main objectives are to 1) assess safety effects of multiple treatments through exploration of the limitations of the current combining methods for multiple CMFs, 2) develop CMFunctions to determine the variation of safety effects of specific single or multiple treatments with different roadway characteristics among treated sites over time, and 3) suggest methodologies to consider the interaction impact of more than one variables and nonlinearity of predictors simultaneously in developing CMFunctions. The detailed objectives will be realized by the following tasks;

Task 1. Exploration and comparison of combined safety effects of multiple treatments. Observational before-after and cross-sectional methods will be applied to estimate CMFs for single and combined treatments. Suggest approaches to estimate more reliable safety effects of multiple treatments.

Task 2. Identify the variation of safety effects of specific treatments through evaluation of CMFs with different roadway characteristics and crash conditions. Determine nonlinear effects of parameters in cross-sectional method to estimate reliable CMFs.

Task 3. Developing simple and full CMFunctions to assess the relationship between CMFs and different roadway characteristics among treated sites over time. Traditional statistical analysis and Bayesian inference techniques will be applied. Moreover, data mining techniques will be adopted to achieve better performance.

Task 4. Suggest alternative implementation strategies to assess combined safety effects of multiple treatments using data mining techniques to overcome the over-estimation problem in developing CMFunctions for combination of multiple roadside treatments.

The first task is analyzing combined safety effects of multiple treatments and it was achieved by the following sub-tasks:

a) Investigating various methods of combining multiple CMFs to estimate the combined safety effects of multiple treatments.
b) Exploring the safety effects of single treatments and the combined treatment using the cross-sectional and observational before-after methods. To conduct the observational before-after with empirical Bayes (EB) method, Florida-specific full SPFs will be developed for different crash types and severity levels. The CMFs will be estimated for various treatments as below:

   - Install shoulder rumble strips

   - Widening shoulder width

   - Install shoulder rumble strips + widening shoulder width

   - Adding a bike lane

   - Lane reduction (Conversion of 4-lane undivided roadways to 3-lane with TWLTL (two-way left-turn lane))

- Road diet (Adding a bike lane + Lane reduction)

c) Calculate the combined CMF by existing combining methods using actual estimated CMFs for two single treatments and compare it with actual estimated CMF for combined treatment.
d) Identifying over- and under-estimation of various existing combining methods for multiple CMFs. Determine the combined effects of multiple treatments based on the location of roadway improvements such as median of roadway and roadside.
e) Determine the difference between (1) multiple treatments on same location, and (2) multiple treatments on different location. Suggest alternative way to improve accuracy of combining multiple CMFs. The task has been achieved in Chapter 3 and Chapter 4.

For the second task, several sub-tasks were carried out as follow:

f) Estimate CMFs for installing roadside barriers for different crash types and severities with different vehicle, driver, weather, time of day conditions using various observational before-after methods. The work is presented in Chapter 5.
g) Evaluate GNMs to assess the safety effects of changing bike lane width with consideration of nonlinear effects (Chapter 6).

The following sub-tasks were conducted for the third task:

h) Develop simple and full CMFunctions for installing bike lanes for different crash types and severities with different roadway and socio-economic characteristics using multiple linear and nonlinear regression models. The task has been achieved and the work is presented in Chapter 7.

i) Develop full CMFunctions for adding a thru lane treatment using Bayesian approach with nonlinearizing link functions to account for the temporal effects on the variation of the safety effects (Chapter 8).

j) Application of data mining technique to develop full CMFunctions for widening shoulder width treatment (Chapter 9).

The final task was achieved by following sub-tasks:

k) Utilize parametric and non-parametric modeling approaches to estimate combined safety effects. The GLM, GNM, and multivariate adaptive regression splines (MARS) models were developed to estimate CMFs in cross-sectional method (Chapter 10). The CMFs were estimated for various roadside treatments as below:

   - Decrease density of driveways
   - Decrease density of roadside poles
   - Increase distance to roadside poles
   - Increase distance to roadside trees
   - Combination of multiple roadside treatments

1.3 Dissertation Organization

The dissertation is organized as follows: Chapter 2, following this chapt er, summarizes the literature on previous CMF and CMFunction related studies. Current CMF development methods (various observational before-after studies and cross-sectional method) are presented. Existing combining methods of multiple CMFs were discussed with their model forms. Moreover, current issues of CMF and CMFunction related researches and their limitations are discussed. Additionally, it will also be explained how to address limitations in these studies. Chapter 3 provides the exploration and comparison of existing combining methods using actual estimated CMFs for single treatments and combination of it. Chapter 4 suggests alternative ways to improve accuracy of combined safety effects using developed adjustment factors and functions. Chapter 5 presents estimated CMFs for different crash types and severities with different vehicle, driver, weather, time of day conditions, and Chapter 6 provides an application of nonlinearizing link function in cross-sectional method to calculate CMFs to reflect the nonlinearity of predictors. Chapter 7 to 9 give a comprehensive analysis about the development simple and full CMFunctions to assess the variation of CMFs with different roadway and socio-economic characteristics among treated sites over time using different modeling techniques. Chapter 7 presents estimation of simple and full CMFunctions process based on assessment of safety effects of adding a bike lane for different crash types and severity levels. Moreover, the effects of including socio-economic parameters in estimating CMFs and developing CMFunctions are presented. Chapter 8 explores the relationship between CMFs and roadway characteristics in developing full CMFunctions for adding a thru lane treatment using Bayesian approach with nonlinearizing link functions to account for the temporal effects. Chapter 9 presents an application of data mining technique in developing full CMFunctions for widening shoulder width treatment to account for the nonlinearity of predictors and interaction impacts between variables at the same time. Chapter 10 offers alternative implementation strategies to assess combined safety effects of multiple treatments using data mining technique to overcome the over-estimation problem in developing CMFunctions for combination of multiple roadside treatments. Finally, Chapter 11 summarizes the dissertation and presents potential improvement for future applications of estimation of CMFs and CMFunctions for multiple treatments.

CHAPTER 2: LITERATURE REVIEW

2.1 Highway Safety Manual and Crash Modification Factors

The HSM published in 2010 perfectly bridge the gap between traffic safety researches and safety improvement applications for the highways. One of the key parts in this manual is the SPF and the CMFs, which can help local agencies and DOTs to discover the hot spots (locations with high crash occurrences) and suggest countermeasures for sites of concern. However, the basic method stated in the HSM was calibrated only based on several states and it need further calibration before applied to a specific area, the calibration factor should be calculated to develop jurisdiction specific models. Researchers are keen to work on the application of HSM in different states. States like Utah (Brimley et al., 2012), Kansas (Howard and Steven, 2012), Oregon (Zhou and Dixon, 2012), Florida (Gan et al., 2012), etc., have already worked on calibrations and modifications of the safety performance functions in the HSM on their own roadways.

Part D of the HSM provides a methodology to evaluate the effects of safety treatments (countermeasures). These can be quantified by CMFs that are expressed as numerical values to identify the percent increase or decrease in crash frequency together with the standard error. A standard error of 0.10 or less indicates that a CMF is sufficiently accurate. CMFs could also be expressed as a function or SPF (equation), graph or combination. CMFs are also known as Collision Modification Factors or Accident Modification Factors (CMFs or AMFs), all of which have exactly the same function. HSM Part D provides CMFs for roadway segments (e.g., roadside elements, alignment, signs, rumble strips, etc.), intersections (e.g., control), interchanges, special facilities (e.g., Hwy-rail crossings), and road networks. CMFs could be applied individually if a single treatment is proposed or multiplicative if multiple treatments are implemented. The proper calibration and validation of CMFs will provide an important tool to practitioners to adopt the most suitable cost effective countermeasure to reduce crashes at hazardous locations. It is expected that the implementation of CMFs will gain more attention after the recent release of the HSM and the 2009 launch of the Clearinghouse website http://www.cmfclearinghouse.org (University of North Carolina Highway Safety Research Center, 2010).

2.1.1 Latest studies related to the HSM and CMFs

Alkhatni et al (2014) examined the effects of presence of weigh stations on injury severity and frequency of crashes on Michigan freeways. The study investigated crash patterns in the vicinity of 12 fixed weigh stations as compared to crash patterns in the vicinity of 65 rest areas and 77 selected comparison segments. Three major influential segments (ISs) were identified: before facility, at facility, and after facility. Comparisons segments with similar traffic and geometric characteristics as the ISs were also identified. The result indicates that presence of fixed weigh station is shown to have positive impact. This indicates that crashes occurring near fixed weigh stations tend to be more severe than those occurring at rest areas and comparison segments.

Chen et al (2014) investigated the safety performance of short left-turn lanes at unsignalized median openings. Six years of crash data were collected from fifty-two median left turn lanes in Houston, Texas, which included forty short lanes and twelve lanes. A Poisson regression model was developed to relate traffic and geometric attributes to the total count of rear-end, sideswipe, and object-motor vehicle crashes at a left-turn lane. CMFs were calculated for future applications in projecting the crash frequency, given a specific change of the lane length. It was statistically evidenced that the difference between actual lane length and the Greenbook recommended length had significant effects on the crash frequency. The CMF is found to be 2.32 if a left-turn lane is 20 percent shorter than what is suggested in the Greenbook.

Dell'Acqua et al (2014) identified the modeling results between HSM and the situation in Italy. This is paper implement the model to assess crash behavior in Italy. To adjust the base predicted crash frequency to meet the current conditions, the accident modification factors (AMFs) calculation for lane width, horizontal curve and vertical grade were identified. Crash types (head-on/side collisions, single-vehicle crashes, rear-end collisions) were investigated based on the vertical grade and the curvature indicator. The result of this paper shows calibration factor is 0.477 when applying to Italy.

Khan et al (2014) assessed the safety effectiveness of shoulder rumble strips in reducing run-off-the-road (ROR) crashes on two-lane rural highways using the observational before and after with EB method. The comprehensive procedure adopted for developing the safety performance function of EB analysis also considers the effects of roadway geometry and paved right shoulder width on the effectiveness of shoulder rumble strips. The results of this study demonstrate the safety benefits of shoulder rumble strips in reducing the ROR crashes on two-lane rural highways using the State of Idaho 2001-2009 crash data. The study finds a 14% reduction in all ROR crashes after the installation of shoulder rumble strips on 178.63-miles of two-lane rural highways in Idaho. The results indicate that shoulder rumble strips were most effective on roads with relatively moderate curvature and right paved shoulder width of 3 feet and more.

Li et al (2014) tried to ensure a high level of road safety based on the best knowledge available of the effects of the road network planning. The authors looked into how changes in road network characteristics affect road casualties. To estimate the safety effectiveness of roadway networking, the Full Bayes (FB) method was conducted. Also the authors applied a panel semi-parametric model to estimate the dose-response function for continuous treatment variables. The result suggests that there are more casualties in the area with a better connectivity and accessibility, where more attention should be paid to the safety countermeasures.

Mohammadi et al (2014) evaluated the changes in motor vehicle crashes that occurred on the Missouri interstate highway system. In this paper, the author applied Empirical Bayesian methods to estimate safety effect as a result of countermeasures. The research associated crashes with traffic and roadway characteristics. Negative binomial (NB) models were developed for the before-after-change conditions. The models developed for the various collision types and crash severities were used to estimate the expected number of crashes at roadway segments in 2008, assuming with and without the implementation. This procedure estimated significant reductions of 10% in the overall number of crashes and a 30% reduction for fatal crashes. Reductions in the number of different collision types were estimated to15 be 18-37%. The results indicate that the policy reduces the number of crashes and decreasing fatalities by reducing the most severe collision types like head-on crashes.

Zeng et al (2014) evaluated evaluate the safety effectiveness of good pavement conditions versus deficient pavement conditions on rural two-lane undivided highways in Virginia. Using the EB method, it was found that good pavements are able to reduce fatal and injury (FI) crashes by 26 percent over deficient pavements, but do not have a statistically significant impact on overall crash frequency. The authors concluded that improving pavement from deficient to good condition can offer a significant safety improvement in terms of reducing crash severity.

Sacchi et al. (2012) studied the transferability of the HSM crash prediction algorithms on two-lane rural roads in Italy. The authors firstly estimated a local baseline model as well as evaluated each CMF based on the Italian data. Homogenous segmentation for the chosen study roads has been performed just to be consistent with the HSM algorithms. In order to quantify the transferability, a calibration factor has been evaluated to represent the difference between the observed number of crashes and the predicted number of crashes by applying HSM algorithm. With a four years crash data, the calibration factor came out to be 0.44 which indicate the HSM model has over predicted the collisions. After investigated the predicted values with observed values by different annual average daily traffic (AADT) levels, the authors concluded that the predicted ability of the HSM model for higher AADT is bad and a constant value of "calibration factor" is not appropriate. This effect was also proved from the comparison between the HSM baseline model and the local calculated baseline model. Furthermore, the authors evaluated CMFs for three main road features (horizontal curve, driveway density and roadside design). The calculation of CMFs has been grouped according to Original CMFs, and results of comparing the calculated CMFs to baseline CMFs indicated that the CMFs are not unsuitable for local Italian roadway characteristics since most of them are not consistent. Finally, several well-known goodness-of-fit measures have been used to assess the recalibrated HSM algorithms as a whole, and the results are consistent as the results mentioned in the split investigation of HSM base model and CMFs. With these facts the authors concluded that the HSM is not suitable to transferable to Italy roads and Europe should orient towards developing local SPFs/CMFs.

Sun et al. (2012) calibrated the SPF for rural multilane highways in the Louisiana State roadway system. The authors investigated how to apply the HSM network screening methods and identified the potential application issues. Firstly the rural multilane highways were divided into sections based on geometric design features and traffic volumes, all the features are distinct within each segment. Then by computing the calibration factor, the authors found out that the average calibration parameter is 0.98 for undivided and 1.25 for divided rural multilane highways. These results turned out that HSM has underestimated the expected crash numbers. Besides the calibration factor evaluation, the authors investigated the network screening methods provided by HSM. 13 methods are promoted in the HSM, each of these methods required different data and data availability issue is the key part of HSM network screening methods application. In the paper, four methods have been adopted: crash frequency, crash rates, excess expected average crash frequency using SPFs (EEACF) and expected average crash frequency with EB Adjustment (EACF). Comparisons between these methods have been done by ranking the most hazardous segments and findings indicate that the easily used crash frequency method produced similar results to the results of the sophisticated models; however, crash rate method could not provide the same thing.

Xie et al. (2011) investigated the calibration of the HSM prediction models for Oregon State Highways. The authors followed the suggested procedures by HSM to calibrate the total crashes in Oregon. In order to calculate the HSM predictive model, the author identified the needed data and came up with difficulties in collecting the pedestrian volumes, the minor road AADT values and the under-represented crash locations. For the pedestrian volume issue, the authors assumed to have “medium” pedestrian when calculate the urban signalized intersections. While for the minor road AADT issue, the authors developed estimation models for the specific roadway types. Then the calibration factors have been defined for the variety types of highways and most of these values are below than 1. These findings indicate an overestimation for the crash numbers by the HSM. However, the authors attribute these results to the current Oregon crash reporting procedures which take a relative high threshold for the Property Damage Only (PDO) crashes. Then for the purpose of proving the crash reporting issue, the authors compared the HSM proportions of different crash severity levels and the Oregon oriented values. Furthermore, calibration factors for fatal and injury crashes have been proved to be higher than the total crash ones, which also demonstrated that Oregon crash reporting system introduce a bias towards the fatal and injury conditions. So the authors concluded that the usages of severity-based calibration factors are more suitable for the Oregon State highways.

Howard and Steven (2012) investigated different aspects of calibrate the predictive method for rural two-lane highways in Kansas State. Two data sets were collected in this study; one data set was used to develop the different model calibration methods and the other one was adopted for evaluating the models accuracy for predicting crashes. At first, the authors developed the baseline HSM crash predictive models and calculated the Observed-Prediction (OP) ratios. Results showed a large range of OP ratios which indicate the baseline method is not very promising in predicting crash numbers. Later on, the author tried alternative ways to improve the model accuracy. Since crashes on Kansas rural highways have a high proportion of animal collision crashes which is nearly five times the default percentage presented in the HSM. The authors tried to come up with a (1) Statewide Calibration factor, (2) Calibration factors by crash types, (3) Calibration using animal crash frequency by county and (4) Calibration utilizing animal crash frequency by section. The observational before-after with EB method was introduced to see whether it would improve the accuracy and also a variety of statistical measures were performed to evaluate the performance. Finally, the authors concluded that the applications of EB method showed consistent improvements in the model prediction accuracy. Moreover, it was suggested that a single statewide calibration of total crashes would be useful for the aggregate analyses while for the project-level analysis, the calibration using animal crash frequency by county is very promising.

Banihashemi (2011) performed a heuristic procedure to develop SPFs and CMFs for rural two-lane highway segments of Washington State and compared the developed models to the HSM model. The author utilized more than 5000 miles of rural two-lane highway data in Washington State and crash data for 2002-2004. Firstly the author proposed an innovative way to develop SPFs and CMFs, incorporating the segment length and AADT. Then CMFs for lane width, shoulder width, curve radius and grade have been developed. After all these procedures, the author came up with two self-developed SPFs and then compared them with the HSM model. The comparison was done at three aggregation levels: (1) consider each data as single observation (no aggregation), (2) segments level with a minimum 10 miles length and (3) aggregated based on geometric and traffic characteristics of highway segments. A variety of statistical measures were introduced to evaluate the performances and the author concluded that mostly the results are comparable, and there is no need to calibrate new models. Finally a sensitivity analysis was conducted to see the influence of data size issue on the calibration factor for the HSM model, and the conclusions indicated that a dataset with at least 150 crashes per year are most preferred for Washington State.

Later on, Banihashemi (2012) conducted a sensitivity analysis for the data size issue for calculating the calibration factors. Mainly five types of highway segment and intersection crash prediction models were investigated; Rural two-lane undivided segments, rural two-lane intersections, rural multilane segments, rural multilane intersections and urban/suburban arterials. Specifically, eight highway segment types were studied. Calibration factors were calculated with different subsets with variety percentages of the entire dataset. Furthermore, the probability that the calibrated factors fall within 5% and 10% range of the ideal calibration factor values were counted. Based on these probabilities, recommendations for the data size issue to calibrate reliable calibration factors for the eight types of highways have been proposed. With the help of these recommendations, the HSM predictive methods can be effectively applied to the local roadway system.

Brimley et al. (2012) evaluated the calibration factor for the HSM SPF for rural two-lane two-way roads in Utah. Firstly, the authors used the SPF model stated in the HSM and found out the calibration factor to be 1.16 which indicate a under estimate of crash frequency by the base model. Later on, under the guidance of the HSM, the authors developed jurisdiction-specific negative binomial (NB) models for the Utah State. More variables like driveway density, passing condition, speed limit and etc. were entered into the models with the p-values threshold of 0.25. Bayesian information criterion (BIC) was selected to evaluate the models and the finally chosen best promising model show that the relationships between crashes and roadway characteristics in Utah may be different from those presented in the HSM.

Zegeer et al. (2012) worked on the validation and application issues of the HSM to analysis of horizontal curves. Three different data sets were employed in this study: all segments, random selection segments and non-random selection segments. Besides, based on the three data sets, calibration factors for curve, tangent and the composite were calculated. Results showed that the curve segments have a relative higher standard deviation than the tangent and composite segments. However, since the development of a calibration factor requires a large amount of data collecting work, a sensitivity analysis of each parameter's influence for the output results for curve segments have been performed. HSM predicted collisions were compared as using the minimum value and the maximum value for each parameter. The most effective variables were AADT, curve radius and length of the curve. Other variables like grade, driveway density won't affect the result much if the mean value were utilized when developing the models. Finally, validation of the calibration factor was performed with an extra data set. Results indicated that the calibrated HSM prediction have no statistical significant difference with the reported collisions.

2.1.2 HSM related research in Florida

State of Florida is among other states that initiated a plan to implement and validate the HSM to its roadways. Figure 1 shows the Florida Department of Transportation (FDOT) timeline of the HSM implementation.

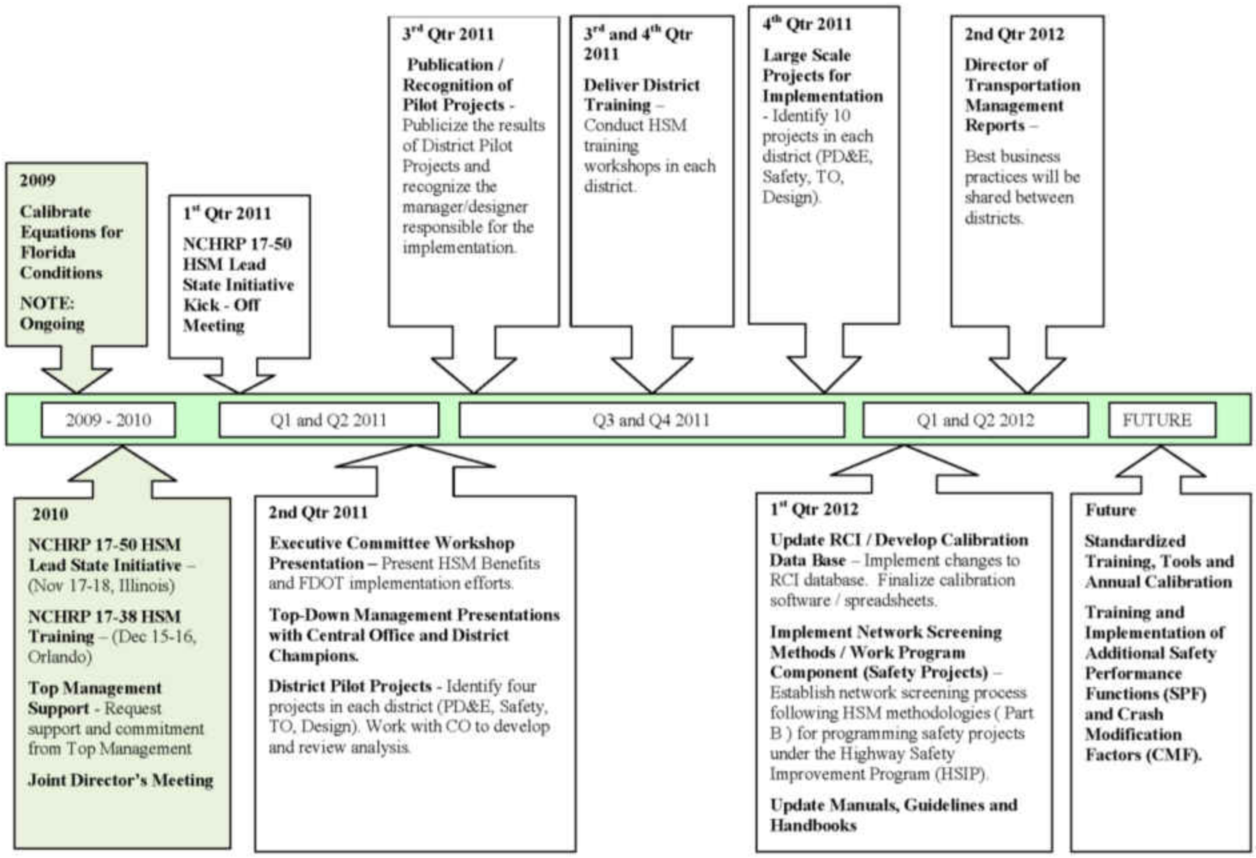


Figure 2-1: FDOT Implementation plan timeline for the HSM (Source: www.dot.state.fl.us)

The HSM is considered a turning point in the approach of analyzing safety data for practitioners and administrators throughout statistically proven quantitative analyses. States and local agencies are still examining ways to implement the HSM. The data requirement for the HSM and *SafetyAnalyst* is the most challenging task that all agencies are still struggling with. Florida has been at the forefront of many states in implementing the HSM and deploying the *SafetyAnalyst*. A research project was sponsored by FDOT and conducted by the University of Florida to develop and calibrate of the HSM equations for Florida conditions. The study provided calibration factors at the segment- and intersection- level safety performance functions from the HSM for Florida conditions or the years 2005 through 2008 (Srinivasan et al., 2011).

Specifically, FDOT has sponsored two projects in its effort to implement *SafetyAnalyst*. The first of these projects was conducted by the University of South Florida (USF) which developed a program to map and convert FDOT's roadway and crash data into the input data format required by *SafetyAnalyst* (Lu et al., 2009).

A second related project was completed recently by Florida International University (FIU). The project successfully developed Florida-based SPFs for different types of segments, ramps, and signalized intersections. These SPFs were then applied to generate high crash locations in *SafetyAnalyst*. Additionally, the project also developed the first known GIS tool for *SafetyAnalyst*. However, the project was unable to develop SPFs, nor generate any *SafetyAnalyst* input files for unsignalized intersections due to the lack of the required data in FDOT's Roadway Inventory Characteristics (RCI). In addition, the SPFs and *SafetyAnalys*t input data files for signalized intersections could only be developed based on very limited data (Gan et al., 2012).

2.2 Crash Modification Factors Development Methods

There are different methods to estimate CMFs, these methods vary from a simple before and after study and before and after study with comparison group to a relatively more complicated methods such EB and FB methods. Also, the cross-sectional method has been commonly used to derive CMFs since it is easier to collect the data compared to before-after methods.

1) The simple (naïve) before and after study

This method compares number of crashes before the treatment and after treatment. The main assumption of this method is that the number of crashes before the treatment would be expected without the treatment. This method tends to overestimate the effect of the treatment because of the regression to the mean (RTM) problem (Hauer, 1997).

2) The before and after study with comparison group

This method is similar to the simple before and after study, however, it uses a comparison group of untreated sites to compensate for the external causal factors that could affect the change in the number of crashes. This method also does not account for the regression to the mean as it does not account for the naturally expected reduction in crashes in the after period for sites with high crash rates.

3) The empirical Bayes before and after study

The EB method can account for the regression to the mean issue by introducing an estimated for the mean crash frequency of similar untreated sites using SPFs. Since the SPFs use AADT and sometimes other characteristics of the site, these SPFs also account for traffic volume changes which provides a true safety effect of the treatment (Hauer, 1997)

4) The full Bayes before and after study

The FB is similar to the EB of using a reference population; however, it uses an expected crash frequency and its variance instead of using point estimate, hence, a distribution of likely values is generated. It is known that the FB method is useful approach since it provides more detailed causal inferences and more flexibility in selecting crash count distributions to account for uncertainty in data used.

5) The cross-sectional method

The cross-sectional studies are useful to estimate CMFs where there are insufficient before and after data for a specific treatment that is actually applied. According to NCHRP project 20-7 (Carter et al. 2012), the CMF can be derived by taking the ratio of the average crash frequency of sites with the feature to the average crash frequency of sites without the feature. This method is also known as safety performance functions or crash prediction models which relate crash frequency with roadway characteristics, length and traffic volume of segments. The CMF can be calculated from the coefficient of the variable associated with treatments – e.g. the exponent of the coefficient when the form of the model is log-linear.

2.2.1 The Simple (Naïve) Before-After Study

The naïve before-after approach is the simplest approach. Crash counts in the before period are used to predict the expected crash rate and, consequently, expected crashes had the treatment not been implemented. This basic Naïve approach assumes that there was no change from the 'before' to the 'after' period that affected the safety of the entity under scrutiny; hence, this approach is unable to account for the passage of time and its effect on other factors such as exposure, maturation, trend and regression-to-the-mean bias. Despite the many drawbacks of the basic Naïve before-after study, it is still quite frequently used in the professional literature because; 1) it is considered as a natural starting point for evaluation, and 2) its easiness of collecting the required data, and 3) its simplicity of calculation. The basic formula for deriving the safety effect of a treatment based on this method is:

$$CMF = \frac{N_a}{N_b} \tag{2-1}$$

where N_a and N_b are the number of crashes at a treated site in the after and before the treatment, respectively. It should be noted that with a simple calculation, the exposure can be taken into account in the Naïve before-after study. The crash rates for both before and after the implementation of a project should be used to estimate the CMFs which can be calculated as:

$$\text{Crash Rate} = \frac{\text{Total Number of Crashes}}{\text{Exposure}} \tag{2-2}$$

where the 'Exposure' is usually calculated in million vehicle miles (MVM) of travel, as indicated in Equation (2-3):

$$\text{Exposure} = \frac{\text{Project Section Length in Miles} \times \text{Mean ADT} \times \text{Number of Years} \times 365 \text{ Days}}{1{,}000{,}000} \tag{2-3}$$

Each crash record would typically include the corresponding average daily traffic (ADT). For each site, the mean ADT can be computed by Equation (2-4):

$$\text{Mean ADT} = \frac{\text{Summation of Individual ADTs Associated with each Crash}}{\text{Total Number of Crashes}} \tag{2-4}$$

2.2.2 The Before-After with Comparison Group Method

To account for the influence of a variety of external causal factors that change with time, the Before-After with comparison group study can be adopted. A comparison group is a group of control sites that remained untreated, and that are similar to the treated sites in trend of crash history, traffic, geometric and geographic characteristics. The crash data at the comparison group are used to estimate the crashes that would have occurred at the treated entities in the ‘after’ period had treatment not been applied. This method can provide more accurate estimates of the safety effect than a naïve before-after study, particularly, if the similarity between treated and comparison sites is high. The before-after with comparison group method is based on two main assumptions (Hauer, 1997): 1) The factors that affect safety have changed in the same manner from the ‘before’ period to ‘after’ period in both treatment and comparison groups, and 2) These changes in the various factors affect the safety of treatment and comparison groups in the same way. Based on these assumptions, it can be assumed that the change in the number of crashes from the ‘before’ period to ‘after’ period at the treated sites, in case of no countermeasures had been implemented, would have been in the same proportion as that for the comparison group. Accordingly, the expected number of crashes for the treated sites that would have occurred in the ‘after’ period had no improvement applied ($N_{expected,T,A}$) follows (Hauer, 1997):

$$N_{expected,T,A} = N_{observed,T,B} \times \frac{N_{observed,C,A}}{N_{observed,C,B}} \tag{2-5}$$

If the similarity between the comparison and the treated sites in the yearly crash trends is ideal, the variance of $N_{expected,T,A}$ can be estimated from Equation (2-6):

$$Var(N_{expected,T,A}) = N_{expected,T,B}^{\ 2}(1/N_{observed,T,B} + 1/N_{observed,C,B} + 1/N_{observed,C,A}) \tag{2-6}$$

It should be noted that a more precise estimate can be obtained in case of using non-ideal comparison group as explained in Hauer (1997), Equation (2-7):

$$\mathrm{Var}(\mathrm{N}_{\mathrm{expected,T,A}}) = \mathrm{N}_{\mathrm{expected,T,B}}{}^{2}(1/\mathrm{N}_{\mathrm{observed,T,B}} + 1/\mathrm{N}_{\mathrm{observed,C,B}} + 1/\mathrm{N}_{\mathrm{observed,C,A}} + \mathrm{Var}(\omega)) \quad 2\text{-}7$$

$$\omega = \frac{r_c}{r_t} \tag{2-8}$$

where $$r_c \cong \frac{N_{\mathrm{exp}ected,c,A}}{N_{\mathrm{exp}ected,c,B}} \tag{2-9}$$

and $$r_t \cong \frac{N_{\mathrm{exp}ected,t,A}}{N_{\mathrm{exp}ected,t,B}} \tag{2-10}$$

And the CMF and its variance can be estimated from Equations (2-11) and (2-12).

$$\mathrm{CMF} = (\mathrm{N}_{\mathrm{observed,T,A}}/\mathrm{N}_{\mathrm{expected,T,A}})/(1 + (\mathrm{Var}(\mathrm{N}_{\mathrm{expected,T,A}})/\mathrm{N}_{\mathrm{expected,T,A}}{}^{2})) \tag{2-11}$$

$$\mathrm{Var}(\mathrm{CMF}) = \frac{\mathrm{CMF}^{2}[(1/\mathrm{N}_{\mathrm{observed,T,A}}) + ((\mathrm{Var}(\mathrm{N}_{\mathrm{expected,T,A}})/\mathrm{N}_{\mathrm{expected,T,A}}{}^{2})]}{[1 + (\mathrm{Var}(\mathrm{N}_{\mathrm{expected,T,A}})/\mathrm{N}_{\mathrm{expected,T,A}}{}^{2}]^{2}} \tag{2-12}$$

Where,

$N_{observed,T,B}$ = the observed number of crashes in the before period for the treatment group.

$N_{observed,T,A}$ = the observed number of crashes in the after period for the treatment group.

$N_{observed,C,B}$ = the observed number of crashes in the before period in the comparison group.

$N_{observed,C,A}$ = the observed number of crashes in the after period in the comparison group.

ω = the ratio of the expected number of crashes in the ‘before’ and ‘after’ for the treatment and the comparison group.

r_c = the ratio of the expected crash count for the comparison group.

r_t = the ratio of the expected crash count for the treatment group.

There are two types of comparison groups with respect to the matching ratio; 1) the before-after study with yoked comparison which involves a one-to-one matching between a treatment site and a comparison site, and 2) a group of matching sites that are few times larger than treatment sites. The size of a comparison group in the second type should be at least five times larger than the treatment sites as suggested by Pendleton (1991). Selecting matching comparison group with similar yearly trend of crash frequencies in the ‘before’ period could be a daunting task. In this study a matching of at least 4:1 comparison group to treatment sites was conducted. Identical length of three years of the before and after periods for the treatment and the comparison group was selected.

2.2.3 The Before-After with Empirical Bayes Method

In the before-after with EB method, the expected crash frequencies at the treatment sites in the ‘after’ period had the countermeasures not been implemented is estimated more precisely using data from the crash history of a treated site, as well as the information of what is known about the safety of reference sites with similar traffic and physical characteristics. The method is based on three fundamental assumptions (Hauer, 1997; Hauer et al. (2002)):

1. The number of crashes at any site follows a Poisson distribution.
2. The means for a population of systems can be approximated by a Gamma distribution.

3. Changes from year to year from sundry factors are similar for all reference sites.

One of the main advantages of the before-after study with EB is that it accurately accounts for changes in crash frequencies in the 'before' and in the 'after' periods at the treatment sites that may be due to regression-to-the-mean bias. It is also a better approach than the comparison group for accounting for influences of traffic volumes and time trends on safety. The estimate of the expected crashes at treatment sites is based on a weighted average of information from treatment and reference sites as given in (Hauer, 1997):

$$\hat{E}_i = (\gamma_i \times y_i \times n) + (1 - \gamma_i)\eta_i \tag{2-13}$$

Where γ_i is a weight factor estimated from the over-dispersion parameter of the negative binomial regression relationship and the expected 'before' period crash frequency for the treatment site as shown in Equation (2-14):

$$\gamma_i = \frac{1}{1 + k \times y_i \times n} \tag{2-14}$$

y_i = Number of average expected crashes of given type per year estimated from the SPF (represents the 'evidence' from the reference sites).

η_i = Observed number of crashes at the treatment site during the 'before' period

n = Number of years in the before period,

k = Over-dispersion parameter

The ‘evidence’ from the reference sites is obtained as output from the SPF. SPF is a regression model which provides an estimate of crash occurrences on a given roadway section. Crash frequency on a roadway section may be estimated using negative binomial regression models (Abdel-Aty and Radwan, 2000; Persaud, 1990), and therefore it is the form of the SPFs for negative binomial model is used to fit the before period crash data of the reference sites with their geometric and traffic parameters. A typical SPF will be of the following form:

$$y_i = e^{(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n)} \tag{2-15}$$

Where β_i*’s* = Regression Parameters,

x_1 and x_2 here are logarithmic values of AADT and section length,

x_i *‘s* $(i > 2)$ = Other traffic and geometric parameters of interest.

Over-dispersion parameter, denoted by k is the parameter which determines how widely the crash frequencies are dispersed around the mean.

And the standard deviation (σ_i) for the estimate in Equation (2-16) is given by:

$$\hat{\sigma}_i = \sqrt{(1-\gamma_i)\times \hat{E}_i} \tag{2-16}$$

It should be noted that the estimates obtained from equation 2-10 are the estimates for number of crashes in the before period. Since, it is required to get the estimated number of crashes at the treatment site in the after period; the estimates obtained from equation (2-10) are to be adjusted for traffic volume changes and different before and after periods (Hauer, 1997; Noyce et al., 2006). The adjustment factors for which are given as below:

Adjustment for AADT (ρ_{AADT}):

$$\rho_{AADT} = \frac{AADT_{after}^{\alpha_1}}{AADT_{before}^{\alpha_1}} \tag{2-17}$$

Where, $AADT_{after}$ = AADT in the after period at the treatment site, and

$AADT_{before}$ = AADT in the before period at the treatment site.

α_1 = Regression coefficient of AADT from the SPF.

Adjustment for different before-after periods (ρ_{time}):

$$\rho_{time} = \frac{m}{n} \tag{2-18}$$

Where, m = Number of years in the after period.

n = Number of years in the before period.

Final estimated number of crashes at the treatment location in the after period ($\hat{\pi}_i$) after adjusting for traffic volume changes and different time periods is given by:

$$\hat{\pi}_i = \hat{E}_i \times \rho_{AADT} \times \rho_{time} \tag{2-19}$$

The index of effectiveness (θ_i) of the treatment is given by:

$$\hat{\theta}_i = \frac{\hat{\lambda}_i / \hat{\pi}_i}{1 + \left(\hat{\sigma}_i^{\,2} \Big/ \hat{\pi}_i^{\,2} \right)} \tag{2-20}$$

Where, $\hat{\lambda}_i$ = Observed number of crashes at the treatment site during the after period.

The percentage reduction (τ_i) in crashes of particular type at each site i is given by:

$$\hat{\tau}_i = (1 - \hat{\theta}_i) \times 100\% \tag{2-21}$$

The Crash Reduction Factor or the safety effectiveness ($\hat{\theta}$) of the treatment averaged over all sites would be given by (Persaud et al., 2004):

$$\hat{\theta} = \frac{\sum_{i=1}^{m} \hat{\lambda}_i \Big/ \sum_{i=1}^{m} \hat{\pi}_i}{1 + \left(\operatorname{var}(\sum_{i=1}^{m} \hat{\pi}_i) \Big/ (\sum_{i=1}^{m} \hat{\pi}_i)^2 \right)} \tag{2-22}$$

Where, m = total number of treated sites, and

$$\operatorname{var}(\sum_{i=1}^{k} \hat{\pi}_i) = \sum_{i=1}^{k} \rho_{AADT}^{\;2} \times \rho_{time}^{\;2} \times \operatorname{var}(\hat{E}_i) \quad \text{(Hauer, 1997)} \tag{2-23}$$

The standard deviation ($\hat{\sigma}$) of the overall effectiveness can be estimated using information on the variance of the estimated and observed crashes, which is given by Equation (2-24).

$$\hat{\sigma} = \sqrt{\frac{\theta^2 \left[\left(\operatorname{var}(\sum_{i=1}^{k} \hat{\pi}_i) \Big/ (\sum_{i=1}^{k} \hat{\pi}_i)^2 \right) + \left(\operatorname{var}(\sum_{i=1}^{k} \hat{\lambda}_i) \Big/ (\sum_{i=1}^{k} \hat{\lambda}_i)^2 \right) \right]}{\left[1 + \left(\operatorname{var}(\sum_{i=1}^{k} \hat{\pi}_i) \Big/ (\sum_{i=1}^{k} \hat{\pi}_i)^2 \right) \right]^2}} \tag{2-24}$$

Where, $\text{var}(\sum_{i=1}^{k}\hat{\lambda}_i) = \sum_{i=1}^{k}\lambda_i$ (Hauer, 1997)

Equation (2-16) is used in the analysis to estimate the expected number of crashes in the after period at the treatment sites, and then the values are compared with the observed number of crashes at the treatment sites in the after period to get the percentage reduction in number of crashes resulting from the treatment.

2.2.4 The Before-After with Full Bayes Method

It is known that the FB approach provided comparable results and might have several advantages over the EB technique as follow: 1) FB models account for the uncertainty associated with parameter estimates and provide exact measures of uncertainty on the posterior distributions of these parameters and hence overcome the maximum likelihood methods' problem of overestimating precision because of ignoring this uncertainty; 2) valid crash models can be estimated using small sample size because of the FB properties, which might be the case of most of road safety benefit analyses; 3) Bayesian inference can effectively avoid the problem of over fitting that occurs when the number of observations is limited and the number of variables is large (3). In the before-after framework, the FB method integrates the EB two-steps into one by calculating the odds ratio and the SPFs into a single step, and hence, integrating any error or variance of the estimated regression coefficient into the final estimates of the safety effectiveness of a treatment. Most importantly, the flexibility of a FB formulation allows for different model specifications which have the capability of accounting for various levels of correlation. Moreover, Persaud et al. (2009) demonstrated that the FB method is useful approach since it provides more detailed causal inferences and more flexibility in selecting crash count distributions to account for uncertainty in data used. In order to assess crash counts data, several studies utilized the Bayesian Poisson-lognormal model (Park and Lord, 2007; Ma et al., 2008; El-Basyouny and Sayed, 2009). In particular, Ma and Kockelman (2006) adopted a multivariate Poisson-lognormal model to simultaneously analyze crash counts with different injury severity levels through the Bayesian paradigm, providing a systematic approach to estimating correlated count data.

In the Bayesian Poisson-lognormal model, the crash frequency Y_{it} has a Poisson distribution conditional on the σ-field generated by the random variables of unobserved heterogeneity (random errors, ε_t) and the set of independent explanatory variables X_{it} (Munkin and Trivedi, 2002). The model can be set up as follows:

$$Y_{it} \sim \text{Poisson } (\lambda_{it} \text{ for i=1,2,...,m and t=1,2,...,n}) \tag{2-25}$$

which, is the observed crash count at segment i in year t with the underlying Poisson mean (i.e. the expected crash frequency) for segment i in year t. The Poisson rate is modeled as a function of the log-link using a log-normal distribution:

$$\log\lambda_{it} = \log e_{it} + X'_{it}\beta + \varepsilon_t \tag{2-26}$$

The random effect ε_t is unknown and therefore has its own prior distribution, p(∅). The joint prior distribution is (Gelman et al., 2004)

$$p(\emptyset,\theta) = p(\emptyset)p(\theta|\emptyset), \tag{2-27}$$

and the joint posterior distribution can be defined as

$$p(\emptyset,\theta|y) \propto p(\emptyset,\theta)p(y|,\emptyset\theta)=p(\emptyset,\theta)p(y|\theta). \quad (2\text{-}28)$$

These posterior distributions were calibrated by Mont Carlo Markov Chain (MCMC) (Gamerman, 2006; Gilks et al, 1996) using all data for the reference sites and the before period data for the treated sites.

The crash reduction factor (CRF) (i.e. 1 - CMF) or the safety effectiveness of the treatment averaged over all sites was calculated as follows (Persaud et al., 2008):

$$CRF = 1 - \frac{\sum_{i=1}^{m}\sum_{t=t_Y}^{t_Y+t_Z} Y_{it}}{\sum_{i=1}^{m}\sum_{t=t_Y}^{t_Y+t_Z} \lambda_{it}} \quad (2\text{-}29)$$

Where m is the total number of treated sites, ty is the first year after treatment, tz is the number of years in the after period, Yit is the actual observed crashes for segment i in year t in the after period, and λ_it is the expected crashes without treatment in the after period for segment i in year t.

2.2.5 The Cross-sectional Method

The cross-sectional studies can be used to estimate the safety effects of certain treatments on specific roadway types (e.g., median width of expressway) since it is difficult to isolate the effect of the treatment from the effects of the other treatments applied at the same time using the before-after methods (Harkey et al., 2008). Moreover, 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. Most cross-sectional studies include principal roadway cross-section attributes such as number of lanes, lane width, shoulder width, surface type, median type, turning lane, vertical grade, and horizontal and vertical curve characteristics, etc. (Shen, 2007). According to the HSM, the CMFs can be estimated by cross-sectional studies when the date of the treatment installation is unknown and the data for the period before treatment installation are not available. The cross-sectional method is generally used for two purposes (Tarko et al., 1998): 1) develop predictive model for the expected number of crashes, and 2) quantify safety impact of highway improvements by CMFs.

As stated by NCHRP project 20-7 (Carter et al., 2012), the CMF can be estimated by taking the ratio of the average crash frequency of sites with the feature to the average crash frequency of sites without the feature. The CMFs can be calculated from the coefficient of the variable associated with treatments as the exponent of the coefficient when the form of the model is log-linear (Lord and Bonneson, 2007). The standard error (SE) of the CMF can be calculated by Equation (2-30) as follows (Harkey et al., 2008):

$$SE = (\exp(\beta_k + SE_{\beta_k}) - \exp(\beta_k - SE_{\beta_k}))/2 \tag{2-30}$$

2.3 Combining Safety Effects of Multiple Treatments

Various methods of combining multiple CMFs for single treatments have been developed to estimate the combined safety effects of multiple treatments. The NCHRP project 17-25 (2008) used a survey to identify the methods of combining multiple CMFs, which have been implemented by different agencies. Table 2-1 summarizes the existing methods for combining multiple CMFs.

Method 1 is a common approach suggested by the HSM for combining multiple CMFs when independence of treatments is assumed. According to Garber and Hoel (2002), this method was first proposed by Roy Jorgensen and Associates for estimation of overall CMF of multiple CMFs. As shown by the equation, CMFs for single treatments are multiplied to estimate combined effects of multiple treatments. However, the assumption of independence cannot account for the potential correlations among multiple treatments.

Method 2 assumes that expected safety effects of the less effective treatment are reduced by a factor in the equation. However, the factor of this equation has no theoretical basis. Therefore, future research is needed to determine this reduction factor. The difference between Method 2 and Method 1 is that Method 2 accounts for difference in effectiveness among multiple treatments.

Method 3 is similar to Method 2 but it has not been used in any studies to estimate the safety effects of combined treatments. According to a survey of the NCHRP 17-25 project, this method was first introduced by Alabama State and the agency practices may have changed since 2003 when the survey was conducted. To the author's best knowledge, there is no clear explanation of this method in the literatures. In particular, it is uncertain which treatment is considered as the first treatment when multiple treatments are applied at the same time. Thus, the authors assume that the treatment with the lowest CMF among all treatments is the first treatment in this study.

Method 4 proposed by Turner (2011) applies a specific weighted factor to the multiplication of CMFs for single treatments. The study determined this weighted factor based on different methods of combining CMFs for single treatments. Since the author applied this method to New Zealand only, the validity of this method for other regions needs to be checked.

Method 5 applies only the lowest CMF (i.e. the CMF for the most effective treatment) among CMFs for multiple treatments according to the survey of the NCHRP 17-25 project. However, this method ignores potential combined effect of multiple treatments. Thus, this method is likely to under-estimate the safety effect of multiple treatments.

Lastly, Method 6 introduced by Bahar (2010) determines a weighted average of multiple CMFs for the same treatment from different studies. Higher weight is applied to the CMF with smaller errors. Gross and Hamidi (2011) compared this method with other methods of combining CMFs.

There are very few studies on combined effects of multiple treatments. Bauer and Harwood (2013) evaluated the safety effect of the combination of horizontal curvature and percent grade on rural two-lane highways. Safety prediction models of five types of horizontal and vertical alignment combinations for fatal-and-injury and PDO crashes were developed and CMFs representing safety performance relative to level tangents were calculated from these models. Elvik (2009) presented an exploratory analysis of models for estimating the combined effects of road safety measures. Based on few studies that have evaluated the effects of multiple road safety measures introduced at the same locations, the paper compares two models. One of the models assumes that the (percentage) effect of a road safety measure remains unchanged when it is combined with other road safety measures. The other model assumes that the most effective measure in a set of measures has a dominant effect that weakens the effects of other road safety measures it is combined with. Evidence from the few studies that were found is consistent with both these models. According to Pitale et al. (2009), the safety effects of paving shoulders, widening paved shoulders (from 2ft to 4ft), and installing shoulder rumble strips on rural two-lane roadways are 16%, 7%, and 15% reductions in crash rates, respectively. Moreover, the result indicated a 37% reduction in crash rates associated with installing shoulder rumble strips + paving shoulders to segments with aggregate shoulders. However, these results were estimated by simply comparing crash rates between the before and after conditions. Gross and Hamidi (2011) applied some of the above methods of combining multiple CMFs to calculate the CMF for shoulder rumble strips + widening shoulder. They combined CMFs for two single treatments (shoulder rumble strips and widening shoulder) from two different sources. They found that the combined CMFs calculated using Methods 1 (HSM) and 2 (systematic reduction of subsequent CMFs) were similar to actual CMFs obtained from two different studies - Pitale et al. (2009) and Hanley et al. (2000). However, CMFs are likely to vary across different study areas even for the same treatment. Thus, combining CMFs obtained from different sources and comparing the combined CMF with actual CMFs from different studies do not clearly identify the best methods of combining multiple CMFs. Also, according to Hanley et al. (2000), some shoulder widening occurred in combination with installation of the rumble strips. However, the range of widening shoulder width was not specified in the study. Thus, there is a need to 1) compare the combined CMF with actual CMF for multiple treatments in the same study area and 2) ensure that roadway geometric conditions (e.g. range of widening shoulder width) are consistent among two treatments and their combination.

In summary, there has been no study that has comprehensively evaluated these existing methods of combining multiple CMFs for single treatments through the comparison with actual CMF for multiple treatments in the same study area.

Table 2-1: Existing methods of combining multiple CMFs (Source: NCHRP project 17-25 (2008), Gross and Hamidi (2011))

No.	Methods	Description
1	$CMF_t = CMF_1 * CMF_2 * \cdots * CMF_n$ CMF_t = CMF for the combined treatments CMF_1 = CMF for the first treatment CMF_2 = CMF for the second treatment CMF_n = CMF for the nth treatment	Assume independence of treatments
2	$CMF_{2,Reduced} = \frac{1-CMF_2}{2} + CMF_2$ $CMF_{combined} = CMF_1 * CMF_{2,Reduced}$ CMF_2 = Less effective CMF than CMF_1	Systematic reduction of safety effects of less effective treatment
3	$CMF_t = CMF_1 - \frac{1-CMF_2}{2} - \cdots - \frac{1-CMF_n}{n}$ CMF_t = CMF for the combined treatments CMF_1 = CMF for the first treatment CMF_2 = CMF for the second treatment CMF_n = CMF for the nth treatment	Safety effects of second treatments is systematically diminished
4	$CMF_{combined}[TurnerMethod] = 1 - [\frac{2}{3}(1 - (CMF_1 * CMF_2))]$	Multiply weighted factor
5	Only the lowest CMF is applied (i.e. treatment with the highest expected crash reduction)	Apply only the most effective CMF
6	$CMF = \frac{\sum_{i=1}^{n} CMF_{unbiased,i}/S_i^2}{\sum_{i=1}^{n} 1/S_i^2}$ $S = \sqrt{\frac{1}{\sum_{i=1}^{n} 1/S_i^2}}$ CMF = combined unbiased CMF value. $CMF_{unbiased,i}$ = unbiased CMF value from study i. S_i = adjusted standard error of the unbiased CMF from study i. n = number of CMFs to be combined. S = estimate of the standard error for the combined CMF	Weighted average of multiple CMFs (Meta-Analysis)

2.4 Estimation of Crash Modification Functions

There are few previous studies that have looked at the variation of CMFs based on different roadway characteristics or different conditions through estimation of CMFunctions. Elvik (2009) provides a framework to evaluate CMFunctions for the same or similar treatment by means of meta-regression analysis (Elvik, 2005) based on multiple studies. He estimated CMFunctions for installation of bypass and converting signalized intersections to roundabouts based on population changes. The results showed that the CMFs increasing with population for both treatments. However, fairly large amounts of data are needed to develop good CMFunctions.

Similar to this study, Elvik (2013) assessed the relationship between safety effects (accident rate) and radius of horizontal curves based on the studies from 10 different countries. The paper evaluates the summary crash modification function to assess the international transferability of national crash modification functions that have been estimated for the relationship between their accident rate and radius of curve. It was found that the estimated crash modification function appears to be a representative summary of these national functions. The results showed that accident rate increases as curve radius decreases and the relationship between accident rate and radius of curve appears to be the same in all countries.

Elvik (2011) applied six linear and non-linear functions to develop CMFunctions for speed enforcement. The CMFunction illustrates the effect of speed enforcement on the injury accidents as a function of the relative change in the level of speed enforcement. The results showed that increasing level of enforcement is associated with a reduction of accidents. The non-linear logarithmic function best fitted the data points from 13 previous studies but the inverse function also fitted the data well.

Park et al. (2014) developed CMFunctions using 5 different linear and non-linear regression functions for two single treatments (installing shoulder rumble strips, widening shoulder width) and combined treatment (installing shoulder rumble strips + widening shoulder width) based on original shoulder width of treated sites. The results show that for the roadway segments with shoulder width of 9 ft or above, only one single treatment can show better safety effects than two treatments. Based on the results of All crashes (KABCO), shoulder rumble strips are more effective in reducing crashes for roadway segments with shoulder width less than 7 ft, whereas widening shoulder width is more effective for roadway segments with shoulder width of 7 ft or above. It was concluded that the CMFunctions can be used to identify general relationships between the CMFs and the roadway characteristics.

Similar to this study, Park and Abdel-Aty (2015a) developed CMFunctions for combination of rumble strips and widening shoulder width treatments. Twelve different types of linear and nonlinear functional forms were compared to find the best fitted function. indicate that the safety effects of two single treatments and combination are higher for the segments with narrower shoulder width. Also, SRS is more safety effective for roadway segments with shoulder width of 10ft or above and 9.5ft or above, whereas WSW is more safety effective for roadway segments with shoulder width less than 10ft and 9.5ft for All crashes (KABCO) and All crashes (KABC). The results also showed that SRS is more safety effective for roadway segments with shoulder width of 7.5ft or above, whereas WSW is more safety effective for roadway segments with shoulder width less than 7.5ft for SVROR (KABCO) crashes. The difference between CMFs of two single treatment and CMFs for multiple treatments is getting larger as shoulder width decreases for both All and SVROR crashes. The results indicate that the safety effects of multiple treatments vary based on characteristics of roadway segments. For the relationship between the CMFs and original shoulder width of treated sites, linear regression and nonlinear regression with power functional form models are the best fitted functions.

Sacchi et al. (2014) also claimed that using a single value of CMF may not be suitable to represent the variation in safety effects of the treatment over time. Thus, the authors developed CMFunctions to incorporate changes over time for the safety effectiveness of treatment. The poisson-lognormal linear intervention and non-linear intervention models were developed and compared to find the best fitted function for the safety effects of the signal head upgrade program. However, the CMFunctions used in this study only account for changes in safety effects over time, but not different roadway characteristics of the treated sites. To overcome this limitation, Sacchi and Sayed (2014) estimated CMFunctions that accounted for AADT changes among treated sites and time trends using the same data for evaluation of the safety effectiveness of the signal head upgrade program.

Park et al. (2015a) estimated CMFunctions using multiple roadway and socio-economic characteristics to assess the safety effects of installation of bike lanes on urban arterials. It was found that CMFunctions with multiple parameters show better model fit than simple models. Also, the results show that the CMFunctions with socio-economic parameters show better model fit than the full CMFunctions without socio-economic parameters for total crashes whereas no socio-economic parameter was significant for injury crashes.

To consider the variation of CMFs over time, Park et al. (2015b) utilized nonlinearizing link functions in developing CMFunctions. The study showed that the CMFs vary across the sites with multiple different roadway characteristics. In particular, the CMFs were lower for the roadways with 1) low LOS level (high AADT per lane) before treatment and high LOS level (low AADT per lane) after treatment and 2) a wide shoulder width. However, the CMFs are relatively higher when the LOS level is the same for the before and after periods. Moreover, the safety effects decrease over time until the third year after treatment and maintained that level after. The CMFunctions also showed the variation of CMFs over time. It was found that CMFunctions with the nonlinear predictor show better model performance than models without the nonlinear predictor. Therefore, it can be concluded that including the nonlinearizing link function in developing CMFunctions improve the goodness of fit of the models, if the variation of CMFs with specific parameters has a nonlinear relationship.

Wang et al. (2015) applied traditional time series regression models to account for temporal effects on the variation of CMFs. The study showed that the model can better predict trends of the CMFs for the signalization and adding red light running cameras (RLCs) when the CMFs are calculated in 90-day moving windows compared to the CMFs calculated in each month. Moving windows was used to compensate the noise due to short sample size. The study also demonstrated that the ARMA time series model can be applied to the prediction of the CMFs in the long term based on historical trend of CMFs over time.

2.5 Roadway Cross-section Elements and Roadside Safety

Evaluating the safety effectiveness of how crash frequency or severity has changed due to a specific improvement or a combination of improvements is a vital step in roadway safety studies. Improvements and countermeasures are mainly motivated by planning, traffic operation and/or safety reasons. Roadway characteristics such as number of lanes, lane width and median types/width are major roadway cross-section elements. Moreover, shoulder rumble strips, shoulder type/width, guardrail and distance between roadside features and roadway are roadside elements.

The widening of roadways with the addition of a through lane is encouraged by certain aspects of traffic planning such as capacity problems or an increase in future traffic demand. Although the relationship between the number of lanes and roadway capacity is well defined in the Highway Capacity Manual (HCM, 2010), which uses the Level of Service (LOS) as a measure to assess the operational performance of roadways with roadway elements, the safety effectiveness of widening urban four-lane roadways to six-lanes is not fully presented. However, since the addition of one through lane in each direction can greatly change cross-sectional elements of roadways, the safety effectiveness of widening urban four-lane roadways to six-lanes has to be fully understood.

Kononov et al. (2008) found that there was a lack of prior studies about the safety effects of the number of lanes on urban freeways. They then estimated the safety performance functions (SPFs) for different number of lanes by the cross-sectional method. By the comparison of the slopes of the SPFs, it was found that an increase in the number of lanes leads to safety improvement.

Also, there are several previous studies that estimate safety effects between two-lane and four-lane rural highways by the cross-sectional method. Four-lane divided roadways were safer than two-lane roadways by a 40 to 60 percent reduction in total crashes in California, Michigan, North Carolina, and Washington State (Council and Stewart, 2000). Fitzpatrick et al. (2005) also found that four-lane divided roadways in Texas show better safety performance when the average daily traffic (ADT) is higher than 10,000. It should be noted that the cross-sectional method was conducted for these studies and there are two major improvements between two-lane and four-lane roadways: addition of a through lane and installation of a raised median.

On the other hand, Abdel-Aty and Radwan (2000) identified that the crash rate increases as the number of lanes on urban roadways increases. Although several previous studies evaluated the safety effectiveness of the change of the number of lanes on roadways, there are no studies that have adopted an observational before-after analysis to estimate the safety effects of widening urban four-lane roadways to six-lanes.

Many researchers have examined the relationship between lane width and crash frequency in the past studies. In general, they found that an increase in lane width reduces crash frequency (Lord and Bonneson, 2007; Yanmaz-Tuzel and Ozbay, 2010; Labi, 2011; Park et al., 2012; Haleem et al., 2013). This is mainly because a wider lane increases the separation between vehicles in adjacent lanes and allows larger deviation of vehicles from the center of the lane (Akgügör and Yıldız, 2007). Larger lane width helps prevent crashes by reducing chances of vehicle encroachment to adjacent lanes. Drivers also feel less pressure as the distance with the other objects in both sides of their vehicles increases (Yang et al., 2013).

The HSM also suggested that crash frequency decreases as lane width increases – i.e. the CMF increases as lane width decreases from 12-ft lane. However, the HSM shows that CMF for a given lane width varies with AADT based on the studies by Zegeer et al. (1988) and Griffin and Mak (1987). More specifically, the CMF is the lowest for AADT < 400 veh/day and the highest for AADT > 2000 veh/day. Based on the expert panel's judgment, the CMF is assumed to increase linearly with AADT for AADT between 400 and 2000 veh/day (Harwood et al., 2000).

For this range of AADT, the CMF is estimated using the CMFunctions which describe the CMF asin a function of AADT.

However, Hauer (2000) suggested that an increase in separation of vehicles on wider lanes tends to increase vehicle speeds and reduce spacing between vehicles. Consequently, an increase in lane width may rather increase crash frequency. In fact, Qin et al. (2004) found that wider lane increased single-vehicle crashes on highway segments in Michigan. Mehta and Lu (2013) also found that crash frequency increased with lane width on rural two-lane roads and rural four-lane divided roads in Alabama. The study accounted for the effects of speed limits and shoulder width in the crash prediction models.

Some studies explained that these opposite effects of increasing lane width are due to the association between lane width and shoulder width, and differences in local conditions. Gross et al. (2009) reported that effects of lane width on crash frequency were neither consistently positive nor negative due to variation in shoulder width. Thus, they suggested that CMFs be determined considering interaction between lane width and shoulder width. Potts et al. (2007) also recommended that narrowing lane width be used as a treatment based on local conditions since the effect of lane width varies by location.

These inconsistent results are also because the relationship between lane width and crash frequency is not linear. Gross and Jovanis (2007) and Gross (2013) found that the odds ratio of crash occurrence increases or decreases depending on ranges of lane width where the base case is 12 ft (= 3.66 m). The odds ratio increases for the ranges of lane width less than 10.5 ft and greater than 12.5 ft but it decreases for lane width of 10.5~12.5 ft. Similarly, Xie et al. (2007)

showed that the relationship between lane width and crash frequency is described in a "concave-downward" polynomial function – crash frequency increases as lane width increases from 9 ft to 10 ft and decreases as lane width increases from 10 ft to 13 ft. This indicates that there is a need to reflect this nonlinear relationship for developing the CMFs to assess safety effects of changing lane width.

Some studies showed that changing lane width is also associated with crash injury severity. Labi (2011) found that increasing lane width reduced higher percentage of fatal/injury crashes but lower percentage of PDO crashes. In particular, wider lanes are more effective in reducing fatal/injury crashes for rural major collectors. Similarly, Wong et al. (2007) reported that a decrease in lane width increases fatal/injury crashes at signalized intersections. However, Park et al. (2012) found that an increase in lane width rather increases fatal/injury crashes at nighttime. Hauer et al. (2004) showed that lane width is associated with PDO crashes, but not injury crashes on four-lane undivided roadway segments. However, differential effects of changing lane width on crash injury severity have not been associated with nonlinear relationship between lane width and crash frequency.

Lee et al. (2015) evaluated safety effects of changing lane width considering nonlinear relationships between lane width and crash rate. It was found that the logarithm of crash rate was the highest for 12-ft lanes and lower for the lane width less than 12 ft or greater than 12 ft. This relationship contradicts some past studies which found that an increase in lane width consistently reduces crash frequency due to a larger separation between vehicles in adjacent lanes. However, a larger separation may rather make drivers feel safer and increase their speeds. This tendency is more likely to be prevalent on the roadway segment with 12-ft lane in Florida due to its higher posted speed limit compared to the segments with wider or narrower lane.

Several studies investigated the safety performance of road diet in urban areas. A road diet involves narrowing or elimination travel lanes on a roadway to make more room for pedestrians and bicyclists (FHWA, 2008). While there can be more than four travel lanes before treatment, road diets are often conversions of four-lane undivided roadways into three-lanes - two travel lanes plus a center turn lane (e.g. TWLTL).

Harkey et al. (2008) used the observational before-after with EB method to evaluate CMF for road diet treatment for total crashes. They found that the CMFs for road diet are 0.53 and 0.81 for Iowa and California/Washington. It was also found that the CMF of road diet for three states is 0.71.

Pawlovich et al. (2006) evaluated the effects of road diet on crashes in Iowa using a Bayesian approach. The study showed that a 25.2% reduction in crash frequency per mile and an 18.8% reduction in crash rate.

Huang et al. (2002) estimated the safety effects of road diet (i.e. conversion of 4-lane to 3-lane with TWLTL) for total and injury crashes. The study includes 12 treated sites and 25 comparison sites in California and Washington. It was found that road diet resulted in an average of 6% crash reduction of total crashes.

2.5.1 Roadside Elements

Roadside elements have been known as one of the most important hazards for roadway safety. Zeng and Schrock (2013) evaluated the safety effects of 10 shoulder design types in winter and non-winter periods. They developed CMFs using cross-sectional methods. The results showed that wider and upgraded shoulders had significantly lower impact on safety in winter periods than non-winter periods.

Wu et al. (2014) proposed an approach to account for the variability in crash severity as a function of geometric design, traffic flow and other roadway features, and tested it by evaluating the safety effects of shoulder rumble strips on reducing crashes. It was found that shoulder rumble strips reduce the total number of crashes, but have no statistically significant effect on reducing the probability of a severe crash outcome.

Turner et al. (2012) found that installation of shoulder rumble strips resulted in an average of 21% reduction of all crashes and 40% reduction of run-off roadway crashes based on their review of 13 studies. Turner et al. (2009) also found from 5 recent studies that shoulder rumble strips reduced injury crashes by around 23%. Jovanis and Gross (2008) estimated safety effects of shoulder width using Case Control and Cohort methods. The results of the two methods showed that crashes decrease as shoulder width increases.

In urban areas, bike lanes are mostly placed in the shoulder of roadways and bicyclists are simultaneously riding next to vehicles. Therefore, there are higher chances of conflicts between bicycles and vehicles. Bike lanes can reduce the number of conflicts by separating bicyclists from vehicles with bicyclists' own designated path. Thus, bike lanes are likely to reduce bike crashes. Abdel-Aty et al. (2014) estimate the safety effectiveness of bike lanes using cross-sectional method and it was found that installation of bike lane has positive safety effects on reducing 4 different crash types and severity levels as follow: total crashes, injury crashes, bike crashes, and bike injury crashes.

Chen et al. (2012) evaluated the safety effects of installation of on-street bicycle lanes in New York City for 5 different crash types and severities as follows: total crashes, bicyclist crashes, pedestrian crashes, multiple-vehicle crashes, and injury or fatal crashes. The Generalized Estimating Equation methodology was conducted to compare the changes in crashes at the treated group and the comparison group before and after periods. The results showed that although the probable increase in the number of bicyclists, installation of bicycle lanes did not lead to an increase in crashes. This may be because vehicular speeds and the number of conflicts between vehicles and bicyclists decreased after the installation.

According to Sadek et al. (2007), based on survey data, the installation of advanced bike lane helps increase awareness of drivers and bicyclists. The responses showed that 75.4% drivers believed that the new bike lane made drivers more aware of the presence of bicyclists. The survey also showed that 76% of bicyclists said that new bike lane had made them more vigilant. However, Jensen (2008) concluded that adding a bike lane increases frequencies of All crashes (KABCO, KABC) and Bike crashes (KABCO) for roadways in Kopenhagen, Denmark. The CMFs of installation of bike lanes were estimated using the observational before-after with comparison group (CG) method in this study. The results showed that the CMFs were 1.30, 1.27, and 1.27 for All crashes (KABCO), All crashes (KABC), and Bike crashes (KABCO), respectively.

On the other hand, Rodegerdts et al. (2004) suggested that adding a bike lane reduces Bike-related crashes (KABCO). The CMF was 0.65 for Bike crashes (KABCO). Nosal and Miranda-Moreno (2012) estimated bicyclists injury risk of bicycle facilities (cycle-tracks, bicycle lanes) and explored the differences in injury risk between different types of bicycle facilities in Montreal, Canada. The study compared injury risk between the treated sites and control streets to assess the impact of bicycle facilities. The results showed that the safety effects of cycle-tracks and bicycle lanes of treated streets were higher than the corresponding control streets. Overall, there was a minimum of 6% to maximum 17% reduction in average injury rates on segments compared to the control streets. Similar to this study, Lusk et al. (2011) also found that relative risk of riding bicycles on the cycle tracks versus on regular streets was 28% reduction in injury rates. However, it is worth to mention that these studies simply compared crash rates between treated sites and comparison sites but didn't find any relationship between roadway characteristics and the safety effects of a bike lane.

Reynolds et al. (2009) reviewed 23 studies that assessed the effect of transportation infrastructure on bicyclist safety. Based on the previous studies that examined impacts of infrastructures at straightaways (e.g. bike lanes or paths) and intersections (e.g. roundabouts, traffic lights), they found that bicycle specific facilities generally reduced crashes and injuries. Additionally, it was reported that street lighting, paved surfaces, and low-angled grades are the factors that can improve bicyclist safety. However, it is worth to note that the 8 papers for bike lanes or paths were published in 90s.

A number of studies addressed the safety effects of guardrails and different types of barriers on roadside and median of roadways. Especially, guardrails and barriers have been widely implemented on roadways during the last several years to improve safety. It is worth to note that addition of barriers might increase the crash frequency, but it might helpful to reduce severe crashes (Elvik, 1995; Miaou et al., 2005; Donnell and Mason, 2006; Tarko et al., 2008; Zou et al., 2014). Moreover, installation of roadside guardrails is found to be effective in reducing crash severity (Michie and Bronstad, 1994; Elvik, 1995; Lee and Mannering, 2002).

On the other hand, Jang et al. (2010) found that installations of median barrier and roadside guardrail can reduce all types of crashes by 77% and 58%. Also, it should be noted that a new chapter for freeway and interchanges is recently added in the HSM. The new chapter contains the CMFs for addition of roadside barriers. However, it is worth to mention that the CMF is representing the safety effects of all types of roadside barriers including concrete and cable barriers, w-beam guardrail, and bridge rail, but not CMF for specific type of roadside barrier.

2.6 Nonlinear Effects in Safety Evaluation

To estimate the CMF using the cross-sectional method, development of SPFs or CPMs is required. Due to its strength of accounting for over-dispersion, GLM with 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 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., 2008b; Li et al., 2011; Lee et al., 2015).

Therefore, researchers have tried to apply different techniques to account for the nonlinearity of variables on crash frequency. For instance, an application of using 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. (2015) 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 12ft whereas wider lanes reduce crashes for lane widths greater than 12ft. 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 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 order to account for both nonlinear effects and interaction impacts between variables, another data mining technique, the MARS, have been used in safety evaluation studies. According to Briand et al. (2004) and Haleem et al. (2013), 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. (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. Harb et al. (2010) applied MARS to assess safety effects of toll-lane processing time.

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.

2.7 Summary (Current Issues)

Considerable researches have been conducted to estimate CMFs for roadway improvements and treatments using various before-after studies and the cross-sectional method. There are several important issues in CMF studies. They are; 1) multiple treatments, 2) variation of CMFs, 3) estimation of CMFunctions, and 4) nonlinear relationship between the safety effects and predictors. First, the HSM suggests that CMFs are multiplied to estimate the combined safety effects of single treatments. However, the HSM cautions that the multiplication of the CMFs may over- or under-estimate combined effects of multiple treatments. Second, since the CMF is a single value which represents average safety effects of the treatment for all treated sites, the heterogeneous effects of roadway characteristics on CMFs among treated sites are ignored. Third, to overcome the limitation of using a fixed value of CMF, crash modification CMFunctions have been developed to predict the variation in CMFs based on the site characteristics. However, although previous studies (Elvik, 2009; Elvik, 2011; Elvik, 2013; Park et al., 2014; Sacchi et al., 2014) assessed the effect of a specific single variable such as AADT on the CMFs, there is a lack of prior studies on variation in the safety effects of specific treatment among treated sites with different multiple roadway characteristics over time. Lastly, the nonlinearity of variables in the cross-sectional method is not discussed in the HSM.

CHAPTER 3: EXPLORATION AND COMPARISON OF CRASH MODIFICATION FACTORS FOR MULTIPLE TREATMENTS

3.1 Introduction

As shown in the literature review, the HSM provides various CMFs for single treatments, but not CMFs for multiple treatments on 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 the CMFs may over- or under-estimate combined effects of multiple treatments.

Moreover, since the CMFs in the first edition of the HSM were determined based on past studies for specific regions, they may not represent a safety impact for other locations and conditions even if roadway characteristics are similar. The objectives of this study are 1) to evaluate safety effects (i.e. CMF) of two single treatments (installing shoulder rumble strips, widening shoulder width) and one combined treatment (installing shoulder rumble strips + widening shoulder width) using before-after studies and cross-sectional studies and 2) to compare the CMFs estimated using the existing methods of combining the CMFs for single treatments with actual CMFs for multiple treatments calculated using before-after studies. From this comparison, the study will show whether the existing methods of combining the CMFs over- or under-estimate actual CMFs. In this study, it is referred to 'All crash types (all severities)' as All crashes (KABCO), 'All crash types (Fatal+Injury)' as All crashes (KABC), 'SVROR (all severities)' as SVROR (KABCO), and 'SVROR (Fatal+Injury)' as SVROR (KABC) for crash types and severity levels.

3.2 Data Preparation

Three sets of data for Florida were used in the study: roadway characteristic inventory (RCI) data for six years (2005-2010), financial project information, and crash data for ten years (2003-2012). In order to identify the treated sites on rural multilane roadways, the RCI data and financial project information were obtained from the RCI historical database and the Financial Management System maintained by the FDOT. The RCI database provides current and historical roadway characteristics data, and reflects features of specific segment for selected dates. Around 200 roadway characteristics are included in the RCI database. The Financial Management System offers a searching system named financial project search. This system provides detailed information on a specific financial project such as district number, status, work type, and year.

Using these two databases, the sites with the two single treatments and the combined treatment, which are installing shoulder rumble strips, widening shoulder width, and shoulder rumble strips + widening shoulder width were identified. Also, comparison group data were collected using the RCI database based on roadway characteristics of the treated group such as functional class, type of road, number of lanes, section ADT, median width, median type, shoulder width, shoulder type, maximum speed limit, and lane width. As suggested by Pendleton (1998), the total length of the comparison group data was set to around five times longer than the total length of the treated group data. A total of 257 and 676 roadway segments were identified for the treated and comparison groups, respectively. The total lengths of the treated and comparison group are 180.722 and 699.092 miles, respectively.

Crash data for these treated and comparison groups in before and after periods were obtained from the Crash Analysis Resource (CAR) database. Due to the difficulty in identifying enough treated sites, all locations that have been treated between 2005 and 2010 were considered for analysis. The crash data was extracted for each site for 2 years before and 2 years after periods. This criterion for crash data was therefore used consistently for the before-after analysis. Once roadway characteristic data and crash data were collected and matched by roadway ID and segment mile point of each site, crashes that occurred in the intersection influence area were manually removed using Google Earth and Transtat-Iview - a GIS searching system offered by FDOT. Table 3-1 summarizes the data.

Table 3-1: Summary of data description

	Treated Group		Comparison Group	
Treatments	Number of Segments	Length (mile)	Number of Segments	Length (mile)
Shoulder Rumble Strip	60	38.684	115	160.621
Widening Shoulder Width	75	102.071	367	361.079
Shoulder Rumble Strip + Widening Shoulder Width	122	39.967	194	177.392
- AADT: 2,000 to 50,000 veh/day - Widening Shoulder Width (0.5 ~ 10 feet)				

3.3 Statistical Method

3.3.1Safety Performance Functions

A SPF that relates the crash frequency to traffic and geometric parameters can be developed using the NB model formulation with the data for the untreated reference sites. Two types of SPFs, which are the Full SPF and the Simple SPF, have been mainly used in the literature. Full SPF relates the frequency of crashes to both traffic and roadway characteristics, whereas Simple SPF consider a traffic parameter only such as AADT as an explanatory variable. It should be noted that CMFs in the HSM are calculated based on the Simple SPF only. However, the Simple

SPF is an over-simplified function since crash frequency is not only affected by the traffic volume. In this study, the Full SPF was used for calculating CMFs in the EB method. The functional form of SPF for fitting the NB regression models is as follows:

$$N_{predicted} = \exp(\beta_o + \beta_1 \ln(AADT) + \beta_2 L + \beta_3 ST + \beta_4 SW) \qquad (3\text{-}1)$$

Where,

$N_{predicted}$= Predicted crash frequency,

β_i = coefficients,

$AADT$= Annual Average Daily Traffic of segment (veh/day),

L = length of segment (mi),

ST = shoulder type (1 = shoulder with rumble strip, 0 = shoulder without rumble strip),

SW = shoulder width (ft).

Four SPFs were developed using the NB model for reference sites of rural multilane roadways based on crash types and severity levels using GENMOD procedure in SAS program (2009). A total of 360 roadway segments were identified as reference sites. These segments have similar roadway characteristics to the treated sites in the before period. Roadway characteristics and matched crash data were collected from RCI and CAR databases, respectively. The Full SPFs were developed for the following four combinations of crash type and severity level: 1) All crashes (KABCO), 2) All crashes (KABC), 3) SVROR (KABCO), and 4) SVROR (KABC). Table 3 shows the results of the calibrated Florida-specific Full SPFs. As shown in the results, crash frequency is higher for road segments without shoulder rumble strip and shorter shoulder width.

Table 3-2: Summary of data description Florida specific calibrated SPFs for rural multilane roadways by crash types and severity levels

		Coefficient										Dispersion (K)	Deviance
		α Intercept		β_1 Log(ADT)		β_2 Segment Length		β_3 Shoulder Type		β_4 Shoulder Width			
Crash Type	Severity	Estimate	P-Value	Estimate	P-Value	Estimate	P-Value	Estimate	P-Value	Estimate	P-Value		
All types	KABCO	-8.6554	<.0001	2.5858	<.0001	0.4800	<.0001	-0.4247	0.0015	-0.0885	<.0001	0.6812	406.04
	KABC	-9.4049	<.0001	2.5362	<.0001	0.5375	<.0001	-0.4994	0.0006	-0.0724	0.0002	0.5923	390.06
SVROR	KABCO	-4.9732	<.0001	1.5589	<.0001	0.3076	<.0001	-0.3439	0.0223	-0.1544	<.0001	0.1494	358.39
	KABC	-5.0920	<.0001	1.4552	<.0001	0.3171	<.0001	-0.6441	<.0001	-0.1589	<.0001	0.1121	317.98

3.3.2 Negative Binomial Models

The NB model has been most frequently used model in crash count model (Maycock and Hall, 1984; Hauer et al,. 1988; Miaou, 1994; Shankar et al., 1995; Poch and Mannering, 1996; Milton and Mannering, 1998; Karlaftis and Tarko, 1998; Persaud and Nguyen, 1998; Abdel-Aty and Radwan, 2000; Carson and Mannering, 2001; Miaou and Lord, 2003; Amoros et al., 2003; De Guervara et al., 2004; Hirst et al., 2004; Abbas, 2004; Lord et al., 2005; Wang and Abdel-Aty, 2006; El-Basyouny and Sayed, 2006; Lord, 2006; Kim and Washington, 2006; Lord and Bonneson, 2007; Lord et al., 2010; Malyshkina and Mannering, 2010; Daniels et al., 2010; Cafiso et al., 2010; Naderan and Shashi, 2010; Abdel-Aty et al., 2011; Ukkusuri et al., 2011; Lee at al., 2013; and Park et al., 2014). Crash data have a gamma-distributed mean for a population of systems, allowing the variance of the crash data to be more than its mean (Shen, 2007). Suppose that the count of crashes on a roadway section is Poisson distributed with a mean λ, which itself is a random variable and is gamma distributed, then the distribution of frequency of crashes in a population of roadway sections follows a negative binomial probability distribution (Hauer, 1997).

yi|λi ~ Poisson (λi)

λ ~ Gamma (a,b)

Then, P(yi) ~ Negbin (λi, k)

$$= \frac{\Gamma(1/k + y_i)}{y_i!\Gamma(1/k)}\left(\frac{k\lambda_i}{1+k\lambda_i}\right)^{y_i}\left(\frac{1}{1+k\lambda_i}\right)^{1/k} \tag{3-2}$$

Where, y = number of crashes on a roadway section per period,

λ= Expected number of crashes per period on the roadway section, and

k= over-dispersion parameter.

The expected number of crashes on a given roadway section per period can be estimated by Equation (3-3).

$$\lambda = \exp(\beta^T X + \varepsilon) \tag{3-3}$$

Where, β is a vector of regression of parameter estimates, and

X is a vector of explanatory variables, and

$\exp(\varepsilon)$ is a gamma distributed error term with mean one and variance k.

Because of the error term the variance is not equal to the mean, and is given by Equation (3-4).

$$\mathrm{var}(y) = \lambda + k\lambda^2 \tag{3-4}$$

As k = 0, the negative binomial distribution approaches Poisson distribution with mean λ. The parameter estimates of the binomial regression model and the dispersion parameter are estimated by maximizing the likelihood function given in Equation (3-5).

$$l(\beta,k)=\prod_i \frac{\Gamma(1/k+y_i)}{y_i!\Gamma(1/k)}\left(\frac{k\lambda_i}{1+k\lambda_i}\right)^{y_i}\left(\frac{1}{1+k\lambda_i}\right)^{1/k} \tag{3-5}$$

Using the above methodology negative binomial regression models were developed and were used to estimate the number of crashes at the treated sites.

3.4 Results

3.4.1 Estimation and Comparison of CMFs

Table 4 presents the CMFs for two single treatments and the combined treatment estimated using the cross-sectional and the before-after CG and EB methods. The cross-sectional method is also known as safety performance function as mentioned in the previous section. Thus, the CMFs can be estimated using the calculated Full SPFs as described in Table 3. The CMF for adding shoulder rumble strips was calculated as exp(β_3). It is worth to note that the CMFs for widening shoulder width by the cross-sectional method can be described in CMFunctions with the shoulder width as a continuous variable (i.e. CMF = exp($\beta_4\times$ shoulder width)) as shown in Figure 3-1. The figure shows that CMFs gradually decrease as shoulder width increases. This indicates that should rumble strips and widening shoulder width have positive effects on road safety. In particular, shoulder rumble strips have higher effects on All crashes than SVROR based on larger difference between the two CMF curves for widening shoulder width and shoulder rumble strips + widening shoulder width. This may be because rumble strips are typically installed on both inside and outside shoulders of rural multilane highways with high speed limits in Florida, and the model also captured the safety effects of inside shoulder rumble strips on reducing crashes in the median. However, the presence of inside rumble strips could not be verified due to insufficient information in the RCI database.

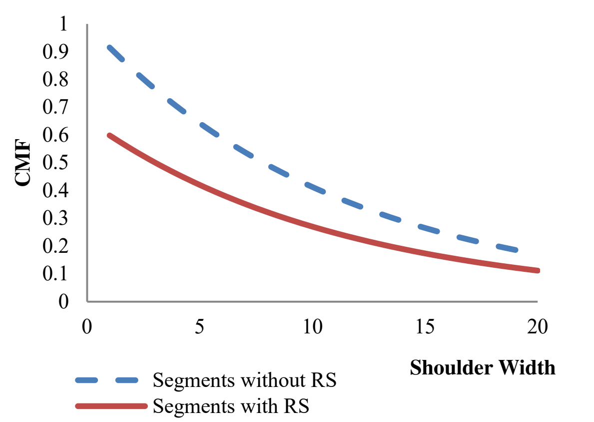
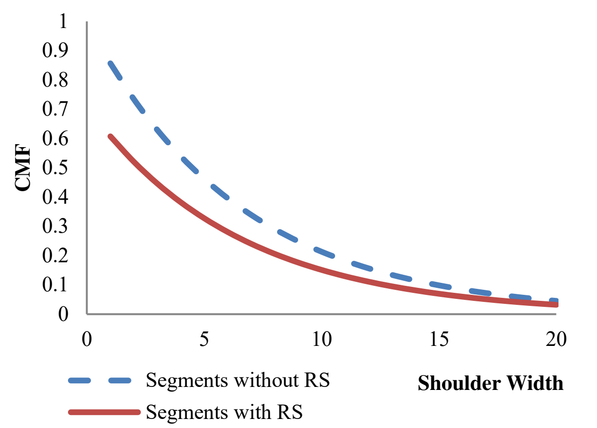
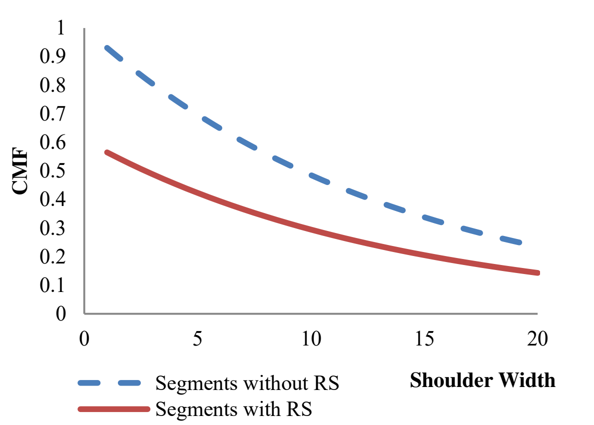
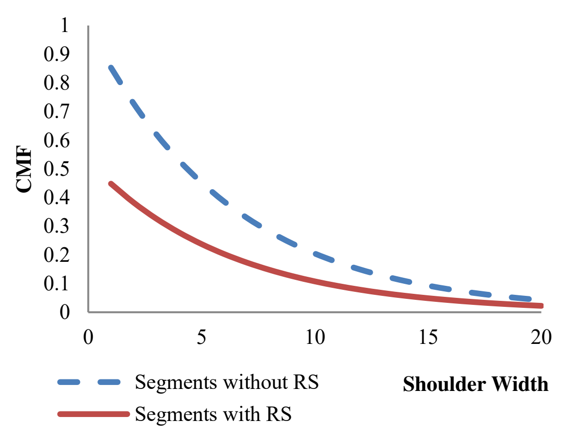
All Crash	
KABCO	KABC
CMF for adding shoulder rumble strips =0.654	CMF for adding shoulder rumble strips= 0.607
	
SVROR	
KABCO	KABC
CMF for adding shoulder rumble strips= 0.709	CMF for adding shoulder rumble strips = 0.525
	

Figure 3-2: Evaluated CMFs using cross-sectional method

However, the CMFs for widening shoulder width are not directly comparable between the cross-sectional and before-after methods because the CMF is described as a function of shoulder width in the cross-sectional method whereas the CMF is fixed in the before-after method for a given crash type and severity level. In general, both cross-sectional and before-after methods consistently show that the safety effects of all the treatments are positive (i.e. CMF < 1) except for the safety effects of widening shoulder width in reducing the SVROR crashes estimated using the CG method. The insignificance of CMFs for this case is mainly due to relatively lower proportion of SVROR in the total crashes associated with widening shoulder width. However, since the EB method uses the predicted crash frequency for estimation of the expected crash frequency based on untreated conditions, the CMFs for widening shoulder width are significant in spite of insufficient SVROR counts for this treatment.

The results of before-after methods for all severity levels (KABCO) show that the CMFs for shoulder rumble strips + widening shoulder width are lower than the CMFs for shoulder rumble strips or widening shoulder width. This indicates that the safety effects increase when multiple treatments are applied instead of only single treatment. Thus, this validates the multiplication of CMFs for single treatments for estimating combined effects of multiple treatments as suggested by the existing methods. However, for injury crashes (KABC), the CMFs for shoulder rumble strips + widening shoulder width is higher than the CMFs for shoulder rumble strips. There is only 6% difference in the CMF estimates between CG and EB methods when only the statistically significant results are considered. Also, the CMFs estimated by both methods show comparable trend for All crashes and SVROR – higher safety effects of the treatments in reducing injury crashes (KABC) than all crashes (KABCO). Moreover, the results of CG method are similar to the EB method with slightly higher standard error except for SVROR (KABCO) for shoulder rumble strips, All crashes (KABCO) for widening shoulder width, and All crashes (KABC) for shoulder rumble strips + widening shoulder width. However, EB method generally provides more reliable estimates of CMFs (i.e. lower standard error) than the CG method.

In comparison of the cross-sectional and before-after methods, it was found that a trend of the CMFs for shoulder rumble strips was generally similar for both methods - higher safety effects of shoulder rumble strips on reducing injury crashes (KABC) than all crashes (KABCO). Also, there was an 8% difference in the CMFs between the cross-sectional and before-after methods when only the best estimate of CMF between the CG and EB methods (i.e. CMF with lower standard error) was considered. This indicates that the cross-sectional study is also a suitable method to estimate CMFs when before-after studies are not feasible due to limitation of data. It is worth noting that the most reliable CMF for the before-after studies was selected in Table 3-3 based on lower standard errors. The CMFs with lower standard error was used for 1) comparison of the CMFs estimated from cross-sectional and before-after methods, 2) calculation of combined CMFs for multiple treatments using the existing methods of combining CMFs and 3) comparison of the actual combined CMFs and estimated combined CMFs.

Table 3-3: Evaluated CMFs of the two treatments and the combined treatment on rural multilane highways

		Cross-Sectional		Observational Before-After Studies			
				CG method		EB method	
Treatment Types	Crash Type (Severity)	CMF	S.E	CMF	S.E	CMF	S.E
Shoulder Rumble Strips	All crashes (KABCO)	0.654**	0.088	0.728**	0.067	**0.763****	**0.056**
	All crashes (KABC)	0.607**	0.088	0.626**	0.089	**0.643****	**0.074**
	SVROR (KABCO)	0.709**	0.107	**0.651****	**0.077**	0.790*	0.112
	SVROR (KABC)	0.525**	0.112	**0.625****	**0.117**	0.695**	0.125
Widening Shoulder Width	All crashes (KABCO)	-	-	0.815**	0.087	**0.771****	**0.053**
	All crashes (KABC)	-	-	0.783**	0.110	**0.688****	**0.064**
	SVROR (KABCO)	-	-	1.105	0.149	**0.607****	**0.164**
	SVROR (KABC)	-	-	1.195	0.207	**0.566****	**0.191**
Shoulder Rumble Strips + Widening Shoulder Width	All crashes (KABCO)	-	-	0.498**	0.063	**0.608****	**0.059**
	All crashes (KABC)	-	-	**0.660****	**0.112**	0.710**	0.120
	SVROR (KABCO)	-	-	0.563**	0.088	**0.541****	**0.085**
	SVROR (KABC)	-	-	**0.611****	**0.147**	0.646**	0.150

**: significant at a 95% confidence level, *: significant at a 90% confidence level,

Note: Values in bold denote the most reliable CMFs among before-after studies.

3.4.2 Comparison of CMFs among segments with different shoulder width

The safety effects of shoulder rumble strips, widening shoulder width, and shoulder rumble strips + widening shoulder width were assessed for the treated sites with different original shoulder width in the before period. The observational before-after studies were applied to the treated sites with shoulder width of 1) 4ft ~ 6ft and 2) 8ft ~ 12ft. These two levels of shoulder width were selected such that there are sufficient samples at each level. Due to low frequency of SVROR crashes, the CMFs with different shoulder width were calculated for All crashes only. The most reliable method between the CG and EB methods (i.e. the CMF with lower standard error) was conducted to estimate the CMFs. Table 3-4 presents the CMFs with different original shoulder width for the two single treatments and the combined treatment estimated.

In general, the results show that the safety effects of all the treatments with different shoulder width are positive and significant at 95% level except for the safety effects of shoulder rumble strips + widening shoulder width on the roadway segments with 8 ft ~ 12 ft shoulder. Moreover, the results show that the CMFs for the roadway segments with 4 ft ~ 6 ft shoulder width are notably lower than the CMFs for 8 ft ~ 12 ft shoulder width. These results imply that the safety effects of the three treatments decrease when they are applied to roadway segments with wider shoulder width.

Based on the results of All crashes (KABCO), multiple treatments are more effective for the roadway segments with 4ft ~ 6ft shoulder width than single treatments, whereas the safety effects of all the treatments for roadway segments with 8 ft ~ 12 ft shoulder width are similar. It is worth to note that for All crashes (KABC), the CMF for shoulder rumble strips + widening shoulder width is rather higher than the CMFs for single treatments for the road segments with 8 ft ~ 12 ft shoulder width. However, the CMFs are not comparable since the CMF for shoulder rumble strips + widening shoulder width is not statistically significant. The result indicates that shoulder rumble strips are more effective than widening shoulder width for the roadway segments with 8 ft ~ 12 ft shoulder width, whereas widening shoulder width is more effective for the roadway segments with 4 ft ~ 6 ft shoulders.

Table 3-4: Evaluated CMFs for the treated sites with different original shoulder width in the before period

			4ft ≤ shoulder width ≤ 6ft			8ft ≤ shoulder width ≤ 12ft		
Treatment Types	Crash Type	Severity	# of segments	CMF	S.E	# of segments	CMF	S.E
Shoulder Rumble Strips	All crashes	KABCO	24	0.614**	0.103	36	0.792**	0.064
		KABC		0.565**	0.137		0.659**	0.086
Widening Shoulder Width	All crashes	KABCO	44	0.617**	0.078	31	0.817**	0.068
		KABC		0.500**	0.084		0.814**	0.067
Shoulder Rumble Strips + Widening Shoulder Width	All crashes	KABCO	75	0.351**	0.062	47	0.807**	0.096
		KABC		0.451**	0.109		0.839	0.142

**: significant at a 95% confidence level

3.4.3 Estimation and Comparison of Evaluated CMFs and Combined CMFs

One of the objectives of this study is to evaluate CMFs of different combination of treatments for equivalent roadway conditions and offer a comparison of evaluated CMFs and combined CMFs using existing methods for combining multiple CMFs. Table 3-5 compares the CMFs estimated using the six different methods of combining CMFs for single treatments (presented in Table 1) to the actual calculated CMFs of shoulder rumble strips, widening shoulder width, and shoulder rumble strips + widening shoulder width. Moreover, over- and under-estimation of actual calculated CMFs by the six existing methods (Table 2-1) for combining multiple CMFs were summarized. Note that the actual calculated CMF for shoulder rumble strips + widening shoulder width for All crashes (KABC) was not statistically significant for the roadway segments with 8ft ~ 12ft shoulder width.

From the comparison between the actual calculated CMFs and the combined CMFs, Methods 1, 2, 5, and 6 produced the combined CMFs closest to the actual calculated CMFs for multiple treatments. More specifically, the best methods are Method 1 for All crashes (KABCO), Method 2 for SVROR (KABCO), and Method 6 for KABC for the roadway segments with 4ft ~12ft shoulder width. For the roadway segments with 4ft ~6ft shoulder width, Method 1 for All crashes (KABCO) and Method 5 for All crashes (KABC) are the best methods. Lastly, Method 6 is the best for All crashes (KABCO) for the roadway segments with 8ft ~12ft shoulder width. The ratio of actual calculated CMF to the best estimate of combined CMF closer to 1 indicates that these methods can estimate the combined effects of multiple treatments at a reasonable accuracy.

In general, for most methods of combining CMFs for single treatments, the combined CMFs of All crashes (KABCO) for the segments with 4 ft ~ 12 ft shoulder width were under-estimated, whereas the combined CMFs of All crashes (KABC) for the segments with 4 ft ~ 12 ft shoulder width were over-estimated. It can also be seen that the combined CMFs for SVROR (KABCO) for the segments with 4 ft ~ 12 ft shoulder width estimated by Methods 1, 2, and 3 were over-estimated, whereas the combined CMFs for SVROR (KABCO) for the segments with 4 ft ~12 ft shoulder width by Methods 4, 5, and 6 are under-estimated. The combined CMFs for SVROR (KABC) were all over-estimated. For the segments with 4 ft ~ 6ft shoulder width, the combined CMFs of All crashes (KABCO) were all under-estimated. It can also be seen that for the segments with 4 ft ~ 6 ft shoulder width, the combined CMFs of All crashes (KABC) estimated by Methods 1, 2, and 3 were over-estimated, whereas the combined CMFs of All crashes (KABC) by Methods 4, 5, and 6 are under-estimated. For segments with 8 ft ~ 12 ft shoulder width, the combined CMFs of All crashes (KABCO) were all over-estimated. This indicates that the over- and under-estimation of actual CMF for multiple treatments depends on the type of crash, severity level, and original geometric characteristics of segments.

Lastly, in order to estimate more reliable combined CMFs, adjustment approaches (averaging and weighting) of the existing methods were attempted. It was found that averaging the CMFs from the best two methods produced better results than using the CMF from only one specific best method. The average of differences between actual calculated CMFs and averages of the combined CMFs from the best two methods was 1.6%, whereas the average of differences between actual calculated CMFs and combined CMFs from only one specific best method was 2.2%. However, the average of differences between actual calculated CMFs and averages of the combined CMFs from the best three methods was 3.3% which is even higher than the average of differences for only one specific best method. The results indicate that it is better not to rely on only one specific existing method of combining CMFs for predicting CMF for multiple treatments.

Table 3-5: Results of actual calculated CMFs and Combined CMFs by existing methods

Crash type (Severity)	Actual calculated CMFs: Rumble Strips (SE)	Actual calculated CMFs: Widening Shoulder (SE)	Actual calculated CMFs: Rumble + Widening Shoulder (SE)	Combined CMFs using Existing Methods (Shoulder Rumble Strips+ Shoulder Widening): Method 1	Method 2	Method 3	Method 4	Method 5	Method 6 (SE)	Average of combined CMFs from the best two methods
Original Shoulder Width: 4ft ~ 12ft										
All crashes (KABCO)	0.763 (0.056)	0.771 (0.053)	0.608 (0.059)	**0.588***	0.680	0.653	0.726	0.763	0.767 (0.038)	(0.588+0.653)/2 = 0.621
				⊕*	⊖	⊖	⊖	⊖	⊖	
All crashes (KABC)	0.643 (0.074)	0.688 (0.064)	0.660 (0.112)	0.442	0.565	0.510	0.628	0.643	**0.669*** (0.048)	(0.669+0.643)/2 = 0.656
				⊕	⊕	⊕	⊕	⊕	⊖*	
SVROR (KABCO)	0.651 (0.077)	0.607 (0.164)	0.541 (0.085)	0.395	**0.501***	0.433	0.597	0.607	0.643 (0.070)	(0.501+0.597)/2 = 0.549
				⊕	⊕*	⊕	⊖	⊖	⊖	
SVROR (KABC)	0.625 (0.117)	0.566 (0.191)	0.611 (0.147)	0.354	0.460	0.379	0.569	0.566	**0.609*** (0.100)	(0.609+0.569)/2 = 0.589
				⊕	⊕	⊕	⊕	⊕	⊕*	
Original Shoulder Width: 4ft ~ 6ft										
All crashes (KABCO)	0.614 (0.103)	0.617 (0.078)	0.351 (0.062)	**0.379***	0.498	0.424	0.586	0.614	0.616 (0.062)	(0.379+0.424)/2 = 0.402
				⊖*	⊖	⊖	⊖	⊖	⊖	
All crashes (KABC)	0.565 (0.137)	0.500 (0.084)	0.451 (0.109)	0.283	0.391	0.283	0.522	**0.500***	0.518 (0.072)	(0.500+0.391)/2 = 0.446
				⊕	⊕	⊕	⊖	⊖*	⊖	
Original Shoulder Width: 8ft ~ 12ft										
All crashes (KABCO)	0.792 (0.064)	0.817 (0.068)	0.807 (0.096)	0.647	0.732	0.647	0.765	0.792	**0.804*** (0.047)	(0.804+0.792)/2 = 0.798
				⊕	⊕	⊕	⊕	⊕	⊕*	

*Best estimate of CMF for multiple treatments compared to actual calculated CMF
⊕ Over-estimated, ⊖ Under-estimated

3.5 Conclusion

While the HSM and related studies caution that the assumption of independence of different treatments can lead to over- or under- estimation of actual CMFs, there was a lack of studies that estimate the combined safety effects of multiple treatments. Thus, the main objective of this study is to comprehensively evaluate the safety effects of two single treatments (shoulder rumble strips and widening shoulder width) and combined treatment (shoulder rumble strips + widening shoulder width) on rural multilane roadways in Florida. The study calculated actual CMFs for shoulder rumble strips + widening shoulder width and also estimated CMFs using six existing methods of combining CMFs for single treatments. CMFs were calculated using two observational before-after studies and cross-sectional studies. The main findings of this study are summarized as follows:

The results of cross-sectional studies show that the CMFs are lower for the roadway segments with shoulder rumble strips and wider shoulder width. This indicates that shoulder rumble strips and widening shoulder width will reduce crash frequencies. The CMFs for shoulder rumble strips estimated using cross-sectional method and before-after studies were similar (only 8% difference) for All crashes and SVROR.

The results of before-after studies show that the safety effects of the two single treatments and the combined treatment were higher for the roadway segments which originally had shorter shoulder width (4 ft ~ 6 ft) in the before period. For All crashes (KABCO), the safety effects of multiple treatments was higher than the effects of single treatments for the segments with 4 ft ~ 6 ft original shoulder width, whereas the safety effects of multiple and single treatments were similar for the segments with 8 ft ~ 12 ft original shoulder width.

The safety effects of the combined treatment are different for different crash types, severity levels and original shoulder width. For all crashes (KABC), shoulder rumble strips were more effective than widening shoulder width for the roadway segments with 8 ft ~ 12 ft original shoulder width, whereas widening shoulder width was more effective for the roadway segments with 4 ft ~ 6 ft shoulder width. Although multiple treatments have generally higher safety effects than single treatments, their combined effects on injury crashes (KABC) were not significantly higher than the effects of single treatments for the segments with 8 ft ~ 12 ft original shoulder width.

Among the six existing methods of combining CMFs for single treatments, the HSM, Systematic Reduction of Subsequent CMFs, Apply only the most effective CMF, and Weighted average of multiple CMFs (Meta-Analysis) provide the most accurate estimates of the combined CMFs for multiple treatments. However, in general, the combined CMFs were under-estimated for all crashes (KABCO) whereas they were over-estimated for injury crashes (KABC). Moreover, it can be concluded that the caution in the HSM about over-estimation of safety effects of multiplying multiple CMFs is valid since the results of Method 1 were mostly over-estimated.

While the results of this study provide empirical evidence of the combined safety effects of multiple treatments, more work is required to further develop the CMFs, CMFuntions, and alternative combining methods. In particular, sufficient sample size and low variances in safety effects of each single treatment are critical for determining reliable CMFs for multiple treatments. As demonstrated in this study, it is recommended that the safety effects of multiple treatments be separately estimated for different crash types, severity levels, and roadway characteristics. Further investigation is needed to identify the reason why the existing methods of combining CMFs for single treatments consistently under- or over-estimate actual CMFs for multiple treatments for a given crash type and severity level. Finally, more in-depth analysis is needed to determine the geometric conditions where multiple treatments are more safety effective than single treatments.

CHAPTER 4: DEVELOPMENT OF ADJUSTMENT FACTORS AND FUNCTIONS TO ASSESS COMBINED SAFETY EFFECTS

4.1 Introduction

In the previous chapter, it was suggested to average the best two existing combining methods to estimate more reliable combined safety effects. Although the estimated combined effects from averaging can improve accuracy, there is still difference between combined and actual safety effects for multiple treatments.

Thus, the objectives of this chapter are 1) to evaluate safety effects of four single treatments (adding bike lanes, conversion 4-lane to 3-lane, installing shoulder rumble strips, widening shoulder width) and two combined treatment (adding bike lanes + conversion 4-lane to 3-lane (i.e. road diet), installing shoulder rumble strips + widening shoulder width) using before-after with EB and cross-sectional studies, 2) to develop adjustment factors by comparison of the combined safety effects of multiple treatments using the HSM combining method with actual calculated CMFs for multiple treatments, and 3) develop the adjustment functions to assess the combined safety effects of multiple treatments. From this comparison, the study will show whether the existing HSM combining method for multiple treatments over- or under-estimates actual CMFs based on different crash types and severities.

In this study, crash types and severity levels are referred to ‘All crash types (all severities)’ as All crashes (KABCO), ‘All crash types (Fatal+Injury)’ as All crashes (KABC), ‘Single vehicle run-off roadways crashes (all severities)’ as SVROR (KABCO), and ‘Single vehicle run-off roadways crashes (Fatal+Injury)’ as SVROR (KABC). Moreover, the treatments are categorized as follow: 'installing shoulder rumble strips' as SRS, 'widening (1~9 ft) shoulder width' as WSW, 'installing shoulder rumble strips + widening (1~9 ft) shoulder width' as SRS+WSW, 'adding bike lanes' as Bike lane, 'conversion 4-lane to 3-lane roadways with TWLTL' as (Lane reduction) and 'adding bike lanes + conversion 4-lane to 3-lane roadways with TWLTL' as (Road diet).

4.2 Data Preparation

For the analysis of using before-after EB method, the road geometry data for roadway segments were identified for 8 years (2004-2011), and for consistency of all treated sites, crash records were collected for 2 years (2004-2005) for before period and 2 years (2010-2011) for after period from multiple sources maintained by the FDOT. These include the RCI and CARS database. The RCI database provides current and historical roadway characteristics data and reflects features of specific segments for the selected dates.

The three types of treatments, which are SRS, WSW and combination of the two treatments (SRS+WSW), were identified from the RCI roadway segments data for locations which have been treated in the years between 2006 and 2009 to ensure sufficient sample size. In this study, each roadway segment has uniform geometric characteristics in before and after periods except three types of treatments and AADT. A segment is represented by roadway identification numbers and beginning and end mile points. An average of AADT in 2004-2005 and 2010-2011 was used for analysis. The total lengths of treated rural two-lane segments for SRS, WSW and SRS+WSW were 61.274, 180.259, and 30.465 miles long, respectively. The total numbers of treated segments for SRS, WSW, and SRS+WSW were 70, 243 and 68, respectively. Also, the reference sites that have similar roadway characteristics to the treated sites in the before period were identified using the RCI database. A total of 2745 roadway segments with 1915.451 miles in length were identified as reference sites.

The crash records in the CARS for the 2-year before and 2-year after periods were linked to the RCI and the averaged AADT data. Many previous studies have found that traffic crashes and economic status or income levels are correlated (Noland, 2003; Romano et al., 2006; Males, 2009; Huang et al., 2010; Abdel-Aty et al., 2013) and the studies suggested using demographic and socio-economic parameters to determine their effects on traffic crashes. However, since the main purpose of this study is to estimate the safety effects of single and multiple treatments, crash data for years of 2006 to 2009 was not used in the analysis 1) to overcome a limitation of reflecting the economic changes due to the economy's slow down in the U.S. during this period, and 2) to avoid the immediate periods before and after the treatments.

For the analysis of using cross-sectional method, the road geometry data and crash records for roadway segments were collected for 10 years (2003-2012) from RCI and CARS database. Table 4-1 summarizes the data for the analysis using EB and cross-sectional methods. The AADT ranges of roadway segments are '1,200 ~ 25,000 veh/day' and '2,000 ~ 50,000 veh/day' for rural two-lane roadways and urban four-lane arterials, respectively. Distributions of each variable among the treated segments for EB analysis are summarized in Table 4-2.

Table 4-1: Summary of data description for EB and cross-sectional methods

Roadway Type	Treatment	Crash Records		Treated Sites		Reference Sites for SPFs	
		Before	After	Number of Sites	Total Length (mile)	Number of Sites	Total Length (mile)
Rural 2-lane roadways	SRS	2004~2005	2010~2011	70	61.274	2745	1915.451
	WSW			243	180.259		
	SRS+WSW			68	30.465		
Roadway Type	Treatment	Crash Records		Treated Sites		Reference Sites for SPFs	
Urban 4-lane undivided arterials	Bike lane	2010~2012		98	11.787	344	104.864
	Lane reduction			219	77.032		
	Road diet			31	11.97		

Table 4-2: Descriptive statistics of treated segments for EB analysis

(a) Shoulder Rumble Strips (SRS)

	Crash frequency in before period				Crash frequency in after period			
Variable	Mean	S.D.	Min.	Max.	Mean	S.D.	Min.	Max.
Number of All (KABCO) crashes	3.686	6.502	0	31	2.814	5.234	0	28
Number of All (KABC) crashes	3.529	6.152	0	29	2.543	4.784	0	26
Number of SVROR (KABCO) crashes	0.929	1.697	0	8	0.600	1.082	0	5
Number of SVROR (KABC) crashes	0.814	1.582	0	8	0.500	0.913	0	4

Variables related to traffic and roadway geometric characteristics				
Variable	Mean	S.D.	Min.	Max.
AADT (veh/day) in before period	6901	4326	2286	19100
AADT (veh/day) in after period	7246	4121	3086	18500
Length (mile)	0.875	1.132	0.107	4.904
Surface width (ft)	24	0.341	22	26
Maximum speed limit (mph)	56.5	4.842	35	60
Original shoulder width	2ft = 6sites, 4ft = 19sites, 6ft = 24sites, 8ft = 7sites, 10ft = 7sites, 12ft = 7sites			

(b) Widening Shoulder Width (WSW)

	Crash frequency in before period				Crash frequency in after period			
Variable	Mean	S.D.	Min.	Max.	Mean	S.D.	Min.	Max.
Number of All (KABCO) crashes	2.414	5.035	0	31	1.729	3.878	0	24
Number of All (KABC) crashes	2.157	4.732	0	29	1.529	3.622	0	23
Number of SVROR (KABCO) crashes	0.429	1.303	0	9	0.257	0.695	0	4
Number of SVROR (KABC) crashes	0.357	1.155	0	8	0.200	0.628	0	4

Variables related to traffic and roadway geometric characteristics				
Variable	Mean	S.D.	Min.	Max.
AADT (veh/day) in before period	5896	3882	1200	17500
AADT (veh/day) in after period	6140	4258	1600	18500
Length (mile)	0.673	0.907	0. 130	4.240
Surface width (ft)	23.771	0.935	18	24
Maximum speed limit (mph)	48.929	7.889	30	60
Original shoulder width	2ft = 9sites, 4ft = 8sites, 6ft = 33sites, 8ft = 43sites, 10ft = 96sites, 12ft = 54sites			

(C) Shoulder Rumble Strips + Widening Shoulder Width (SRS+WSW)

	Crash frequency in before period				Crash frequency in after period			
Variable	Mean	S.D.	Min.	Max.	Mean	S.D.	Min.	Max.
Number of All (KABCO) crashes	1.882	2.657	0	11	1.235	1.838	0	10
Number of All (KABC) crashes	1.750	2.588	0	11	1.088	1.646	0	9
Number of SVROR (KABCO) crashes	0.529	0.872	0	4	0.294	0.459	0	1
Number of SVROR (KABC) crashes	0.441	0.780	0	3	0.221	0.418	0	1
Variables related to traffic and roadway geometric characteristics								
Variable	Mean		S.D.		Min.		Max.	
AADT (veh/day) in before period	7566		5350		1650		23500	
AADT (veh/day) in after period	7145		5308		1350		25000	
Length (mile)	0.448		0.744		0.120		4.690	
Surface width (ft)	23.882		1.420		20		32	
Maximum speed limit (mph)	53.529		10.653		30		65	
Original shoulder width	2ft = 7sites, 4ft = 8sites, 6ft = 6sites, 8ft = 12sites, 10ft = 7sites, 12ft = 28sites							

4.3 Methodology

4.3.1Safety Performance Functions

Four full SPFs were developed using the NB model for four combinations of crash type and severity levels: 1) All crashes (KABCO), 2) All crashes (KABC), 3) SVROR (KABCO), and 4) SVROR (KABC) using 2-year before and 2-year after crash data. The SPFs were developed for reference sites of rural two-lane roadways in Florida using the NLMIXED procedure in the SAS program (SAS Institute, 2009) as shown in Table 4-3. To reflect the nonlinear relationship between AADT and crash frequency, logarithm of AADT was used instead of AADT (Wong et al. 2007; Abdel-Aty and Haleem, 2011; Park et al., 2014). In general, the results of four full SPFs show that crash frequency is higher for the roadway segments with higher AADT and longer length. It is worth noting that the crash frequency in the after period is lower than the before period for both All and SVROR crashes and this trend is consistent with the declining trend of traffic crashes over the last eight years (2004~2011) in the United States (NHTSA, 2013). Since this declining trend of traffic crashes is not only based on AADT, one explanatory variable (i.e. Time Difference) is included in the model to account for time difference between before and after periods. For example, the difference between predicted crash counts for before and after periods are mostly based on AADT changes even when simple or full SPF is applied since we assume there is no geometric changes (i.e. treatment) during before and after periods except AADT. According to Schick (2009), some factors such as economic changes and driver behavior are related to crash frequency. In particular, economy is changing as time changes. Thus, the declining trend of traffic crashes based on time changes might not be captured using SPF without the time difference term when 1) AADT of before and after periods are similar and 2) the time gap between before and after periods is larger. In this study, AADT changes of before and after periods for two single treatments and combination are similar, and there is four years time gap between before and after periods to ensure enough sample size of treated sites.

Table 4-3: Calibrated SPFs for rural two-lane roadways by crash types and severities

Coefficient						AIC
	α Intercept	β_1 Log (ADT)	β_2 Time Difference (Before Period)	β_3 Surface Width (Total Lane Width)	c Dispersion coefficient	
Crash Type (Severity)	Estimate (P-Value)	Estimate (P-Value)	Estimate (P-Value)	Estimate (P-Value)	Estimate (P-Value)	
All (KABCO)	-16.0913 (<.0001)	0.9309 (<.0001)	0.1078 (0.0571)	0.3702 (<.0001)	-0.7693 (<.0001)	13944
All (KABC)	-16.6181 (<.0001)	0.8693 (<.0001)	0.1269 (0.0274)	0.3896 (<.0001)	-0.5623 (<.0001)	10722
SVROR (KABCO)	-14.2772 (<.0001)	0.3758 (<.0001)	0.1324 (0.0884)	0.4182 (<.0001)	-0.7034 (<.0001)	5139.9
SVROR (KABC)	-13.6972 (<.0001)	0.2740 (<.0001)	0.1832 (0.0549)	0.4114 (<.0001)	-1.1174 (<.0001)	3831.4

4.4 Results

4.4.1Evaluated CMFs and Developed Adjustment Factors

In order to estimate CMFs using cross-sectional method, a NB regression model for urban roadways was evaluated as shown in Table 4-4. The CMFs estimated using the observational before-after with EB and cross-sectional methods were presented in Table 4-5. The CMFs for Bike lane, Lane reduction and Road diet were calculated as $\exp(\beta_3)$, $\exp(\beta_4)$ and $\exp(\beta_5)$. It is worth to mention that the analyses for KABC severity level and other crash type (e.g. bike crashes) were also performed but the results of NB regression models were not significant due to low crash frequency. Therefore, the CMFs for Bike lane, Lane reduction and Road diet were calculated using cross-sectional method for All crashes (KABCO) only. Since the coefficient for Bike lane is significant only at 85%, it is recommended to use the estimated CMF for Bike lane to check general safety impact of treatment with statistically large variation.

Generally, the safety effects of SRS, WSW, SRS+WSW, Bike lane, Lane reduction, and Road diet were positive for All and SVROR crashes. Also, the safety effects of two combined treatments were higher than single treatments. Moreover, the CMFs for SVROR (KABCO) crashes are notably lower than the CMFs for All (KABCO) crashes for SRS, WSW and SRS+WSW. These results indicate that SRS, WSW and SRS+WSW are more effective in reducing SVROR crashes. It is worth to note that due to the low frequency of SVROR (KABC) crashes, the estimated CMFs are not significant at 90% confidence level. Although the CMFs that are not significant at 90% confidence level may not represent reliable safety effects of treatment statistically, it can be suggested to use of the insignificant CMFs to check the general impact of treatments with relatively large variation. It is worth to note that for SRS, WSW and

SRS+WSW are more effective to reduce KABCO than KABC crashes. To estimate adjustment factors to modify the combined safety effects of multiple treatments, the actual calculated CMFs of SRS+WSW were divided by the combined CMFs using the HSM procedure (multiply single CMFs to estimate combined safety effectiveness), as shown in Table 4. In general, the combined safety effects using the HSM procedure were over-estimated by 4 to 10 percent for SRS and WSW whereas there was over-estimation by 2 percent for Bike lane and Lane reduction. This may be because SRS and WSW are implemented on same location (i.e. roadside) whereas Bike lane and Lane reduction are installed on different location (i.e. roadside and mainline). Moreover, the results imply that the adjustment factors can vary based on different crash types and severity levels. The results also indicate that it is better not to rely on the HSM combining method to predict CMF for multiple treatments, particularly when multiple treatments are implemented on same location. Thus, it can be recommended to develop adjustment factors to predict the combined safety effects of multiple treatments based on different 1) crash types and severity levels, and 2) implemented location of treatments.

Table 4-4: NB crash prediction model for urban arterials

	Coefficient						Dispersion (K)	Goodness of Fit	
	α Intercept	β_1 Log(AADT)	β_2 Segment Length	β_3 Bike Lane	β_4 Lane Reduction	β_5 Road Diet		Deviance	AIC
Crash Type (Severity)	Estimate (P-Value)	Estimate (P-Value)	Estimate (P-Value)	Estimate (P-Value)	Estimate (P-Value)	Estimate (P-Value)			
All Crashes (KABCO)	-7.9851 (<.0001)	1.0161 (<.0001)	1.0006 (<.0001)	-0.2473 (0.1489)	-0.6768 (<.0001)	-0.8889 (0.0025)	1.7902	754.6141	3922

Table 4-5: Evaluated CMFs and developed adjustment factors

	Shoulder Rumble Strips (SRS)		Widening Shoulder Width (WSW)		Shoulder Rumble Strips + Widening Shoulder Width (SRS+WSW) [a]		SRS × WSW (HSM method) [b]	Adjustment Factor (a/b)
Crash Type (Severity)	CMF	S.E	CMF	S.E	CMF	S.E	Combined CMF	
All (KABCO)	0.83**	0.07	0.87**	0.05	0.75**	0.10	0.72	1.05
All (KABC)	0.84*	0.08	0.89**	0.06	0.78*	0.11	0.75	1.04
SVROR (KABCO)	0.75*	0.14	0.82*	0.10	0.68*	0.17	0.62	1.10
SVROR (KABC)	0.80	0.16	0.87	0.12	0.75	0.21	0.70	1.08
	Bike Lane		Lane Reduction		Road Diet (Bike Lane + Lane Reduction) [a]		Bike Lane × Lane Reduction (HSM method) [b]	Adjustment Factor (a/b)
All (KABCO)	0.78*	0.04	0.51**	0.07	0.41**	0.12	0.40	1.02

**: significant at a 95% confidence level, *: significant at a 90% confidence level

4.4.2 Developed CMFunctions

Generally, the variation of CMFs with different roadway characteristics among treated sites is ignored because the CMF is a fixed value that represents overall safety effects of the treatment for all treated sites. Thus, the crash modification functions (CMFunctions) have been utilized to determine the relationship between the safety effects and roadway characteristics (Elvik, 2005; 2009; 2011; Park et al, 2014; Sacchi et al, 2014; Park et al, 2015). The CMFunctions of SRS, WSW and SRS+WSW were also developed in order to observe the general relationships between CMFs and the original shoulder width of roadway segments in the before period. The CMFs were estimated for the treated sites with different shoulder widths and used to develop CMFunctions. The range of standard errors of CMFs for different shoulder width was 0.05 to 0.3, but the standard errors were less than 0.2 for most of CMFs. The HSM suggests that a standard error of 0.1 or less indicates that the CMF value is sufficiently accurate, precise, and stable. Also, for treatments that have CMFs with a standard error of 0.1 or less, 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. Due to low frequency of SVROR (KABC) crashes, the CMFuntions were developed for All crashes and SVROR (KABCO). Twelve linear and nonlinear regression functions (Table 4-6) were compared and the best fitted function was identified based on the adjusted R-squared value. To ensure that the CMF value from CMFunction cannot be negative estimate, log form of linear and nonlinear models were utilized (Sacchi and Sayed, 2014). It was found that linear and two nonlinear functional forms (power, power 2) are the best fitted functions for this relationship.

Table 4-6: Log linear and nonlinear functional forms

Function Name	Equation
Linear	$Ln(Y) = A + (B_1 \cdot X)$
Inverse	$Ln(Y) = A + (B_1 / X)$
Exponential	$Ln(Y) = A + exp(B_1 \cdot X)$
Log	$Ln(Y) = A + (B_1 \cdot log\ X)$
Power	$Ln(Y) = A + (X^{B_1})$
Power 2	$Ln(Y) = A + (X^{B_1}) + (X^{B_2})$
Quadratic	$Ln(Y) = A + (B_1 \cdot X) + (B_2 \cdot X^2)$
Polynomial	$Ln(Y) = \{(B_1 \cdot X) + (B_2 \cdot X^2) + (B_3 \cdot X^3)\} \times \exp(B_4 \cdot X)$
Polynomial 2	$Ln(Y) = \{A + (B_1 \cdot X) + (B_2 \cdot X^2)\} \times \exp(B_4 \cdot X)$
Power_Exponential	$Ln(Y) = \{(B_1 \cdot X) + (X^{B_2})\} \times \exp(B_4 \cdot X)$
Power_Exponential 2	$Ln(Y) = \{A + (X^{B_1})\} \times \exp(B_2 \cdot X)$
Power_Exponential 3	$Ln(Y) = \{A + (X^{B_1}) + (X^{B_2})\} \times \exp(B_3 \cdot X)$

Tables 4-7, 4-8, and 4-9 present the developed CMFunctions of SRS, WSW and SRS+WSW for All (KABCO), All (KABC) and SVROR (KABCO), respectively. In this study, the CMFunction is defined as the function of original shoulder width of roadway segments for the CMF. In other words, Y and X represent the CMF and original shoulder width in each CMFunction. The relationship between CMFs and the original shoulder width indicates that the safety effects of two single treatments and combination are higher for the segments with narrower shoulder width. In other words, crash frequencies are more likely to decrease if the treatment is applied to the segments with narrower shoulder width. Moreover, for both All (KABCO) and All (KABC) crashes, SRS is more safety effective for roadway segments with shoulder width of 10ft or above and 9.5ft or above, whereas WSW is more safety effective for roadway segments with shoulder width less than 10ft and 9.5ft. Park et al (2014) found similar trends for the two single treatments and combination on rural multilane roadways for All (KABCO). The study reported that for All crashes (KABCO), widening shoulder width is more effective for roadway segments with shoulder width less than 7ft, whereas shoulder rumble strips are more effective for roadway segments with shoulder width of 7ft or above. It was also found that for SVROR (KABCO) crashes, SRS is more safety effective for roadway segments with shoulder width of 7.5ft or above, whereas WSW is more safety effective for roadway segments with shoulder width less than 7.5ft. It is worth to note that the difference between CMFs of two single treatment and CMFs for multiple treatments is getting larger as shoulder width decreases for both All and SVROR crashes. The results indicate that the safety effects of multiple treatments vary based on characteristics of roadway segments. Figures 4-1, 4-2, and 4-3 provide the comparison of CMFunctions of each treatment for All (KABCO), All (KABC) and SVROR (KABCO), respectively.

Table 4-7: Developed CMFunctions for All crashes (KABCO)

(a) Shoulder Rumble Strips (SRS)

Functional Form = Power				
Parameter	Coefficient	Standard error	t-value	p-value
A	-1.3469	0.0186	-72.29	<.0001
B_1	0.0782	0.0084	9.36	0.0007
Root Mean Squared Error (Root_MSE) = 0.0158 R-Square = 0.9450 Adj. R-Square = 0.9313				

(b) Widening Shoulder Width (WSW)

Functional Form = Linear				
Parameter	Coefficient	Standard error	t-value	p-value
A	-0.4223	0.0272	-15.55	<.0001
B_1	0.0275	0.0035	7.90	0.0014
Root Mean Squared Error (Root_MSE) = 0.0292 R-Square = 0.9398 Adj. R-Square = 0.9247				

(c) Shoulder Rumble Strips + Widening Shoulder Width (SRS+WSW)

Functional Form = Power				
Parameter	Coefficient	Standard error	t-value	p-value
A	-1.7575	0.0397	-44.23	<.0001
B_1	0.1902	0.0140	13.60	0.0002
Root Mean Squared Error (Root_MSE) = 0.0370 R-Square = 0.9639 Adj. R-Square = 0.9549				

Table 4-8: Developed CMFunctions for All crashes (KABC)

(a) Shoulder Rumble Strips (SRS)

Functional Form = Power 2				
Parameter	Coefficient	Standard error	t-value	p-value
A	-2.2562	0.0169	-133.75	<.0001
B_1	0.1780	0.0097	18.35	0.0004
B_2	-0.2080	0.0337	-6.16	0.0086
Root Mean Squared Error (Root_MSE) = 0.0054 R-Square = 0.9951 Adj. R-Square = 0.9918				

(b) Widening Shoulder Width (WSW)

Functional Form = Linear				
Parameter	Coefficient	Standard error	t-value	p-value
A	-0.4917	0.0375	-13.11	0.0002
B_1	0.0370	0.0048	7.68	0.0015
Root Mean Squared Error (Root_MSE) = 0.0403 R-Square = 0.9365 Adj. R-Square = 0.9206				

(c) Shoulder Rumble Strips + Widening Shoulder Width (SRS+WSW)

Functional Form = Power				
Parameter	Coefficient	Standard error	t-value	p-value
A	-1.8010	0.0475	-37.94	<.0001
B_1	0.2093	0.0160	13.05	0.0002
Root Mean Squared Error (Root_MSE) = 0.0449 R-Square = 0.9589 Adj. R-Square = 0.9487				

Table 4-9: Developed CMFunctions for SVROR crashes (KABCO)

(a) Shoulder Rumble Strips (SRS)

Functional Form = Power				
Parameter	Coefficient	Standard error	t-value	p-value
A	-1.5106	0.0182	-83.06	<.0001
B_1	0.1110	0.0076	14.61	0.0001
Root Mean Squared Error (Root_MSE) = 0.0159 R-Square = 0.9746 Adj. R-Square = 0.9682				

(b) Widening Shoulder Width (WSW)

Functional Form = Linear				
Parameter	Coefficient	Standard error	t-value	p-value
A	-0.5390	0.0344	-15.67	<.0001
B_1	0.0362	0.0044	8.20	0.0012
Root Mean Squared Error (Root_MSE) = 0.0369 R-Square = 0.9439 Adj. R-Square = 0.9298				

(c) Shoulder Rumble Strips + Widening Shoulder Width (SRS+WSW)

Functional Form = Power				
Parameter	Coefficient	Standard error	t-value	p-value
A	-2.0666	0.0505	-40.96	<.0001
B_1	0.2467	0.0157	15.70	<.0001
Root Mean Squared Error (Root_MSE) = 0.0490 R-Square = 0.9684 Adj. R-Square = 0.9605				

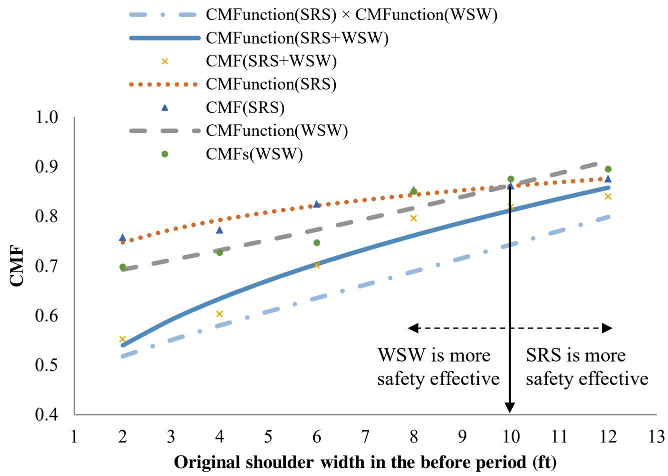


Figure 4-1: Comparison of CMFunctions for SRS, WSW, and SRS+WSW for All crashes (KABCO) with different original shoulder width in the before period

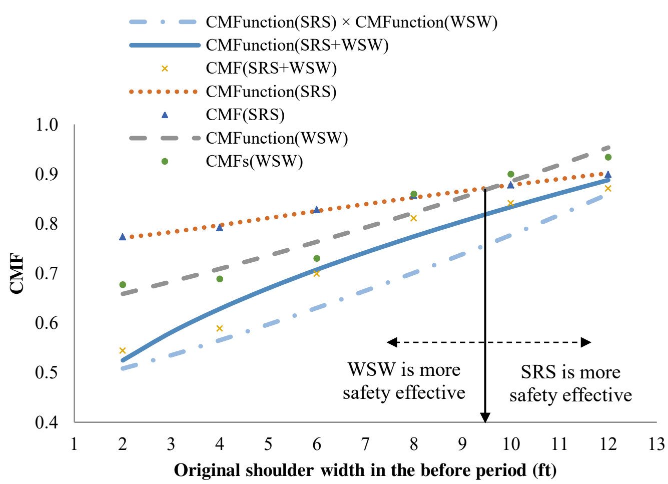


Figure 4-2: Comparison of CMFunctions for SRS, WSW, and SRS+WSW for All crashes (KABC) with different original shoulder width in the before period

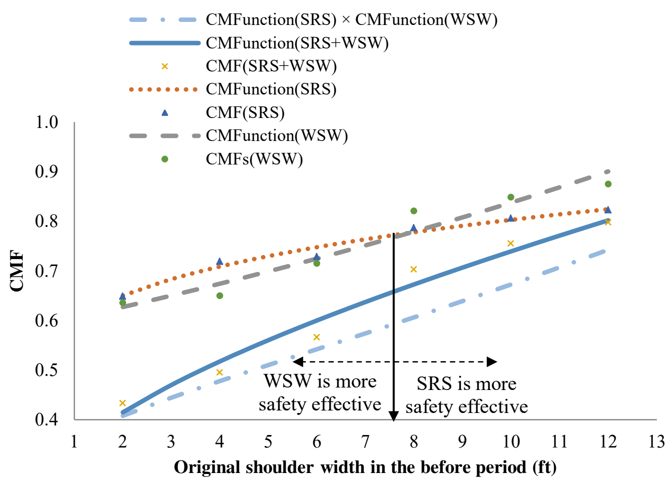


Figure 4-3: Comparison of CMFunctions for SRS, WSW, and SRS+WSW for SVROR crashes (KABCO) with different original shoulder width in the before period

4.4.3 Development of Adjustment Functions

In Figures 4-1, 4-2, and 4-3, the combined safety effects of SRS and WSW (i.e. CMFunction of SRS × CMFunction of WSW) are presented for All (KABCO), All (KABC) and SVROR (KABCO). It is worth to note that the combined safety effects are mostly over-estimated compared to actual CMFunction of SRS+WSW. Moreover, the difference between combined safety effects and actual CMFunction of SRS+WSW (i.e. adjustment factors) shows nonlinear relationship as original shoulder width changes. In particular, the results of All crashes showed that the difference between combined safety effects and actual estimated CMFs are larger as the shoulder width increases for roadway segments with shoulder width less than 6ft. However, the opposite effects were found as the difference between combined safety effects and actual estimated CMFs are smaller as the shoulder width increases for roadway segments with shoulder width of 6ft or above. The results also showed that for SVROR (KABCO) crashes, the difference between combined safety effects and actual estimated CMFs are larger as shoulder width increases for roadway segments with shoulder width less than 7ft, whereas the difference between combined safety effects and actual estimated CMFs are smaller as shoulder width increases for roadway segments with shoulder width of 7ft or above. Therefore, the adjustment functions were developed for All crashes and SVROR (KABCO) to determine this nonlinear relationship. Nonlinear regression functions from Table 4 were compared and the best fitted function was identified based on adjusted R-squared value. It was found that polynomial nonlinear regression models are the best fitted functions for this relationship. Table 4-10 presents the developed nonlinear adjustment functions to modify combined safety effects of SRS and WSW for different crash types and severities. In this study, the adjustment function is defined as the function of original shoulder width of roadway segments for the adjustment factor. In other words, Y and X represent the adjustment factor and original shoulder width in each adjustment function, respectively.

Table 4-10: Estimated nonlinear adjustment functions to modify combined effect of SRS and WSW

(a) All crashes (KABCO)

Functional Form = Polynomial 2				
Parameter	Coefficient	Standard error	t-value	p-value
A	-0.0547	0.0050	-10.96	<.0001
B_1	0.0594	0.0025	24.15	<.0001
B_2	-0.0023	0.0002	-12.61	<.0001
B_3	-0.1242	0.0118	-10.52	<.0001
Root Mean Squared Error (Root_MSE) = 0.0010 R-Square = 0.9883 Adj. R-Square = 0.9876				

(b) All crashes (KABC)

Functional Form = Polynomial				
Parameter	Coefficient	Standard error	t-value	p-value
B_1	-0.0388	0.0162	-2.40	0.0476
B_2	0.0409	0.0082	5.00	0.0016
B_3	-0.0023	0.0003	-8.77	<.0001
B_4	-0.3043	0.0371	-8.19	<.0001
Root Mean Squared Error (Root_MSE) = 0.0046 R-Square = 0.9849 Adj. R-Square = 0.9785				

(c) SVROR crashes (KABCO)

Functional Form = Polynomial 2				
Parameter	Coefficient	Standard error	t-value	p-value
A	-0.1047	0.0056	-18.67	<.0001
B_1	0.0692	0.0023	29.86	<.0001
B_2	-0.0027	0.0002	-12.27	<.0001
B_3	-0.1225	0.0115	-10.64	<.0001
Root Mean Squared Error (Root_MSE) = 0.0010 R-Square = 0.9889 Adj. R-Square = 0.9885				

4.5 Conclusion

Although the HSM caution that the assumption of independence of different treatments can lead to over- or under- estimation of actual safety impact of multiple CMFs, there was a lack of studies that assess the combined safety effects of multiple treatments. Therefore, the main objective of this study is to comprehensively evaluate the safety effects of four single treatments and two combined treatments based on location of treatment and roadway types. The study calculated actual CMFs for SRS+WSW and Bike lane + Lane reduction and also estimated combined CMFs using HSM procedure. The CMFs were calculated using observational before-after with EB and cross-sectional methods.

The results of estimated CMFs indicate that four single treatments and two combined treatments will reduce crash frequencies. In particular, the estimated CMFs show higher safety effects on KABCO crashes than KABC. Moreover, the CMFs for SVROR (KABCO) crashes are notably lower than the CMFs for All (KABCO) crashes. These results indicate that SRS, WSW and SRS+WSW are more effective in reducing SVROR crashes. Also, it is worth noting that the safety effects of two combined treatments were higher than single treatments.

In order to adjust the combined CMFs for multiple treatments by the HSM combining procedure, the adjustment factors were estimated by comparison of actual calculated CMFs and the combined CMFs for SRS + WSW and Bike lane + Lane reduction. Generally, the combined safety effects using the HSM procedure were over-estimated by 4 to 10 percent for SRS and WSW, and 2 percent for Bike lane and Lane reduction. This may be because SRS and WSW are implemented on same location (i.e. roadside) whereas Bike lane and Lane reduction are installed on different location (i.e. roadside and mainline).

Moreover, the results indicate that the adjustment factors can vary based on different crash types and severity levels. Therefore, it is recommended to develop and apply adjustment factors to predict the combined safety effects of multiple treatments based on 1) different crash types and severity levels, and 2) implemented location of treatments. In particular, the combined safety effects need to be adjusted when multiple treatments are implemented on same location. It can be concluded that the caution in the HSM about over-estimation of safety effects of multiplying multiple CMFs is valid since the results of combined CMFs were over-estimated in this study.

The results of developed CMFunctions indicate that the safety effects of two single treatments and combination are higher for the segments with narrower shoulder width. Also, SRS is more safety effective for roadway segments with shoulder width of 10ft or above and 9.5ft or above, whereas WSW is more safety effective for roadway segments with shoulder width less than 10ft and 9.5ft for All crashes (KABCO) and All crashes (KABC). The results also showed that SRS is more safety effective for roadway segments with shoulder width of 7.5ft or above, whereas WSW is more safety effective for roadway segments with shoulder width less than 7.5ft for SVROR (KABCO) crashes. The difference between CMFs of two single treatment and CMFs for multiple treatments is getting larger as shoulder width decreases for both All and SVROR crashes. The results indicate that the safety effects of multiple treatments vary based on characteristics of roadway segments. For the relationship between the CMFs and original shoulder width of treated sites, linear regression and nonlinear regression with power functional form models are the best fitted functions.

In this study, to determine the nonlinear relationship of the difference between combined safety effects and actual estimated CMFs, the adjustment functions were developed using nonlinear regression models. Generally, the combined safety effects are over-estimated compared to actual estimated CMFs for multiple treatments. It is worth to point out that the amount of over-estimation showed nonlinear shape for both All and SVROR crashes. In particular, it was found that for All crashes, the difference between the combined safety effect and the actual estimated CMFs are larger as shoulder width increases for roadway segments with shoulder width less than 6ft, whereas the difference between combined safety effects and actual estimated CMFs are smaller as shoulder width increases for roadway segments with shoulder width of 6ft or above. It was also found that for SVROR (KABCO) crashes, the difference between combined safety effect and actual estimated CMFs are larger as shoulder width increases for roadway segments with shoulder width less than 7ft, whereas the difference between the combined safety effect and actual estimated CMFs are smaller as shoulder width increases for roadway segments with shoulder width of 7ft or above. It was found that nonlinear regression models with polynomial functional form are the best fitted functions to adjust the combined safety effects of multiple treatments.

Although the results of this study provide empirical evidence of the combined safety effects of multiple treatments, the study has some limitations and more work is required to further develop alternative way to adjust combined safety effects. In particular, sufficient sample size and low variances in safety effects of each single treatment are critical for determining reliable CMFs and CMFunctions for multiple treatments. Also, including multiple target areas (e.g. more states, countries) in the analysis may produce more generalized conclusions. More in-depth analysis is also needed to determine the geometric conditions where multiple treatments are more safety effective than single treatments. Further investigation is needed to identify the reason why the HSM method of combining CMFs mostly over-estimates actual CMFs for multiple treatments 1) for different combination of treatments, 2) for a given crash type and severity level, and 3) for a location of treatments.

As the HSM provides various CMFs from previous studies using data of specific states or locations, the results of this study may be applicable to other states or countries. However, it is recommended to check the similarity of the target state or location to Florida conditions. In particular, the characteristics of roadways (e.g. AADT range, roadway type, shoulder width range, etc.) and crash data (crash types, severity levels and scales, etc.) of the target state or location need to be similar to the characteristics of Florida. Lastly, since this study focuses on specific treatments (i.e. SRS, WSW, SRS+WSW), the estimated CMFunctions and adjustment functions may not be generalizable to other treatments.

CHAPTER 5: EVALUATE VARIATION OF CRASH MODIFICATION FACTORS FOR DIFFERENT CRASH CONDITIONS

5.1 Introduction

From the previous chapters, it was shown that the safety effects of specific treatments have variations based on different roadway characteristics among treated sites. In this chapter, the CMFs were developed for different crash types and severities with different crash conditions to identify changes of the safety effects. The main objectives of this study are 1) to estimate CMFs for the installation of different types of roadside barriers, and 2) to determine the changes of safety effects for different crash types and severities based on different vehicle, driver, weather and time of day information. Two observational before-after analyses (i.e. EB and FB approaches) were utilized in this study to estimate CMFs. To consider the variation of safety effects based on different vehicle, driver, weather, and time of day information, the crashes were categorized based on vehicle size (passenger and heavy), driver age (young, middle, and old), weather condition (normal and rain), and time difference (day time and night time). It is known that the EB approach has been the most common and rigorous approach to perform observational before-after evaluations in the last two decades (Gross et al., 2010; Ahmed et al., 2015). On the other hand, with the advancement in statistical modeling techniques and computing capabilities, adopting the FB approach has been utilized recently (Aul and Davis, 2006; Pawlovich et al., 2006; Li et al., 2008a; Lan et al., 2009; Persaud et al., 2009; El-Basyouny and Sayed, 2010; 2011; 2012a; 2012b). In this chapter, crash types and severity levels are referred to 'All crash types' as All crashes' and 'run-off roadways crashes' as ROR crashes.

5.2 Data Preparation

The road geometry data for roadway segments were obtained for 9 years (2003-2011) from the database of RCI. In order to identify the treated sites on freeways, the financial management system was used. The financial management system offers a searching system named financial project search. This system provides detailed information on a specific financial project such as district number, status, work type, and year.

A total of 147 freeway segments totaling 68.168 miles were identified as treated sites with installation of roadside barriers during 2007. A segment is represented by roadway identification numbers, and beginning and end mile points. It was found that among the 147 treated sites, w-beam guardrails were implemented on 127 sites and concrete barriers were installed on 20 sites. The crash records were obtained from the CARS for the 4-year before (2003-2006) and 4-year after (2008-2011) periods. Also, the reference sites were identified using the RCI database. A total of 328 roadway segments with 119.899 miles in length were identified as reference sites. It is to be noted that reference sites are different than the comparison group; the reference sites are broader than the comparison group with more variation in AADT, roadway characteristics, and crash history in order to correct for the regression-to-the-mean threat. The FB approach integrates the EB two-step into one and hence, FB utilizes information from a reference group of sites and the before information from the treated sites to estimate the long-term expected crash frequency. Table 5-1 presents a summary of distributions of each variable for the treated segments along with crash frequency.

Table 5-1: Descriptive statistics of treated segments

(a) Roadway characteristics

Variables related to traffic and roadway geometric characteristics				
Variable	Mean	S.D.	Min.	Max.
AADT (veh/day) in before period	59,834.014	15,436.665	36,500	104,600
AADT (veh/day) in after period	56,636.735	14,903.484	35,000	104,200
Length (mile)	0.464	0.398	0.103	3.007
Numbers of lane	2.265	0.645	2	5
Surface width (ft)	27.184	7.734	24	60
Shoulder width (ft)	10.122	1.517	4	20
Median width (ft)	34.293	10.619	20	65
Curvature (Radius/5730ft)	0.468	0.802	0	3.05
Maximum speed limit (mph)	66.224	5.692	50	70
Distance to roadside barriers	13.272	3.493	9	30
Roadside barrier type	W-beam guardrails = 127sites, Concrete barrier = 20sites			

(b) Crash frequency

		Crash frequency in before period					Crash frequency in after period				
Crash Type	Severity	Mean	S.D.	Min.	Max.	Total	Mean	S.D.	Min.	Max.	Total
All crashes	KABCO	17.415	17.462	0	84	2,560	16.048	16.046	0	80	2,359
	KABC	8.497	8.803	0	48	1,249	7.204	7.544	0	43	1,059
	KAB	4.286	4.509	0	26	630	3.184	3.643	0	26	468
ROR crashes	KABCO	5.367	6.058	0	36	789	4.544	5.262	0	26	668
	KABC	2.925	3.302	0	17	430	2.231	2.669	0	14	328
	KAB	1.612	2.015	0	12	237	1.088	1.380	0	7	160

5.3 Methodology

5.3.1 Safety Performance Functions

In order to estimate the Florida-specific full SPFs for freeways, crash data of both before and after periods for the reference sites were used with a time difference term. However, the variable of time difference was not significant which indicates that there is no significant difference between the before and after periods under no treatment condition. Also, it is worth to note that the SPFs were evaluated using segment length as an offset. However, the SPFs using segment length as a variable show better model fitness. The SPFs were developed for different crash types and severity levels. Also, the SPFs were developed based on different vehicle, driver, weather, and time information. To consider the variation of safety effects based on different information, the crashes were categorized based on vehicle size (passenger and heavy), driver age (young, middle, and old), weather condition (normal and rain), and time difference (day time and night time).

5.3.2 Full Bayes Method

Generally, it is known that the FB approach provided comparable results and might have several advantages over the EB technique as follow: 1) FB models account for the uncertainty associated with parameter estimates and provide exact measures of uncertainty on the posterior distributions of these parameters and hence overcome the maximum likelihood methods' problem of overestimating precision because of ignoring this uncertainty; 2) valid crash models can be estimated using small sample size because of the FB properties, which might be the case of most of road safety benefit analyses; 3) Bayesian inference can effectively avoid the problem of over fitting that occurs when the number of observations is limited and the number of variables is large (3). In the before-after framework, the FB method integrates the EB two-steps into one by calculating the odds ratio and the SPFs into a single step, and hence, integrating any error or variance of the estimated regression coefficient into the final estimates of the safety effectiveness of a treatment. Most importantly, the flexibility of a FB formulation allows for different model specifications which have the capability of accounting for various levels of correlation.

5.4 Results

5.4.1 CMFs for Different Crash Types and Severities using EB and FB Methods

In order to estimate CMFs using the observational before-after with EB method, six full SPFs were developed by the NB model as shown in Table 5-2. Moreover, Table 5-3 presents the evaluated Bayesian Poisson-lognormal models for FB analyses. In general, the results of the full SPFs and the developed Bayesian Poisson-lognormal models show that crash frequency is higher for the roadway segments with higher AADT and longer length. The results also show that the crash frequency is lower for the roadways with wider shoulder and median widths.

Table 5-2: Estimated parameters of SPFs by NB method for All and ROR crashes

Crash Type	Severity	Intercept (p-value)	Segment length (p-value)	Log AADT (p-value)	Shoulder width (p-value)	Median width (p-value)	Maximum Speed (p-value)	Dispersion (k)	Deviance	AIC
All crashes	KABCO	-13.9584 (<.0001)	1.6937 (<.0001)	1.6798 (<.0001)	-0.0360 (0.0304)	-0.0034 (0.0010)	-0.0364 (0.0014)	0.4408	716.4	4086.9
	KABC	-16.8558 (<.0001)	1.6259 (<.0001)	1.6796 (<.0001)	-0.0405 (0.0237)	-0.0029 (0.0066)	-	0.4102	719.1	3448.7
	KAB	-14.9333 (<.0001)	1.5983 (<.0001)	1.4368 (<.0001)	-0.0446 (0.0284)	-	-	0.3918	699.4	2760.6
ROR crashes	KABCO	-13.7554 (<.0001)	1.3730 (<.0001)	1.3902 (<.0001)	-0.0915 (<.0001)	-0.0039 (0.0756)	-	0.4697	705.7	2696.8
	KABC	-13.8629 (<.0001)	1.3806 (<.0001)	1.3738 (<.0001)	-0.1013 (<.0001)	-0.0044 (0.0013)	-	0.4345	683.0	2284.0
	KAB	-14.5482 (<.0001)	1.4380 (<.0001)	1.3503 (<.0001)	-0.0932 (0.0004)	-	-	0.4341	646.5	1733.3

Table 5-3: Estimated parameters of Bayesian Poisson-lognormal models for All and ROR crashes

(a) All crashes

	KABCO			KABC			KAB		
	Mean (S.D)	Interval 2.5%	Interval 97.5%	Mean (S.D)	Interval 2.5%	Interval 97.5%	Mean (S.D)	Interval 2.5%	Interval 97.5%
Intercept	-12.1 (3.223)	-17.38	-5.741	-14.87 (1.655)	-17.02	-10.63	-15.01 (1.328)	-17.72	-12.68
Log AADT	1.308 (0.275)	0.7634	1.747	1.496 (0.141)	1.154	1.685	1.428 (0.1164)	1.237	1.666
Segment length	1.388 (0.1079)	1.169	1.589	1.424 (0.08565)	1.255	1.592	1.449 (0.08938)	1.279	1.629
Shoulder width	-0.06071 (0.02325)	-0.1088	-0.02302	-0.0485 (0.01833)	-0.0847	-0.01362	-0.03811 (0.02091)	-0.07888	0.00386
Median width	-0.00376 (0.00151)	-0.00697	-0.00103	-0.00275 (0.00123)	-0.00531	-0.00044	-	-	-
τ	1.914 (0.2287)	1.44	2.33	2.374 (0.2171)	1.969	2.821	2.527 (0.2817)	2.016	3.126
DIC	3599.54			3155.17			2609.43		

(b) ROR crashes

	KABCO			KABC			KAB		
	Mean (S.D)	Interval 2.5%	Interval 97.5%	Mean (S.D)	Interval 2.5%	Interval 97.5%	Mean (S.D)	Interval 2.5%	Interval 97.5%
Intercept	-13.83 (0.8021)	-15.14	-12.0	-13.73 (1.165)	-15.49	-10.81	-14.28 (1.528)	-17.21	-11.49
Log AADT	1.373 (0.07084)	1.213	1.498	1.342 (0.09969)	1.089	1.492	1.307 (0.1342)	1.06	1.558
Segment length	1.301 (0.09071)	1.119	1.476	1.309 (0.09571)	1.126	1.5	1.358 (0.1069)	1.151	1.569
Shoulder width	-0.08455 (0.0225)	-0.1278	-0.04032	-0.09776 (0.02398)	-0.1453	-0.05139	-0.0886 (0.02675)	-0.1399	-0.0364
Median width	-0.00383 (0.00132)	-0.00642	-0.00122	-0.00441 (0.00142)	-0.00722	-0.00168	-	-	-
τ	2.167 (0.242)	1.733	2.682	2.358 (0.3032)	1.825	3.005	2.476 (0.4538)	1.743	3.512
DIC	2524.65			2180.12			1692.16		

The CMFs estimated for different crash types and severity levels using the EB and FB methods were presented in Table 5-4. It should be noted that the CMFs were estimated for all types of roadside barriers (i.e. w-beam guardrails + concrete barriers) and w-beam guardrails only. Due to the low sample size of treated sites with concrete barriers, it was not possible to calculate the

CMFs for concrete barriers only. Generally, the safety effects of roadside barriers are positive and statistically significant for KAB severity level for both All and ROR crashes. The results show that roadside barriers are safety effective to reduce ROR (KABC) crashes whereas the CMFs are not statistically significant for All (KABC) crashes. Also, the estimated CMFs are statistically insignificant for KABCO except the CMF for w-beam guardrail from the EB method. The results show that the safety effectiveness of w-beam guardrails for All (KABCO) crashes is negative and this result is consistent with the HSM. This indicates that an addition of w-beam guardrails on roadside might increase crash frequency but reduce crash severity.

Overall, there are no big differences between the results of EB and FB methods. In particular, the standard errors of estimated CMFs by EB and FB methods are almost similar. This indicates that the results from the EB method are comparable to the FB method and this result is consistent with Persaud et al. (2009) and Ahmed et al. (2015). It is worth to mention that for the CMFs for installation of W-bean guardrails only, the result from EB method produces slightly better estimates (i.e. lower standard error) for ROR crashes. This indicates that although the FB method has several statistical advantages over the EB approach, the EB method might show more reliable estimates when 1) sufficient sample size of reference sites was obtained and used to calculate full SPFs, and 2) there are enough crash frequencies for both treated and reference sites. FB might have been advantageous if the sample size was smaller.

Table 5-4: Evaluated CMFs for all and ROR crashes using EB and FB methods

Crash type	Severity	CMFs from the EB method				CMFs from the FB method			
		Roadside Barriers (W-Beam + Concrete)		W-Beam Guardrail Only		Roadside Barriers (W-Beam + Concrete)		W-Beam Guardrail Only	
		CMF	S.E	CMF	S.E	CMF	S.E	CMF	S.E
All crashes	KABCO	1.04	0.03	1.09**	0.03	1.01	0.03	1.06	0.03
	KABC	0.96	0.04	1.01	0.04	0.94	0.04	0.99	0.04
	KAB	0.82**	0.05	0.85**	0.05	0.82**	0.05	0.84*	0.05
ROR crashes	KABCO	0.95	0.05	1.01	0.05	0.93	0.05	1.01	0.06
	KABC	0.84**	0.06	0.88*	0.06	0.84**	0.06	0.89	0.07
	KAB	0.74**	0.07	0.75**	0.08	0.73**	0.07	0.74*	0.08

**: significant at 95% confidence level, *: significant at 90% confidence level

5.4.2 Variation of CMFs with Different Crash Conditions

In order to identify the changes of CMFs, the full SPFs were developed for ROR crashes based on different vehicle, driver, weather, and time information as shown in Table 5-5. It should be noted that the CMFs with different information were calculated for ROR crashes only since roadside barriers were found to be more effective in reducing ROR crash frequency and severity than all crashes in the previous section. Moreover, the EB method was conducted due to its better estimates for analysis of ROR crashes in the previous section.

Table 5-5: Estimated parameters of SPFs by NB method for ROR crashes with different crash conditions

Crash Type	Severity	Intercept (p-value)	Segment length (p-value)	Log AADT (p-value)	Shoulder width (p-value)	Median width (p-value)	Maximum Speed (p-value)	Curve (R/5730ft) (p-value)	Dispersion (k)	Deviance	AIC
ROR passenger vehicle crashes	KABCO	-19.3427 (<.0001)	1.3188 (<.0001)	1.6311 (<.0001)	-0.0980 (<.0001)	-0.0027 (0.0649)	0.0391 (0.0710)	0.1566 (0.0311)	0.5230	697.8	2392.4
	KABC	-24.3237 (<.0001)	1.2537 (<.0001)	1.7642 (<.0001)	-0.0933 (0.0002)	-	0.0847 (0.0030)	-	0.4906	668.2	2005.9
	KAB	-26.3205 (<.0001)	1.2697 (<.0001)	1.7710 (<.0001)	-0.0611 (0.0399)	-	0.0992 (0.0065)	-	0.4239	607.1	1471.9
ROR heavy vehicle crashes	KABCO	-11.3263 (<.0001)	1.2216 (<.0001)	1.0493 (<.0001)	-0.0692 (0.0224)	-0.0072 (0.0002)	-	-	0.5076	600.9	1497.2
	KABC	-12.6849 (<.0001)	1.3048 (<.0001)	1.1699 (<.0001)	-0.1129 (0.0011)	-0.0066 (0.0035)	-	-	0.5639	526.7	1217.6
	KAB	-24.9431 (0.0007)	1.1369 (<.0001)	1.3792 (<.0001)	-0.1845 (<.0001)	-0.0053 (0.1030)	0.1513 (0.0185)	-	0.5658	423.4	841.3
ROR young age driver (15~24 years old) crashes	KABCO	-14.1884 (<.0001)	1.1546 (<.0001)	1.3293 (<.0001)	-0.1049 (<.0001)	-	-	-	0.2424	658.3	1629.5
	KABC	-26.8371 (<.0001)	1.0761 (<.0001)	1.6896 (<.0001)	-0.1114 (<.0001)	-	0.1264 (0.0010)	0.1630 (0.0817)	0.1758	608.7	1348.6
	KAB	-24.3044 (<.0001)	1.0713 (<.0001)	1.5270 (<.0001)	-0.0903 (0.0091)	-0.0039 (0.1132)	0.1073 (0.0272)	-	0.1036	541.9	985.7
ROR middle age driver (25~64 years old) crashes	KABCO	-14.9349 (<.0001)	1.3714 (<.0001)	1.4501 (<.0001)	-0.0885 (0.0003)	-0.0042 (0.0039)	-	-	0.5154	674.4	2204.8
	KABC	-22.2459 (<.0001)	1.3210 (<.0001)	1.6751 (<.0001)	-0.0954 (0.0004)	-0.0039 (0.0212)	0.0682 (0.0189)	-	0.5265	630.0	1843.5
	KAB	-15.5379 (<.0001)	1.4118 (<.0001)	1.3861 (<.0001)	-0.0856 (0.0101)	-	-	-	0.5887	561.7	1337.2
ROR old age driver (≥ 65 years old) crashes	KABCO	-21.3009 (<.0001)	1.3154 (<.0001)	1.7774 (<.0001)	-	-0.0133 (0.0003)	-	0.4557 (0.0014)	0.8739	359.3	730.8
	KABC	-25.1901 (<.0001)	1.5886 (<.0001)	2.0357 (<.0001)	-	-0.0094 (0.0530)	-	0.5391 (0.0038)	1.3116	244.8	475.7
	KAB	-30.3211 (<.0001)	1.3519 (<.0001)	2.4284 (<.0001)	-	-	-	-	0.6200	192.5	308.3
ROR crashes in day time	KABCO	-13.8290 (<.0001)	1.2474 (<.0001)	1.3459 (<.0001)	-0.0733 (0.0016)	-0.0030 (0.0293)	-	-	0.4836	700.5	2317.6
	KABC	-21.5279 (<.0001)	1.2149 (<.0001)	1.5952 (<.0001)	-0.0766 (0.0018)	-	0.0676 (0.0085)	-	0.3973	659.9	1941.4
	KAB	-20.9055 (<.0001)	1.1509 (<.0001)	1.4021 (<.0001)	-0.0471 (0.1067)	-	0.0767 (0.0173)	-	0.2364	622.3	1407.4
ROR crashes in night time	KABCO	-17.9102 (<.0001)	1.4484 (<.0001)	1.6618 (<.0001)	-0.1108 (<.0001)	-	-	-	0.5273	619.4	1672.5
	KABC	-22.4477 (<.0001)	1.3075 (<.0001)	1.7175 (<.0001)	-0.1238 (<.0001)	-0.0065 (0.0023)	0.0601 (0.1101)	-	0.3783	561.5	1315.9
	KAB	-20.7547 (<.0001)	1.4888 (<.0001)	1.8584 (<.0001)	-0.1529 (<.0001)	-	-	-	0.4710	464.6	959.7
ROR crashes in normal weather condition	KABCO	-19.5112 (<.0001)	1.3168 (<.0001)	1.4868 (<.0001)	-0.0552 (0.0124)	-0.0055 (0.0002)	0.0584 (0.0098)	-	0.3625	685.7	2107.0
	KABC	-22.2356 (<.0001)	1.3074 (<.0001)	1.5724 (<.0001)	-0.0683 (0.0054)	-0.0047 (0.0051)	0.0811 (0.0041)	-	0.3677	642.8	1781.8
	KAB	-25.5861 (<.0001)	1.3186 (<.0001)	1.6583 (<.0001)	-0.0745 (0.0135)	-	0.1071 (0.0038)	-	0.4104	571.9	1392.0
ROR crashes in rain condition	KABCO	-16.6552 (<.0001)	1.1959 (<.0001)	1.5939 (<.0001)	-0.1278 (<.0001)	-	-	0.1491 (0.0763)	0.7166	633.2	1933.5
	KABC	-16.8452 (<.0001)	1.1699 (<.0001)	1.5809 (<.0001)	-0.1329 (<.0001)	-	-	-	0.6279	590.1	1556.8
	KAB	-15.3647 (<.0001)	1.1892 (<.0001)	1.3730 (<.0001)	-0.1102 (0.0036)	-0.0047 (0.0583)	-	-	0.3730	500.2	995.6

To determine the variation of CMFs with vehicle, driver, weather, and time information, the CMFs were estimated based on different vehicle size (passenger and heavy), driver age (young, middle, and old), weather condition (normal and rain), and time period (day time and night time). Table 5-6 presents the estimated CMFs with different vehicle types. ROR crashes are categorized in two vehicle types which are passenger and heavy vehicles. Passenger vehicle is representing small cars such as sedan, coupe, etc. Heavy vehicle is including truck, bus, van, and recreational vehicles (RV). In general, roadside barriers were safety effective in reducing KAB crashes for both passenger and heavy vehicles. However, it is worth to mention that roadside barriers are more effective for heavy vehicles KAB crashes than passenger vehicles. Moreover, for KABC crashes, the CMFs for heavy vehicles are statistically significant and lower than the CMFs for passenger vehicle. The result also shows that an addition of w-beam guardrails can increase KABCO crashes for passenger vehicles.

Table 5-6: Evaluated CMFs for ROR crashes with different vehicle types

Crash type	Severity	CMFs from the EB method			
		Roadside Barriers (W-Beam + Concrete)		W-Beam Guardrail Only	
		CMF	S.E	CMF	S.E
ROR passenger vehicle crashes	KABCO	1.03	0.08	1.15*	0.08
	KABC	0.92	0.08	0.98	0.09
	KAB	0.81*	0.10	0.81*	0.11
ROR heavy vehicle crashes	KABCO	0.90	0.08	0.93	0.09
	KABC	0.72**	0.10	0.75**	0.11
	KAB	0.66**	0.12	0.65**	0.13

**: significant at 95% confidence level, *: significant at 90% confidence level

The evaluated CMFs with different ranges of driver age are presented in Table 5-7. ROR crashes were divided into three driver age groups (young age: 15-24 years of age, middle age: 25-64 years of age, old age: 65 years of age and older) (Liu et al., 2007). Although, most of estimated

CMFs are not statistically significant, we can still check general variation of safety effects based on driver age groups. Generally, the safety effects of roadside barriers were positive for KABC and KAB crashes for middle and old age drivers. Moreover, it was found that w-beam guardrails are more safety effective to reduce KAB crashes for old age drivers than middle age drivers. It was also found that all CMFs for young age drivers were insignificant. The results indicate that installation of roadside barriers might not be safety effective for young age drivers. This may be because young age drivers tend to drive at higher speed than middle and old age drivers.

Table 5-7: Evaluated CMFs for ROR crashes with different ranges of driver age

Crash type	Severity	CMFs from the EB method			
		Roadside Barriers (W-Beam + Concrete)		W-Beam Guardrail Only	
		CMF	S.E	CMF	S.E
ROR young age driver (15~24 years old) crashes	KABCO	1.06	0.10	1.12	0.11
	KABC	1.06	0.14	1.11	0.15
	KAB	0.91	0.16	0.95	0.18
ROR middle age driver (25~64 years old) crashes	KABCO	0.93	0.06	1.05	0.08
	KABC	0.79**	0.07	0.85*	0.08
	KAB	0.69**	0.09	0.70**	0.10
ROR old age driver (more than 64 years old) crashes	KABCO	0.91	0.15	0.93	0.17
	KABC	0.80	0.23	0.80	0.25
	KAB	0.62	0.25	0.58*	0.25

**: significant at 95% confidence level, *: significant at 90% confidence level

Table 5-8 shows the estimated CMFs for ROR crashes in different weather conditions. ROR crashes in rain condition on roadways with wet surface were identified and grouped. Also, ROR crashes in normal weather condition on roadways with dry surface were grouped for the analysis. It is worth to note that ROR crashes in other weather conditions such as fog were excluded in the analysis. The results show that roadside barriers are more safety effective in reducing KAB crashes in the rain condition than the normal weather condition whereas the opposite was found for KABC crashes. In the rain condition, relatively more ROR crashes are expected due to the slippery roadway surface. Therefore, the safety effects for the possible injury (C) and property damage only (O) severity levels might be lower in the rain condition than normal weather condition since the barriers can also be perceived and considered as a roadside obstacle (Ben-Bassat and Shinar, 2011). However, for more severe ROR crashes, roadside barriers can prevent the serious impact between roadside hazard (e.g. trees, poles, ditch, etc.) and uncontrollable vehicle in slippery condition through colliding with energy absorbing barriers.

Table 5-8: Evaluated CMFs for ROR crashes with different weather conditions

Crash type	Severity	CMFs from the EB method			
		Roadside Barriers (W-Beam + Concrete)		W-Beam Guardrail Only	
		CMF	S.E	CMF	S.E
ROR crashes in normal weather	KABCO	0.92	0.06	0.95	0.72
	KABC	0.82**	0.08	0.87	0.09
	KAB	0.76**	0.10	0.79*	0.11
ROR crashes in rain and wet surface condition	KABCO	0.92	0.08	1.12	0.09
	KABC	0.90	0.10	0.96	0.11
	KAB	0.75**	0.12	0.75*	0.13

**: significant at 95% confidence level, *: significant at 90% confidence level

The CMFs were estimated for ROR crashes based on time difference as show in Table 5-9. ROR crashes were categorized as day time and night time crashes using crash records in CARS. It was found that roadside barriers are more effective to reduce KABC and KAB crashes in night time than day time. This may be because ROR crashes in night time tend to be more severe due to low visibility and high driving speed. Also, roadside barriers might be more helpful during night time to prevent impacts with roadside hazards.

Table 5-9: Evaluated CMFs for ROR crashes with different time of day

Crash type	Severity	CMFs from the EB method			
		Roadside Barriers (W-Beam + Concrete)		W-Beam Guardrail Only	
		CMF	S.E	CMF	S.E
ROR crashes in day time	KABCO	0.96	0.06	1.05	0.07
	KABC	0.94	0.08	1.01	0.09
	KAB	0.84*	0.10	0.89	0.12
ROR crashes in night time	KABCO	0.92	0.09	0.98	0.10
	KABC	0.71**	0.09	0.73**	0.10
	KAB	0.60**	0.11	0.53**	0.11

**: significant at 95% confidence level, *: significant at 90% confidence level

5.5 Conclusion

Since a CMF represents the overall safety performance of specific treatments among treated sites by a fixed value, there is a need to explore the changes of safety effects with different vehicle, driver, weather, and time information. Thus, the main objective of this study is to evaluate safety effects of adding specific type and combination of roadside barriers on freeways for different crash types and severity levels based on different ranges of vehicle size (passenger and heavy vehicles), driver age (young, middle, and old), weather condition (normal and rain), and time difference (day time and night time). The study calculated CMFs using the observational before-after with EB and FB methods. The finding from this study indicated that the FB provides comparable results to the EB method. The before-after with FB might be a promising technique to obtain a reliable estimate of the expected crashes at specific group of treated sites, especially when relatively scarce information about the treated sites are available, in case of low traffic volumes, or if only few years of crash data are available. However, the EB method might show more reliable estimates when 1) sufficient sample size of reference sites was obtained and used to calculate full SPFs, and 2) there are enough crash frequencies for both treated and reference sites.

The results of estimated CMFs for different crash types and severity levels indicate that roadside barriers are safety effective to reduce ROR (KABC) crashes whereas the CMFs are not statistically significant for all (KABC) crashes. The results also show that the safety effects of roadside barriers are positive and statistically significant for KAB severity level for both all and ROR crashes. It was found that installation of w-beam guardrails might increase crash frequency but reduce crash severity.

From the estimation of CMFs for ROR crashes with different vehicle, driver, weather and time information, it was found that the safety effects vary based on different ranges of vehicle size (passenger and heavy vehicles), driver age (young, middle, and old), weather condition (normal and rain), and time difference (day time and night time). The results show that guardrails are more safety effective in reducing injury and severe ROR crashes for middle and old age drivers than young age drivers. It was found that the CMFs for injury and severe ROR crashes were lower for heavy vehicles than passenger cars. It was also found that the safety effects of treatment were higher for injury and severe ROR crashes in night time than day time. Lastly, the CMFs were lower for severe ROR crashes in rain condition than normal weather condition.

As demonstrated in this study, it is recommended that the CMFs be separately estimated for different crash types and severity levels, and different vehicle types, driver age, weather condition, and time of day. It might be worth to investigate more variations of safety effects based on other characteristics such as pavement conditions, seasonal difference, regional difference, etc.

CHAPTER 6: APPLICATION OF GENERALIZED NONLINEAR MODELS IN CROSS-SECTIONAL ANALYSIS

6.1 Introduction

The CMF can be estimated by observational before-after studies or the cross-sectional method (Gross et al., 2010; Carter et al., 2012). It is known that observational before-after studies with EB and CG methods are the more common approaches among the various before-after studies (Gross et al., 2010; Abdel-Aty et al., 2014). The cross-sectional method has been commonly applied to calculate CMFs due to its easiness with obtaining data compared to the before-after approaches. According to Harkey et al. (2008), the cross-sectional method can also be used to estimate CMFs since 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. For this reason, CMFs have been evaluated using the cross-sectional method (Lord and Bonneson, 2007; Stamatiadis et al., 2009; Li et al., 2011; Abdel-Aty et al., 2014; Park et al., 2014).

It is required to develop SPFs to estimate CMFs using the cross-sectional method and the GLM with NB distribution has been commonly used to develop SPFs to account for over-dispersion (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 CMF (Stamatiadis et al., 2009; Carter et al., 2012). 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 fixed value in the GLM (Lee et al., 2015).

For this reason, an application of using GNM for crash analysis has been recommended (Lao et al., 2013; Lee et al., 2015; Park et al., 2015b; Park and Abdel-Aty, 2015b). Therefore, the objective of this study is to assess the safety effectiveness of installation of bike lane with different bike lane width through 1) evaluation and comparison of GLMs and GNMs, and 2) estimation of CMFs using cross-sectional analysis. In this chapter, crash types and severities are categorized as follow: all crash types with all severities (KABCO) as ‘Total crashes’, all crash types with KABC severity levels as ‘Injury crashes’, and bike related crashes with KABCO severity levels as ‘Bike crashes’.

6.2 Data Preparation

Three sets of data for Florida were used in this study: RCI data for five years (2008-2012), socio-economic parameters from the U.S. Census (U.S. Census Bureau, 1994) and crash data for five years (2008-2012). A segment is represented by roadway identification numbers and beginning and end mile points. The total 256 roadway segments with 51.262 miles in length were identified for the analysis, respectively. In addition to these traffic and roadway geometric characteristics, socio-economic parameters were collected from the U.S. Census Bureau website using PLANSAFE Census Tool (Washington et al., 2010) for each site. This census information was aggregated for the geographic entity (Block Groups) using the same tool. Distributions of each variable among these treated segments are summarized in Table 6-1.

Table 6-1: Descritive statistics of target segments

Variable	Mean	S.D.	Min.	Max.
Crash frequency				
Total crashes	7.055	8.156	0	30
Injury crashes	4.24	4.89	0	22
Bike crashes	0.236	0.700	0	4
Variables related to traffic and roadway characteristics				
Natural logarithm of AADT (veh/day)	10.206	0.493	7.972	10.994
Length (mile)	0.202	0.216	0.05	2.203
Lane width (ft)	12.650	3.109	10	24
Median width (ft)	25.268	15.480	0	80
Median Type (2= median with barrier, 1= median with no barrier, 0=no median)	2= 98sites, 1= 130sites, 0= 28sites			
Shoulder width (ft)	3.167	1.564	2	9
Bike lane width (ft)	4.581	1.428	2	10
Bike lane (1= bike lane, 0= regular shoulder)	1= 55%, 0= 55%			
Demographic and socio-economic variables				
Log of population density (per square mile)	7.265	0.869	4.722	3.003
Log of median household income of each zone (US Dollars)	10.884	0.438	9.719	11.860
Proportion of people with education level less than high school	0.122	0.106	0	0.444
Proportion of commuters by public transport in total commuters	0.007	0.018	0	0.087
Proportion of commuters by bicycle in total commuters	0.005	0.011	0	0.051
Proportion of commuters by walk in total commuters	0.010	0.018	0	0.070

6.3 Methodology

6.3.1 Generalized Nonlinear Model

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., 2015). 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 Equation (6-1) as follow:

$$N_{predicted,i} = \exp(\beta_0 + \beta_1 ln(AADT_i) + \beta_k(X_{ki}) + \gamma_l(U(y_{li}))) \tag{6-1}$$

where,

$N_{predicted,\ i}$=Predicted crash frequency on segment i,

β_k = coefficients for the variable k,

$AADT_i$=Annual Average Daily Traffic of segment i (veh/day),

X_{ki} = Linear predictor k of segment i.

γ_l = coefficients for the nonlinear predictor l,

y_{li} = Nonlinear predictor l of segment i.

The standard error (SE) of the CMF can be calculated by Equation (6-2) as follows (Harkey et al., 2008):

$$SE = \frac{\exp\left(\beta_k+SE_{\beta_k}\right)-\exp\left(\beta_k-SE_{\beta_k}\right)}{2} \tag{6-2}$$

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(β_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., 2015).

6.4 Results

6.4.1 Developed Nonlinearizing Link Function

The nonlinearizing link function was developed to reflect the nonlinearity of bike lane width on crashes as shown in Figure 6-1. The relationship between the logarithm of crash rates (ln(CR)) and bike lane width was plotted to determine the form of nonlinearizing link function. Crash rate was defined as the number of crashes per mile. To identify the best fitted function, eleven nonlinear regression functions (Park and Abdel-Aty, 2015) were compared. It was found that quadratic nonlinear functional form was the best fitted for the relationship between crash rates and bike lane width. A linear regression line was also fitted to the observed data but it does not clearly reflect this nonlinear relationship between the logarithm of crash rates and bike lane width. The developed nonlinearizing link function can be used as a nonlinear predictor in analysis to improve model fit (Lao et al., 2013; Lee et al., 2015).

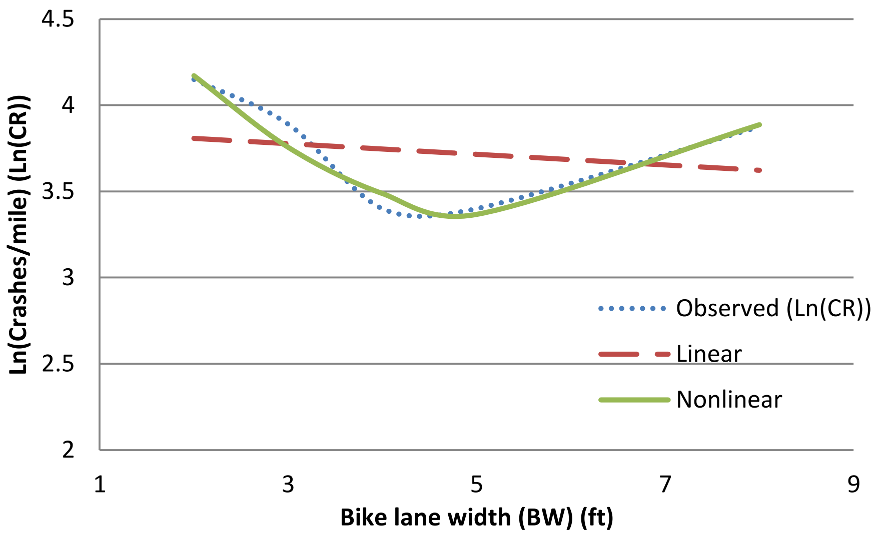


Figure 6-1: Development of nonlinearizing link function for bike lane width

The developed nonlinearizing link function is summarized by Equation (6-3) as follows:

$$U_{BW} = 5.4438 - 0.7834 \times BW + 0.0736 \times BW^2 \tag{6-3}$$

where,

U_{BW} = Nonlinearizing link function for bike lane width.

6.4.2 Estimation of Crash Modification Factors

GNMs for total, injury, and bike crashes were developed using the nonlinearizing link function as shown in Table 6-2. In order to compare model performance, GLMs were also developed. All the models fit the data well since the ratios of deviance to degrees of freedom are close to 1 except the models for bike crashes due to the low crash frequency. In general, the estimated parameters were statistically significant at a 95% confidence level except two cases (U_{BW} of GNM for injury crashes and bike lane width of GLM for bike crashes). The GNMs generally provided better model fits (i.e. smaller AIC value) than the GLMs. This indicates that the inclusion of nonlinearizing link function improved the model fit.

Overall, the results of both GLMs and GNMs show that an increase of bike lane width reduces crash frequency. However, in the GNMs, it was found that crash rates decreases as the bike lane width increases until 5 ft width and it increases as the bike lane width exceeds 5 ft. It was also found that for total and injury crashes, the safety effects decrease as the proportion of people with education level less than high school increases. This may be because education level is correlated with the other socio-economic factors such as income level and employment rate, and these factors can contribute to the higher crash risk (Huang et al., 2010; Abdel-Aty et al., 2013; Park et al., 2015a). Many studies have already found a correlation between traffic crashes and economic status (e.g. income) or education level (Noland, 2003; Huang et al., 2010; Abdel-Aty et al., 2013).

Table 6-2: Estimated parameters of GLM and GNM for different crash types

(a) GLM

	Total crashes			Injury crashes			Bike crashes		
Parameter	Coeffi-cient	Standard error	p-value	Coeffi-cient	Standard error	p-value	Coeffi-cient	Standard error	p-value
Intercept	-9.1165	1.5663	<.0001	-8.5313	1.5801	<.0001	-17.5820	4.4308	<.0001
Log(AADT)	1.0439	0.1504	<.0001	0.9335	0.1518	<.0001	1.5560	0.4265	0.0003
Bike lane width	-0.0689	0.0293	0.0186	-0.0596	0.0295	0.0434	-0.1090	0.0713	0.1264
Proportion of education level less than high school	1.8476	0.6601	0.0051	2.1561	0.6626	0.0011	-	-	-
Dispersion	1.0452	0.1107		0.9443	0.1168		2.9220	1.0596	
Deviance	288.2841			288.0907			115.6905		
Log likelihood	2518.1859			863.6802			-115.8662		
AIC	1499.2615			1274.4299			283.0744		

(b) GNM

	Total crashes			Injury crashes			Bike crashes		
Parameter	Coeffi-cient	Standard error	p-value	Coeffi-cient	Standard error	p-value	Coeffi-cient	Standard error	p-value
Intercept	-9.6167	1.5700	<.0001	-8.9525	1.5838	<.0001	-19.4228	4.5064	<.0001
Log(AADT)	1.0145	0.1521	<.0001	0.9154	0.1532	<.0001	1.5058	0.4194	0.0003
U_{BW}	0.1468	0.0729	0.0440	0.1081	0.0729	0.1380	0.4553	0.1986	0.0218
Proportion of education level less than high school	1.6360	0.6542	0.0124	1.9691	0.6576	0.0028	-	-	-
Dispersion	1.0490	0.1117		0.9482	0.1181		2.7094	1.0027	
Deviance	285.9362			285.5978			115.8271		
Log likelihood	2441.5881			829.3560			-114.3635		
AIC	1481.5439			1259.0178			280.0690		

Table 6-3 presents the estimated CMFs for changing bike lane width using the cross-sectional method. All CMFs were significant at a 90% confidence interval. Note that segments with no bike lane were selected as the base line (i.e. CMF=1). The CMFs from the GLMs show that the safety effects of bike lane consistently decreased as bike lane width increased. On the other hand, the developed CMFs using the GNMs indicate that the safety effects decreased until certain point (5 ft bike lane width) and it increased after this point. This may be because drivers tend to regard a bike lane as a normal vehicle lane or parking area when the bike lane width is similar to the width of vehicle travel lane and adequate marking or signs are not correctly used (Toole, 2010). Also, drivers may be less cautious when they perceive that there are enough spaces in the bike lane for bicycles and they are unlikely to have conflicts with bicyclists. Similarly, bicyclists may not be aware of vehicles when they are using a wide bike lane (Park et al., 2015a). Thus, this indicates that estimated CMFs using the GLMs may misrepresent actual safety effects of changing bike lane width. The results also show that bike lane is more safety effective in reducing bike crashes than total and injury crashes.

Table 6-3: Estimated CMFs for installation of bike lane with different width

Bike lane width		GLM			GNM		
		Total crashes	Injury crashes	Bike crashes	Total crashes	Injury crashes	Bike crashes
No bike lane (Base condition)	CMF	1.000	1.000	1.000	1.000	1.000	1.000
	S.E	-	-	-	-	-	-
2ft	CMF	0.871	0.888	0.804	0.830	0.871	0.560
	S.E	0.051	0.052	0.115	0.077	0.081	0.143
3 ft	CMF	0.813	0.836	0.721	0.781	0.833	0.464
	S.E	0.072	0.074	0.155	0.096	0.103	0.158
4 ft	CMF	0.759	0.788	0.647	0.750	0.809	0.410
	S.E	0.089	0.093	0.187	0.107	0.116	0.163
5 ft	CMF	0.709	0.742	0.580	0.737	0.799	0.388
	S.E	0.104	0.110	0.211	0.112	0.121	0.165
6 ft	CMF	0.661	0.699	0.520	0.740	0.801	0.393
	S.E	0.117	0.124	0.229	0.111	0.120	0.165
7 ft	CMF	0.617	0.659	0.466	0.759	0.816	0.425
	S.E	0.128	0.137	0.242	0.104	0.112	0.162
8 ft	CMF	0.576	0.621	0.418	0.796	0.845	0.492
	S.E	0.136	0.148	0.252	0.090	0.096	0.155

6.5 Conclusions

The GNMs were developed to account for the nonlinear relationship between crash rates and bike lane widths. For this purpose, the developed nonlinearizing link function was used in the analysis. The CMFs were calculated for total, injury and bike crashes using the cross-sectional method. Socio-economic characteristics of the sites collected from the U.S. Census were also considered to reflect the effect of the factors associated with bike use. The main findings of this study are summarized as follows:

The nonlinearizing link function was developed to reflect the nonlinear relationship between the crash rates and bike lane width. It was found that the quadratic nonlinear functional form was the best fitted for this relationship. A linear regression line was also fitted to the observed data but it does not clearly reflect this nonlinear relationship between the logarithm of crash rates and bike lane width. The developed nonlinearizing link function was used in the GNMs to account for the nonlinear effects of changes of bike lane width. The results show that the GNMs generally provided better model fits than the GLMs. Therefore, it can be concluded that including the nonlinearizing link function in GNMs improve the goodness of fit of the models, if the crash rates have a nonlinear relationship with specific parameters.

The results of estimated CMFs using the GLMs indicate that the safety effects of bike lane consistently decreased as bike lane width increased. However, the developed CMFs using the GNMs indicate that the safety effects decreased until 5 ft bike lane width and it increased after this point. It was also found that bike lane is more safety effective in reducing bike crashes than total and injury crashes.

In future work, it is required to further improve the GNMs by increasing sample size and including additional roadway and socio-economic characteristics. It is also recommended to investigate nonlinear relationships between the other treatments and crash rate to reflect nonlinear variation of CMFs using GNMs.

CHAPTER 7: DEVELOPMENT OF SIMPLE AND FULL CRASH MODIFICATION FUNCTIONS USING REGRESSION MODELS

7.1 Introduction

As stated in the previous chapters, a CMF is a multiplicative factor that represents potential changes in the expected number of crashes as a result of implementing a specific treatment (or countermeasure) in a fixed value. Since the CMF is a single value which represents average safety effects of the treatment for all treated sites, the heterogeneous effects of roadway characteristics on CMFs among treated sites are ignored. To overcome this limitation, CMFunctions have been developed to predict the variation in CMFs based on the site characteristics. However, although several previous studies assessed the effect of a specific single variable such as AADT on the CMFs, there is a lack of prior studies on variation in the safety effects of adding a bike lane among treated sites with different multiple roadway characteristics.

Thus, the objective of this chapter is to determine relationship between the safety effects of adding a bike lane and the site characteristics through 1) estimation of CMFs for adding a bike lane using before-after with EB and cross-sectional methods and 2) development of simple and full CMFunctions based on different roadway and socio-economic characteristics of the treated sites to account for the heterogeneous effects. Also, although socio-economic characteristics such as population density and bike commuter rate of the treated sites are potentially associated with bike travel patterns, their effects on crashes have not been investigated. In this study, demographic and socio-economic parameters were used in the analysis to explore their effects.

In this study, it is referred to all crash types with all severities as ‘All crashes (KABCO)’, all crash types with KABC severity levels as ‘All crashes (KABC)’, bike crashes with all severities as ‘Bike (KABCO)’, and bike crashes with KABC severity levels as ‘Bike (KABC)’.

7.2 Data Preparation

Four sets of data for Florida were used in this study: RCI data for ten years (2003-2012), financial project information, socio-economic parameters from the U.S. Census Bureau and crash data for ten years (2003-2012). The RCI data and financial projects information were obtained from the RCI historical database and the Financial Management System maintained by the FDOT to identify the treated sites on urban arterials. The RCI database provides current and historical roadway characteristics data, and reflects features of specific segment for selected dates. The Financial Management System offers a searching system named financial project search. This system provides detailed information on a specific financial project such as district number, status, work type, period and year. Using these two databases, the sites with treatment (adding a bike lane) were identified. The total length of the treated urban arterials is 37.671 miles long and the total number of the treated segments is 227. Also, the reference sites that have similar roadway characteristics to the treated sites in the before period were identified using the RCI database. The reference sites were selected from the same region as the treated sites to improve comparability between the reference and treated sites. Transtat-Iview and Google Earth were used to verify and modify the RCI and financial project information data, if there were any missing values.

In addition to these traffic and roadway geometric characteristics, socio-economic parameters were collected for each site. According to Schick (2009), traffic accidents are related to three factors (Environment, Vehicle, and Human) and transportation politics, socio-demographic characteristics, and sociological factors are one of the factors that can represent the human factor. The socio-economic and demographic parameters were collected from the U.S. Census Bureau website using PLANSAFE Census Tool (Washington et al., 2010). Moreover, this census information was aggregated for the geographic entity (Block Groups) using the same tool. There are two types of geographic entity (Block Groups and Census Tracts) in the U.S. Census and the Block Groups are smaller zone units than the Census Tracts. According to Levine et al. (1995), choosing relatively small spatial zone units can associate characteristics of the zone with crashes and avoid the biases caused by aggregation. Moreover, the zone size of urban areas is much smaller than rural areas, and therefore each zone in the urban areas has relatively small number of roadway segments. Thus, socio-economic parameters in each zone with small spatial units can be more accurately reflected on the roadway segments in urban areas.

Table 7-1 presents the descriptive statistics of the variables for the treated sites. From the comparison of crash frequencies between the before and after periods, it was found that after adding a bike lane, average numbers of crashes were reduced by 22% for All crashes (KABCO) and 62% for Bike (KABCO). Similarly, average numbers of crashes were reduced by 22% for All crashes (KABCO) and 67% for Bike (KABC). This indicates that adding a bike lane is more effective in reducing Bike crashes than All crashes. Moreover, it is worth to mention that proportion of PDO crashes was much higher for Bike crashes than All crashes. However, this may be because of low frequency of Bike crashes.

The crash data were obtained from the CARS maintained by FDOT for these treated and reference sites in before and after periods. All segments that have been treated in the years between 2006 and 2009 were selected for analysis to ensure sufficient sample size. The crash data was extracted for each site for 3-year before (2003-2005) and 3-year after periods (2010-2012). This criterion for crash data was used consistently for the before-after analysis. The intersection- related crashes were removed.

Table 7-1: Descriptive statistics of the variables for treated sites

Variable Name	Definition	Mean	S.D.	Min.	Max.
Crash frequency in before period					
All (KABCO)	Number of crashes for all crash types and all severity levels	6.1171	7.4186	0	35
All (KABC)	Number of crashes for all crash types and KABC severity levels	3.7098	4.6828	0	24
Bike (KABCO)	Number of bike crashes for all severity levels	0.1410	0.4773	0	3
Bike (KABC)	Number of bike crashes for KABC severity levels	0.0264	0.1608	0	1
Crash frequency in after period					
All (KABCO)	Number of crashes for all crash types and all severity levels	4.7818	6.0438	0	30
All (KABC)	Number of crashes for all crash types and KABC severity levels	2.8933	4.2455	0	24
Bike (KABCO)	Number of bike crashes for all severity levels	0.0529	0.2772	0	2
Bike (KABC)	Number of bike crashes for KABC severity levels	0.0088	0.0937	0	1
Variables related to traffic and roadway geometric characteristics					
AADT	Annual Average Daily Traffic (veh/day)	35,262	17,880	10,845	76,500
No_Lanes	Number of lanes (2 lanes = 49 sites, 4 lanes = 97 sites, 6 lanes = 50 sites, 8 lanes = 31 sites)				
AADT_Lanes	AADT per lane (veh/day/lane)	7,708	1,988	3,200	12,750
Length	Segment length (mile)	0.1565	0.1777	0.11	0.97
Surf_width	Total surface width of roadway (ft)	55.63	21.5	22	96
Bike_width	Width of paved bike lane (ft)	4.9339	1.9048	3	10
Med_width	Median width (ft)	26.427	14.215	0	46
Lane_width	Width of vehicle travel lane (ft)	11.805	0.472	10.667	13.333
Med_type	Type of median (1 = with barrier, 0 = no barrier)	1 = 25.55%, 0 = 74.45%			
Sidewalk	Sidewalk for pedestrian (1 = yes, 0 = no)	1 = 39.65%, 0 = 60.35%			
Demographic and socio-economic variables					
Log_Pop_Den	Log of population density (per square mile)	7.3547	0.7539	4.5074	9.1965
Log_Med_Inc	Log of median household income of each zone (US Dollars)	10.8222	0.4297	9.7193	11.86
P_High_edu	Proportion of people with education level less than high school	0.1223	0.1025	0	0.4436
P_Pub_Comm	Proportion of commuters by public transport in total commuters	0.0048	0.013	0	0.0867
P_Bike_Comm	Proportion in total commuters of commuters by bicycle in total commuters	0.0067	0.0151	0	0.0879
P_Walk_Comm	Proportion of commuters by walk in total commuters	0.0074	0.02	0	0.1797
Avg_Const_Yr	Average construction year of structures (1 = average construction year of structures is before 1987, 0 = average construction year of structures is after 1987)	1 = 62.11%, 2 = 37.89%			

7.3 Statistical Method

7.3.1 Safety Performance Functions

Four full SPFs were developed using the NB model for reference sites of urban arterials. The SPFs were developed for different crash types and severity levels as shown in Table 7-2. All variables are significant at a 90% confidence level, respectively. In general, the results of four full SPFs show that crash frequency is higher for the roadway segments with higher AADT and longer length. It is worth noting that crash frequency decreases as median household income increases. This may be because income level is correlated with the other socio-factors such as education level and employment rate, and these factors can contribute to the higher crash risk (Huang et al. (2010); Abdel-Aty et al. (2013)).

Table 7-2: Florida-specific full SPFs for urban arterials

	Coefficient				Dispersion (k)	Goodness of Fit	
	α Intercept	β_1 Ln (AADT)	β_2 Segment Length	β_3 Ln (Median Household Income)		Deviance	AIC
Crash Type (Severity)	Estimate (P-Value)	Estimate (P-Value)	Estimate (P-Value)	Estimate (P-Value)			
All (KABCO)	-3.3762 (0.0851)	1.0823 (<.0001)	2.9507 (<.0001)	-0.5513 (<.0001)	1.6224	587.3420	3293.5609
All (KABC)	-3.7374 (0.0546)	1.0374 (<.0001)	3.1437 (<.0001)	-0.5350 (<.0001)	1.5218	567.5066	2744.9946
Bike (KABCO)	-8.7589 (0.0210)	1.4849 (<.0001)	2.7948 (<.0001)	-0.7553 (0.0027)	1.6357	291.5820	705.3721
Bike (KABC)	-7.6940 (0.0456)	1.1417 (<.0001)	2.7827 (<.0001)	-0.8555 (0.0010)	1.6834	281.7257	680.2444

7.3.2 Multiple Linear Regression with Data Mining Technique

Multivariate regression method was conducted to develop full CMFunction to observe the heterogeneous effects of multiple roadway characteristics among treated sites for the safety effectiveness of treatment using SAS Enterprise Miner program (SAS Institute, Inc., 2014). Figure 7-1 presents processing flow diagram in SAS Enterprise Miner program.

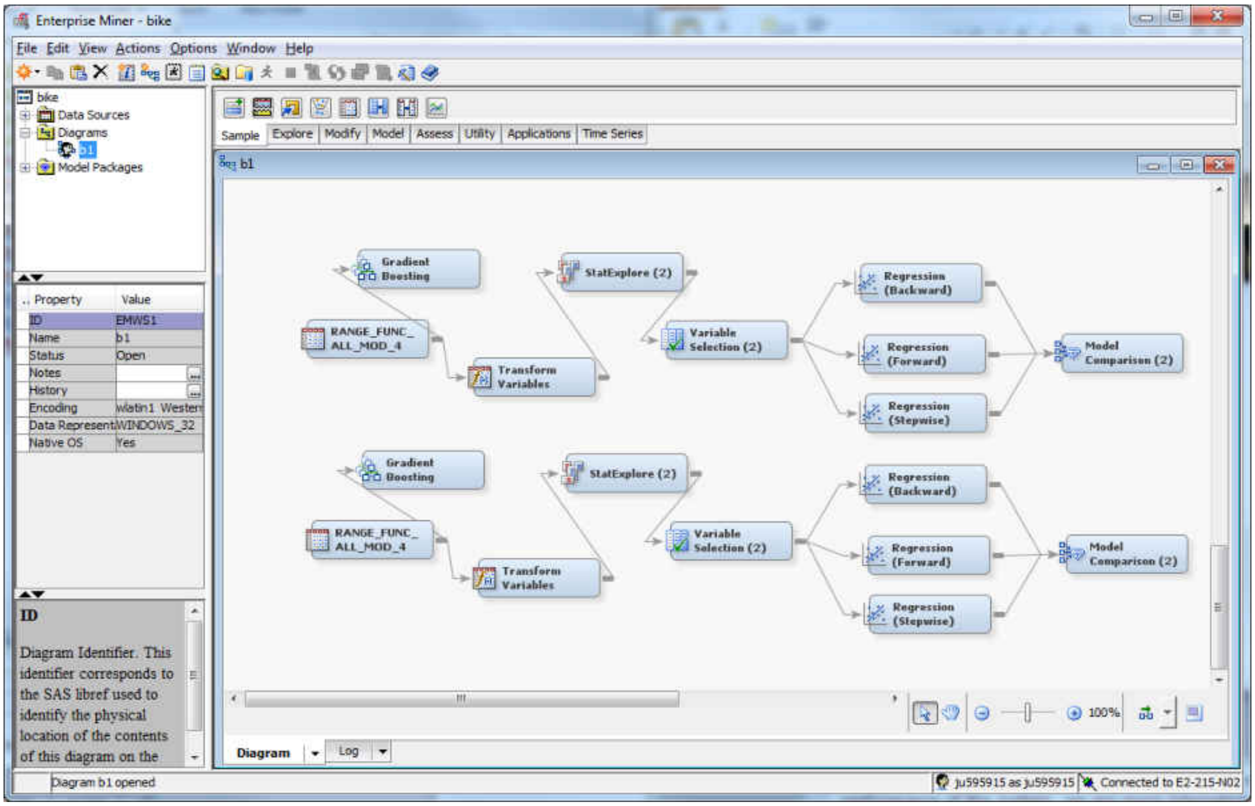


Figure 7-1: Enterpirse Miner Diagram

Variable selection node and gradient boosting node with 50 iterations were used to identify correlation among variables and importance of each variable. Variable transformation node was used to identify the variables that need to be transformed. Two variables (AADT and AADT per lane) were log transformed since they showed high skewness. Three different selection criteria options (backward, forward, stepwise) were applied and the best fitted model was found using regression node and model comparison node. In order to evaluate the advantage of including socio-economic parameters in CMFunctions, the full CMFunctions were estimated using 1) traffic and roadway geometric parameters and 2) traffic, roadway geometric and socio-economic parameters, separately.

7.4 Results

The CMFs for adding a bike lane were calculated using the observational before-after with EB and cross-sectional methods. In case of evaluation of CMFunctions, the CMFs for each treated site were calculated using the before-after with EB method. Lastly, two types of CMFunctions (simple and full) were developed for observing variation and relationship between the CMFs and different roadway characteristics.

7.4.1Estimated CMFs using Cross-Sectional and Before-After with EB Method

The CMFs estimated using the observational before-after with EB and cross-sectional methods were presented in Table 7-3. In the cross-sectional method, the CMFs were estimated using the coefficient of the variable associated with adding a bike lane (i.e. exp(β_3)). The coefficients of all variables in the NB crash prediction models are shown in Table 7-4.

In general, both cross-sectional and before-after with EB methods show that the safety effects of adding a bike lane are positive (i.e. CMF < 1). Also, there was an 8% difference in the CMFs between the cross-sectional and before-after methods. The suggested CMF between the before-after with EB and cross-sectional studies was selected based on lower standard errors. The CMF for Bike (KABC) estimated using the before-after with EB method was not significant due to lower number of bike injury crashes. Therefore, the CMF using cross-sectional method was selected as the suggested CMF for Bike (KABC). It is worth to note that the CMFs for Bike crashes are notably lower than the CMFs for All crashes. These results imply that adding a bike lane is more effective in reducing Bike crashes.

Table 7-3: Evaluated CMFs of adding a bike lane by cross-sectional and before-after with EB methods on urban arterials

Calculation Method	Crash Modification Factor (Standard Error)			
	All crashes (KABCO)	All crashes (KABC)	Bike (KABCO)	Bike (KABC)
Before-After with EB	**0.829*** (0.029)**	**0.804*** (0.039)**	**0.439*** (0.083)**	-
Cross-sectional	0.680*** (0.083)	0.726*** (0.089)	0.422*** (0.096)	**0.398*** (0.093)**

***: significant at a 95% confidence level

Note: Values in bold denote the suggested CMFs between cross-sectional and before-after studies.

Table 7-4: Estimated parameters of crash prediction models by negative binomial regression method

	Coefficient					Dispersion (K)	Goodness of Fit	
	α Intercept	β_1 Log (AADT)	β_2 Segment Length	β_3 Bike Lane	β_4 Surface Width		Deviance	AIC
Crash Type (Severity)	Estimate (P-Value)	Estimate (P-Value)	Estimate (P-Value)	Estimate (P-Value)	Estimate (P-Value)			
All Crashes (KABCO)	-5.6584 (0.0009)	0.6567 (0.0003)	3.0304 (<.0001)	-0.3861 (0.0015)	0.0139 (0.0026)	1.6478	587.4475	3300.6216
All Crashes (KABC)	-6.5465 (0.0001)	0.6972 (0.0001)	3.1861 (<.0001)	-0.3196 (0.0086)	0.0107 (0.0194)	1.5603	567.6695	2754.8602
Bike (KABCO)	-13.6638 (<.0001)	1.1077 (0.0014)	2.5895 (<.0001)	-0.8623 (0.0001)	0.0138 (0.0785)	1.6979	293.8709	711.2364
Bike (KABC)	-13.2241 (0.0001)	1.0530 (0.0028)	2.5632 (<.0001)	-0.9205 (<.0001)	0.0155 (0.0529)	1.7699	284.2315	687.4210

7.4.2 Comparison of CMFs among Segments with Different Roadway Characteristics

Due to low frequency of Bike crashes, the CMFs with different roadway characteristics were calculated for All crashes only. The safety effects of adding a bike lane were assessed for the treated sites with different roadway characteristics for three types of severity levels. The observational before-after with EB method was applied to the treated sites with different levels of 1) AADT per lane, 2) median width, 3) lane width, and 4) bike lane width. Each roadway characteristic has different levels such that there are sufficient samples and the CMF is significant at 85% level at each level. It is worth to note that the CMFs significant at an 85% confidence level might introduce systematic type I-errors. Thus, the CMFs significant at 90% and 95% confidence levels were recommended to use. Moreover, it is suggested to use the CMFs significant at an 85% confidence level to check general impact of treatment with relatively large variation. For the comparison of statistical differences between CMFs, confidence interval of each CMF based on the significant level was also presented.

The CMFs with different ranges of AADT per lane were estimated as shown in Table 7-5. It was found that the CMF for adding a bike lane consistently increases as AADT per lane increases for all of the two severity levels. The results indicate that adding a bike lane has higher safety effects on urban roadways with lower AADT per lane. Moreover, it is worth to note that the safety effects of adding a bike lane are higher for injury crashes (KABC) than all severities (KABCO).

Table 7-5: Evaluated CMFs for the treated sites with different ranges of AADT per lane

	3,200 ≤ AADT per Lane ≤ 5,750		6,000 ≤ AADT per Lane ≤ 7,500		7,625 ≤ AADT per Lane ≤ 9,300		9,417 ≤ AADT per Lane ≤ 12,750	
	47 Segments		63 Segments		58 Segments		59 Segments	
Crash Type (Severity)	Confidence Interval	CMF (S.E)	Confidence Interval	CMF (S.E)	Confidence Interval	CMF (S.E)	Confidence Interval	CMF (S.E)
All crashes (KABCO)	0.618 ~ 0.888	0.753*** (0.069)	0.694 ~ 0.918	0.806*** (0.057)	0.705 ~ 0.955	0.830*** (0.064)	0.843 ~ 0.998	0.921* (0.054)
All crashes (KABC)	0.513 ~ 0.877	0.695*** (0.093)	0.638 ~ 0.964	0.801*** (0.083)	0.667 ~ 0.977	0.822*** (0.079)	0.694 ~ 0.928	0.881** (0.071)

***: significant at a 95% confidence level, **: significant at a 90% confidence level, *: significant at a 85% confidence level

Table 7-6 presents the estimated CMFs with different median widths. The results show that the safety effects are higher for roadway segments with narrow median width (i.e. median width ≤ 16ft). This may be because wide medians are typically installed on the roadways with high traffic volume and speed limits. Thus, higher median width indirectly reflects higher chances of conflicts between 1) vehicles and vehicles and 2) vehicles and bicycles.

Table 7-6: Evaluated CMFs for the treated sites with different median width

	Median Width ≥ 17ft(a+b)		Median Width ≤ 16ft	
	148 Segments		79 Segments	
Crash Type (Severity)	Confidence Interval	CMF (S.E)	Confidence Interval	CMF (S.E)
All crashes (KABCO)	0.834 ~ 0.986	0.910*** (0.039)	0.628 ~ 0.816	0.722*** (0.048)
All crashes (KABC)	0.777 ~ 0.981	0.879*** (0.052)	0.529 ~ 0.745	0.637*** (0.055)
	Median Width ≥ 40ft(a)		17ft ≤ Median Width ≤ 36ft (b)	
	85 Segments		63 Segments	
	Confidence Interval	CMF (S.E)	Confidence Interval	CMF (S.E)
All crashes (KABCO)	0.832 ~ 0.994	0.913** (0.056)	0.819 ~ 0.993	0.906** (0.053)
All crashes (KABC)	0.742 ~ 0.986	0.864** (0.074)	0.785 ~ 0.999	0.892* (0.074)

***: significant at a 95% confidence level, **: significant at a 90% confidence level, *: significant at a 85% confidence level

Table 7-7 presents the estimated CMFs with different lane widths for adding a bike lane. It was found that the CMFs are lower for lane width less than or greater than 12 ft. Thus, lane width has a nonlinear effect on CMFs. In particular, CMFs were the lowest for narrow lane width of 10.5~11.5 ft. This may be because drivers are more aware of bicyclists on the bike lane (Sadek et al., 2007) and drive more cautiously to avoid collision with bicyclists when the lane width is narrower. In fact, the safety effects of the roadways with narrow lane width can be higher than the roadways with wide lane width for specific roadway conditions (Mehta and Lu, 2003; Gross et al., 2009).

Table 7-7: Evaluated CMFs for the treated sites with different lane width

	Lane Width ≥ 12ft [(a+b)]		10.5ft ≤ Lane Width ≤ 11.5ft	
	172 Segments		55 Segments	
Crash Type (Severity)	Confidence Interval	CMF (S.E)	Confidence Interval	CMF (S.E)
All crashes (KABCO)	0.809 ~ 0.947	0.878*** (0.035)	0.672 ~ 0.892	0.782*** (0.056)
All crashes (KABC)	0.763 ~ 0.959	0.861*** (0.050)	0.552 ~ 0.806	0.679*** (0.065)
	Lane Width > 12ft [(a)]		Lane Width = 12ft [(b)]	
	10 Segments		162 Segments	
	Confidence Interval	CMF (S.E)	Confidence Interval	CMF (S.E)
All crashes (KABCO)	-	0.869 (0.103)	0.808 ~ 0.960	0.884*** (0.039)
All crashes (KABC)	-	0.827 (0.135)	0.760 ~ 0.968	0.864*** (0.053)

***: significant at a 95% confidence level

The CMFs for different bike lane width were estimated as shown in Table 7-8. The results showed that the safety effects of adding a bike lane are generally positive except one case: 3 ft width of bike lane for All crashes (KABCO). However, it is worth to mention that the CMFs for roadways with 3 ft width of bike lane are not statistically significant and standard errors are relatively higher than the other cases. Therefore, the CMFs for roadways with 3 ft width of bike lane may not represent the actual safety effects of treatment. Also, the roadways with 10 ft width of bike lane are mostly sharing roadways for bike lane and parking area. Thus, it can be concluded that the safety effects for 10 ft width of bike lane are lower than 4ft to 5ft width of bike lane because of potential conflict between a parking vehicle and a bike. The results also showed that the safety effects of adding a bike lane are relatively higher for the roadways with 4ft to 5ft width of bike lane. Thus, it can be concluded that the urban roadways with 4 ft to 5 ft width of bike lane are safer than the roadways with the other bike lane width when a bike lane is added. According to AASHTO (1999), the minimum width of bike lane is 3 ft and the recommended width of bike lane is 4ft ~ 5ft.

Table 7-8: Evaluated CMFs for the treated sites with different bike lane width

	3 ft ≤ Bike Lane Width ≤ 4 ft [(a+b)]				5 ft ≤ Bike Lane Width ≤ 10 ft [(c+d)]			
	146 Segments				81 Segments			
Crash Type (Severity)	Confidence Interval		CMF (S.E)		Confidence Interval		CMF (S.E)	
All crashes (KABCO)	0.757 ~ 0.913		0.835*** (0.040)		0.733 ~ 0.921		0.827*** (0.048)	
All crashes (KABC)	0.676 ~ 0.872		0.774*** (0.050)		0.732 ~ 0.978		0.855*** (0.063)	
	Bike Lane Width = 3 ft (a)		Bike Lane Width = 4 ft (b)		Bike Lane Width = 5 ft (c)		8 ft ≤ Bike Lane Width ≤ 10 ft [(d)]	
	12 Segments		134 Segments		43 Segments		38 Segments	
	Confidence Interval	CMF (S.E)	Confidence Interval	CMF (S.E)	Confidence Interval	CMF (S.E)	Confidence Interval	CMF (S.E)
All crashes (KABCO)	-	1.031 (0.154)	0.740 ~ 0.904	0.822*** (0.042)	0.605 ~ 0.871	0.738*** (0.068)	0.734 ~ 0.984	0.859*** (0.064)
All crashes (KABC)	-	0.955 (0.180)	0.648 ~ 0.848	0.748*** (0.051)	0.598 ~ 0.946	0.772*** (0.089)	-	0.917 (0.088)

***: significant at a 95% confidence level

7.4.3 Estimation of Simple CMFunctions with Single Roadway Characteristics

The simple CMFunctions for adding a bike lane were developed in order to observe the variation of CMFs with different roadway characteristics. In this study, the simple CMFunction is defined as the function of any single explanatory variable, not only AADT. The effectiveness of adding a bike lane in reducing crashes by severity level was assessed for each treated site. Figure 3 presents the simple CMFunctions with five different roadway characteristics for two severity levels. Due to low frequency of Bike crashes, the CMFuntions were developed for All crashes only. Also, due to poor model fit, the CMFunctions for KABC crashes were not shown for median width and bike lane width in Figure 7-2.

A total of 227 roadway segments with the same roadway characteristics and roadway ID were grouped into 67 data points to remove observations with zero crash count. Since the simple CMFunction need to be fitted with one continuous variable, five different continuous roadway characteristics were used to estimate each CMFunction: 1) log of AADT per lane, 2) log of AADT, 3) log of population density, 4) median width and 5) bike lane width. Based on previous study by Elvik (2011), five linear and non-linear functions - Linear, Inverse, Quadratic, Power, and Exponential - were compared and the best fitted function was identified based on the R-squared value. It was found that Inverse ($y = a + b_1/x$), Quadratic ($y = a + b_1 \cdot x + b_2 \cdot x^2$), and Exponential ($y = a \cdot \exp(b_1 \cdot x)$) non-linear regression models were the best fitted functions for different roadway characteristics.

In general, the relationship between CMFs and roadway characteristics shows that the safety effects of adding a bike lane are higher for All crashes (KABC) than All crashes (KABCO). It is worth to mention that based on the relationship between CMFs and AADT per lane, the CMFs for All crashes (KABC) are notably higher than the CMFs for All crashes (KABCO) when AADT per lane is lower than 9000 veh/day whereas the CMFs for All crashes (KABC) are similar to the CMFs for All crashes (KABCO) when AADT per lane is 9000 veh/day or above. This indicates that adding a bike lane can be more effective to reduce injury crashes (KABC) for roadway segments with lower AADT.

Similar to the relationship between CMFs and AADT per lane, the result of simple CMFunction for population density shows that the CMF increases as population density increases. Since the spatial units with higher population density have more frequent interaction among vehicles, bicyclists and pedestrians in unit area, crash risk is likely to be higher in these spatial units (Huang et al., 2010). Therefore, population density can be used to reflect the variation in effects of safety treatment among different urban arterials.

Moreover, it is worth to note that the simple CMFunctions for different median width and bike lane width show non-linear relationship. The results show that the CMF decreases as the bike lane width increases until 8 ft width and it increases as the lane width exceeds 8 ft. This may be because drivers tend to regard a bike lane as a normal vehicle lane or parking area when the bike lane width is similar to the width of vehicle travel lane and adequate marking or signs are not correctly used (Toole, 2010). Also, drivers may be less cautious when they perceive that there are enough spaces in the bike lane for bicycles and they are unlikely to have conflicts with bicyclists. Similarly, bicyclists may not be aware of vehicles when they are using a wide bike lane. In particular, a bike lane has higher safety effects on the urban roadways with 4 ft ~ 8 ft width. Simple CMFunctions for different median widths, the variation of CMFs is relatively small and it shows linear relationship when undivided segments are omitted in the analysis.

Usually, undivided roadways have a higher likelihood of crash occurrence than divided roadways. The R-squared values of each non-linear regression model except two cases (CMFucntions with AADT per lane for KABCO and KABC) are relatively low due to insufficient sample size of segments with different roadway characteristics. Therefore, it is recommended that the simple CMFunctions be used to identify general relationships between the CMFs and the roadway characteristics, if the size of sample is not sufficient and the R-squared value of the estimated model is very low.

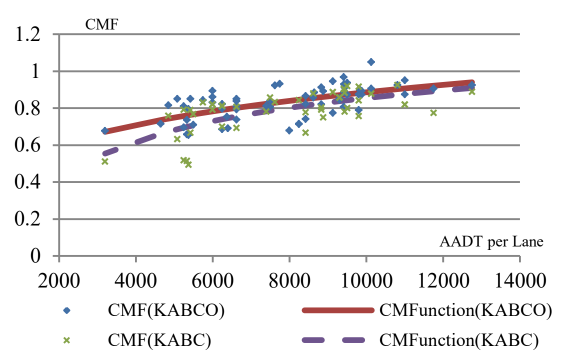
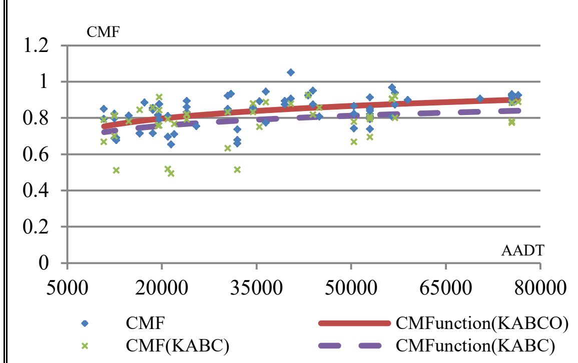
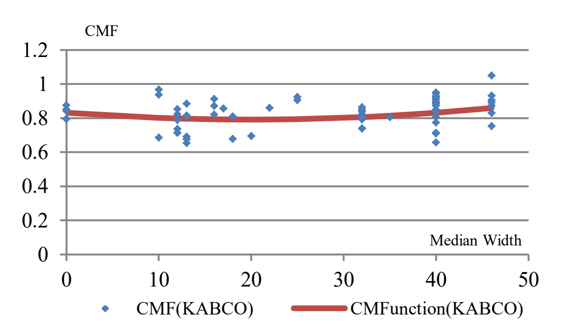
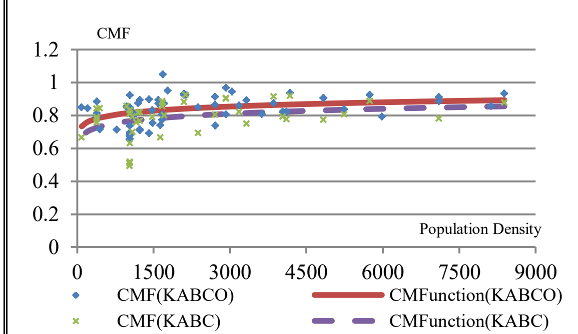
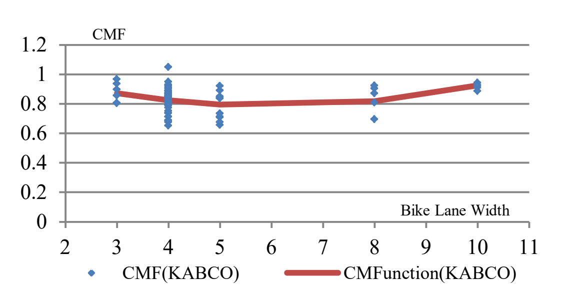
Crash Type (Severity)	Log of AADT per Lane					Log of AADT			
	Function	Coefficients A (P-value)	Coefficients B_1 (P-value)	Coefficients B_2 (P-value)	r^2 (Adj r^2)	Function	Coefficients A (P-value)	Coefficients B_1 (P-value)	r^2 (Adj r^2)
All crashes (KABCO)	Exponential	0.0948 (0.0044)	0.2427 (<.0001)	-	0.3965 (0.3872)	Exponential	0.3233 (<.0001)	0.0911 (<.0001)	0.2392 (0.2275)
All crashes (KABC)	Inverse	2.9821 (<.0001)	-19.5920 (<.0001)	-	0.4506 (0.4378)	Exponential	0.3513 (0.0090)	0.0775 (0.0329)	0.1020 (0.0812)
Graph									
	Median Width					Log of Population Density			
All crashes (KABCO)	Quadratic	0.8316 (<.0001)	-0.0040 (0.1755)	0.0001 (0.0523)	0.1321 (0.1050)	Exponential	0.6036 (<.0001)	0.0433 (0.0027)	0.1286 (0.1152)
All crashes (KABC)	-	-	-	-	-	Exponential	0.5298 (<.0001)	0.0530 (0.0268)	0.1095 (0.0888)
Graph									
	Bike Lane Width								
All crashes (KABCO)	Quadratic	1.1250 (<.0001)	-0.1120 (0.0097)	0.0092 (0.0051)	0.1491 (0.1225)				
Graph									

Figure 7-2: Developed simple CMFunctions for adding a bike lane with different roadway characteristics among treated sites

7.4.4 Estimation of Full CMFunctions with Multiple Roadway Characteristics

Since it was found that CMFs are likely to vary with roadway characteristics, the relationship between CMFs and multiple roadway characteristics was also examined. Multivariate regression models were developed to observe the variation of CMFs with multiple roadway characteristics among treated sites. It was found that the multivariate regression models with backward and stepwise selections were the best fitted full CMFunctions.

Table 15 presents the full CMFunctions for adding a bike lane for All crashes (KABCO). It can be seen that the CMFs increase as AADT per lane increases. Also, it was found that adding a bike lane has higher safety effects for the roadways with narrow median width. This may be because the roadways with wider median width are generally representing higher roadway classification level with higher speed limit, higher traffic volume and more number of lanes. Due to these roadway characteristics, the roadways in higher functional classification level have higher crash risk due to more conflicts and lane changes. Since the simple CMFunctions show a non-linear relationship between the CMF and bike lane width, bike lane width was categorized as a binary variable (= 1 for 4 ft to 8 ft, = 0 otherwise). The results of the full CMFunction without socio-economic parameters show that the safety effects of adding a bike lane are higher for bike lanes with 4 ft to 8 ft width. On the other hand, the full CMFunction with socio-economic parameters captured the variation of CMFs with additional two socio-economic characteristics (bike commuter rates and average construction year of structures). The average construction year of structures was calculated based on the construction year of structures variable from the U.S. Census that represent average construction year of structures in each spatial unit. Based on the median year (i.e. 1987) of all observations, the median year of structures variable was set as a binary parameter (1 = structures were constructed before 1987, 0

= structures were constructed after 1987). Therefore, adding a bike lane has higher safety effects for the roadways in the zone with structures constructed before the median year. All selected variables are significant at 85% for the full CMFunction without socio-economic parameters and significant at 90% level for the full CMFunction with socio-economic parameters.

Table 7-9: Multivariate (Full) CMFunction for adding a bike lane for All crashes (KABCO)

a) All Crashes and KABCO without Socio-economic Parameters

Selection Option: Stepwise							
Analysis Of Variance							
Source	DF	Sum of Squares	Mean Square	F Value	Pr> F	R-Square	Adjusted R-Square
Model	3	0.2148	0.0716	16.75	<.0001	0.4437	0.4172
Error	63	0.2693	0.0043				
Corrected Total	66	0.4842					

Analysis of Maximum Likelihood Estimates				
Variable	Parameter Estimate	Standard Error	T Value	Pr>|T|
Intercept	-0.7373	0.2798	-2.64	0.0106
Log AADT per Lane	0.1740	0.0312	5.58	<.0001
Width of Bike Lane (= 1 for 4 ft to 8 ft, = 0 otherwise)	-0.0168	0.0114	-1.48	0.1447
Median Width	0.0009	0.0005	1.70	0.0932

b) All Crashes and KABCO with Socio-economic Parameters

Selection Option: Backward							
Analysis Of Variance							
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr> F**	**R-Square**	**Adjusted R-Square**
Model	4	0.2328	0.0582	14.35	<.0001	0.4808	0.4473
Error	62	0.2514	0.0041				
Corrected Total	66	0.4842					

Analysis of Maximum Likelihood Estimates				
Variable	Parameter Estimate	Standard Error	T Value	Pr>|T|
Intercept	-1.1217	0.2799	-4.01	0.0002
Log AADT per Lane	0.2130	0.0312	6.82	<.0001
Median Width	0.0014	0.0006	2.60	0.0116
Bike Commuter Rate	1.3573	0.5579	2.43	0.0179
Average Const. Year (1 = structures were constructed before 1987, 0 = structures were constructed after 1987)	-0.0160	0.0089	-1.79	0.0781

The full CMFunction for All crashes (KABC) were developed as shown in Table 7-10. However, no socio-economic parameter was significant. The result of full CMFunction shows that the

CMFs are lower for bike lane with 4 ft to 8 ft width. It can be seen that the CMFs vary with number of lanes. All selected variables are significant at 90% level for the full CMFunction.

Table 7-10: Multivariate (Full) CMFunction for adding a bike lane for All crashes (KABC)

<table>
<tr><th colspan="8">Selection Option: Backward</th></tr>
<tr><th colspan="8">Analysis Of Variance</th></tr>
<tr><th>Source</th><th>DF</th><th>Sum of Squares</th><th>Mean Square</th><th>F Value</th><th>Pr> F</th><th>R-Square</th><th>Adjusted R-Square</th></tr>
<tr><td>Model</td><td>5</td><td>0.2792</td><td>0.0558</td><td>8.56</td><td><.0001</td><td>0.5232</td><td>0.4621</td></tr>
<tr><td>Error</td><td>39</td><td>0.2544</td><td>0.0065</td><td></td><td></td><td></td><td></td></tr>
<tr><td>Corrected Total</td><td>44</td><td>0.5336</td><td></td><td></td><td></td><td></td><td></td></tr>
</table>

<table>
<tr><th colspan="6">Analysis of Maximum Likelihood Estimates</th></tr>
<tr><th colspan="2">Variable</th><th>Parameter Estimate</th><th>Standard Error</th><th>T Value</th><th>Pr>|T|</th></tr>
<tr><td colspan="2">Intercept</td><td>-1.6928</td><td>0.4659</td><td>-3.63</td><td>0.0008</td></tr>
<tr><td colspan="2">Log AADT</td><td>0.2402</td><td>0.0445</td><td>5.40</td><td><.0001</td></tr>
<tr><td rowspan="3">Number of Lanes (Base: 8 lanes)</td><td>2</td><td>0.2253</td><td>0.0417</td><td>5.40</td><td><.0001</td></tr>
<tr><td>4</td><td>0.0446</td><td>0.0224</td><td>1.99</td><td>0.0534</td></tr>
<tr><td>6</td><td>-0.0977</td><td>0.0270</td><td>-3.62</td><td>0.0008</td></tr>
<tr><td colspan="2">Width of Bike Lane (= 1 for 4 ft to 8 ft, = 0 otherwise)</td><td>-0.0427</td><td>0.0189</td><td>-2.26</td><td>0.0293</td></tr>
</table>

It was found that both full CMFunctions with and without socio-economic parameters for the two severity levels show better model fit than any simple CMFunctions. This indicates that the CMFs vary with multiple roadway conditions. It was also found that the full CMFunction with socio-economic parameters show better model fit than the full CMFunction without socio-economic parameters for All crashes (KABCO). Therefore, it is recommended to use the full CMFunction with socio-economic parameters for All crashes (KABCO) to estimate the safety effectiveness of adding a bike lane on urban arterials, if data is available. On the other hand, socio-economic parameters were not significant in the full CMFunction for All crashes (KABC). This implies that socio-economic parameters can improve CMFunctions only for specific crash types and severity levels. Thus, it is recommended to develop multivariate regression models to predict the variation in the safety effects of treatments among the treated sites with multiple roadway characteristics. Table 7-11 presents a summary of the estimated simple and full CMFunctions for adding a bike lane for different severity levels.

Table 7-11: Summary of simple and full CMFunctions for adding a bike lane for All Crashes with different severity levels

<table>
<tr><th></th><th colspan="6">Simple CMFuntions</th></tr>
<tr><th>Crash Type (Severity)</th><th colspan="2">By AADT per Lane</th><th>By AADT</th><th>By Median Width (ft)</th><th>By Bike Lane Width (ft)</th><th>By Population Density (per Sq Mile)</th></tr>
<tr><td>All crashes (KABCO)</td><td colspan="2">$CMF = 0.0948 \times EXP(0.2427 \cdot Log(AADT\ per\ Lane))$</td><td>$CMF = 0.3233 \times EXP(0.0911 \cdot Log(AADT))$</td><td>$CMF = 0.8316 - 0.0040 \cdot Median\ Width + 0.0001 \cdot Median\ Width^2$</td><td>$CMF = 1.1250 - 0.1120 \cdot Bike\ Lane\ Width + 0.0092 \cdot Bike\ Lane\ Width^2$</td><td>$CMF = 0.6036 \times EXP(0.0433 \cdot Log(Population\ Density))$</td></tr>
<tr><td>All crashes (KABC)</td><td colspan="2">$CMF = 2.9821 + \frac{-19.5920}{Log(AADT\ per\ Lane)}$</td><td>$CMF = 0.3513 \times EXP[0.0775 \cdot Log(AADT)]$</td><td>-</td><td>-</td><td>$CMF = 0.5298 \times EXP(0.0530 \cdot Log(Population\ Density))$</td></tr>
<tr><th></th><th colspan="6">Full CMFunctions</th></tr>
<tr><th></th><th># of Lanes</th><th colspan="3">Without Socio-economic Parameters</th><th colspan="2">With Socio-economic Parameters</th></tr>
<tr><td>All crashes (KABCO)</td><td>All</td><td colspan="3">$CMF = -0.7373 + 0.1740 \cdot Log(AADT\ per\ Lane) + 0.0009 \cdot Median\ Width - 0.0168 \cdot Width\ of\ Bike\ Lane$</td><td colspan="2">$CMF = -1.1217 + 0.2130 \cdot Log(AADT\ per\ Lane) + 0.0014 \cdot Median\ Width + 1.3573 \cdot Bike\ Comm\ Rate - 0.0160 \cdot Average\ Const\ Year$</td></tr>
<tr><td rowspan="4">All crashes (KABC)</td><td>2</td><td colspan="5">$CMF = -1.6928 + 0.2402 \cdot Log(AADT) + 0.2253 - 0.0427 \cdot Width\ of\ Bike\ Lane$</td></tr>
<tr><td>4</td><td colspan="5">$CMF = -1.6928 + 0.2402 \cdot Log(AADT) + 0.0446 - 0.0427 \cdot Width\ of\ Bike\ Lane$</td></tr>
<tr><td>6</td><td colspan="5">$CMF = -1.6928 + 0.2402 \cdot Log(AADT) - 0.0977 - 0.0427 \cdot Width\ of\ Bike\ Lane$</td></tr>
<tr><td>8 (base)</td><td colspan="5">$CMF = -1.6928 + 0.2402 \cdot Log(AADT) - 0.0427 \cdot Width\ of\ Bike\ Lane$</td></tr>
</table>

7.5 Conclusion

The main objective of this chapter is to evaluate the safety effectiveness of adding a bike lane in Florida based on the heterogeneous effects of multiple roadway characteristics among treated sites. The CMFs were calculated for All crashes and Bike crashes using the cross-sectional and observational before-after with EB methods. The simple and full CMFunctions were developed to observe relationships between the CMFs and different roadway characteristics. Socio-economic characteristics of the sites collected from the U.S. Census were also considered to reflect the effect of the factors associated with bike use. The main findings of this study are summarized as follows:

The results of CMFs using the cross-sectional and observational before-after with EB methods show that the safety effects of adding a bike lane are high for All crashes and Bike crashes on urban arterials. In particular, adding a bike lane is more effective in reducing Bike crashes than All crashes. There was an 8% difference in the CMFs between the cross-sectional and before-after with EB methods. The most reliable CMFs between the cross-sectional and before-after methods were selected based on lower standard errors.

The CMFs with different roadway characteristics were estimated using the observational before-after with EB method. The CMFs with different roadway characteristics were calculated for All crashes only due to low frequency of Bike crashes. In general, the CMFs were likely to vary with roadway characteristic. In particular, the safety effects were higher for the roadways with 1) low AADT per lane, 2) narrow median width, 3) narrow lane width, and 4) 4 ft to 5 ft width of bike lane. This indicates that a bike lane is more effective in reducing crashes for specific road geometric and traffic conditions.

The results of simple CMFunctions show that Inverse, Quadratic, and Exponential non-linear regression models were the best fitted functions for different roadway characteristics. The relationship between CMFs and roadway characteristics indicates that the safety effects of adding a bike lane for injury crashes (KABC) are higher than all severities (KABCO). The results of simple CMFunctions with AADT per lane show that the safety effects for All crashes (KABC) were significantly higher than All crashes (KABCO) when AADT per lane is less than 9000 veh/day whereas the safety effects for All crashes (KABC) were similar to All crashes (KABCO) when AADT per lane is 9000 veh/day or above. In case of the simple CMFunctions with bike lane width, the safety effects were higher for the roadway segments with 4 ft ~ 8 ft width of a bike lane. This implies that a bike lane is effective in reducing more severe crashes. This is because a bike lane is likely to increase driver's awareness of bicyclists on roadways and can reduce bike crashes where bicyclists are more likely to be severely injured.

The full CMFunctions were also developed to observe the variation of CMFs with multiple roadway characteristics in this study. The results show that the multivariate regression models with backward and stepwise subset selections were the best fitted for multiple roadway characteristics. It was found that both full CMFunctions with and without socio-economic parameters show better model fit (i.e. higher adjusted R-squared value) than all simple CMFunctions. It implies that the safety effects of adding a bike lane vary with multiple roadway characteristics. Also, the results show that the full CMFunctions with socio-economic parameters show better model fit than the full CMFunctions without socio-economic parameters for All crashes (KABCO) whereas no socio-economic parameter was significant for All crashes (KABC). Therefore, it can be concluded that socio-economic parameters improve the goodness-of-fit of the CMFunctions.

Based on the findings in this study, it is recommended to use 4 ft to 8 ft width for a bike lane and add a bike lane at the sites with narrower median (where traffic volume and speed limit are potentially lower). These treatments are likely to increase the effect of bike lanes in reducing crashes.

Since only the data for Florida was used in this study, the safety effects of adding a bike lane might be different for the other states in the U.S. or the other countries. However, a variety of variables including socio-economic parameters were considered in this study to capture the safety effects of treatment with different roadway conditions. Also, it is worth to note that some CMFs in the HSM were recommended to be applied to the U.S condition. Thus, it can be concluded that the findings from this study can provide more reliable effects of safety treatment based on different roadway characteristics in the U.S.

This chapter demonstrates that the safety effects of adding a bike lane can be better predicted using CMFunctions for the treated sites with different roadway and socio-economic characteristics. More work is required to further improve the CMFunctions by including additional roadway and socio-economic characteristics such as horizontal and vertical alignment, actual volume of bicyclists and population of young age group. It is also recommended that multivariate regression models with different options of variable selection be developed to identify key factors affecting safety effects of adding a bike lane more effectively. Moreover, developing full CMFunctions with different roadway characteristics to incorporate changes in safety effects of treatment over time can be an alternative way of estimation of CMFunction.

CHAPTER 8: DEVELOPMENT OF CRASH MODIFICATION FUNCTIONS USING BAYESIAN APPROACH WITH NONLINEARIZING LINK FUNCTION

8.1 Introduction

In the previous chapter, various simple and full CMFunctions were developed using multiple linear regression models. Although traditional statistical models have been utilized in most of data analysis fields, Bayesian models are gaining momentum with the advancement in statistical modeling techniques and computing capabilities. In this chapter, Bayesian regression models with nonlinearizing link function were adopted to develop the CMFunctions considering nonlinear temporal effect.

The widening of roadways with the addition of a through lane is encouraged by certain aspects of traffic planning such as capacity problems or an increase in future traffic demand. Although the relationship between the number of lanes and roadway capacity is well defined in the HCM, which uses the Level of Service (LOS) as a measure to assess the operational performance of roadways, the safety effectiveness of widening urban four-lane roadways to six-lanes is not presented. However, since the addition of one through lane in each direction can greatly change the capacity and cross-sectional elements of roadways, the safety effectiveness of widening urban four-lane roadways to six-lanes has to be fully understood.

Due to the limitations of the HCM on the safety aspects and the demand of safety analysis of specific roadway elements, the HSM was developed to introduce a science-based technical approach for safety analysis. The HSM presents analytical methods to determine and quantify the safety effectiveness of treatments or improvements in transportation fields. However, it is worth noting that there is no CMF in the HSM for widening urban four-lane roadways to six-lanes.

In this chapter, the safety effectiveness of widening urban four-lane roadways to six-lanes was evaluated using the observational before-after EB method. The CMFs with different roadway conditions were also estimated to check the variation of the effects among treated sites. Moreover, the CMFs for each aggregated site were calculated and used for estimation of the CMFunctions. A nonlinearizing link function was also defined to represent the effect of time changes, and it was applied in developing the CMFunctions. Lastly, the CMFunctions with and without the non-linearizing link function were developed to determine the relationship between the safety effects of adding a through lane and the roadway characteristics at different time periods using the Bayesian regression method. Crash types and severity levels are referred to ‘All crash types (KABCO)’ as total crashes and ‘All crash types (KABC)’ as injury crashes.

8.2 Data Preparation

In this study, three sets of data for Florida from the FDOT were used: RCI data for ten years (2003-2012), financial project information, and crash data for ten years (2003-2012). The RCI data was obtained from the RCI historical database, and the financial projects information was identified using Financial Management System. The RCI database provides current and historical roadway characteristics data and reflects the features of specific segments for selected dates. The Financial Management System offers a searching system named financial project search. This system provides detailed information on a specific financial project such as district number, status, work type, costs, period, and year. Treated sites with urban four-lane roadways widened to six-lanes were identified using these two databases. The total length of the treated urban arterials was 46.908 miles long and the total number of the treated segments was 138. Also, the reference sites that have similar roadway characteristics to the treated sites in the before period were identified using the RCI database. In order to obtain the reference sites, untreated roadway segments under same roadway ID as a treated segment were identified since segments in one roadway ID mostly have similar roadway characteristics (e.g. AADT, number of lanes, lane width, etc.). If all segments for one roadway ID have been treated, the reference sites that have similar roadway characteristics as the treated roadway within the same city or county level were selected. A total of 177 roadway segments with 125.432 mile in length were identified as reference sites. Moreover, any missing values or errors of data were verified and corrected or removed using Transtat-Iview (a GIS searching system offered by FDOT) and Google Earth.

The crash data was obtained from the CARS database for these treated and reference sites in before and after periods. All segments that have been treated in the years between 2006 and 2008 were selected for analysis to ensure sufficient sample size. The crash data was extracted for each site for the 3-year before period (2003-2005) and the 4-year after period (2009-2012). Roadway characteristics data from the RCI system for the treated and reference sites were matched with crash data by roadway ID and segment mile point for each site.

The descriptive statistics of the parameters for the treated sites are presented in Table 8-1. It is worth mentioning that shoulder width and median width were narrower after treatment for 17.14% and 40.00% of treated sites, respectively. This may have been because of right of way restriction for widening roadways as in many cases of urban areas. To consider AADT changes before and after the treatment in terms of operational performance, the treated sites were grouped into 3 categories based on LOS changes (TRB, 2010). The total crashes in the before and after periods are 287 and 245, and the numbers of injury crashes in the before and after periods are 162 and 131, respectively.

Table 8-1: Descriptive statistics of the variables for treated sites

Variable Name	Definition	Mean	S.D.	Min.	Max.
Crash frequency in before period					
Total	Number of crashes for all crash types and all severity levels	8.2010	4.7938	2	24
Injury	Number of crashes for all crash types and KABC severity levels	7.0069	3.7643	1	15
Crash frequency in after period					
Total	Number of crashes for all crash types and all severity levels	4.6297	2.6775	0	12
Injury	Number of crashes for all crash types and KABC severity levels	3.7456	2.0609	0	8
Variables related to traffic and roadway geometric characteristics					
AADT_Before	Annual Average Daily Traffic (veh/day) in before period	41,073	8,361	20,500	60,683
AADT_After	Annual Average Daily Traffic (veh/day) in after period	40,960	8,020	25,500	57,979
LOS_Category	LOS E of 4-lane to LOS C of 6-lane = 53 sites, LOS E of 4-lane to LOS D of 6-lane = 37 sites, LOS D of 4-lane to LOS D of 6-lane = 48 sites				
Shld_Width_Before	Width of shoulder lane in before period (ft)	5.7714	2.5677	2	12
Shld_Width_After	Width of shoulder lane in after period (ft)	5.0857	1.9759	2	10
Narrowing_Shld_Width	1= Shoulder width was narrowed , 0=No changes	1 = 17.14%, 0 = 82.86%			
Med_Width_Before	Width of median in before period (ft)	29.8	11.844	6	48
Med_Width_After	Width of median in after period (ft)	23.371	8.5305	6	43
Narrowing_Med_Width	1= Median width was narrowed , 0=No changes	1 = 40.00%, 0 = 60.00%			
Max_Speed	Maximum Speed Limit (mph)	49.571	5.7358	40	60
Lane_width	Width of vehicle travel lane (ft)	11.805	0.472	10.667	13.333
Shld_Type	Type of shoulder (1 = paved, 0 = no)	1 = 77.14%, 0 = 22.86%			
Med_Type	Type of median (1 = with barrier, 0 = no barrier)	1 = 37.14%, 0 = 62.86%			

8.3 Methodology

8.3.1 Safety Performance Functions

Table 8-2 presents the results of the full SPF models for the total and injury crashes per year. In order to estimate the full SPFs, crash data of both before and after periods for the reference sites was used with time difference term. However, the variable of time difference was not significant which indicates that there is no significant difference between the before and after periods under no treatment condition. Moreover, the full SPFs were developed using crash data for the before period and after periods separately. It was found that the full SPFs using crash data for the after period show better model fitness than the model with crash data of before period. Thus, in this study, the full SPFs were developed using the recent 4-year crash data (2009-2012), and all variables are significant at a 95% confidence level.

Table 8-2: Estimated parameters of SPFs by NB method for urban 4-lane roadways

	Coefficient					Dispersion (K)	Goodness of Fit	
	α Intercept	β_1 Ln (AADT)	β_2 Segment Length	β_3 Shoulder Type	β_4 Median Width		Deviance	AIC
Crash Type	Estimate (P-Value)	Estimate (P-Value)	Estimate (P-Value)	Estimate (P-Value)	Estimate (P-Value)			
Total	-8.7362 (<.0001)	1.0717 (<.0001)	0.3443 (<.0001)	-0.7047 (<.0001)	-0.0142 (0.0119)	0.5214	187.1956	979.8421
Injury	-8.3552 (<.0001)	0.9767 (<.0001)	0.3428 (<.0001)	-0.5577 (0.0004)	-0.0168 (0.0030)	0.4043	182.2309	791.9376

8.3.2 Bayesian Regression

Bayesian analysis is the process of fitting a probability model to a set of data and summarizing the posterior probability distribution on the model parameters and on unobserved quantities. Bayesian methods use the posterior probability to measure uncertainty in inferences based on the statistical analysis. Specifically, Bayesian inference generates a multivariate posterior distribution across all parameters of interest, whereas the traditional statistical approaches offer only the model values of parameters. The advantages of Bayesian estimation methods over classical approaches in both philosophical and practical aspects for transportation applications are well described in Washington et al. (2005).

In Bayesian analysis, Markov Chain Monte Carlo (MCMC) methods (Gilks et al., 1996) using Gibbs sampler are broadly utilized to generate a large number of samples from posterior distribution, since the summary of posterior distributions of model parameters may not be tractable algebraically. In this study, a random parameter regression model was fitted assuming explanatory parameters as non-informative with zero mean and a large variance, i.e., Normal($0,10^3$) (Gelman et al., 2004; Gelman, 2006; Sacchi and Sayed, 2014). The WinBUGS software was used to run three Markov chains for each parameter for 30,000 iterations. The first 10,000 iterations in each chain were discarded as burn-in runs. The Deviance Information Criteria (DIC) value was used to compare the models with and without nonlinearizing link function (Spiegelhalter et al., 2005).

8.4 Results

The CMFs were estimated by the observational before-after analysis with EB method using Florida-specific full SPFs for total and injury crashes. The CMFs were also calculated for different roadway conditions over time. Nonlinearizing link functions for time trend was plotted as nonlinear power functional forms and used in developing the CMFunctions. In the case of the evaluation of the CMFunctions, the CMFs for each aggregated treated site were estimated. The CMFunctions with and without nonlinearizing link functions were developed using Bayesian regression method. Lastly, the advantage of using nonlinearizing link functions in developing CMFunctions was determined by the comparison of different models.

8.4.1 Estimated CMFs for Different Time Periods and Roadway Conditions

Table 8-3 presents the estimated CMFs using the observational before-after analysis with the EB method for total and injury crashes for different time periods. Generally, the safety effects of widening urban four-lane roadways to six-lane roadways were positive for both total and injury crashes. It is worth noting that the CMFs decrease over time until the third year after treatment. The differences between the safety effects of the third year and fourth year periods after treatment are only 0.4% and 0.6% for total and injury crashes, respectively. This indicates that drivers are impacted by the change in roadway elements over time and that the safety impact might be consistent after certain time after treatment.

Table 8-3: Estimated CMFs of widening urban 4-lane to 6-lane roadways for different time periods

		CMF (S.E)			
Crash Type	Time Periods	1st year after treated	2nd year after treated	3rd year after treated	4th year after treated
Total	One year term	0.901 (0.074)	0.847** (0.068)	0.798** (0.066)	0.802** (0.066)
	All years	0.850** (0.073)			
Fatal + Injury	One year term	0.841* (0.092)	0.755** (0.088)	0.696** (0.083)	0.702** (0.084)
	All years	0.761** (0.088)			

**: significant at a 95% confidence level, *: significant at a 90% confidence level

The CMFs estimated for the treated sites with different roadway characteristics (LOS changes and shoulder widths) are presented in Table 8-4 and 8-5, respectively. Since widening roadways can greatly change the roadway cross-sectional elements and the change is triggered mainly by operational issues, the LOS levels of each treated site in the periods before and after the treatment were determined and categorized into three groups. Although the CMFs that are not significant at 90% confidence level may not represent statistically reliable safety effects of the treatment, it can be suggested to use these CMFs to check the general impact of widening of the four-lane roadway to six-lanes with relatively large variation. The HSM suggests that a standard error of 0.1 or less indicates that the CMF value is sufficiently accurate, precise, and stable. Also, for treatments that have CMFs with a standard error of 0.1 or less, 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.

The results show that the safety effects are higher for roadway segments with low LOS level (high AADT per lane) in the period before the treatment and high LOS level (low AADT per lane) after. This may be because higher AADT per lane is significantly correlated with crash risk (Abdel-Aty and Radwan, 2000). It was also found that the CMFs are higher for shoulder widths less than or equal to 4 ft after treatment. Moreover, it is worth noting that the safety effects of conversion of urban four-lane roadways to six-lanes are higher for injury crashes than for total crashes.

Table 8-4: Estimated CMFs of widening urban 4-lane to 6-lane roadways for different LOS changes

	LOS Changes in before and after periods					
	LOS E of 4-lanes → LOS C of 6-lanes		LOS E of 4-lanes → LOS D of 6-lanes		LOS D of 4-lanes → LOS D of 6-lanes	
	53 Segments		37 Segments		48 Segments	
Crash Type	CMF	S.E	CMF	S.E	CMF	S.E
Total	0.809**	0.079	0.853*	0.100	0.918	0.096
Fatal + Injury	0.657**	0.121	0.742*	0.157	0.868	0.175

**: significant at a 95% confidence level, *: significant at a 90% confidence level

Table 8-5: Estimated CMFs of widening urban 4-lane to 6-lane roadways for different shoulder width

	Shoulder Width in after period (ft)			
	≤ 4		≥ 6	
	38 Segments		100 Segments	
Crash Type	CMF	S.E	CMF	S.E
Total	0.916	0.098	0.737**	0.106
Fatal + Injury	0.807*	0.111	0.702**	0.147

**: significant at a 95% confidence level, *: significant at a 90% confidence level

8.4.2 Developed Nonlinearizing Link Function over Time

The nonlinearizing link function for total ($U_{yr(total)}$) and injury ($U_{yr(injury)}$) crashes was developed as shown in Figure 8-1 since the safety effects of widening urban four-lane roadways to six-lanes showed a nonlinear relationship with time after treatment (Table 8-3). The relationship between the safety effects (ln(CMF)) and time trend (i.e. years after treatment) was plotted to determine the form of nonlinearizing link function. Nonlinear models with log form were assessed to estimate non-negative CMF value from the link functions (Sacchi and Sayed, 2014; Park and

Abdel-Aty, 2015a). It was found that the observed CMFs initially decreased over time but it was consistent after certain amount of time after treatment for both total and injury crashes. Linear regression lines were also fitted but it did not reflect the nonlinear trend of CMFs over time clearly. Eleven nonlinear regression functions (Park and Abdel-Aty, 2015a) were compared to identify the best fitted function.

The results show that double power and single power nonlinear functions were best fitted for total and injury crashes, respectively. The developed nonlinearizing link functions can be used as a nonlinear predictor in analysis to improve model fit (Lao et al., 2013; Lee et al., 2015). It is worth noting that interaction effects between the CMFs and other explanatory variables were also investigated, but nonlinear effects were not found from any other parameters.

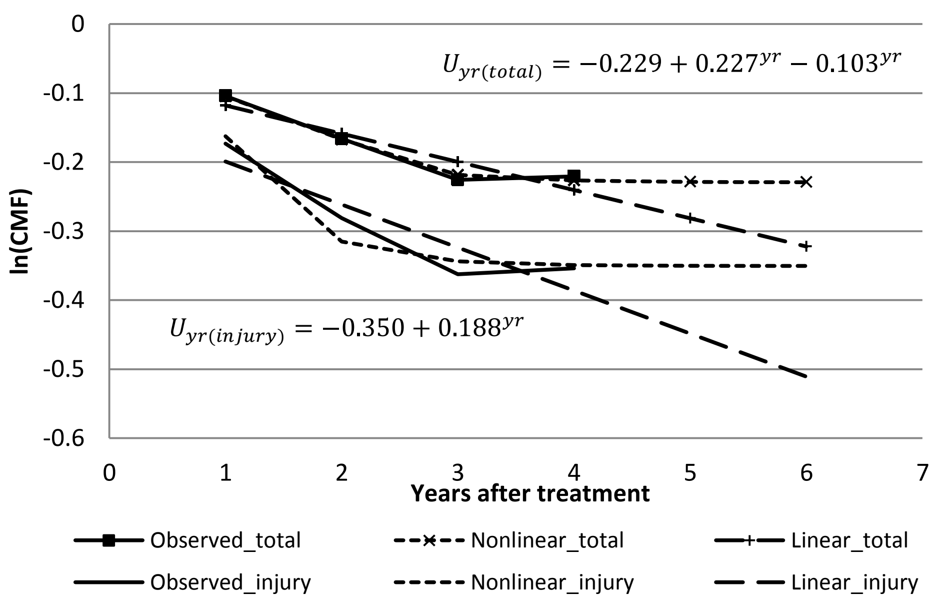


Figure 8-1: Development of nonlinearizing link functions in different time periods for total and injury crashes

8.4.3 CMFunctions by the Bayesian Regression Method

The CMFunctions for conversion of urban four-lane roadways to six-lanes were developed in order to identify the variation of CMFs with different multiple roadway characteristics. The CMFunctions with and without the nonlinearizing link function using Bayesian regression model were utilized to identify the advantages of using nonlinear predictors in analysis. Basically, the nonlinear predictors were used to reflect nonlinear relationship between the observed CMFs and time trend (i.e. years after treatment) in developing CMFunction with nonlinearizing link function. On the other hands, a continuous variable for time trend was used to evaluate the CMFunction without nonlinearizing link function. It is worth to note that the time trend was treated as a categorical variable with dummy variables in developing CMFunction. However, some variables were not significant at a 90% confidence level. Thus, it was not able to identify statistically significant nonlinear effect of changes of CMFs over time.

Tables 8-6 and 8-7 present the estimated CMFunctions with and without the nonlinear predictor for widening urban four-lane roadways to six-lane for total and injury crashes, respectively. To ensure that the CMF value from CMFunction cannot be negative estimate, log form of models were utilized (Sacchi and Sayed, 2014; Park and Abdel-Aty, 2015a).

In general, both CMFunctions for total and injury crashes provide similar inferences. The CMFs decrease with a low LOS level (i.e. LOS E) before treatment as LOS level is higher afterwards when urban four-lane roadways are widened to provide an additional one through lane in each direction. However, the safety effects are relatively lower when the LOS levels of before and after periods are same. The results also show that narrowing shoulder width has negative safety effects on urban roadways. Moreover, it was found that narrowing median width has negative safety effects but the effects are smaller than narrowing the shoulder width for total crashes.

On the other hand, there is no significant difference between the effects of narrowing shoulder width and narrowing median width for injury crashes. It can be recommended that for reducing total crashes, narrowing median width is preferable to make space for widening urban four-lane roadways than narrowing the shoulder width, if the roadways have to be widened and there is not enough right of way. It is worth noting that according to the CMFunction without the nonlinearizing link function, the CMFs decreased in value over time. However, the observed CMFs were consistent after certain amount of time after treatment based on the result of CMFunction with the nonlinear predictor. It is worth noting also that the effect of original shoulder width of treated sites was determined in CMFunctions for total crashes, whereas it was not identified in CMFunctions for injury crashes. The results show that the safety effects are higher as original shoulder width increases. According to the DIC guideline (Spiegelhalter et al., 2005), differences of more than 10 might rule out the model with the higher DIC value. Also, the differences of DIC value more than 5 and less than 10 generally can be used to identify reasonable improvement of model fit. Therefore, it can be concluded that using the nonlinearizing link function in developing CMFunctions can increase model fit significantly since the DIC values of the models with the nonlinear predictor for total and injury crashes are 9.07 and 6.37 lower than the models without the nonlinear predictor, respectively. All selected variables for both models are significant at 95%.

Table 8-6: Estimated CMFunctions by Bayesian models with and without nonlinearizing link function for total crashes

		CMFunction without Nonlinear predictor				CMFunction with Nonlinear predictor			
Variable		Estimate	SD	Interval 5.00%	Interval 95.00%	Estimate	SD	Interval 5.00%	Interval 95.00%
Intercept		0.0159	0.0208	-0.01839	0.05017	0.07742	0.02326	0.03893	0.1155
Years after treatment		-0.06086	0.005091	-0.06925	-0.05249	-	-	-	-
$U_{yr(total)}$ (Time Changes)		-	-	-	-	1.009	0.07904	0.8796	1.139
Narrowing Shoulder Width (1=Yes, 0=No)		0.1066	0.01858	0.07581	0.1373	0.1066	0.01818	0.07659	0.1364
Narrowing Median Width (1=Yes, 0=No)		0.02322	0.01211	0.003348	0.04318	0.02328	0.01189	0.003736	0.04279
LOS Changes Category (Base: LOS E to LOS D)	LOS D to LOS D	0.03756	0.008573	0.02348	0.05164	0.03748	0.008412	0.02358	0.05129
	LOS E to LOS C	-0.03357	0.008326	-0.04729	-0.01992	-0.0336	0.008199	-0.04712	-0.02022
Original Shoulder Width (ft)		-0.01809	0.002694	-0.02249	-0.01365	-0.0181	0.002634	-0.02244	-0.01375
DIC		-110.694				-119.767			

Table 8-7: Estimated CMFunctions by Bayesian models with and without nonlinearizing link function for injury crashes

		CMFunction without Nonlinear predictor				CMFunction with Nonlinear predictor			
Variable		Estimate	SD	Interval 5.00%	Interval 95.00%	Estimate	SD	Interval 5.00%	Interval 95.00%
Intercept		-0.2224	0.02326	-0.2607	-0.1842	-0.09047	0.03393	-0.1463	-0.03485
Years after treatment		-0.05933	0.007427	-0.07152	-0.04712	-	-	-	-
$U_{yr(injury)}$ (Time Changes)		-	-	-	-	0.9579	0.1061	0.7836	1.133
Narrowing Shoulder Width (1=Yes, 0=No)		0.06487	0.02365	0.02576	0.1035	0.06492	0.02309	0.02699	0.103
Narrowing Median Width (1=Yes, 0=No)		0.06972	0.01755	0.04081	0.0985	0.06969	0.01713	0.04154	0.09782
LOS Changes Category (Base: LOS E to LOS D)	LOS D to LOS D	0.04709	0.0124	0.02672	0.06744	0.04708	0.01216	0.02715	0.06716
LOS Changes Category (Base: LOS E to LOS D)	LOS E to LOS C	-0.04563	0.01205	-0.06549	-0.02582	-0.04559	0.01179	-0.06499	-0.02623
DIC		-9.201				-15.575			

Table 8-8 presents a summary of equations for the developed CMFunctions with nonlinearizing link functions to estimate the safety effects (i.e. CMFs) of widening urban roadways with different additional treatments based on different LOS changes over time.

Table 8-8: Summary of developed CMFunctions

		Combination of treatments			
Crash Type	LOS Changes	Widening urban roadways (WUR) only	WUR + Narrowing shoulder width (NSW)	WUR + Narrowing median width (NMW)	WUR + NSW + NMW
Total	LOS E to LOS D	$\exp\{0.0774 - 0.0181 * shld.width + 1.009 * U_{yr(total)}\}$	$\exp\{0.184 - 0.0181 * shld.width + 1.009 * U_{yr(total)}\}$	$\exp\{0.1007 - 0.0181 * shld.width + 1.009 * U_{yr(total)}\}$	$\exp\{0.2073 - 0.0181 * shld.width + 1.009 * U_{yr(total)}\}$
	LOS D to LOS D	$\exp\{0.1149 - 0.0181 * shld.width + 1.009 * U_{yr(total)}\}$	$\exp\{0.2215 - 0.0181 * shld.width + 1.009 * U_{yr(total)}\}$	$\exp\{0.1382 - 0.0181 * shld.width + 1.009 * U_{yr(total)}\}$	$\exp\{0.2448 - 0.0181 * shld.width + 1.009 * U_{yr(total)}\}$
	LOS E to LOS C	$\exp\{0.0438 - 0.0181 * shld.width + 1.009 * U_{yr(total)}\}$	$\exp\{0.1504 - 0.0181 * shld.width + 1.009 * U_{yr(total)}\}$	$\exp\{0.0671 - 0.0181 * shld.width + 1.009 * U_{yr(total)}\}$	$\exp\{0.1737 - 0.0181 * shld.width + 1.009 * U_{yr(total)}\}$
Injury	LOS E to LOS D	$\exp\{-0.0905 + 0.9579 * U_{yr(injury)}\}$	$\exp\{-0.0256 + 0.9579 * U_{yr(injury)}\}$	$\exp\{-0.0208 + 0.9579 * U_{yr(injury)}\}$	$\exp\{0.0441 + 0.9579 * U_{yr(injury)}\}$
	LOS D to LOS D	$\exp\{-0.0434 + 0.9579 * U_{yr(injury)}\}$	$\exp\{0.0215 + 0.9579 * U_{yr(injury)}\}$	$\exp\{0.0263 + 0.9579 * U_{yr(injury)}\}$	$\exp\{0.0912 + 0.9579 * U_{yr(injury)}\}$
	LOS E to LOS C	$\exp\{-0.1361 + 0.9579 * U_{yr(injury)}\}$	$\exp\{-0.0712 + 0.9579 * U_{yr(injury)}\}$	$\exp\{-0.0664 + 0.9579 * U_{yr(injury)}\}$	$\exp\{-0.0015 + 0.9579 * U_{yr(injury)}\}$

8.5 Conclusion

Roadway safety is a major concern for the public, and it is an important component of roadway management strategy. Therefore, a number of CMFs have been estimated for various roadway improvements and treatments (or countermeasures). Also, the CMFunctions for specific single roadway characteristics and or time trends have been developed by only a few previous studies. However, since a CMF represents the overall safety performance of specific treatments in a fixed value, there is a need to explore the variation of CMFs with different multiple roadway characteristics and time trends among treated sites.

This chapter proposed an approach to determine the relationship between safety effects of treatments and multiple roadway characteristics at different time periods through evaluation of the safety effectiveness of widening urban four-lane roadways to six-lanes. This study also identified the advantages of using nonlinearizing link functions in developing CMFunctions to achieve better model performance.

The results of CMFs using the observational before-after analysis with the EB method show that conversion of urban four-lane roadways to six-lane roadways is safety effective for both total and injury crashes. It was also found that the safety effects vary across the sites with different roadway characteristics. In particular, the CMFs were lower for the roadways with 1) low LOS level (high AADT per lane) before treatment and high LOS level (low AADT per lane) after treatment and 2) a wide shoulder width. However, the CMFs are relatively higher when the LOS level is the same for the before and after periods. Moreover, the safety effects decrease over time until the third year after treatment and maintained that level after.

The results of the estimated CMFunctions show that the CMFs vary across the sites with multiple different roadway characteristics. The CMFunctions also showed the variation of CMFs over time. It was found that CMFunctions with the nonlinear predictor show better model performance (i.e., lower DIC values) than models without the nonlinear predictor. Therefore, it can be concluded that including the nonlinearizing link function in developing CMFunctions improve the goodness of fit of the models, if the variation of CMFs with specific parameters has a nonlinear relationship.

It is suggested that more work is required to further improve the CMFunctions by including additional roadway and possibly socio-economic characteristics. Also, a more general relationship could be observed if a longer after period is considered.

CHAPTER 9: UTILIZATION OF MULTIVARIATE ADAPTIVE REGRESSION SPLINES MODEL IN ASSESSING VARIATION OF SAFETY EFFECTS

9.1 Introduction

While the introduced nonlinear regression approaches in previous chapters can reflect the nonlinear effects on the safety performance, interaction impacts between predictors are not considered. In this chapter, an application of using MARS model is conducted to determine the variation of CMFs.

This study first evaluates the CMFs for widening shoulder widths on rural multilane roadways using the observational before-after with the EB method to check the overall safety effects. Secondly, the CMFs were calculated for each aggregated break points based on different roadway characteristics such as the original shoulder widths of treated sites in the before period and the actual widened widths. Lastly, the CMFunctions were developed using multiple linear regression and MARS models to determine the variation of CMFs. The MARS is one of the promising data mining techniques due to its ability to consider the interaction impact of more than one variable and nonlinearity of predictors simultaneously.

In this chapter, crash types and severities are categorized as follow: all crash types with all severities (or total crashes) as 'All (KABCO)', all crash types with KABC severity levels (or injury crashes) as 'All (KABC)', all crash types with KAB severity levels (or severe crashes) as 'All (KAB)', run-off roadways crashes with all severities as 'ROR (KABCO)', ROR crashes with KABC severity levels as ‘ROR (KABC)’, and ROR crashes with KAB severity levels as ‘ROR (KAB)’.

9.2 Data Preparation

In this study, more detailed roadway information and additional treated locations were obtained in addition to previously used dataset in the Chapter 3. Three sets of data maintained by FDOT were used in this study: RCI data for eight years (2004-2011), financial project information and CARS database. Treated sites were identified from the financial project information and the RCI dataset.

All segments that have been treated in the years between end of 2006 and beginning of 2009 were selected for analysis to ensure sufficient sample size. Crash records were collected for 2 years (2004-2005) for before period and 2 years (2010-2011) for after period from CARS. Crash records for 2006 and 2009 were not included in the analysis to account for several data issues (e.g. initial period to prepare roadway construction, finalizing period of construction, stable time for drivers to get used to the new roadway conditions, etc.). In this study, each roadway segment has uniform geometric characteristics in before and after periods except changes of shoulder width and annual average daily traffic (AADT). The total 241 treated roadway segments with 185.822 miles long and 1796 reference sites with 881.882 miles in length were identified, respectively. Distributions of each variable among these treated segments are summarized in Table 9-1.

Table 9-1: Descriptive statistics of treated segments

	Crash frequency in before period				Crash frequency in after period			
Variable	Mean	S.D.	Min.	Max.	Mean	S.D.	Min.	Max.
Number of All (KABCO) crashes	4.037	6.773	0	57	3.249	5.148	0	33
Number of All (KABC) crashes	2.398	3.850	0	24	1.680	2.750	0	19
Number of All (KAB) crashes	1.506	2.467	0	13	0.942	1.687	0	11
Number of ROR (KABCO) crashes	0.950	2.041	0	22	0.622	1.487	0	12
Number of ROR (KABC) crashes	0.577	1.253	0	10	0.344	0.881	0	7
Number of ROR (KAB) crashes	0.407	0.909	0	6	0.203	0.581	0	5

Variables related to traffic and roadway geometric characteristics				
Variable	Mean	S.D.	Min.	Max.
AADT (veh/day) in before period	20548.02	13491.79	4200	60500
AADT (veh/day) in after period	20272.82	12987.71	4100	51500
Length (mile)	0.771	1.000	0.1	4.634
Lane width (ft)	11.975	0.156	11	12
Median width (ft)	46.232	18.718	10	130
Maximum speed limit (mph)	59.274	9.519	40	70
Number of lanes	4 lanes = 226 sites, 6 lanes = 17 sites			
Original shoulder width	2~4ft = 8sites, 5~6ft = 9sites, 7~8ft = 39sites, 9~10ft = 75sites, 11~12ft = 110sites			
Actual widened width	1ft=50sites, 2ft=32sites, 3ft=35sites, 4ft=15sites, 5ft=20sites, 6ft=69sites, 7~8ft=15sites, 9~10ft=5sites			

9.3 Methodology

9.3.1 Safety Performance Functions

In this study, six full SPFs were developed using the NB model for combinations of crash type and severity levels using 2-year before and 2-year after crash data. The SPFs were developed for reference sites of rural multilane roadways in Florida shown in Table 9-2. Also, it is worth to note that the SPFs were evaluated using segment length as an offset. However, the SPFs using segment length as a variable show better model fitness.

In general, the results of six full SPFs show that crash frequency is higher for the roadway segments with higher AADT and longer length. The results also show that the crash frequency is lower for the roadways with wider median widths and lower speed limits. For All (KABCO) crashes, the results indicate that an increase in lane width can increase crash frequency. In order to account for trend of crash frequency based on time changes, a binary variable (i.e. before period) was included to represent the 2-year before period. It is worth noting that the model with categorical variable for each year was assessed but it was not statistically significant. The results indicate that the crash frequency in the after period is lower than the before period for both All and ROR crashes and this trend is consistent with the declining trend of traffic crashes over the last eight years (2004~2011) in the United States (NHTSA, 2013). Since this decline trend on crashes might affect the evaluation of safety effects of treatment, it is better to capture the time changes in the SPFs to account for the trend of crash frequency in the EB analysis.

Table 9-2: Florida specific calibrated SPFs for rural multilane roadways by crash type and severity level

	Estimated Coefficient (p-value)							Dispersion	Deviance	AIC
Crash types	Constant	Ln.AADT	Length	Before period (2004~2005)	Maximum speed limit	Median width	Lane width			
All (KABCO)	-13.9082 (<.0001)	1.3072 (<.0001)	1.0244 (<.0001)	0.0718 (0.1445)	-	-0.0047 (0.0011)	0.0953 (0.0535)	1.4801	3507.5	13191.2
All (KABC)	-14.2983 (<.0001)	1.3374 (<.0001)	1.0163 (<.0001)	0.1122 (0.0344)	0.0125 (0.0029)	-0.0053 (0.0038)	-	1.3581	3166.6	10000.7
All (KAB)	-13.3037 (<.0001)	1.1501 (<.0001)	1.0093 (<.0001)	0.1755 (0.0027)	0.0184 (<.0001)	-0.0058 (0.0054)	-	1.1965	2802.8	7443.2
ROR (KABCO)	-11.8034 (<.0001)	0.8311 (<.0001)	0.8701 (<.0001)	0.1459 (0.0888)	0.0299 (<.0001)	-	-	1.5529	1857.8	3952.5
ROR (KABC)	-12.2116 (<.0001)	0.7835 (<.0001)	0.8644 (<.0001)	0.1734 (0.0992)	0.0357 (<.0001)	-	-	1.3286	1431.5	2681.4
ROR (KAB)	-11.6202 (<.0001)	0.6718 (<.0001)	0.8292 (<.0001)	0.2513 (0.0428)	0.0419 (<.0001)	-0.0079 (0.0937)	-	1.0601	1167.6	1988.2

9.3.2 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 Equation (9-1) (Put et al., 2004). It is worth to note that log form of MARS model was fitted to develop CMFs in this study.

$$\hat{y} = \exp(b_0 + \sum_{m=1}^{M} b_m B_m(x)) \tag{9-1}$$

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 Equation (9-2). 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).

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

$$C(M) = M + dM \tag{9-2}$$

where,

y_i = response for observation i,

n = number of observations,

$C(M)$ = complexity penalty function,

d = defined cost for each basis function optimization.

9.4 Results

9.4.1 Estimation of CMFs using EB method

Table 9-3 presents the estimated CMFs using the observational before-after analysis with the EB method. In general, the safety effects of widening shoulder width were positive for both All and ROR crashes. It is worth to note that the CMFs for ROR crashes are lower than the CMFs for All crashes. These results indicate that widening shoulder width is more effective in reducing ROR than All crashes. Moreover, it was found that safety effects are higher for more severe crashes.

To identify changes of CMFs based on site characteristics, the safety effects of widening shoulder width were calculated for the treated sites with different original shoulder widths and actual widened widths. The results show that the safety effects are higher for roadway segments with narrow original shoulder width (i.e. 2 ~ 8 ft shoulder width) for both All and ROR crashes. The results also show that the safety effects of widening shoulder width are higher as actual widened width increases. Thus, it can be concluded that the safety effects vary based on the different original shoulder widths and actual widened widths among treated sites. It is worth to note that some CMFs are not significant at a 90% confidence level. Although the CMFs that are not significant at the 90% confidence level may not represent reliable safety effects of treatments statistically, it can be suggested to use the insignificant CMFs to check the general impact of treatments with relatively large variation. The HSM suggests that a standard error of 0.1 or less indicates that the CMF value is sufficiently accurate, precise, and stable. Also, for treatments that have CMFs with a standard error of 0.1 or less, other related CMFs with standard errors of 0.2 to 0.3 may also be included and considered to account for the effects of the same treatment on other facilities, other crash types or other severities.

Table 9-3: Estimated CMFs of widening shoulder width for different original shoulder widths and actual widened widths

	Overall Safety Effects		Different Original Shoulder Width				Different Actual Widened Width			
			2 ~ 8 ft		9 ~ 12 ft		1 ~ 4 ft		5 ~ 10 ft	
Crash Type (Severity)	CMF	S.E	CMF	S.E	CMF	S.E	CMF	S.E	CMF	S.E
All (KABCO)	0.88**	0.04	0.72**	0.07	0.94	0.05	0.94	0.07	0.85**	0.05
All (KABC)	0.82**	0.05	0.73**	0.09	0.84**	0.06	0.85*	0.09	0.80**	0.06
All (KAB)	0.79**	0.06	0.69**	0.12	0.82**	0.08	0.84	0.12	0.77**	0.08
ROR (KABCO)	0.75*	0.08	0.66**	0.15	0.77**	0.09	0.77*	0.14	0.74**	0.09
ROR (KABC)	0.72*	0.10	0.62**	0.18	0.74**	0.11	0.73	0.17	0.71**	0.12
ROR (KAB)	0.69**	0.11	0.57**	0.19	0.73*	0.14	0.71	0.21	0.68**	0.13

**: significant at a 95% confidence level, *: significant at a 90% confidence level

9.4.2 Development of CMFunctions

The CMFunctions were developed to determine the variation of CMFs with different site characteristics among treated segments as shown in Tables 9-4 and 9-5. Due to low frequency of All (KAB) and ROR crashes, the CMFunctions were evaluated for All (KABCO) and All (KABC) crashes only. A total of 241 roadway segments with the same roadway characteristics and roadway ID were grouped into 24 data points based on different original shoulder width and actual widened width. As suggested by Sacchi and Sayed (2014) and Park et al. (2015b), log form of models were utilized to ensure that the CMF value from CMFunction cannot be negative estimate. The CMFunctions were developed using multiple linear regression and MARS models. In this study, the ADAPTIVEREG procedure in the SAS program (SAS Institute Inc., 2012) was used to fit a MARS model and 2-way maximum order of interactions was used consistently for the different crash severities. Moreover, the basis functions were constructed for each severity level since the rate of changes can vary within the range for different severities. According to the

Park and Abdel-Aty (2015b), it is recommended to use a MARS model to examine the nonlinearity and interaction impacts between variables.

Overall, the results show that the CMFs increase as original shoulder width increases for both All (KABCO) and All (KABC) crashes. In other words, widening shoulder width has higher safety effects for the roadways with narrow shoulder width. To evaluate more reliable estimates, the variables for actual widened width and median width were transformed as binary variables. The results show that widening shoulder width has lower CMFs for the roadways with narrower median width. This may be because the safety treatments are generally more safety effective when they are implemented for the hazardous roadway conditions (e.g. narrower shoulder and median widths, higher traffic volumes in each lane, more roadside obstacles, etc.). According the developed SPFs in Table 9-2, the roadways with wide median width have less crashes and this indicates that narrower median width represents hazardous roadway condition. Therefore, it might be more safety effective to widen right shoulder width for the roadways with narrower median width than the roadways with wide median width. It should be noted that the treatment is still effective in reducing crashes in general. Also, it was found that the CMFs decrease as actual widened shoulder width increases.

In the MARS models, the estimated parameters of basis functions were statistically significant at a 90% confidence level. The basis functions are constructed by using truncated power functions based on knot values. 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 10 – original shoulder width when original shoulder width is lower than 10, and is 0 for otherwise (where the knot value is 10). Other basis functions are constructed in a similar manner by using different knot values. It is worth to note that various interaction impacts among variables under different ranges based on knot values were found from MARS whereas no interaction impact was found in the linear regression models. Moreover, two variables (i.e. AADT and maximum speed limit) that were not captured in the regression model were found to be significant in MARS. The results also show that the MARS models generally provide better model fits than the regression models. This may be because MARS can account for both nonlinear effects and interaction impacts between variables.

Table 9-4: Estimated CMFunctions of widening shoulder width using regression model

	All (KABCO)			All (KABC)		
Parameter	Estimate	SE	p-value	Estimate	SE	p-value
Constant	-0.5170	0.0486	<.0001	-0.5394	0.0867	<.0001
Original Shoulder Width in Before Period (ft)	0.0258	0.0041	<.0001	0.0246	0.0072	0.0028
Actual Widened Shoulder Width Indicator (1:Sites with 1~4ft shoulder width widened, 0: Sites with 5~10ft shoulder width widened)	0.1648	0.0205	<.0001	0.1729	0.0365	0.0001
Median Width Indicator (1: Sites with less than 40ft median width, 0: Sites with 40ft or more than 40ft median width)	-0.0599	0.0250	0.0265	-0.0653	0.0446	0.1587
MSE	0.0024			0.0077		
R-squared	0.8826			0.7084		
Adj. R-squared	0.8649			0.6647		

Table 9-5: Estimated CMFunctions of widening shoulder width using MARS model

(a) MARS model for All (KABCO) Crashes

Basis Function	Basis Function Information	Estimate	SE	p-value
BF0	Constant	-0.2257	0.0163	<.0001
BF1	MAX (10 – Original shoulder width, 0)	-0.0151	0.0083	0.0874
BF2	MAX (Original shoulder width – 10, 0)	-	-	-
BF3	Actual Widened Shoulder Width Indicator (1:Sites with 1~4ft shoulder width widened, 0: Sites with 5~10ft shoulder width widened)	0.1726	0.0174	<.0001
BF4	Median Width Indicator (1: Sites with less than 40ft median width, 0: Sites with 40 ft or more than 40ft median width)	-0.1720	0.0479	0.0021
BF5	BF2 × MAX (10.02127– Ln. AADT, 0)	-0.0371	0.0170	0.0426
BF6	BF4 × MAX (Original shoulder width – 6, 0)	0.0247	0.0101	0.0252
MSE= 0.0014				
R-squared= 0.9385				
Adj. R-squared= 0.9215				

(b) MARS model for All (KABC) Crashes

Basis Function	Basis Function Information	Estimate	SE	p-value
BF0	Constant	-0.5535	0.0502	<.0001
BF1	MAX (Original shoulder width – 4, 0)	0.1001	0.0318	0.0055
BF2	Actual Widened Shoulder Width Indicator (1:Sites with 1~4ft shoulder width widened, 0: Sites with 5~10ft shoulder width widened)	0.1765	0.0324	<.0001
BF3	MAX (Original shoulder width – 6, 0)	-0.0888	0.0390	0.0354
BF4	Median Width Indicator (1: Sites with less than 40ft median width, 0: Sites with 4 0ft or more than 40ft median width)	-	-	-
BF5	BF4 × MAX (Maximum speed limit– 65, 0)	-0.0439	0.0149	0.0086
BF6	BF4 × MAX (10.16585 – Ln. AADT, 0)	-0.0565	0.0502	0.1027
MSE= 0.0049				
R-squared= 0.8329				
Adj. R-squared= 0.7865				

9.5 Conclusion

The study assesses safety effectiveness of widening shoulder widths on rural multilane roadways considering the variation of CMFs with different site characteristics. In order to determine this variation, the CMFunctions were developed using different statistical approaches. In particular, MARS modeling approach was applied to quantify the changes of CMFs based on varying influential factors due to its strength to account for nonlinearity and interaction impacts between variables.

The results of estimated CMFs indicate that widening shoulder width will reduce crash frequencies. In particular, the estimated CMFs show higher safety effects on severe crashes. Moreover, the CMFs for ROR crashes are lower than the CMFs for All crashes. The CMFs were also estimated based on different ranges of original shoulder width and actual widened width. It was found that CMFs estimated separately for different ranges of original shoulder width and actual widened width can better capture the effects of interactions between safety effects and site characteristics.

The CMFunctions were derived based on this observed relationship. The results of CMFunctions show that the CMFs increase as original shoulder width increases for both All (KABCO) and All (KABC) crashes. Moreover, it was found that the CMFs decrease as actual widened shoulder width increases. The results also show that widening shoulder width has higher safety effects for the roadways with narrower median width. The study demonstrates that the developed CMFunctions using MARS model can better reflect variations in safety effects of widening shoulder width than the CMFunctions using the multiple linear regression.

It is recommended to include multiple target areas (e.g. more states) in the analysis to produce more generalized results. Moreover, it might be worth to investigate more variations of safety effects based on other characteristics such as seasonal difference, regional difference, different crash conditions, etc.

CHAPTER 10: SAFETY ASSESSMENT OF MULTIPLE TREATMENTS USING PARAMETRIC AND NONPARAMETRIC APPROACHES

10.1 Introduction

This chapter offers alternative implementation strategies to assess combined safety effects of multiple treatments using data mining technique to overcome the over-estimation problem in developing CMFunctions for combination of multiple roadside treatments. 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 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 GLMs in 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., 2015). 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. Therefore, researchers have tried to apply different techniques to account for the nonlinearity of variables on crash frequency as follow: 1) GNM (Lao et al., 2013; Lee et al., 2015; Park et al., 2015b), 2) GAM (Li et al., 2011; Zhang et al., 2012), and 3) Random parameter modeling approach (Eluru et al., 2008; Anastasopoulos and Mannering, 2009; Venkataraman et al., 2013; Xu and Huang, 2015). However, most studies investigated only the main effect of each variable, but not the effects of interaction between variables. Moreover, 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 MARS, have been used in safety evaluation studies (Harb et al., 2010; Haleem et al., 2010; 2013; Park and Abdel-Aty, 2015b).

In this chapter, 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.

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; Wu et al., 2014; Park et al., 2014; Park and Abdel-Aty, 2015a). 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.

In this study, crash types and severities are referred 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.

10.2 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 FDOT. These include RCI and 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 because crash rates (= crash frequency per mile) may be exceptionally high on these segments even for a small number of crashes. 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 (Buddhavarapu et al., 2013; Li et al., 2013; Lee et al., 2015). However, since the RCI data for roadway pavement condition has some missing values and it was difficult to verify and collect manually through Google Earth images, it was not used in the analysis. Distributions of each variable among these treated segments are summarized in Table 10-1.

Table 10-1: Descriptive statistics of treated sites

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 = 194sites			
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

10.3 Methodology

10.3.1 Cross-sectional Method

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 Equation (10-1).

$$CMF = \exp\{\beta_k \times (x_{kt} - x_{kb})\} \quad (10\text{-}1)$$

where,

x_{kt} = linear predictor k of treated sites,

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

If a geometric characteristic is expressed in a binary variable (e.g. treatment (= 1) or no treatment (= 0)), the CMF will be exp(β_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.

10.4 Results

10.4.1 Developed Nonlinearizing Link Functions

The nonlinearizing link functions were developed to reflect the nonlinearity of AADT and driveway density on crashes as shown in Figure 10-1 and Figure 10-2. 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., 2015). 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.

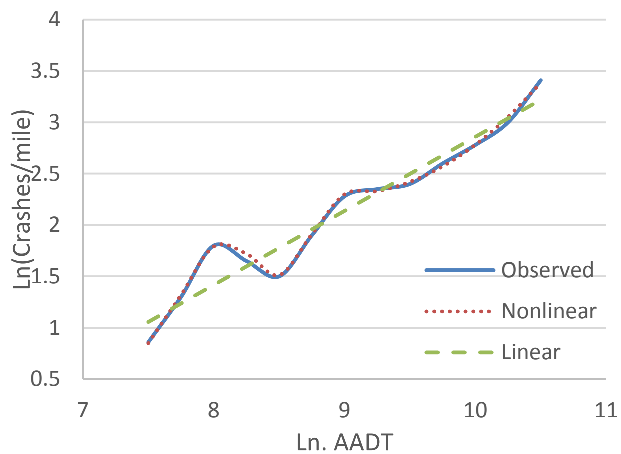
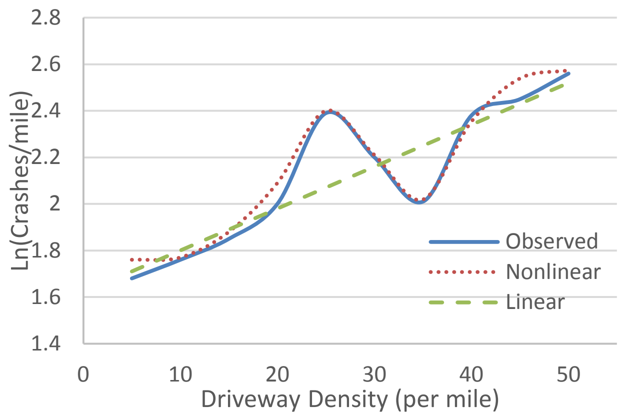


Figure 10-1: Development of nonlinearizing link functions for AADT

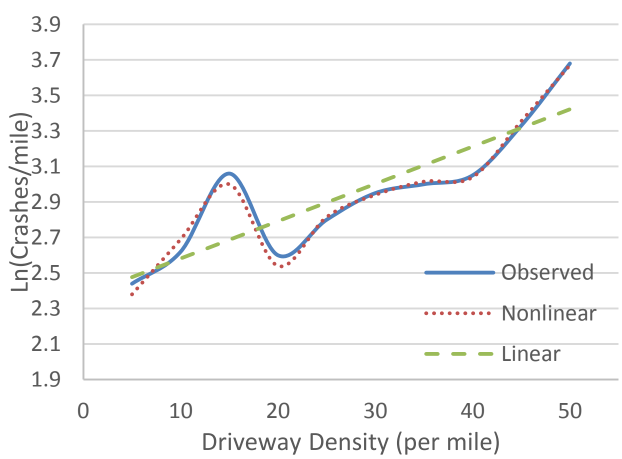
The nonlinearizing link functions for AADT are summarized as shown in Equation (10-2) as follows:

$$U_{AADT} \begin{cases} = 1.79 + 1.880(Ln.AADT - 8) & Ln.AADT \leq 8 \\ = 1.79 - 1.108(Ln.AADT - 8)^2 & 8 \leq Ln.AADT \leq 8.5 \\ = 2.3 + 1.560(Ln.AADT - 9) & 8.5 < Ln.AADT \leq 9 \\ = 2.3 + 0.482(Ln.AADT - 9)^2 & 9 \leq Ln.AADT \end{cases} \tag{10-2}$$

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.



(a) Ln. AADT ≤ 9.8



(b) Ln. AADT > 9.8

Figure 10-2: Development of nonlinearizing link functions for driveway density with 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 Equation (10-3).

a) $Ln.AADT \leq 9.8$

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

b) $Ln.AADT > 9.8$

$$U_{Driveway,AADT}\begin{cases} = 3.0 + 0.062(Drwy.Den - 15) & Drwy.Den \leq 15 \\ = 3.0 - 0.092(Drwy.Den - 15) & 15 \leq Drwy.Den < 20 \\ = 3.04 - 0.001(Drwy.Den - 40)^2 & 20 \leq Drwy.Den \leq 40 \\ = 3.04 + 0.063(Drwy.Den - 40) & 40 \leq Drwy.Den \end{cases} \tag{10-3}$$

10.4.2 Generalized Linear and Nonlinear Models

The GNMs with U_{AADT} and both U_{AADT} and $U_{Driveway,AADT}$ for total, injury, severe, and ROR crashes were developed using the nonlinearizing link functions as shown in Table 10-2. 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 $U_{Driveway,AADT}$ 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 $U_{Driveway,AADT}$. This indicates that the effects of inclusion of nonlinearizing link functions in the developing crash prediction models can vary based on different crash types.

Table 10-2: Estimated parameters of GLMs and GNMs

(a) NB (GLM)

	Total crashes			KABC crashes			KAB crashes			ROR crashes		
Parameter	Coeffi-cient	SE	p-value	Coeffi-cient	SE	p-value	Coeffi-cient	SE	p-value	Coeffi-cient	SE	p-value
Constant	-10.2411	1.6393	<.0001	-9.2788	1.5748	<.0001	-10.7040	1.7656	<.0001	-17.0584	3.6675	<.0001
Ln(AADT)	1.0127	0.1668	0.0032	0.8047	0.1650	<.0001	0.8210	0.1896	<.0001	1.4405	0.3880	0.0002
Driveway Density × Ln(AADT)	0.0024	0.0008	<.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	Coeffi-cient	SE	p-value	Coeffi-cient	SE	p-value	Coeffi-cient	SE	p-value	Coeffi-cient	SE	p-value
Constant	-4.2188	0.7411	<.0001	-4.5657	0.6657	<.0001	-5.7501	0.6686	<.0001	-8.0212	1.3603	<.0001
U_{AADT}	1.4852	0.2443	<.0001	1.2146	0.2374	<.0001	1.1948	0.2642	<.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			0.7360			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	Coeffi-cient	SE	p-value	Coeffi-cient	SE	p-value	Coeffi-cient	SE	p-value	Coeffi-cient	SE	p-value
Constant	-5.7366	1.0149	<.0001	-5.8520	0.8890	<.0001	-6.7111	0.8483	<.0001	-9.5796	1.5788	<.0001
U_{AADT}	1.5417	0.2460	<.0001	1.2424	0.2354	<.0001	1.2367	0.2615	<.0001	1.9385	0.4936	<.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	0.4980	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		

10.4.3 Development of MARS models

In this study, the ADAPTIVEREG procedure in the SAS program (SAS Institute, 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 10-3 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% 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.

Table 10-3: 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	<.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	<.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)	N/S	N/S	N/S
BF8	MAX (25.237 - Driveways Density, 0)	-0.0753	0.0170	<.0001
BF9	Curve (1 if exists; 0 otherwise)	N/S	N/S	N/S
BF10	BF6 × MAX (Driveways Density – 51.565, 0)	0.0680	0.0159	<.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	<.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	<.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)	N/S	N/S	N/S
BF2	MAX (8.501 – Ln. AADT, 0)	-2.0676	0.5329	0.0001
BF3	MAX (Poles Density – 93.75, 0)	N/S	N/S	N/S
BF4	MAX (93.75 - Poles Density, 0)	N/S	N/S	N/S
BF5	BF3 × MAX (Driveways Density – 56.497, 0)	0.9660	0.2270	<.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	<.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	<.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	<.0001
BF17	MAX (Distance to Trees – 5, 0)	N/S	N/S	N/S
BF18	MAX (5 – Distance to Trees, 0)	N/S	N/S	N/S
BF19	BF1 × MAX (42.357 - Driveways Density, 0)	0.1606	0.0377	<.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	<.0001
BF3	MAX (Poles Density – 93.645, 0)	N/S	N/S	N/S
BF4	MAX (93. 645- Poles Density, 0)	N/S	N/S	N/S
BF5	BF3 × MAX (Driveways Density – 51.565, 0)	0.2382	0.0509	<.0001
BF6	Curve (1 if exists; 0 otherwise)	1.1727	0.2471	<.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
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				

10.4.4 Estimation of Crash Modification Factors

Table 10-4 presents a summary of the CMFunctions to estimate the safety effects 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.

Table 10-4: Summary of CMFunctions for different crash types

(a) Total Crashes

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

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

(b) Injury Crashes

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

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

(c) Severe Crashes

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

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

(d) ROR Crashes

	GLM		MARS	
Treatment	**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)	-	-	-	-

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 (2015a) 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 10-5, 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 1ft 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 1ft 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 (2015a). 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 10ft, 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 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.

Table 10-5: 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)

10.5 Conclusion

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 analyzes 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 over-estimated 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. 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.

CHAPTER 11: CONCLUSIONS

11.1 Summary

The dissertation focuses on exploration and development of CMFs and CMFunctions for multiple treatments. The main objective of this study are to 1) assess safety effects of multiple treatments through exploration of the limitations of combining methods for multiple CMFs, 2) develop CMFunctions to determine the variation of safety effects of specific single or multiple treatments with different roadway characteristics among treated sites over time, and 3) suggest methodologies to consider simulataneously the interaction impact of more than one variables and nonlinearity of predictors in developing CMFunctions. Based on the evaluation results, corresponding improvement suggestions have been made.

In Chapter 3, it was attempted to comprehensively estimate the safety effects of two single treatments (shoulder rumble strips and widening shoulder width) and combined treatment (shoulder rumble strips + widening shoulder width) on rural multilane roadways. The results of before-after studies showed that the safety effects of the two single treatments and the combined treatment were higher for the roadway segments which originally had shorter shoulder width (4 ft ~ 6 ft) in the before period. It was also found that the safety effects of multiple treatments was higher than the effects of single treatments for the segments with 4 ft ~ 6 ft original shoulder width, whereas the safety effects of multiple and single treatments were similar for the segments with 8 ft ~ 12 ft original shoulder width. Moreover, the accuracy of the combined CMFs for multiple treatments calculated by the existing combining methods based on actual estimated combined CMFs was evaluated. From this evaluation, Chapter 3 showed whether the existing methods of combining the CMFs over- or under-estimate actual CMFs.

Although the estimated combined effects from averaging the best two methods can estimate more reliable combined CMFs, there is still difference between combined and actual safety effects for multiple treatments. Therefore, development of adjustment factors and functions was proposed to improve the accuracy in combining CMFs in Chapter 4. In order to adjust the combined CMFs for multiple treatments by the HSM combining procedure, the adjustment factors were estimated by comparison of actual calculated CMFs and the combined CMFs for adding shoulder rumble strips + widening shoulder width and installing bike lane + lane reduction based on different implemented locations. In Chapter 4, the CMFunctions were also developed for two single treatments (adding shoulder rumble strips, widening shoulder width) and combination to identify the relationship between CMFs and original shoulder width of roadway. It was found that the difference between CMFs of two single treatment and CMFs for multiple treatments is getting larger as shoulder width decreases for both All and SVROR crashes. The results indicated that the safety effects of multiple treatments vary based on the characteristics of the roadway segments. To determine the nonlinear relationship of the difference between combined safety effects and actual estimated CMFs, the adjustment functions were developed using nonlinear regression models in Chapter 4.

In Chapter 5, the CMFs were developed for different crash types and severities with different crash conditions to identify changes of the safety effects for installing different types of roadside barriers. Two observational before-after analyses (i.e. EB and FB approaches) were utilized in Chapter 5 to estimate CMFs. To consider the variation of safety effects based on different vehicle, driver, weather, and time of day conditions, the crashes were categorized based on vehicle size (passenger and heavy), driver age (young, middle, and old), weather condition (normal and rain), and time difference (day time and night time).

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 fixed value in the GLM. For this reason, an application of using GNM in cross-sectional analysis to estimate CMFs considering nonlinear effects of the treatment is proposed in Chapter 6. Both GLMs and GNMs were developed and compared to assess the safety effectiveness of installation of bike lane with different bike lane width in Chapter 6. The nonlinearizing link function was developed to reflect the nonlinear relationship between the crash rates and bike lane width.

In Chapter 7, the CMFs for adding a bike lane on urban arterials were estimated using before-after EB and cross-sectional methods for different crash types and severities. Simple and full CMFunctions were developed based on different roadway and socio-economic characteristics of the treated sites to account for the heterogeneous effects. In order to develop CMFunctions, multiple linear and nonlinear regression models were utilized and data mining techniques were adopted to achieve better model performance. To explore potential association of socio-economic parameters with bike travel partterns and crash rates, various demographic and socio-economic parameters were used in the analysis.

In Chapter 8, Bayesian regression models with nonlinearizing link function were adopted to develop the CMFunctions considering nonlinear temporal effect. Although traditional statistic models have been widely utilized in the traffic safety field, Bayesian models are gaining momentum with the advancement in statistical modeling techniques and computing capabilities. In Chapter 8, the safety effectiveness of widening urban four-lane roadways to six-lanes was evaluated using the observational before-after EB method. The CMFs with different roadway conditions were also estimated to check the variation of the effects among treated sites over time. Moreover, the nonlinearizing link functions were defined to represent the effect of time changes, and they were applied in developing the CMFunctions. Lastly, the CMFunctions with and without the non-linearizing link function were developed and compared.

While the introduced nonlinear regression approaches in previous chapters can reflect the nonlinear effects on the safety performance, interaction impacts between predictors are not considered. In Chapter 9, an application of using MARS model is proposed to determine the variation of CMFs considering the interaction impact of more than one variable and nonlinearity of predictors simultaneously. The CMFs for widening shoulder widths on rural multilane roadways were evaluated using the before-after EB method. Moreover, the CMFunctions were developed using multiple linear regression and MARS models to determine the variation of CMFs.

Chapter 10 offers alternative implementation strategies to assess combined safety effects of multiple treatments using parametric and nonparametric approaches to overcome the over-estimation problem in developing CMFunctions for combination of multiple roadside treatments. 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 difficult to consider the safety effect of single treatment from other multiple treatments implemented at the same time using the observational before-after studies. In Chapter 10, 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.

11.2 Research Implications

The implications from Chapter 3 are as follow: First, the CMFs for adding shoulder rumble strips estimated using cross-sectional method and before-after studies were similar (only 8% difference) and comparable for All crashes and SVROR. Second, among the six existing methods of combining CMFs for multiple treatments, the HSM method (multiplication), Systematic Reduction of Subsequent CMFs, Apply only the most effective CMF, and Weighted average of multiple CMFs (Meta-Analysis) provide the most accurate estimates of the combined CMFs for multiple treatments. However, in general, the combined CMFs were under-estimated for all crashes (KABCO) whereas they were over-estimated for injury crashes (KABC). In Chapter 4, an average of the combined CMFs from the best two methods was closer to the actual CMF than the combined CMF from only one best method. This indicates that it is better not to rely on only one specific existing method of combining CMFs for predicting CMF for multiple treatments. Also, it is recommended that the safety effects of multiple treatments be separately estimated for different crash types, severity levels, and roadway characteristics.

The findings from Chapter 4 may give several implications. In Chapter 4, it was attempted to improve accuracy of combined safety effects through developing adjustment factors and functions for multiple treatments. It is recommended to develop and apply adjustment factors and functions to predict the combined safety effects of multiple treatments based on 1) different crash types and severity levels, and 2) implemented locations (e.g., roadside, mainline, etc.) of treatments. In particular, the combined safety effects need to be adjusted when multiple treatments are implemented at the same location. As the HSM provides various CMFs from previous studies using data of specific states or locations, the results of this study may be applicable to other states or countries. However, it is recommended to check the similarity of the target state or location to Florida conditions. In particular, the characteristics of roadways (e.g. AADT range, roadway type, shoulder width range, etc.) and crash data (crash types, severity levels and scales, etc.) of the target state or location need to be similar to the characteristics of Florida. Lastly, since this study focuses on specific single and combinations of treatments, the estimated CMFunctions and adjustment functions may not be generalizable to other treatments.

Chapter 5 carries several implications for practitioners. The finding from Chapter 5 indicates that the FB provides comparable results to the EB method. From the estimation of CMFs for ROR crashes with different vehicle, driver, weather and time information, it was found that the safety effects vary based on different ranges of vehicle size (passenger and heavy vehicles), driver age (young, middle, and old), weather condition (normal and rain), and time difference (day time and night time). In particular, the results show that guardrails are more safety effective in reducing injury and severe ROR crashes for middle and old age drivers than young age drivers. It was found that the CMFs for injury and severe ROR crashes were lower for heavy vehicles than passenger cars. It was also found that the safety effects of treatment were higher for injury and severe ROR crashes in night time than day time. Lastly, the CMFs were lower for severe ROR crashes in rain condition than normal weather condition. Therefore, it is recommended that the CMFs be separately estimated for different crash types and severity levels, and different vehicle types, driver age, weather condition, and time of day.

The findings from Chapter 6 are useful for researchers and practitioners when the CMF is estimated using the cross-sectional method and there is a nonlinearity of specific predictor. In Chapter 6, it was found that the GNMs with developed nonlinearizing link function generally provided better model fits than the GLMs. Therefore, it can be suggested that the nonlinearizing link function is developed and included in GNMs improve the goodness of fit of the models, if the crash rates have a nonlinear relationship with specific parameters. It is also recommended to investigate nonlinear relationships between the other treatments and crash rate to reflect nonlinear variation of CMFs using GNMs.

Chapter 7 provides important implications for traffic safety analysts. The results of CMFs using the cross-sectional and observational before-after with EB methods show that the safety effects of adding a bike lane are high for All crashes and Bike crashes on urban arterials. In particular, adding a bike lane is more effective in reducing Bike crashes than All crashes. There was an 8% difference in the CMFs between the cross-sectional and before-after with EB methods. Also, the CMFs with different roadway characteristics were estimated. In general, the CMFs were likely to vary with roadway characteristic. In particular, the safety effects were higher for the roadways with 1) low AADT per lane, 2) narrow median width, 3) narrow lane width, and 4) 4 ft to 5 ft width of bike lane. This indicates that a bike lane is more effective in reducing crashes for specific road geometric and traffic conditions. The results of simple CMFunctions show that Inverse, Quadratic, and Exponential non-linear regression models were the best fitted functions for different roadway characteristics. The full CMFunctions were also developed to observe the variation of CMFs with multiple roadway characteristics in Chapter 7. The results show that the multiple regression models with backward and stepwise subset selections were the best fitted for multiple roadway characteristics. It was found that both full CMFunctions with and without socio-economic parameters show better model fit (i.e. higher adjusted R-squared value) than all simple CMFunctions. It implies that the safety effects of adding a bike lane vary with multiple roadway characteristics. Also, the results show that the full CMFunctions with socio-economic parameters show better model fit than the full CMFunctions without socio-economic parameters for All crashes (KABCO) whereas no socio-economic parameter was significant for All crashes (KABC). Therefore, it can be suggested that socio-economic parameters are included to improve the goodness-of-fit of the CMFunctions. Based on the findings in Chapter 7, it is recommended to use 4 ft to 8 ft width for a bike lane and add a bike lane at the sites with narrower median (where traffic volume and speed limit are potentially lower). These treatments are likely to increase the effect of bike lanes in reducing crashes.

Several important implications were found from Chapter 8. An approach to determine the relationship between safety effects of treatments and multiple roadway characteristics at different time periods through evaluation of the safety effectiveness of widening urban four-lane roadways to six-lanes was proposed in Chapter 8. Moreover, the advantages of using nonlinearizing link functions in developing CMFunctions to achieve better model performance were identified. The results of CMFs using the observational before-after analysis with the EB method show that conversion of urban four-lane roadways to six-lane roadways is safety effective for both total and injury crashes. It was also found that the safety effects vary across the sites with different roadway characteristics. In particular, the CMFs were lower for the roadways with 1) low LOS level (high AADT per lane) before treatment and high LOS level (low AADT per lane) after treatment and 2) a wide shoulder width. Moreover, the safety effects decrease over time until the third year after treatment and maintained that level after. The results of the estimated CMFunctions show that the CMFs vary across the sites with multiple different roadway characteristics. The CMFunctions also showed the variation of CMFs over time. It was found that CMFunctions with the nonlinear predictor show better model performance than models without the nonlinear predictor. Similar to the results of Chapter 7, it can be recommended to include the nonlinearizing link function in developing CMFunctions to improve the goodness of fit of the models, if the variation of CMFs with specific parameters has a nonlinear relationship.

Chapter 9 carries out several implications for researchers. The results of estimated CMFs indicate that widening shoulder width will reduce crash frequencies. In particular, the estimated CMFs show higher safety effects on severe crashes. The CMFs were also estimated based on different ranges of original shoulder width and actual widened width. It was found that CMFs estimated separately for different ranges of original shoulder width and actual widened width can better capture the effects of interactions between safety effects and site characteristics. The results of CMFunctions in Chapter 9 show that the CMFs increase as original shoulder width increases for both All (KABCO) and All (KABC) crashes. Moreover, it was found that the CMFs decrease as actual widened shoulder width increases. The results also show that widening shoulder width has higher safety effects for the roadways with narrower median width. It was demonstrated that the developed CMFunctions using MARS model can better reflect variations in safety effects of widening shoulder width than the CMFunctions using the multiple linear regression.

The findings from Chapter 10 suggest very important implications for both researchers and practitioners. 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. Thus, alternative implementation strategies to assess combined safety effects of multiple treatments using parametric and nonparametric approaches to overcome the over-estimation problem in developing CMFunctions was proposed in Chapter 10. 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. From the development and comparison of GLM and GNM for different crash types and severities, it was found that the GNMs generally provided slightly better model fits than the GLMs but the difference was not significant. This may be because the interaction impacts among variables under different ranges were not reflected by the GNMs. It was also found that MARS models generally provide better model fitness than the GLMs and GNMs. Moreover, the combined safety effects over-estimated 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. 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 the model.

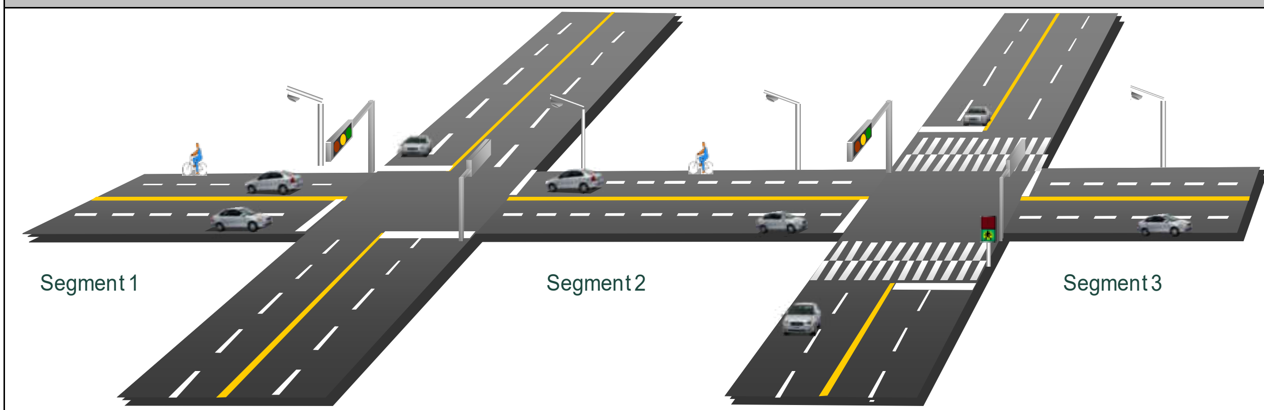
11.3 Implication Scenario

Generally, the variation of CMFs with different roadway characteristics among treated sites is ignored because the CMF is a fixed value that represents overall safety effects of the treatment for all treated sites. Thus, the simple and full CMFunctions can be utilized to determine the relationship between safety effects and roadway characteristics as described in Chapter 7 of this dissertation.

An example of implication of using simple and full CMFunctions of adding a bike lane is presented in Figure 11-1. Three segments are randomly selected from Chapter 7. The average differences between observed crash counts and expected crash counts using fixed CMF, simple CMFunction, and full CMFunction are 20%, 15%, and 10%, repectively. The expected crash counts by CMFs estimated from full CMFunction are more close to the observed crash counts in the after period than using fixed CMF and simple CMFunction. This indicates that using CMFunctions can reflect the variation of safety effects based on different roadway characteristics whereas a fixed CMF only show overall safety effect of treatments among treated sites. In particular, since full CMFunction can reflect multiple roadway characteristics and has better model fit than simple CMFunction, it is suggested to utilize full CMFunction when data is available.

		Segment 1	Segment 2	Segment 3
Observed Injury Crashes (Crashes per year per mile)	Before Treatment	4	10	8
	After Treatment	3	7	4
AADT (veh/day)		21,000	24,000	21,000
Number of Lanes		4	4	4
Width of Bike Lane (ft)		3	4	4
Expected Injury Crashes After Treatment (Crashes per year per mile)	Fixed Value	$CMF = 0.804$		
		3.22	8.04	6.43
	Simple CMFunction	$CMF = 0.3513 \times EXP[0.0775 \cdot Log(AADT)]$		
		$CMF = 0.760$	$CMF = 0.768$	$CMF = 0.760$
		3.04	7.68	6.08
	Full CMFunction	$CMF = -1.6928 + 0.2402 \cdot Log(AADT) + 0.0446(No.of\ lanes{:}\ 4) - 0.0427(Width\ of\ Bike\ Lane{:}\ 4{\sim}8ft)$		
		$CMF = 0.742$	$CMF = 0.732$	$CMF = 0.700$
		2.97	7.32	5.60

Before Treatment



After Treatment (Addition of Bike Lanes)

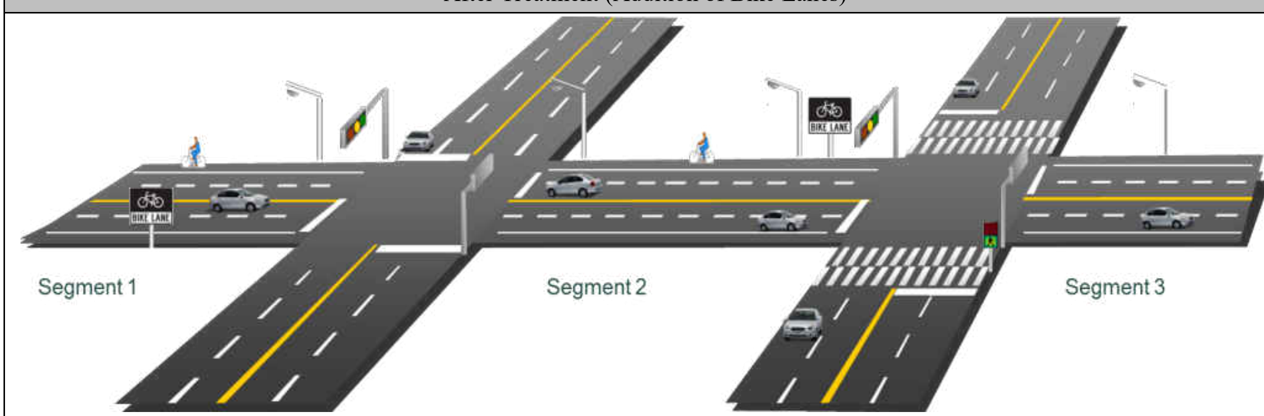


Figure 11-1: Implication scenario of using simple and full CMFunctions

REFERENCES

Abbas, K.A., 2004. Traffic Safety Assessment and Development of Predictive Models for Accidents on Rural Roads in Egypt. Accident Analysis and Prevention 36 (2), pp. 149–163.

Abdel-Aty, M.A., and Radwan, E., 2000. Modeling Traffic Accident Occurrence and Involvement. Accident Analysis and Prevention, Vol. 32, No. 5, pp. 633-642.

Abdel-Aty, M., and Haleem, K., 2011. Analyzing Angle Crashes at Unsignalized Intersection using Machine Learning Techniques. Accident Analysis and Prevention 43, pp. 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.

Abdel-Aty, M.A., Lee, J., Siddiqui, C., Choi, K., 2013. Geographical Unit Based Analysis in the Context of Transportation Safety Planning, to be published in Transportation Research Part A: Policy and Practice.

Abdel-Aty, M. A., Siddiqui, C., and Huang, H., 2011. Integrating Trip and Roadway Characteristics in Managing Safety at Traffic Analysis Zones. Transportation Research Record 2213, pp. 20-28.

Abraham, A., Steinberg, D. and 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., Abdel-Aty, M., and Park, J., 2015. Evaluation of the Safety Effectiveness of the Conversion of Two-Lane Roadways to Four-Lane Divided Roadways: Full Bayes vs. Empirical Bayes. Transportation Research Board 94th Annual Meeting, Washington, D.C.

Akgügör, A. P., Yıldız, O. 2007. Sensitivity analysis of an accident prediction model by the fractional factorial method. Accident Analysis and Prevention 39, 63–68.

Alkhatni, F., Kwigizile, V., Oh, J., 2014. Investigating Crash Frequency and Injury Severity at Freeway Weigh Stations in Michigan. TRB 93rd Annual Meeting.

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

American Association of State Highway Transportation Officials (AASHTO), 1999. Guide for the Development of Bicycle Facilities, Washington, D.C.

Amoros, E., Martin, J.L., Laumon, B., 2003. Comparison of road crashes incidence and severity between some French counties. Accident Analysis and Prevention 35 (4), pp. 537–547.

Anastasopoulos, P.C., Mannering, F.L, 2009. A note on modeling vehicle accident frequencies with random-parameters count models. Accident Analysis and Prevention 41, 153–159.

Aul, N. and Davis, G., 2006. Use of propensity score matching method and hybrid Bayesian method to estimate crash modification factors of signal installation. Transportation Research Record, 1950, 17-23.

Bahar, G., 2010. Methodology for the Development and Inclusion of Crash Modification Factors in the First Edition of the Highway Safety Manual, Transportation Research Circular, No. E-C-142, Transportation Research Board of the National Academies of Science, Washington D.C.

Banihashemi, M., 2012. Highway safety manual, calibration dataset sensitivity analysis. In: Proceedings of the TRB 2012 Annual Meeting, Washington, D.C.

Banihashemi, M., 2011. Highway safety manual, new model parameters vs. Calibration of crash prediction models. In: Proceedings of the TRB 2011 Annual Meeting, Washington, D.C.

Bateman, M., Howard, I., Johnson, A., Walton, J., 1998. Model of the performance of a roadway safety fence and its use for design. Transportation Research Record 1647, 122–129.

Bauer, K. M., and Harwood, D. W., 2013. Safety Effects of Horizontal Curve and Grade Combinations on Rural Two-Lane Roads. Presented at Presented at 92nd Annual Meeting of the Transportation Research Board, Washington, D.C.

Ben-Bassat, T., and Shinar, D., 2011. Effect of shoulder width, guardrail and roadway geometry on driver perception and behavior. Accident Analysis and Prevention 43, pp.2142–2152.

Briand, L., Freimut, B., Vollei, F., 2004. Using multiple adaptive regression splines to support decision making in code inspections. Journal of Systems and Software 73, 205–217.

Brimley, B., Saito, M., Schultz, G.G., 2012. Calibration of highway safety manual safety performance function and development of new models for rural two-lane two-way highways. TRB 2012 Annual Meeting, Washington D.C.

Buddhavarapu, P., Banerjee, A., Prozzi, J.A., 2013. Influence of pavement condition on horizontal curve safety. Accident Analysis and Prevention 52, 9–18

Cafiso, S., Di Silvestro, G., Persaud, B., Begum, M.A., 2010. Revisiting The Variability Of The Dispersion Parameter of Safety Performance Functions Using Data for Two-Lane Rural Roads.Transportation Research Record 2148, pp. 38-46.

Carrasco, O., McFadden, J., Chandhok, P., and 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.

Carson, J., Mannering, F., 2001. The Effect of Ice Warning Signs on Accident Frequencies and Severities. Accident Analysis and Prevention 33 (1), pp. 99–109.

Carter, D., Srinivasan, R., Gross, F., Council, F., 2012. Recommended Protocols for Developing Crash Modification Factors. Final Report NCHRP 20-7(314).

Chen, L., Chen, C., Srinivasan, R., McKnight, E., Ewing, R., Roe, M., 2012. Evaluating the Safety Effects of Bicycle Lanes in New York City, American Jourbal of Public Health 102(6), pp.1120-1127.

Chen, X., Qi, Y., Lu, Y., 2014. Safety Impacts of Using Short Left-Turn Lanes at Unsignalized Median Openings. TRB 93rd Annual Meeting.

Council, F.M., and Stewart, J.R., 2000. Safety Effects of the Conversion of Rural Two-Lane to Four-Lane Roadways Based on Cross-Sectional Models. FHWA-990327. FHWA, U.S. Department of Transportation.

Daniels, S., Brijs, T., Nuyts, E., Wets, G., 2010. Explaining Variation in Safety Performance of Roundabouts. Accident Analysis and Prevention 42(2), pp. 393-402.

De Guevara, F. L., Washington, S. P., and Oh, J., 2004. Forecasting Crashes at the Planning Level: Simultaneous Negative Binomial Crash Model Applied in Tucson, Arizona. Transportation Research Record 1897, pp. 191-199.

Dell'Acqua, G., Busiello, M., Russo, F., Biancardo, S., 2014. Highway Safety Manual Transferability to Italian Low-Volume Roads. TRB 93rd Annual Meeting.

Donnell, E.T., Mason, J.M., 2006. Predicting the frequency of median barrier crashes on Pennsylvania interstate highways. Accident Analysis and Prevention 38, pp.590–599.

El-Basyouny, K., and Sayed, T., 2012a. Measuring Direct and Indirect Treatment Effects Using Safety Performance Intervention Functions. Safety Science, Vol. 50 (4), 2012, pp. 1125-1132.

El-Basyouny, K., and Sayed, T., 2012b. Measuring Safety Treatment Effects Using Full Bayes NonLinear Safety Performance Intervention Functions. Accident Analysis and Prevention, Vol. 45, pp. 152-163.

El-Basyouny, K., and Sayed, T., 2011. A full Bayes multivariate intervention model with random parameters among matched pairs for before–after safety evaluation. Accident Analysis and Prevention, Vol. 43(1), pp. 87–94.

El-Basyouny, K. and Sayed, T., 2010. Full Bayes Approach to Before-and-After Safety Evaluation with Matched Comparisons. Transportation Research Record: Journal of the Transportation Research Board, Vol. 2148, pp. 1-8.

El-Basyouny, K., and Sayed, T., 2009. Collision prediction models using multivariate Poisson-lognormal regression Accident Analysis & Prevention 41, pp. 820–828

El-Basyouny, K., and Sayed, T., 2006. Comparison of Two Negative Binomial Regression Techniques in Developing Accident Prediction Models. Transportation Research Record 1950, pp. 9–16.

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. Accident Analysis and Prevention 40 (3), 1033–1054.

Elvik, R., 2013. International Transferability of Accident Modification Functions for Horizontal Curves. Accident Analysis and Prevention 59, pp.487-496.

Elvik, R., 2011. Developing an Accident Modification Function for Speed Enforcement. Safety Science 49, pp.920-925.

Elvik, R., 2009. An Exploratory Analysis of Models for Estimating the Combined Effects of Road Safety Measures. Accident Analysis and Prevention 41, pp.876–880.

Elvik, R., 2009. Developing Accident Modification Functions. Transportation Research Record 2103, pp. 18-24.

Elvik, R., 2005. Introductory Guide to Sytematic Reviews and Meta-Analysis. Transportation Research Record 1908, pp.230-235.

Elvik, R., 1995. The safety value of guardrails and crash cushions: a meta-analysis of evidence from evaluation studies. Accident Analysis and Prevention 27 (4), pp.523–549.

Elvik, R., Hoye, A., Vaa, T., Sorensen, M., 2009. The handbook of road safety measures, 2nd edition.

Friedman, J., 1991. Multivariate adaptive regression splines. Annals of Statistics 19, 1–141.

Federal Highway Administration Office of Safety, 2008. A resident's guide for creating safe and walkable communities, FHWA-SA-07-016, Washington, DC.

Fitzpatrick, K., Schneider, W.H., and Park, E.S., 2005. Comparisons of Crashes on Rural Two-lane and Four-lane Highways in Texas. Report-046181, Texas Department of Transportation.

Friedman, J., 1991. Multivariate adaptive regression splines. Annals of Statistics 19, 1–141.

Gamerman, D., 2006. Markov Chain Monte Carlo Stochastic Simulation for Bayesian inference. Taylor & Francis, Boca Raton.

Gan, A., Haleem, K., Alluri, P., Lu, J., Wang, T., Ma, M., Diaz, C., 2012. Preparing Florida for Deployment of SafetyAnalyst for All Roads, Final Report BDK80 977-07.

Garber, N. J. and Hoel, L. A., 2002. Traffic and Highway Engineering: ThirdEdition, Brooks/Cole Publishing Company, Pacific Grove, CA.

Gattis, J., Varghese, J., Toothaker, L., 1993. Analysis of guardrail-end accidents in Oklahoma. Transportation Research Record 1419, 52–62.

Gelman, A., Carlin, J.B., Stern, H.S., and Rubin, D.B., 2004. Bayesian Data Analysis, Seconded, Champman and Hall, London.

Gelman, A., 2006. Prior Distribution for Variance Parameters in Hierarchical Models. Bayesian Analysis 1.3, pp. 515-534.

Griffin, L. I., Mak, K. K., 1987. The Benefits to be Achieved from Widening Rural, Two-Lane Farm-to-Market Roads in Texas, Report No. IAC(86-87) - 1039, Texas Transportation Institute, College Station, Texas.

Griffith, M.S., 1999. Safety Evaluation of Rolled-in Continuous Shoulder Rumble Strips Installed on Freeways. Transportation Research Record, No. 1665, Transportation Research Board, National Research Council, Washington, D.C., pp. 28-34.

Gilks, W.R., Richardson, S., Speigelhalter, D.J., 1996. Markov Chain Monte Carlo in Practice. Chapman and Hall, London

Good, M., Fox, J., Joubert, P., 1987. An in-depth study of accidents involving collisions with utility poles. Accident Analysis and Prevention 19 (5), 397–413.

Gross, F., 2013. Case–control analysis in highway safety: Accounting for sites with multiple crashes. Accident Analysis and Prevention 61, 87-96.

Gross, F., and Hamidi, A., 2011. Investigation of Existing and Alternative Methods for Combining Multiple CMFs, Highway Safety Improvement Program Technical Support, T-06-013.

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: Journal of the Transportation Research Board, No. 2103, 42-49.

Gross, F., Jovanis, P. P., 2007. Estimation of the safety effectiveness of lane and shoulder width: Case-control approach. Journal of Transportation Engineering, Vol.133, 362-369.

Hadi, M. A., J. Aruldhas, L. Chow, and J. Wattleworth., 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 No. 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. Accident Analysis and Prevention 55, 12-21.

Hanley, K. E., Gibby, A. R., and Ferrara, T. C., 2000. Analysis of Accident Reduction Factors on California State Highways, Transportation Research Record, No. 1717, Transportation Research Board, National Research Council, Washington, D.C., pp. 37-45.

Harb, R., Radwan, E., Su, X., 2010. Empirical analysis of toll-lane processing times using proportional odds augmented MARS. ASCE: Journal of Transportation Engineering 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.

Hauer, E., Council, F. M., Mohammedshah, Y., 2004. Safety models for urban four-lane undivided road segments. Transportation Research Record 1897, 96-105.

Hauer, E., Harwood, D., Council, F., and Griffith, M., 2002. Estimating safety by the empirical bayes method: a tutorial. Transportation Research Record, No. 1784, Transportation Research Board, National Research Council, Washington, D.C., pp. 126-131.

Hauer, E., Ng, J. C. N., and Lovell, J., 1988. Estimation of Safety at Signalized Intersections. Transportation Research Record 1185, pp. 48–61.

Hauer, E., 2000. Lane width and Safety. http://ezrahauer.files.wordpress.com/2012/08/lane-width-and-safety.pdf. Accessed March 2014.

Hauer, E., 1997. Observational Before–after Studies in Road Safety: Estimating the Effect of Highway and Traffic Engineering Measures on Road Safety, Pergamon Press, Elsevier Science Ltd., Oxford, UK.

Highway Capacity Manual (HCM), 2010. Transportation Research Board, National Research Council, Washington, D.C.

Hirst, W.M., Mountain, L.J., Maher, M.J., 2004. Sources of error in road safety scheme evaluation: a method to deal with outdated accident prediction models. Accident Analysis and Prevention 36 (5), pp. 717–727.

Howard, L., Steven, S., 2012. Calibration of highway safety manual prediction method for rural Kansas highways. TRB 2012 Annual Meeting, Washington, D.C.

Huang, H., Abdel-Aty, M., and Darwiche, A. L., 2010. County-level Crash Risks Analysis in Florida: Bayesian Spatial Modeling. Transportation Research Record 2148, pp. 27-37.

Huang, H.F., Stewart, J.R., and Zegeer, C.V., 2002. Evaluation of lane reduction 'road diet' measures on crashes and injuries, Transportation Research Record: Journal of the Transportation Research Board, No. 1784, Transportation Research Board of the National Academies, Washington, D.C., pp. 80–90.

Jang, H., Lee, S., and Kim, S. W., 2010. Bayesian Analysis for Zero-Inflated Regression Models with the Power Prior: Applications to Road Safety Countermeasures. Accident Analysis and Prevention 42, pp. 540-547.

Karlaftis, M. G., and Tarko, A. P., 1998. Heterogeneity Considerations in Accident Modeling, Accident Analysis and Prevention 30(4), pp .425-433.

Kennedy, J., 1997. Effect of light poles on vehicle impacts with roadside barriers. Transportation Research Record 1599, 32–39.

Khan, M., Abdel-Rahim, A., Williams, C., 2014. Potential Crash Reduction Benefits of Shoulder Rumble Strips in Two-Lane Rural Highways. TRB 93rd Annual Meeting.

Kim, D., Washington, S., 2006. The Significance Of Endogeneity Problems in Crash Models: An Examination of Left-Turn Lanes in Intersection Crash Models. Accident Analysis and Prevention 38(6), pp. 1094–1100.

Kononov, J., Bailey, B., and Allery, B., 2008. Exploratory Analysis of Relationship between the Number of Lanes and Saftey on Urban Freeways. Transportation Research Board 87th Annual Meeting, Washington, D.C.

Kuhfeld, W.F., and Cai, W., 2013. Introducing the New ADAPTIVEREG Procedure for Adaptive Regression. SAS Global Forum: Statistics and Data Analysis, Paper 457-2013.

Jensen, S. U., 2008. Bicycle Tracks and Lanes: a Before-After Study. The 87th TRB Annual Meeting Compendium of Papers CD-ROM. Washington, D.C.

Jovanis, P., and Gross, F., 2008. Estimation of Safety Effectiveness of Changes in Shoulder Width with Case Control and Cohort Methods. Transportation Research Record, No. 2019, Transportation Research Board, National Research Council, Washington, D.C., pp. 237-245.

Labi, S., 2011. Efficacies of roadway safety improvements across functional subclasses of rural two-lane highways. Journal of Safety Research 42, 231-239.

Lan, B., Persaud, B., Lyon, C. and Bhim, R., 2009. Validation of a full Bayes methodology for observational before-after road safety studies and application to evaluation of rural signal conversions. Accidents Analysis and Prevention, 41 (3), 574–580.

Lao, Y, Zhang, G., Wang, Y., Milton, J., 2013. Generalized nonlinear models for rear-end crash risk analysis. Accident Analysis and Prevention 62, 9-16.

Lee, C., Abdel-Aty, M., Park, J., and Wang, J., 2015. Development of crash modification factors for changing lane width on roadway segments using generalized nonlinear models. Accident Analysis and Prevention 75, pp.83-91.

Lee, J., Abdel-Aty, M A.., Siddiqui, C., and Choi, K., 2013. Analysis of Residence Characteristics of Drivers, Pedestrians and Bicyclists Involved in Traffic Crashes. Presented in Transportation Research Board 92nd Annual Meeting.

Lee, J., and Mannering, F., 2002. Impact of roadside features on the frequency and severity of run-off-roadway accidents: an empirical analysis. Accident Analysis and Prevention 34, pp.149–161.

Lee, J., Nam, B., Abdel-Aty, M., 2015. Effects of Pavement Surface Conditions on Traffic Crash Severity. Journal of Transportation Engineering, 10.1061/(ASCE)TE.1943-5436.0000785 , 04015020.

Levine, N., Kim, K. E., and Nitz, L. H., 1995. Spatial Analysis of Honolulu Motor Vehicle Crashes: II. Zonal Generators. Accident Analysis and Prevention 27, pp. 675-685.

Li, H., Graham, D.J., Majumdar, A., 2014. Effects of Changes in Road Networks Characteristics on Road Casualties: An Application of Full Bayes Models Using Panel Data, in: Transportation Research Board 93rd Annual Meeting.

Li, W., Carriquiry, A., Pawlovich, M. and Welch, T., 2008a. The choice of statistical models in road safety countermeasures effectiveness studies in Iowa. Accident Analysis and Prevention, 40 (4), 1531–1542.

Li, X., Lord, D., Zhang, Y., Xie, Y., 2008b. Predicting motor vehicle crashes using support vector machine models. Accident Analysis & Prevention 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. Journal of Transportation Engineering 137, 74-83.

Li, Y., Liu, C., Ding, L., 2013. Impact of pavement conditions on crash severity. Accident Analysis and Prevention59, 399–406

Liu, C., Utter, D., Chen, C.L., 2007. Characteristics of Crash Injuries Among Young, Middle-Aged, and Older Drivers, National Highway Traffic Safety Administration: Technical Report.

Lord, D., 2006. Modeling Motor Vehicle Crashes Using Poisson-Gamma Models: Examining the Effects of Low Sample Mean Values and Small Sample Size on the Estimation of the Fixed Dispersion Parameter. Accident Analysis and Prevention 38(4), pp. 751–766.

Lord, D., Bonneson, J. A., 2007. Development of Accident Modification Factors for Rural Frontage Road Segments in Texas. Transportation Research Record: Journal of the Transportation Research Board, No. 2023, pp. 20-27.

Lord, D., Geedipally, S.R., and Guikema, S., 2010. Extension of The Application Of Conway–Maxwell–Poisson Models: Analyzing Traffic Crash Data Exhibiting Underdispersion. Risk Analysis 30(8), pp.1268-1276.

Lord, D., Manar, A., and Vizioli, A., 2005. Modeling Crash-Flow-Density and Crash-Flow-V/C Ratio for Rural and Urban Freeway Segments. Accident Analysis and Prevention 37 (1), pp. 185–199.

Lord, D., Mannering, F.L., 2010. The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives. Transportation Research Part A: Policy and Practice, vol 44 (5), 291-305.

Lu, L., Lu, J., Lin, P., Wang, Z., and Chen, H., 2009. Development of an Interface between FDOT's Crash Analysis Reporting System and the Safety Analyst. Final Report BDK84 977-01.

Lusk, A.C., Furth, P.G., Morency, P., Miranda-Moreno, L.F., Willet, W.C., Dennerlein, J.T., 2011. Risk of Injury for Bicycling on Cycle Rracks versus in the Street. Injury Prevention 17, pp.131-135.

Ma, J., Kockelman, K., Damien, P., 2008. A multivariate Poisson-lognormal regression model for prediction of crash counts by severity, using Bayesian methods Accident Analysis & Prevention 40, pp. 964–975

Ma, J., Kockelman, K., 2006. Bayesian multivariate Poisson regression for models of injury count, by severity Transportation Research Record: Journal of the Transportation Research Board 1950, pp. 24–34

Males, M. A., 2009. Poverty as a Determinant of Young Drivers' Fatal Crash Risks. Journal of Safety Research 40, pp. 443-448.

Malyshkina, N., and Mannering, F., 2010. Empirical Assessment of the Impact of Highway Design Exceptions on the Frequency and Severity of Vehicle Accidents. Accident Analysis and Prevention 42(1), pp. 131–139.

Maycock, G., and Hall, R. D., 1984. Accidents at 4-Arm Roundabouts. TRRL Laboratory Report 1120, Transportation and Road Research Laboratory, Crowthorne, UK.

Mehta, G., Lu, Y., 2013. Calibration and development of safety performance functions for Alabama: Two-lane, two-way rural roads and four-lane divided highways. Transportation Research Record: Journal of the Transportation Research Board, No. 2398, 75-82.

Miaou, S.-P., 1994. The Relationship between Truck Accidents and Geometric Design of Road Sections: Poisson versus Negative Binomial Regressions. Accident Analysis and Prevention 26(4), pp. 471–482.

Miaou, S.-P., and Lord, D., 2003. Modeling Traffic Crash-Flow Relationships for Intersections: Dispersion Parameter, Functional Form, and Bayes versus Empirical Bayes. Transportation Research Record 1840, pp. 31–40.

Miaou, S.P., Bligh, R.P., Lord, D., 2005. Developing guidelines for median barrier installation: Benefit-cost analysis with Texas data. Transportation Research Record: Journal of the Transportation Research Board 1904, pp.3-19.

Milton, J., Mannering, F., 1998. The Relationship Among Highway Geometrics, Traffic-Related Elements and Motor Vehicle Accident Frequencies. Transportation 25 (4), pp. 395–413.

Michie, J.D., and Bronstad, M.E., 1994. Highway guardrails: safety feature or roadside hazard?. Transportation Research Record: Journal of the Transportation Research Board 1468, pp.1-9.

Mohammadi, M., Samaranayake, V., Bham, G., 2014. Safety Effect of Missouri's Strategic Highway Safety Plan-Missouri's Blueprint for Safer Roadways. TRB 93rd Annul Meeting.

Munkin, M., Trivedi, P., 2002. Simulated maximum likelihood estimation of multivariate mixed-poisson regression models, with application. Econometrics Journal 2, 29–48.

Naderan, A., and Shahi, J., 2010. Aggregate Crash Prediction Models: Introducing Crash Generation Concept. Accident Analysis and Prevention 42, pp. 339-346.

National Cooperative Highway Research Program (NCHRP), 2012. Recommended Protocols for Developing Crash Modification Factors. Final Report NCHRP 20-7(314).

NHTSA, 2013. 2012 Motor Vehicle Crashes: Overview, U.S. Department of Transportation, National Highway Traffic Safety Administration.

Noland, R. B, 2003. Traffic Fatalities and Injuries: The Effect of Changes in Infrastructure and Other Trends. Accident Analysis and Prevention 35, pp. 599-611.

Nosal, T., Miranda-Moreno, L.F., 2012. Cycle-tracks, bicycle lanes & on-street cycling in Montreal: a preliminary comparison of the cyclist injury risk. Presented at the 91st Annual Meeting of the Transportation Research Board, Washington, D.C.

Park, E. S., Carlson, P. J., Porter, R. J., Andersen, C. K., 2012. Safety effects of wider edge lines on rural, two-lane highways. Accident Analysis and Prevention 48, 317– 325.

Park, E.S., Lord, D., 2007. Multivariate poisson-lognormal models for jointly modeling crash frequency by severity Transportation Research Record: Journal of the Transportation Research Board 2019, pp. 1–6.

Park, J., Abdel-Aty, M., Lee, C., 2014. Exploration and Comparison of Crash Modification Factors for Multiple Treatments on Rural Multilane Roadways. Accident Analysis and Prevention 70, pp.167-177.

Park, J., Abdel-Aty, M., Lee, J., and 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. Accident Analysis and Prevention 74, pp.179-191.

Park, J., Abdel-Aty, M., Wang, J., and Lee, C., 2015b. Assessment of Safety Effects for Widening Urban Roadways in Developing Crash Modification Functions using Nonlinearizing Link Functions. Accident Analysis and Prevention 79, pp.80-87.

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 and 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 and Prevention, In-press.

Patel, R.B., Council, F.M., and 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.

Pawlovich, M., Wen, L., Carriquiry, A., and Welch, T., 2006. Iowa's experience with road diet measures: use of Bayesian approach to assess impacts on crash frequencies and crash rates, Transportation Research Record: Journal of the Transportation Research Board, No. 1953, Transportation Research Board of the National Academies, Washington, D.C., pp. 163–171.

Pendleton, O., 1998. Application of New Crash Analysis Methodologies: Volume 1 - General Methodology, Texas Transportation Institute. The Texas A&M University System, College Station.

Persaud, B., Lan, B., Lyon, C. and Bhim, R., 2009. Comparison of empirical Bayes and full Bayes approaches for before-after road safety evaluations. Presented at the 89th Annual Meeting of the Transportation Research Board, Washington, D.C.

Persaud, B., and Lyon, C., 2007. Empirical Bayes before–after safety studies: Lessons learned from two decades of experience and future directions. Accident Analysis and Prevention 39, pp.546-555.

Persaud, B., Retting, R., Lyon, C., 2004. Crash Reduction Following Installation of Centerline Rumble Strips on Rural Two-lane Roads. Accident Analysis and Prevention 36, 1073-1079.

Persaud, B.P., and Nguyen, T., 1998. Disaggregate Safety Performance Models for Signalized Intersections on Ontario Provincial Roads. Transportation Research Record 1635, pp. 113–120.

Poch, M., Mannering, F., 1996. Negative Binomial Analysis of Intersection-Accident Frequencies. Journal of Transportation Engineering 122 (2), pp. 105–113.

Potts, I. G., Harwood, D. W., Richard, K. R., 2007. Relationship of lane width to safety for urban and suburban arterials. Transportation Research Record, 2023, 63-82.

Pitale, J.T., Shankwitz, C., Preston, H., and Barry, M., 2009. Benefit-Cost Analysis of In-Vehicle Technologies and Infrastructure Modifications as a Means to Prevent Crashes Along Curves and Shoulders. Minnesota Department of Transportation.

Put, R., Xu, Q., Massart, D. and Heyden, Y., 2004. Multivariate adaptive regression splines (MARS) in chromatographic quantitative structure–retention relationship studies. Journal of Chromatography A, 1055, pp. 11–19.

Qin, X., Ivan, J. N., Ravishanker, N., 2004. Selecting exposure measures in crash rate prediction for two-lane highway segments. Accident Analysis and Prevention 36, 183–191.

Ray, M., 1999. Impact conditions in side-impact collisions with fixed roadside objects. Accident Analysis and Prevention 31 (1), 21–30.

Reid, J., Sicking, D., Faller, R., Pfeifer, B., 1997. Development of a new guardrail system. Transportation Research Record 1599, 72–80.

Reynolds, C., Harris, M., Teschke, K., Cripton, P., Winters, M., 2009. The Impact of Transportation Infrastructure on Bicycling Injuries and Crashes: A Review of the Literature. Environmental Health 8(1), pp.47-66.

Rodegerdts, L. A., Nevers, B., Robinson, B., 2004. Signalized Intersections: Informational Guide. FHWA-HRT-04-091. Federal Highway Administration Turner-Fairbank Highway Research Center, McLean, VA.

Romano, E., Tippetts, S., Blackman, K., and Voas, R., 2006. Language, Income, Education and Alcohol-Related Fatal Motor Vehicle Crashes. Journal of Ethnicity in Substance Abuse 5(2), pp. 119-137.

Sacchi, E., Persaud, B., Bassani, M., 2012. Assessing international transferability of the highway safety manual crash prediction algorithm and its components. In: Proceedings of the 2012 TRB Annual Meeting, Washington, D.C.

Sacchi, E., and Sayed, T., 2014. Accounting for Heterogeneity among Treatment Sites and Time Trends in Developing Crash Modification Functions. Accident Analysis and Prevention 72, pp.116-126.

Sacchi, E., Sayed, T., El-Basyouny, K., 2014. Collision Modification Functions: Incorporating changes over time. Accident Analysis and Prevention 70, pp.46-54.

Sadek, W. A., Dickason, A., and Kaplan, J., 2007. Effectiveness of a Green, High-Visibility Bike Lane and Crossing Treatment. Presented at 86th Annual Meeting of the Transportation Research Board, Washington, D.C.

SAS Institute Inc., 2014. SAS Enterprise Miner Workstation 7.1 Product Document. http://support.sas.com/documentation/onlinedoc/miner/#miner71

SAS Institute Inc. SAS/STAT® 9.2 User's Guide, Second Edition. Cary, NC, 2009. http://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/viewer.htm#titlepage.htm (Accessed in April, 2014)

Schick S., 2009. Accident Related Factors. Traffic Accident Causation in Europe (TRACE): Project No. 027763.

Shankar, V. N., Albin, R. B., Milton, J.C., and Mannering, F.L., 1998. Evaluating Median Cross-Oover Likelihoods with Clustered Accident Counts: An Empirical Inquiry Using Random Effects Negative Binomial Model. Transportation Research Record 1635, pp. 44–48.

Shen, Q., 2007. Development of Safety Performance Functions For Empirical Bayes Estimation of Crash Reduction Factors. (Ph. D Thesis): Florida International University, 206 p.

Snedecor, G.W., Cochran, W.G., 1980, Statistical Methods 7th edition, Iowa State University Press.

Spiegelhalter, D.J., Best, N.G., Carlin, B.P., and Lunn, D., 2005. WinBUGS User Manual. MRC Biostatistics Unit, Cambridge, Available from http://www.mrc-cam.ac.uk/bugs

Srinivasan, S., Haas, P., Dhakar, N., Hormel, R., Torbic, D., and Harwood, D., 2011. Development and Calibration of Highway Safety Manual Equations for Florida Conditions. Final Report BDK77 977-06.

Stamatiadis, N., Pigman. J., Sacksteder, J., Ruff, W., Lord, D., 2009. NCHRP Report 633: Impact of Shoulder Width and Median Width on Safety, Transportation Research Board of the National Academies, Washington, D.C.

Sun, X., Magri, D., H.Shirazi, H., Gillella, S., Li, L., 2012. Application of the highway safety manual: Louisiana experience with rural multilane highways. TRB 2012 Annual Meeting. Washington, D.C.

Tarko, A., S. Eranky, and K. Sinha, Methodological Considerations in the Development and Use of Crash Reduction Factors. Paper Presented at the 77th Annual Meeting of the Transportation Research Board, Washington, D.C., 1998.

Tarko, A.P., Villwock, N.M., Blond, N., 2008. Effect of Median Design on Rural Freeway Safety: Flush Medians with Concrete Barriers and Depressed Medians. Transportation Research Record: Journal of the Transportation Research Board 2060, pp.29-37.

The ADAPTIVEREG Procedure, 2012. SAS Institute Inc., Cary, North Carolina, USA.

The NLMIXED Procedure., 2009. SAS Institute Inc., Cary, North Carolina, USA, http://support.sas.com/onlinedoc/913/docMainpage.jsp Accessed April 2014.

Turner, B., Affum, J., Tziotis, M. and Jurewicz, C., 2009. Review of iRAP Risk Parameters. ARRB Group Contract Report for iRAP.

Turner, B. Steinmetz, L., Lim, A. and Walsh, K., 2012. Effectiveness of Road Safety Engineering Treatments. APR422-12. Austroads Project No: ST1571.

Turner, B., 2011. Estimating the Safety Benefits when Using Multiple Road Engineering Treatments, Road Safety Risk Reporter, 11.

Turner, D., 1984. Prediction of bridge accident rates. Journal of Transportation Engineering 110 (1), 45–54.

Ukkusuri, S., Hasan S., and Aziz, M. H. A., 2011. Random Parameter Model Used to Explain Effects of Built-Environment Characteristics on Pedestrian Crash Frequency, Transportation Research Record 2237, pp. 98-106.

University of North Carolina Highway Safety Research Center, 2010. Crash Modification Factors Clearinghouse. (U.S. Department of Transportation Federal Highway Administration) Retrieved 2014, from http://www.cmfclearinghouse.org

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. Accident Analysis and Prevention 59, 309–318.

Viner, J., 1995. Rollovers on sideslopes and ditches. Accident Analysis and Prevention 27 (4), 483–491.

Wang, J., Abdel-Aty, M., Park, J., Lee, C., and Kuo, P., 2015. Estimating safety performance trends over time for treatments at intersections in Florida. Accident Analysis and Prevention 80, pp.37-47.

Wang, X., and Abdel-Aty, M. A., 2006. Temporal and Spatial Analyses of Rear-End Crashes at Signalized Intersections. Accident Analysis and Prevention 38(6), pp. 1137–1150.

Washington, S., Congdon, P., Karlaftis, M., and Mannering, F., 2005. Bayesian Multinomial Logit Models: Exploratory Assessment of Transportation Applications. Transportation Research Board 84th Annual Meeting, Washington, D.C.

Washington, S.P., Karlaftis, M.G., Mannering, F.L., 2010. Statistical and Econometric Methods for Transportation Data Analysis. Second Edition, Chapman Hall/CRC, Boca Raton, FL.

Washington, S., Schalkwyk, I. V., You, D., Shin, K., and Samuelson, J. P., 2010. PLANSAFE User Manual: Forecasting the Safety Impacts of Socio-Demographic Changes and Safety Countermeasures. NCHRP 8-44(2), http://www.trb.org/Main/Blurbs/PLANSAFE_Forecasting_the_Safety_Impacts_of_SocioDe_163790.aspx.

Wong, S.C., Sze, N.N., and Li, Y.C., 2007. Contributory Factors to Traffic Crashes at Signalized Intersections in Hong Kong. Accident Analysis and Prevention 39, pp. 1107-1113.

Wu, K.F., Donnell, E.T., and Aguero-Valverde, J., 2014. Relating Crash Frequency and Severity: Evaluating the Effectiveness of Shoulder Rumble Strips on Reducing Fatal and Major Injury Crashes. Accident Analysis and Prevention 67, pp.86-95.

Xie, F., Gladhill, K., K.Dixon, K., M.Monsere, C., 2011. Calibrating the highway safety manual predictive models for oregon state highways. In: Proceedings of the TRB 2011 Annual Meeting, Washington, D.C.

Xie, Y., Lord, D., Zhang, Y., 2007. Predicting motor vehicle collisions using Bayesian neural network models: An empirical analysis. Accident Analysis and Prevention 39, 922–933.

Xu, P., Huang, H., 2015. Modeling crash spatial heterogeneity: Random parameter versus geographically weighting. Accident Analysis and Prevention75, 16–25.

Yang, Q., Overton, R., Han, L., Yan, X., Richards, S. H., 2013. The influence of curbs on driver behaviors in four-lane rural highways—A driving simulator based study. Accident Analysis and Prevention 50, 1289– 1297.

Yanmaz-Tuzel, O., Ozbay, K., 2010. A comparative Full Bayesian before-and-after analysis and application to urban road safety countermeasures in New Jersey. Accident Analysis and Prevention 42, 2099–2107.

Zegeer, C., Council, F., 1995. Safety relationships associated with cross-sectional roadway elements. Transportation Research Record 1512, 29–36.

Zegeer, C. V., Reinfurt, D. W., Hummer, J., Herf, L., Hunter, W., 1988. Safety effects of cross-section design for two-lane roads. Transportation Research Record, 1195, 20-32.

Zegeer, C.V., Sundstrom, C.A., Hummer, J.E., Rasdorf, W., Findley, D.J., 2012. Suggestions on how agencies should apply the highway safety manual to two-lane road curves. In: Proceedings of the TRB 2012 Annual Meeting, Washington, D.C.

Zeng, h., and Schrock, S., 2013. Safety-Effectiveness of Various Types of Shoulders on Rural Two-Lane Roads in Winter and Non-winter Periods. Presented at 92nd Annual Meeting of the Transportation Research Board, Washington, D.C.

Zeng, H., Fontaine, M.D., Smith, B.L., 2014. Estimation of the Safety Effect of Pavement Condition on Rural Two-Lane Highways. TRB 93rd Annual Meeting.

Zhang, Y., Xie, Y., Li, L., 2012. Crash frequency analysis of different types of urban roadway segments using generalized additive model. Journal of Safety Research 43, 107-114.

Zhou, Y., Dixon, K.K., 2012. Comparing highway safety manual predictive method to traditional ranking methods: Case study of intersections in Corvallis, Oregon. TRB 2012 Annual Meeting, Washington D.C.

Zou, Y., Tarko, A.P., Chen, E., Romero, M.A., 2014. Effectiveness of cable barriers, guardrails, and concrete barrier walls in reducing the risk of injury. Accident Analysis and Prevention 72, pp.55–65.