Safety Effects for Widening Shoulder Width on Rural Multi-lane Roadways in Developing Crash Modification Functions using Multivariate Adaptive Regression Splines

Juneyoung Park, Ph.D *
Mohamed Abdel-Aty, Ph.D., P.E.

Department of Civil, Environmental and Construction Engineering
University of Central Florida
Orlando, Florida 32816-2450
(407) 823-0300
jypark@knights.ucf.edu

* Corresponding Author

November 2015

Prepared for Presentation at the 95th Transportation Research Board Annual Meeting and Possible Publication in the Transportation Research Record

Word count: 4,935 words + 6 tables = 6,435 equivalent words
ABSTRACT

This study assesses the safety effectiveness of widening shoulder widths by developing crash modification factors (CMFs) and crash modification functions (CMFunctions). This study first evaluates the CMFs for widening shoulder widths on rural multilane roadways using the observational before-after with the empirical Bayes (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 multivariate adaptive regression splines (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. Crash and roadway characteristics data were collected for roadways in Florida, and the Florida-specific full SPFs for different crash types and severity levels were developed, and used for the observational before-after with EB estimation.

The results indicated that widening shoulder widths resulted in a crash reduction of 12 percent for total crashes (all severities), 18 percent for injury crashes, and 21 percent for severe crashes on rural multilane roadways for all types of crashes. For run-off roadway (ROR) crashes, the overall crash reduction of the treatment was 25 percent for total crashes, 28 percent for injury crashes, and 31 percent for severe crashes. Moreover, the results showed that the safety effects vary across the sites with different roadway characteristics. In particular, the original shoulder widths of treated sites in the before period and actual widened width are significant parameters that affect the variation of CMFs. Moreover, it was found that the MARS models show more reliable estimates than normal regression, if the variation of CMFs with specific parameters has nonlinear relationships and interaction effects. The proposed CMFunctions provide general insights into roadway design and selection of sites for widening shoulder width for reducing crashes.

Key words: Safety Effectiveness; Crash Modification Factors and Functions; Multivariate Adaptive Regression Splines; Observational Before-After with EB Method; Widening Shoulder Width
INTRODUCTION

Part D of the Highway Safety Manual (HSM) (1) presents a variety of crash modification factors (CMFs) for safety treatments on roadway segments and at intersections. A CMF can be used to compute the expected number of crashes after implementing a given countermeasure at a specific site. CMFs have been estimated using observational before-after studies or the cross-sectional methods. The before-after method evaluates the safety effects of treatments by comparing crash frequencies during the time periods before and after implementation of the treatments (2). On the other hand, the cross-sectional method is used when the before-after study cannot be used due to the following conditions: 1) the date of the treatment installation is unknown, 2) the data for the period before treatment installation are not available, and 3) the effects of other factors on crash frequency must be controlled for creating a crash modification function (CMFunction) (3; 4). The cross-sectional method is also known as safety performance functions (SPFs) or crash prediction models.

Although the HSM provides a large number of estimates of safety effects (i.e. CMFs) of roadway treatments, the effect is stated in terms of a fixed single factor. Since this single value represents an average safety effect of the treatment for all treated sites, the heterogeneous effects of roadway characteristics on CMFs among treated sites are ignored. In order to overcome this limitation, CMFunctions have been developed to predict the variation in CMFs based on the site characteristics. Elvik (5) proposed an approach to estimate CMFunctions for the same or similar treatments by means of meta-regression analysis (6). By using this framework, Elvik (7) estimated CMFunctions for the installation of bypass and converting signalized intersections to roundabouts based on population changes. The results showed that the CMFs increase with population for both treatments. However, in order to develop good CMFunctions, fairly large amounts of data are needed. Elvik (8) also identified the variation of safety effectiveness of horizontal curves with the lengths of their radii. He estimated the summary CMFunction to assess the international transferability of national CMFunctions from 10 different countries. The results showed that the estimated CMFunction can be a summary of these national functions. It was found that the safety effects decrease as the radius of the curve decreases and that the variations in safety performance appear to be same for all 10 countries.

In order to reflect the nonlinear relationship between CMFs and roadway characteristics, Elvik (9) applied five 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. It was found that higher levels of enforcement are associated with a reduction in crash frequencies. The non-linear logarithmic function best fitted the data points from 13 previous studies but the inverse function also fitted the data well. Similar to this study, Park et al. (10) applied one linear and four non-linear functions to develop CMFunctions for adding shoulder rumble strips, widening shoulder width, and combination to check the relationship between CMFs and original shoulder width of treated sites. It was found that the exponential nonlinear regression was the best fitted function to develop CMFunctions.
However, other traffic and roadway characteristics (e.g. AADT, actual widened width, etc.) were not considered in developing CMFunctions. To consider the variation of CMFs over time, Park et al. (11) utilized nonlinearizing link functions in developing CMFunctions. Similarly, Sacchi et al. (12) developed CMFunctions to incorporate the effectiveness of a safety treatment (signal head upgrade program) with changes over time using poisson-lognormal linear intervention and non-linear intervention models. Sacchi and Sayed (13) 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. Although they identified the variation of safety effects with both AADT changes and time trends, only single roadway characteristic was used to develop the CMFunction. Park et al. (14) applied one linear and four non-linear functions to develop CMFunctions for adding a bike lane on urban arterials. The results of CMFunctions show that inverse, quadratic, and exponential nonlinear regression models were the best fitted functions for different single roadway characteristics. It was found that CMFunctions with multiple parameters show better model fit than simple models. The study also showed that the multiple regression models with backward and stepwise subset selections were the best fitted for different multiple roadway characteristics. However, there is a lack of prior studies on variation in the safety effects of widening shoulder widths among treated sites with different original shoulder width and actual widened width.

Thus, the objective of this study is to identify the relationship between the safety effects of widening shoulder width and roadway characteristics through 1) evaluation of CMFs using the observational before-after with EB method and 2) development of CMFunctions based on different site characteristics and actual widened width to reflect variation of the safety effects. The remainder of this study is organized as follows. The second section reviews the past studies on the safety effects of shoulder width. The third section describes data collection and preparation. The fourth section describes methodologies of estimating the CMFs and CMFunctions. The fifth section presents and discusses the results. The final section draws conclusions. In this paper, we define 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)’. 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).

SAFETY EFFECTS OF SHOULDER WIDTH

Roadside elements have been known as one of the most important cross-sectional roadway characteristics for safety. In particular, widening shoulder width is one of the roadside safety countermeasures that have been known to be effective in reducing crashes. Hadi et al. (15)
developed negative binomial models to analyze safety effects of shoulder width on rural multilane highways. The study shows that a small reduction in crashes (1 ~ 3%) could be attained if the unpaved shoulder is widened by 1 ft. Jovanis and Gross (16) estimated safety effects of shoulder width using case control and cohort methods. The results of the two methods showed that crashes decrease as the shoulder width increases. Moreover, Harkey et al. (17) developed CMFs for shoulder width on rural multilane roadways with greater than 2,000 vehicles per day. It was found that increasing shoulder width is safety effective in reducing crashes. Zeng and Schrock (18) 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. Park et al. (10) assessed the safety effects of shoulder rumble strips and widening shoulder width on rural multilane roadways using the observational before-after with CG and EB methods, and the cross-sectional study. The study showed that two single treatments and combination are generally safety effective for total and SVROR crashes. The results also showed that for the roadway segments with shoulder widths of 9 ft or above, only one single treatment can show better safety effects than two treatments. Based on the results of total crashes (KABCO), shoulder rumble strips are more effective in reducing crashes for roadway segments with shoulder widths less than 7 ft, whereas widening shoulder width is more effective for roadway segments with shoulder widths of 7 ft or above. Park and Abdel-Aty (19) found similar results for the safety effects of shoulder rumble strips and widening shoulder width on rural two lane roadways.

However, according to Stamatiadis et al. (20), wider shoulders may encourage higher operating speeds because they may communicate to the driver the presence of wider space for correcting errors. Some studies explained associated interaction effects between lane width and shoulder width due to the differences in local conditions. Gross et al. (21) reported that the effects of lane width on crash frequency were neither consistently positive nor negative due to the variation in shoulder widths. Thus, they suggested that CMFs be determined considering interaction between lane width and shoulder width. Potts et al. (22) also recommended that narrowing lane width be used as a treatment based on local roadway characteristics such as shoulder width since the effect of lane width varies with location. However, it is worth to note that the variation of safety effects of shoulder width based on the original shoulder width and actual widened width has not been investigated.

DATA PREPARATION

In this study, more detailed roadway information and additional treated locations were obtained in addition to previously used dataset by the authors (10). Three sets of data maintained by Florida DOT were used in this study: roadway characteristic inventory (RCI) data for eight years (2004-2011), financial project information and crash analysis reporting system (CARS) database. The RCI database provides current and historical roadway characteristics data and reflects features of
specific segments for the selected dates. 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. There were no other geometric changes between 2006 and 2009 for the identified treated sites except shoulder width and AADT, respectively. 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 1.

### Table 1: Descriptive statistics of treated segments

<table>
<thead>
<tr>
<th>Variable</th>
<th>Crash frequency in before period</th>
<th>Crash frequency in after period</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>S.D.</td>
</tr>
<tr>
<td>Number of All (KABCO) crashes</td>
<td>4.037</td>
<td>6.773</td>
</tr>
<tr>
<td>Number of All (KABC) crashes</td>
<td>2.398</td>
<td>3.850</td>
</tr>
<tr>
<td>Number of All (KAB) crashes</td>
<td>1.506</td>
<td>2.467</td>
</tr>
<tr>
<td>Number of ROR (KABCO) crashes</td>
<td>0.950</td>
<td>2.041</td>
</tr>
<tr>
<td>Number of ROR (KABC) crashes</td>
<td>0.577</td>
<td>1.253</td>
</tr>
<tr>
<td>Number of ROR (KAB) crashes</td>
<td>0.407</td>
<td>0.909</td>
</tr>
</tbody>
</table>

Variables related to traffic and roadway geometric characteristics

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

### METHODOLOGY

#### Safety Performance Functions

Generally, a SPF relates the crash frequency to traffic and roadway characteristics. The negative binomial (NB) model is most commonly used to develop a SPF since the function can account for over-dispersion. A SPF can be estimated 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. The full SPF
relates the frequency of crashes to both traffic and roadway characteristics, whereas the simple SPF considers a traffic parameter only such as Annual Average Daily Traffic (AADT) as an explanatory variable. It is worth to note that the HSM provides the CMFs calculated based on the simple SPF only. However, the simple SPF is an over-simplified function to reflect the relationship between crash frequency and roadway characteristics since crash frequency is not only affected by the traffic volume (14). Thus, in this study, the full SPF was used for calculating CMFs in the before–after with EB method. The functional form of SPF for fitting the NB regression models is shown in Equation 1 as follows:

\[ N_{predicted,i} = \exp(\beta_0 + \beta_1 \ln(AADT_i) + \cdots + \beta_k(X_{ki})) \]

where,
- \( N_{predicted,i} \) = Predicted crash frequency on segment \( i \),
- \( \beta_k \) = coefficients for the variable \( k \),
- \( AADT_i \) = Annual Average Daily Traffic of segment \( i \) (veh/day),
- \( X_{ki} \) = Roadway characteristic \( k \) of segment \( i \) \((k>2)\). 

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 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. Generally, it has been known that wider lane width is effective in reducing crashes (23 - 25). On the other hand, some studies found opposite effects (4, 19, 26, 27). In particular, Hauer (28) 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 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 (29). 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 2: Florida specific calibrated SPFs for rural multilane roadways by crash type and severity level

<table>
<thead>
<tr>
<th>Crash types</th>
<th>Constant</th>
<th>Ln.AADT</th>
<th>Length</th>
<th>Before period (2004-2005)</th>
<th>Maximum speed limit</th>
<th>Median width</th>
<th>Lane width</th>
<th>Dispersion</th>
<th>Deviance</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>All (KABCO)</td>
<td>-13.9082</td>
<td>1.3072</td>
<td>1.0244</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1.4801</td>
<td>3507.5</td>
<td>13191.2</td>
</tr>
<tr>
<td>All (KABC)</td>
<td>-14.2983</td>
<td>1.3374</td>
<td>1.0163</td>
<td>(0.1445)</td>
<td>0.0125</td>
<td>0.0047</td>
<td>0.0953</td>
<td>1.3581</td>
<td>3166.6</td>
<td>10000.7</td>
</tr>
<tr>
<td>All (KAB)</td>
<td>-13.3037</td>
<td>1.1501</td>
<td>1.0093</td>
<td>(0.0344)</td>
<td>(0.0029)</td>
<td>0.0054</td>
<td>-</td>
<td>1.1965</td>
<td>2802.8</td>
<td>7443.2</td>
</tr>
<tr>
<td>ROR (KABCO)</td>
<td>-11.8034</td>
<td>0.8311</td>
<td>0.8701</td>
<td>(0.0027)</td>
<td>(0.0001)</td>
<td>-</td>
<td>-</td>
<td>1.5529</td>
<td>1857.8</td>
<td>3952.5</td>
</tr>
<tr>
<td>ROR (KABC)</td>
<td>-12.2116</td>
<td>0.7835</td>
<td>0.8644</td>
<td>(0.0888)</td>
<td>(0.0001)</td>
<td>-</td>
<td>-</td>
<td>1.3286</td>
<td>1431.5</td>
<td>2681.4</td>
</tr>
<tr>
<td>ROR (KAB)</td>
<td>-11.6202</td>
<td>0.6718</td>
<td>0.8292</td>
<td>(0.0992)</td>
<td>(0.0001)</td>
<td>-</td>
<td>-</td>
<td>1.0601</td>
<td>1167.6</td>
<td>1988.2</td>
</tr>
</tbody>
</table>

**Before and After with Empirical Bayes (EB) 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 yearly traffic trend, physical characteristics, and land use. The before-after with EB method, which is based on the research by Hauer (30), is a well-accepted approach to evaluate safety effects of treatments due to its statistical strength. The safety effectiveness of a treatment is calculated by comparing the observed number of crashes to the expected number of crashes in the after period. One of the main advantages of the before-after study with empirical Bayes 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 detailed procedure of EB method can be found from Hauer (30) and Hauer et al. (31). In the before-after with EB method, the expected number of crashes without treatment \(N_{expected,B}\) can be estimated using Equation 2.

\[
N_{expected,B} = w(N_{predicted,B}) + (1 - w)(N_{observed,B})
\]

where,

- \(N_{expected,B}\) = expected crash frequency in the before period,
- \(N_{predicted,B}\) = predicted crash frequency estimated using the SPF in the before period,
- \(N_{observed,B}\) = observed crash frequency in the before period,
$w = a \text{ weight factor estimated using over-dispersion parameter from the negative binomial model (SPF) and the predicted crash frequency in the before period for the treated site as shown in Equation 3.}$

$$w = \frac{1}{1 + k \times N_{\text{predicted,B}}}$$

where,

$k = \text{over-dispersion parameter.}$

The expected crash frequency in the after period can be calculated as a product of the expected crash frequency in the before period and the ratio of the predicted crash frequency in the after period to the predicted crash frequency in the before period. According to Persaud and Lyon (32) and Gross et al. (2), the CMF can be estimated as shown in the following Equation 4.

$$CMF = \frac{\left[\frac{N_{\text{observed,A}}}{N_{\text{expected,A}}}\right]}{1 + \frac{\text{Var}(N_{\text{expected,A}})}{N_{\text{expected,A}}^2}}$$

where,

$N_{\text{expected,A}} = \text{expected crash frequency in the after period,}$

$N_{\text{observed,A}} = \text{observed crash frequency in the after period,}$

$N_{\text{predicted,A}} = \text{predicted crash frequency estimated using the SPF in the after period}$

$\text{Var}(N_{\text{expected,A}}) = \text{variance of expected crash frequency in the after period.}$

The standard error of the overall safety effectiveness can be calculated by taking square root of the variance of CMF. The variance of CMF can be calculated using Equation 5.

$$\text{Var}(CMF) = \frac{CMF^2 \left[ \frac{1}{(N_{\text{observed,A}})} + \left( \frac{\text{Var}(N_{\text{expected,A}})}{N_{\text{expected,A}}^2} \right) \right]}{\left[ 1 + \frac{\text{Var}(N_{\text{expected,A}})}{N_{\text{expected,A}}^2} \right]^2}$$

where,

$\text{Var}(CMF) = \text{variance of crash modification factor.}$

**Multivariate Adaptive Regression Splines**

The MARS analysis can be used to model complex relationships using a series of basis functions (33). According to the Abraham et al. (34), 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 (35). 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 (33). The MARS model is defined as shown in Equation 6 (36). It is worth to note that log form of MARS model was fitted to develop CMFunctions in this study.

$$\hat{y} = \exp(b_0 + \sum_{m=1}^M b_mB_m(x))$$

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 (36 - 38). 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 each iteration. The second step (pruning phase) performs backward deletion procedure to eliminate the least contributing basis functions. Generalized cross-validation (GCV) criterion is generally used in this pruning step to find the best model. The GCV criterion can be estimated by Equation 7. 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.

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

$$C(M) = M + dM$$

where,

- $y_i$ = response for observation $i$,
- $n$ = number of observations,
- $C(M)$ = complexity penalty function,
- $d$ = defined cost for each basis function optimization.
RESULTS AND DISCUSSION

Estimation of CMFs using EB method

Table 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 (1).

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

<table>
<thead>
<tr>
<th>Crash Type (Severity)</th>
<th>Overall Safety Effects</th>
<th>Different Original Shoulder Width</th>
<th>Different Actual Widened Width</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>2 ~ 8 ft</td>
<td>9 ~ 12 ft</td>
</tr>
<tr>
<td>All (KABC)</td>
<td>0.88**</td>
<td>0.72**</td>
<td>0.07</td>
</tr>
<tr>
<td>All (KAB)</td>
<td>0.82**</td>
<td>0.73**</td>
<td>0.09</td>
</tr>
<tr>
<td>All (KABCO)</td>
<td>0.79**</td>
<td>0.69**</td>
<td>0.12</td>
</tr>
<tr>
<td>ROR (KABC)</td>
<td>0.75*</td>
<td>0.66**</td>
<td>0.15</td>
</tr>
<tr>
<td>ROR (KABCO)</td>
<td>0.72*</td>
<td>0.62**</td>
<td>0.18</td>
</tr>
<tr>
<td>ROR (KAB)</td>
<td>0.69**</td>
<td>0.57**</td>
<td>0.19</td>
</tr>
</tbody>
</table>

*: significant at a 95% confidence level, **: significant at a 90% confidence level
Development of CMFunctions

The CMFunctions were developed to determine the variation of CMFs with different site characteristics among treated segments as shown in Tables 4 and 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 by keeping most important variables (different original shoulder width, actual widened width, AADT, median width and maximum speed limit) to have a reasonable number of individual segments for each data point (average 10 segments for each point). Moreover, As suggested by Sacchi and Sayed (13) and Park et al. (19), 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 (39) was used to fit a MARS model. It was found that there is 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 the predictive power for highly structured data, the applicability of the 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. According to the Park and Abdel-Aty (38), 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.). As we found from the developed SPF in Table 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 (40). 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 $10 - \text{original shoulder width}$ 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 4: Estimated CMFunctions of widening shoulder width using regression model

<table>
<thead>
<tr>
<th>Parameter</th>
<th>All (KABCO)</th>
<th>All (KABC)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
<td>SE</td>
</tr>
<tr>
<td>Constant</td>
<td>-0.5170</td>
<td>0.0486</td>
</tr>
<tr>
<td>Original Shoulder Width in Before Period (ft)</td>
<td>0.0258</td>
<td>0.0041</td>
</tr>
<tr>
<td>Actual Widened Shoulder Width Indicator (1: Sites with 1<del>4ft shoulder width widened, 0: Sites with 5</del>10ft shoulder width widened)</td>
<td>0.1648</td>
<td>0.0205</td>
</tr>
<tr>
<td>Median Width Indicator (1: Sites with less than 40ft median width, 0: Sites with 40ft or more than 40ft median width)</td>
<td>-0.0599</td>
<td>0.0250</td>
</tr>
<tr>
<td>MSE</td>
<td>0.0024</td>
<td></td>
</tr>
<tr>
<td>R-squared</td>
<td>0.8826</td>
<td></td>
</tr>
<tr>
<td>Adj. R-squared</td>
<td>0.8649</td>
<td></td>
</tr>
</tbody>
</table>
Table 5: Estimated CMFunctions of widening shoulder width using MARS model

(a) MARS model for All (KABCO) Crashes

<table>
<thead>
<tr>
<th>Basis Function</th>
<th>Basis Function Information</th>
<th>Estimate</th>
<th>SE</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>BF0</td>
<td>Constant</td>
<td>-0.2257</td>
<td>0.0163</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>BF1</td>
<td>MAX (10 – Original shoulder width, 0)</td>
<td>-0.0151</td>
<td>0.0083</td>
<td>0.0874</td>
</tr>
<tr>
<td>BF2</td>
<td>MAX (Original shoulder width – 10, 0)</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>BF3</td>
<td>Actual Widened Shoulder Width Indicator (1:Sites with 1<del>4ft shoulder width widened, 0: Sites with 5</del>10ft shoulder width widened)</td>
<td>0.1726</td>
<td>0.0174</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>BF4</td>
<td>Median Width Indicator (1: Sites with less than 40ft median width, 0: Sites with 40ft or more than 40ft median width)</td>
<td>-0.1720</td>
<td>0.0479</td>
<td>0.0021</td>
</tr>
<tr>
<td>BF5</td>
<td>BF2 × MAX (10.02127 – Ln. AADT, 0)</td>
<td>-0.0371</td>
<td>0.0170</td>
<td>0.0426</td>
</tr>
<tr>
<td>BF6</td>
<td>BF4 × MAX (Original shoulder width – 6, 0)</td>
<td>0.0247</td>
<td>0.0101</td>
<td>0.0252</td>
</tr>
</tbody>
</table>

MSE = 0.0014
R-squared = 0.9385
Adj. R-squared = 0.9215

(b) MARS model for All (KABC) Crashes

<table>
<thead>
<tr>
<th>Basis Function</th>
<th>Basis Function Information</th>
<th>Estimate</th>
<th>SE</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>BF0</td>
<td>Constant</td>
<td>-0.5535</td>
<td>0.0502</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>BF1</td>
<td>MAX (Original shoulder width – 4, 0)</td>
<td>0.1001</td>
<td>0.0318</td>
<td>0.0055</td>
</tr>
<tr>
<td>BF2</td>
<td>Actual Widened Shoulder Width Indicator (1:Sites with 1<del>4ft shoulder width widened, 0: Sites with 5</del>10ft shoulder width widened)</td>
<td>0.1765</td>
<td>0.0324</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>BF3</td>
<td>MAX (Original shoulder width – 6, 0)</td>
<td>-0.0888</td>
<td>0.0390</td>
<td>0.0354</td>
</tr>
<tr>
<td>BF4</td>
<td>Median Width Indicator (1: Sites with less than 40ft median width, 0: Sites with 40ft or more than 40ft median width)</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>BF5</td>
<td>BF4 × MAX (Maximum speed limit – 65, 0)</td>
<td>-0.0439</td>
<td>0.0149</td>
<td>0.0086</td>
</tr>
<tr>
<td>BF6</td>
<td>BF4 × MAX (10.16585 – Ln. AADT, 0)</td>
<td>-0.0565</td>
<td>0.0502</td>
<td>0.1027</td>
</tr>
</tbody>
</table>

MSE = 0.0049
R-squared = 0.8329
Adj. R-squared = 0.7865

In order to check the capability of selected break points for the basis functions in MARS models, regression models with dummy variables based on the break points from MARS were developed as presented in Table 6. Although the developed models show better model fitness than the normal regression models in Table 4, MARS models still produce more reliable estimates since it can account for interaction impacts among multiple variables under different ranges.
Table 6: Estimated CMFunctions of widening shoulder width using regression model with dummy variables based on the breakpoints from MARS

(a) All (KABCO)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>SE</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-0.2525</td>
<td>0.0155</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Original Shoulder Width in Before Period (Base: Original Shoulder Width ≥ 10ft)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Original Shoulder Width &lt; 6ft</td>
<td>-0.2332</td>
<td>0.0256</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>6ft ≤ Original Shoulder Width &lt; 10ft</td>
<td>-0.1305</td>
<td>0.0229</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Actual Widened Shoulder Width Indicator (1: Sites with 1<del>4ft shoulder width widened, 0: Sites with 5</del>10ft shoulder width widened)</td>
<td>0.1552</td>
<td>0.0195</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Median Width Indicator (1: Sites with less than 40ft median width, 0: Sites with 40ft or more than 40ft median width)</td>
<td>NS</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>log AADT Indicator (1: Sites with less than 10.02127, 0: Sites with 10.02127 or more than 10.02127)</td>
<td>NS</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>MSE</td>
<td>0.0020</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R-squared</td>
<td>0.9021</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adj. R-squared</td>
<td>0.8874</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(b) All (KABC)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>SE</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-0.2770</td>
<td>0.0333</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Original Shoulder Width in Before Period (Base: Original Shoulder Width ≥ 10ft)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Original Shoulder Width &lt; 6ft</td>
<td>-0.1969</td>
<td>0.0848</td>
<td>0.0315</td>
</tr>
<tr>
<td>6ft ≤ Original Shoulder Width &lt; 10ft</td>
<td>-0.2055</td>
<td>0.0507</td>
<td>0.0007</td>
</tr>
<tr>
<td>Actual Widened Shoulder Width Indicator (1: Sites with 1<del>4ft shoulder width widened, 0: Sites with 5</del>10ft shoulder width widened)</td>
<td>0.1928</td>
<td>0.0339</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Median Width Indicator (1: Sites with less than 40ft median width, 0: Sites with 40ft or more than 40ft median width)</td>
<td>-0.0873</td>
<td>0.0371</td>
<td>0.0296</td>
</tr>
<tr>
<td>log AADT Indicator (1: Sites with less than 10.02127, 0: Sites with 10.02127 or more than 10.02127)</td>
<td>NS</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Maximum Speed Limit Indicator (1: Sites with less than 65mph, 0: Sites with 65mph or more than 65mph)</td>
<td>NS</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>MSE</td>
<td>0.0063</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R-squared</td>
<td>0.7759</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adj. R-squared</td>
<td>0.7287</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

NS: not significant

CONCLUSIONS

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.

Although the study provides empirical evidence of the variation of safety effects based on different site characteristics, there are some limitations in this study. The MARS may not clearly consider specific nonlinear trend because the rate of change is assumed to be fixed within a given range of a variable although the rate can vary within the range. Generalized nonlinear models (GNMs) can be possibly adopted to overcome this limitation. More general relationship between safety effects and site characteristics could not be observed due to a lack of enough samples in this study. Thus, it can be suggested to use the developed CMFunctions to check the overall impact of widening shoulder width with relatively large variation. Also, including multiple target areas (e.g. more states) in the analysis may produce more generalized conclusions.

Acknowledgements

The authors thank the Florida Department of Transportation (FDOT) for providing the data used in this study and funding this research. All opinions, findings, and conclusions are solely those of the authors.

REFERENCES


