Developing Crash Modification Factors for Horizontal Curves on Rural Two-Lane Undivided Highways using a Cross-Sectional Study

By

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

Dominique Lord, Ph.D.
Professor, Zachry Department of Civil Engineering
Texas A&M University, 3136 TAMU
College Station, Texas 77843-3136
Phone: (979) 458-3949, fax: (979) 845-6481
Email: d-lord@tamu.edu

and

Srinivas R. Geedipally, Ph.D., P.E.
Texas A&M Transportation Institute
Texas A&M University System
110 N. Davis Dr., Suite 101
Arlington, Texas 76013
Phone: (817) 462-0519, fax (817) 461-1239
Email: Srinivas-G@tti.tamu.edu

Word count: 6,671 Words (4,921 Text + 4 Tables + 3 Figures)

August 1, 2016

*Corresponding author
ABSTRACT

Horizontal curves have been identified as experiencing more crashes than tangent sections on roadways, especially on rural two-lane highways. The first edition of the *Highway Safety Manual (HSM)* provides crash modification function (CM-Function) for curves on rural two-lane highways. The CM-Function proposed in the *HSM* may suffer from both outdated data and analysis technique. Before-after studies are usually the preferred method for estimating the safety effects of treatments. Unfortunately, this method is not feasible for curves. Previous studies have frequently used regression models for developing CM-Functions for horizontal curves. As recently documented in the literature, some potential problems exist with using regression models for developing crash modification factors (CMFs). This study utilized cross-sectional study to develop curvature CM-Function. Curves located on Texas rural two-lane undivided highways were divided into a number of bins based on the curve radius. The safety was predicted with the assumption that these curves had been tangents. The observed number of crashes that occurred on the curves was compared with the “dummy” tangents and for different bins. The results showed that horizontal curve radius has a significant role in the risk of a crash. Based on these results, a new CM-Function was developed. The prediction performance of the *HSM* CM-Function and another function that was recently proposed in the literature were compared with the new CM-Function in this study. It was found that the new CM-Function documented in this study outperformed both.
INTRODUCTION

Horizontal curves are essential elements of all highway systems. At the same time, curves have also been consistently identified as safety concerns by transportation agencies. Each year, about 25 percent of fatal crashes occur at horizontal curves in the United States, causing approximately 10,000 fatalities (1; 2). The total mileage of curve segments accounts for a relatively low percentage of the entire roadway network. However, crashes are over-represented at horizontal curves, and the average crash rate for curves is much higher than that of tangents (3). Crash statistics have shown that an obvious higher percentage of fatal curve-related crashes occur on rural roads, particularly on two-lane roadways. Specifically, nearly 70 percent of the fatal curve-related crashes occurred on rural two-lane highways (1).

The first edition of Highway Safety Manual (HSM) (4) provides a safety performance function (SPF) for rural two-lane highways as well as a crash modification function (CM-Function) for horizontal curves, which can be used to predict the number of crashes for curves. However, the HSM CM-Function for horizontal curve was derived based on studies conducted over 25 years ago (5). The crash dataset analyzed in the studies was collected in the 1980’s. The results may suffer from both outdated data and crash analysis tools, since the road characteristics, vehicle performance, driving behavior as well as safety evaluation methods have evolved significantly over the last two or three decades. As a result, it is necessary to assess whether or not the CM-Function provided in the HSM can adequately capture the safety effects of horizontal curves. If not, developing more reliable crash modification factors (CMFs) or CM-Functions for horizontal curves may be necessary.

The safety effect for a treatment is ideally evaluated through before-after studies, in particular when it is based on the empirical Bayes (EB) method (6). However, it is nearly impossible to assess the safety effects of horizontal curves using such approach in practice, since changing the alignment of a curve is both costly and time consuming. More importantly, when a curve’s alignment is improved, other features associated with the curve (e.g., pavement type and width, side slope, superelevation, etc.) will also change simultaneously. Thus, the before-after study is not feasible. As an alternative, safety analysts have been using cross-sectional studies, particularly regression models or SPFs, to evaluate the safety effects of some treatments or highway design features, although these models have limitations, such as the omitted variable bias, misspecification in functional form, and independence assumption among others (7; 8). Some researchers have also criticized the use of regression models for developing CMFs, since SPFs cannot capture the cause-effect relationship between variables (6; 9).

Recently, Banihashemi (10) examined the effect of horizontal curves on urban arterials using cross-sectional methods. The estimated CM-Functions improved the safety prediction significantly. Until now, such method has not been used to develop the CMF for curves on rural two-lane highways. Thus, the main objective of this study is to develop a CMF for horizontal curves on rural two-lane undivided highways using the cross-sectional approach and to compare the results with that included in HSM and other recently developed CMFs.

BACKGROUND

Due to the higher crash occurrence on horizontal curves, efforts on curve safety have been made by researchers for decades. Previous studies indicate that curve radii affect crash occurrence
significantly. In general, shaper curves are associated with higher crash frequency (5). For example, the probability of experiencing a crash at a curve with a radius of 500 ft is about two times of that of a tangent segment (11).

Harwood et al. (3) developed the CM-Function for horizontal curves on rural two-lane highways based on the previous study conducted by Zegeer et al. (5). The CMF for a specific curve is a function of the length of the curve, its radius and the absence of spiral transition, as shown in Equation 1. As mentioned above, this CM-Function has been included in the first edition of HSM (4).

\[ CMF = \frac{1.55L + \frac{0.2}{R} - 0.012S}{1.55L} \] (1)

Where,
- \( CMF \) = specific CMF for a horizontal curve on rural two-lane highways;
- \( L \) = length of the curve including length of spiral transitions, if present (mi);
- \( R \) = radius of curvature (ft); and,
- \( S = 1 \) if spiral transition curves are present; 0.5 for only one spiral transition curve; and 0 for none.

Meanwhile, the HSM has also provided the SPF for segments with base conditions on rural two-lane highways, shown in Equation (2).

\[ N_{pre} = AADT \times L \times 365 \times 10^{-6} \times e^{-0.312} \] (2)

Where,
- \( N_{pre} \) = predicted annual crash number;
- \( AADT \) = annual average daily traffic (veh/day); and,
- \( L \) = length of the segment (mi).

For a given curve, with length equal to \( L \) mi, and radius equal to \( R \) ft, without spiral transition, the annual predicted number of crashes will be

\[ N_{pre} = k \times L \times \left( \frac{1.55L + \frac{0.2}{R}}{1.55L} \right) \] (3a)

Or equivalently,

\[ N_{pre} = k \times L + k \times \frac{51.7}{R} \] (3b)

Where,
- \( k \) = a constant (given AADT), it equals to \( AADT \times 365 \times 10^{-6} \times e^{-0.312} \).

Note that the first term in Equation 3b is the predicted number of crashes of a tangent segment having the same length as the curve. The second term contains curve radius only. This means when comparing a horizontal curve with an adjacent tangent of the same length, the increase in crashes at the curve is only related to its radius, but not with the curve length. In other words, the risk of crashing in a curve is related to how the driver enters and leaves the curve. Despite that, when the driver is driving on the curve proper, it is as safe as driving on the
adjacent tangent. It can be shown that the situation is the same for curves with spiral transition curve(s).

Recently, Gooch et al. (12) quantified the safety performance of horizontal curves on two-lane rural highways using the propensity scores-potential outcome framework. Data on about 10,000 miles of highways in Pennsylvania with eight years of crash records were analyzed. The results indicate that the presence of a horizontal curve and its degree of curvature are the most significant variables associated with crash frequency. CM-Function for horizontal curves as a function of the degree of curvature was developed, as shown in Equations 4 (denoted as Gooch CM-Function hereafter).

\[
CMF = \exp(0.053 \times HC + 0.054 \times D)
\]  (4)

Where,

\[
CMF = \text{CMF for a horizontal curve (total crash)};
\]

\[
HC = \text{presence of a horizontal curve}; \text{ and}
\]

\[
D = \text{degree of curvature} (D = 5729.47/R, R \text{ in ft}).
\]

Although not for rural two-lane highways, Fitzpatrick et al. (13) developed the CM-Function for horizontal curves on rural four-lane highways, as shown in Equation 5.

\[
CMF = e^{0.0831 \times D}
\]  (5a)

Or equivalently,

\[
CMF = e^{145.2/R}
\]  (5b)

Bauer and Harwood (14) analyzed the safety effects of different combinations of horizontal curves and vertical grades on rural two-lane highways. The CMF for curves varies when designed together with different types of vertical alignments (e.g., crest or sag vertical curves). Nevertheless, the expected crash frequency increases as the horizontal curve becomes sharper.

Saleem and Persaud (15) developed crash prediction models for curve sections on rural two-lane highways. Considering the effects of increased tangent lengths for sharper curves when the deflection angle is given (16), the authors estimated CMFs for flattening a horizontal curve from the minimum radius by a couple of factors (i.e., 1.10, 1.25, 1.50, and 2.00). Their results indicate that the safety effect of horizontal curve flattening lies within the same range at different design speeds and deflection angles. The greater the curve flattening factor is, the higher the reduction in crashes.

There are some other studies on the safety of horizontal curves on rural two-lane highways, e.g., (17), (18). They are not directly related to this study, thus not presented here. It can be seen that nearly all the CM-Functions for horizontal curves found in the literature were developed using regression models. That is some distribution model (commonly negative binomial or NB distribution) of the crash counts and a functional form between crash mean and variables of interests (e.g., segment length, traffic volume, curve radius, etc.) are assumed. The model is then fitted, and estimated parameters are used to develop CMFs for horizontal curves. With this kind of methods, the functional form plays a vital role in developing CMFs (9). If a
class of equation $f(Radius)$ has been assumed in the regression model, the developed CM-Function can only have the same form as $f(Radius)$.

**METHODOLOGY**

The cross-sectional approach to develop the safety effect of curves used in this study is similar to the one used by Banihashemi (10; 19), but necessary changes were made to fulfill the objective of this study. The approaches for developing and assessing CMFs are described below.

**Development of CMFs**

The core concept of the cross-sectional approach for developing CMFs is to compare the safety performance of two or several groups of entities. One group contains entities with one kind of roadway feature, while the other does not. The difference in safety between the two or multiple groups is assumed to be linked to the particular roadway feature. For horizontal curves in this study, one group contains curve segments with specific radius categories, while the other include tangents or curves with different radii. The safety effect is assessed through comparing the safety (i.e., crashes) between groups. Particularly, the crashes assigned to a curve with certain range of radii are observed in real word. While, that of tangents are estimated using the SPFs from the HSM by assuming these segments have the same characteristics as the curves except that the radii are equal to infinity. So, there is one-to-one curve-tangent pair in the two comparison groups. The specific procedure is illustrated below:

**Step 1: Divide Curves into Bins**

Horizontal curves are divided into a number (say $m$) of bins according to curve radii. Note that Banihashemi (10) divided the curves such that the total length of curves of each bin was approximately equal. In this study, curves were divided based on the frequency of curves, such that each bin contained approximately similar number of curves. The reason for dividing curves in this way is that the length of curves correlates with radii in some degree. Sharper curves (i.e., smaller radius) generally tend to be shorter. In the meanwhile, sharper curves accounts for less of entire curves compared to flatter ones. Dividing them by frequency instead of total length can better balance the sample size between bins.

Calculate total number of observed crashes and the mean of radii within each of the $m$ bins. The total number of observed crashes represents the “safety” of curves in that bin, and the mean of radii can be viewed as the “feature” of the curves within the bin. Also note here, Banihashemi (10) used the mid-point of each radius bin to present its “feature.” However, radius is usually not equally or symmetrically distributed. Using the mean instead of mid-point might better represent the curve feature, since the mean is the asymptotic expectation of the radii.

**Step 2: Predict Crash Number if the Curves Had Been Tangents**

Predict the number of crashes for each curve if it had been a tangent using the base SPF provided by HSM. To predict the number of crashes on the “tangents”, this study used the basic SPF for rural two-lane highways, shown below as Equation (6).

$$N_{pre} = AADT \times L \times 365 \times 10^{-6} \times e^{-0.312} \times C$$  \hspace{1cm} (6)
Where,

\[ C = \text{calibration factor}. \]

The calibration factor was assumed to be 1.0 in this step (this will not affect the results in Step 4, which is of interest). It is worth mentioning that no CMF was considered in this step, because these “tangents” are homogeneous segments (note that the characteristics, e.g., lane and shoulder widths, of the curves are all the same or similar except radius, more detailed information is documented in Section Data Description below).

Calculate the predicted number of crashes of each “tangent” bin. This number represents what the “safety” of each curve bin would be if they had been tangents.

**Step 3: Calculate the Initial Ratios of Observed to Predicted “Safety” and Normalize Them**

Calculate the initial ratio for each bin by dividing the observed “safety” by its predicted if the curves had been tangents.

Choose one bin as base condition (usually the bin with the greatest radii), divide the initial ratios by that of the base bin. The new ratios are now normalized. The normalized ratio of each bin represents the safety effect of its feature (i.e., curve radii) compared to the base bin.

**Step 4: Develop CM-Function of CMF**

If the normalized ratios are significantly different from 1.0, this means the feature (i.e., horizontal curve) affects crash risk. Otherwise, the feature may not be recognized as a safety-related factor. If it is, plot the normalized ratio against the mean radius. Fit them if there is an obvious pattern between the two. The fitted curve serves as the tentative CM-Function for horizontal curve radius. If there is no obvious pattern, discrete pairs between mean radius and normalized ratio can be used as CMFs for specific highway curves.

The protocol for developing CMFs/CM-Functions is illustrated in Figure 1.
Assessment of CMFs

To assess the performance of the proposed CM-Function, it is compared with HSM CM-Function and the one recently developed by Gooch et al. (12). The comparison is mainly based on the observed crash counts versus that of predicted with each CM-Function. The process for predicting crashes is the same as that introduced in HSM. It is described below.

Step 1: Split the Validation Data

Divide the validation data into several bins. This is similar to that used to develop CMFs for horizontal curves. However, the number of bins and thresholds may not necessary be the same. Calculate the total observed number of crashes for each bin.

Step 2: Initial Prediction with Given CM-Function

Predict the initial number of crashes for each curve using the base SPF and the given CM-Function (i.e., HSM CM-Function, Gooch CM-Function, and the one proposed in this study, respectively).

FIGURE 1 Protocol for developing CMF for horizontal curve.
**Step 3: Calculate Calibration Factor and Update Prediction**

Calculate the total number of predicted crashes (all bins) and derive the calibration factor by dividing the total predicted by the total observed. The calibration factor is estimated and used such that the total number of predicted crashes equal to the total number of observed (4; 10).

Update the predicted number of crashes for each curve through multiplying its initial prediction by the estimated calibration factor.

Calculate the total number of updated (i.e., calibrated) prediction for each bin. The closer the predicted crash numbers of bins to that of observed, the better that CM-Function is.

**DATA DESCRIPTION**

This study utilized horizontal curves (with radius no greater than 5,000 ft) located on rural two-lane undivided highways in Texas. The curve data were extracted from the Texas Roadway Inventory database. Five years (from 2010 through 2014) of crash data on these curves were extracted from the Crash Records Information System (CRIS). Both data are managed by the Texas Department of Transportation (TxDOT).

The curvature data was pre-examined based on a couple of criteria. First, the characteristics of the curves should be as close as to the base condition defined in the HSM (if they had been tangents). Second, the curves selected should be homogenous as much as possible (except the radii). And last, some outliers (i.e., extreme values) and those with obvious errors should be excluded from the dataset.

Finally, 26,234 horizontal curves, and 3,968 crashes during the five-year period were identified. Data in the first three years (2010 through 2012) were used to develop CMFs, and that in the last two years (2013 and 2014) were used for validation. The summary statistics of the data are shown in Table 1.

**TABLE 1 Summary Statistics of Horizontal Curves**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Min</th>
<th>Max</th>
<th>Mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Curve Length (mile)</td>
<td>0.001</td>
<td>0.599</td>
<td>0.121 (0.084)</td>
</tr>
<tr>
<td>Curve Radius (ft)</td>
<td>63.2</td>
<td>4915.1</td>
<td>1692.8 (921.7)</td>
</tr>
<tr>
<td>ADT (2010-2012)</td>
<td>0</td>
<td>9,333</td>
<td>684.9 (845.3)</td>
</tr>
<tr>
<td>ADT (2013-2014)</td>
<td>8</td>
<td>9,375</td>
<td>702.7 (922.3)</td>
</tr>
<tr>
<td>Crash (3 yr: 2010-2012)</td>
<td>0</td>
<td>14</td>
<td>0.087 (0.375)</td>
</tr>
<tr>
<td>Crash (2 yr: 2013-2014)</td>
<td>0</td>
<td>7</td>
<td>0.054 (0.275)</td>
</tr>
</tbody>
</table>

Note: SD = standard deviation.

**RESULTS**

The following two sections present the analysis results for the data with the methodology described in the previous sections.
**Significance of Horizontal Curve and CM-Function Development**

For the development of curvature CMFs, the curves were divided into eight bins on the basis of curve radius. Following the procedure documented in the previous sections, the predicted number of crashes, the initial ratio were calculated for each bin, as shown in Table 2 (columns 6 and 7, respectively). It can be seen that horizontal curve affects safety significantly without surprise. As the mean radius decreases from about 3,000 ft to about 300 ft, the initial ratio increases from 0.73 to 2.74 continuously. The sharper is the curve, the higher the crash risk is. This is consistent with various previous studies on the effect of horizontal curve on safety (5; 12; 20; 21).

To further quantify the effects of horizontal curves, the initial ratios were normalized. The bin with greatest mean radius (i.e., bin 8 in Table 2) was chosen as the base condition, and the initial ratio of each bin was divided by that of the base condition (i.e., 0.73 in Table 2). The normalized ratios are shown in column 8 in Table 2.

### TABLE 2 Results of Initial and Normalized Ratios for Each Curve Bin

<table>
<thead>
<tr>
<th>Bin</th>
<th>Radius Range (ft)</th>
<th>Mean Radius</th>
<th>Obs.</th>
<th>Pred.</th>
<th>Int. Ratio</th>
<th>Norm. Ratio</th>
<th>Fre. of Curves</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>63.5 - 565.4</td>
<td>331.1</td>
<td>196</td>
<td>71.63</td>
<td>2.74</td>
<td>3.73</td>
<td>2536</td>
</tr>
<tr>
<td>2</td>
<td>572.8 - 708.8</td>
<td>584.3</td>
<td>191</td>
<td>74.05</td>
<td>2.58</td>
<td>3.52</td>
<td>1851</td>
</tr>
<tr>
<td>3</td>
<td>711.7 - 941.8</td>
<td>761.8</td>
<td>187</td>
<td>85.51</td>
<td>2.19</td>
<td>2.98</td>
<td>1640</td>
</tr>
<tr>
<td>4</td>
<td>949.0 - 1130.8</td>
<td>962.3</td>
<td>236</td>
<td>117.06</td>
<td>2.02</td>
<td>2.75</td>
<td>2250</td>
</tr>
<tr>
<td>5</td>
<td>1134.5 - 1403.1</td>
<td>1156.7</td>
<td>373</td>
<td>251.93</td>
<td>1.48</td>
<td>2.02</td>
<td>3532</td>
</tr>
<tr>
<td>6</td>
<td>1420.0 - 1888.8</td>
<td>1457.9</td>
<td>313</td>
<td>246.55</td>
<td>1.27</td>
<td>1.73</td>
<td>3703</td>
</tr>
<tr>
<td>7</td>
<td>1899.2 - 2817.7</td>
<td>1969.5</td>
<td>445</td>
<td>436.11</td>
<td>1.02</td>
<td>1.39</td>
<td>5006</td>
</tr>
<tr>
<td>8</td>
<td>2840.9 - 4915.1</td>
<td>3032.5</td>
<td>331</td>
<td>451.81</td>
<td>0.73</td>
<td>1.00</td>
<td>5716</td>
</tr>
<tr>
<td>Overall</td>
<td>63.5 - 4915.1</td>
<td>2,272</td>
<td>1,735*</td>
<td></td>
<td></td>
<td></td>
<td>26,234</td>
</tr>
</tbody>
</table>

Notes: Obs. = observed crash counts; Pred. = predicted number of crashes, * calibration factor was assumed to be 1.0 and this is the reason that the total observed does not equal to that of predicted, but this does not affect the normalized ratios which are of interest; Int. Ratio = initial ratio; Norm. Ratio = normalized ratio; Fre. of Curves = frequency of curves.

To develop a CM-Function representing the relationship between horizontal curve radius and crash risk, a couple of commonly used functions (i.e., linear, exponential, logarithmic, power, etc.) were tried to fit the points. Power function was found to fit it well, as shown in Figure 2. The fitting results is described in Table 3.
### FIGURE 2 Normalized ratios and developed CMF for horizontal curve.

### TABLE 3 CMF Fitting Results

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coef. Value $^1$</th>
<th>SE $^2$</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Equation Form: $CMF = a \times R^b$ or equivalently $log(CMF) = log(a) + b \log(R)$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$log(a)$</td>
<td>5.28</td>
<td>0.50</td>
<td>$&lt; 2e-16$</td>
</tr>
<tr>
<td>$b$</td>
<td>-0.65</td>
<td>0.07</td>
<td>0.0001</td>
</tr>
</tbody>
</table>


The fitted equation (CM-Function) is shown in Equation 7.

$$CMF = 196.4 \times R^{-0.65}$$  \hspace{1cm} (7)

Where, $R$ = the radius of a horizontal curve (ft); and $CMF$ = the specific CMF for the curve.

Note that, since the data analyzed only contains radii of 5,000 ft or less, the CM-Function only applies to curves with radius no greater than 5,000 ft on rural two-lane highways.

### Validation

To validate the performance of the proposed CM-Function for horizontal curve, the curves were re-divided into nine bins. The number of crashes were predicted using three CM-Functions, i.e., the *HSM* CM-Function, Gooch CM-Function, as well as the one proposed in this study. They were compared with the observed number of crashes in 2013 and 2014. The application of these CM-Functions followed the way that was introduced in *HSM*, and calibration...
factor was calculated for each function. This way, the total predicted number equal to the total observed. The coefficient of variation of the observed crash data was 5.09. The sample size, 26,234, satisfies the sample-size requirement at a 90% level (22). Table 4 lists the curve splitting, observed crash counts as well as crash prediction results with the three CM-Functions.

### TABLE 4 Predicted Number of Crashes with Different CM-Functions

<table>
<thead>
<tr>
<th>Bin</th>
<th>Range (ft) (Min, Max)</th>
<th>Mean Radius</th>
<th>Obs.</th>
<th>HSM</th>
<th>Gooch</th>
<th>Proposed</th>
<th>Fre. of Curves</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>63.5 318</td>
<td>207.5</td>
<td>40</td>
<td>89.5 (123.8%)</td>
<td>81.6 (104.1%)</td>
<td>29.6 (-25.9%)</td>
<td>1014</td>
</tr>
<tr>
<td>2</td>
<td>318 573</td>
<td>413.7</td>
<td>83</td>
<td>91.3 (10.1%)</td>
<td>72.5 (-12.7%)</td>
<td>84.1 (1.3%)</td>
<td>1524</td>
</tr>
<tr>
<td>3</td>
<td>573 716</td>
<td>584.4</td>
<td>106</td>
<td>88.7 (-16.3%)</td>
<td>77 (-27.3%)</td>
<td>104.5 (-1.5%)</td>
<td>1856</td>
</tr>
<tr>
<td>4</td>
<td>716 955</td>
<td>762.8</td>
<td>108</td>
<td>83.3 (-22.9%)</td>
<td>77.3 (-28.4%)</td>
<td>108.0 (0%)</td>
<td>1645</td>
</tr>
<tr>
<td>5</td>
<td>955 1146</td>
<td>962.9</td>
<td>139</td>
<td>96.8 (-30.4%)</td>
<td>96.7 (-30.4%)</td>
<td>132.6 (-4.6%)</td>
<td>2247</td>
</tr>
<tr>
<td>6</td>
<td>1146 1400</td>
<td>1156.4</td>
<td>235</td>
<td>192.1 (-18.2%)</td>
<td>201.8 (-14.1%)</td>
<td>264 (12.3%)</td>
<td>3517</td>
</tr>
<tr>
<td>7</td>
<td>1400 1910</td>
<td>1457.8</td>
<td>215</td>
<td>178.7 (-16.9%)</td>
<td>184.1 (-14.4%)</td>
<td>216.9 (0.9%)</td>
<td>3708</td>
</tr>
<tr>
<td>8</td>
<td>1910 2865</td>
<td>1971.0</td>
<td>264</td>
<td>295.3 (11.9%)</td>
<td>311.6 (18%)</td>
<td>295.7 (12%)</td>
<td>5019</td>
</tr>
<tr>
<td>9</td>
<td>2865 4915</td>
<td>3032.8</td>
<td>227</td>
<td>301.3 (32.7%)</td>
<td>314.4 (38.5%)</td>
<td>181.6 (-20%)</td>
<td>5704</td>
</tr>
<tr>
<td>Overall</td>
<td>63.5 4915</td>
<td>1,417</td>
<td>1,417</td>
<td>1,417</td>
<td>1,417</td>
<td>1,417</td>
<td>26,234</td>
</tr>
</tbody>
</table>

Note: Obs. = observed crash counts; Fre. of Curves = frequency of curves; HSM, Gooch, Proposed = predicted number of crashes with CMFs provided in HSM, developed by Gooch et al. (12), and proposed in this study, respectively; Numbers in parentheses indicate the corresponding prediction bias (i.e., the difference between predicted and observed divided by the observed crash count). A positive number means over prediction, and vice versa.

Obviously, the proposed CM-Function performs better than both the HSM and Gooch CM-Functions. The comparison is also illustrated in Figure 3. It can be seen that the latter two predicted similar number of crashes for each bin. They under-predict the number of crashes when the radius is between about 500 ft and 2,000 ft, but over-predict it when the curve is sharp (radius less than about 500 ft) or flat (radius greater than about 2,000 ft). The bias varies between 10% and 124%, and it becomes significant in the two ends. Overall, the prediction is unacceptable. In contrast, the predicted crash numbers with the proposed CM-Function fit the observed crash counts well when the curve radius is between about 500 ft and 2,000 ft. The crash numbers are slightly underestimated when the radius is less than about 500 ft or greater than about 2,000 ft. At the two ends (sharper or flat curves), all the three CM-Functions produce
higher prediction biases. However, the magnitude of the proposed CM-Function is significantly smaller than the other two (i.e., 25% versus more than 100%).

![Figure 3 Prediction bias comparison between the three CM-Functions.](image)

**FIGURE 3** Prediction bias comparison between the three CM-Functions.

**CONCLUSIONS AND DISCUSSIONS**

This paper has analyzed the safety effect of horizontal curves on rural two-lane undivided highways on safety using cross-sectional study. The curves were divided into a couple of bins based on their radii, and the crash numbers were predicted using HSM method if they had been tangents. The ratios between the observed and predicted number of crashes were analyzed and further developed into a CM-Function as a function of the curve radius. The crash number was predicted with the proposed CM-Function, and compared with that provided in the first edition of HSM as well as the one recently developed by Gooch et al. (12). Their performances were assessed by comparing the predicted number of crashes with the observed crash counts in another two years. The main conclusions are summarized as follows: (1) horizontal curve plays significant roles in terms of crash occurrence on rural two-lane undivided highways. The sharper is the curve, the higher the crash risk is. This is consistent with previous studies; (2) either the HSM CM-Function or the recently developed CM-Function can adequately predict the number of crashes for the horizontal curves used in this study; and (3) the proposed CM-Function performs better than the other two, and in general predicts the number of crashes well.

There is an interesting question that needs further consideration. As has been discussed previously, the HSM CM-Function for curve suffers from both outdated data and analysis.
approach, and it may not predict the safety of curves well, as the validation results have shown. The recently developed CM-Function by Gooch et al. (12) was also unable to predict the number of crashes adequately. It is worth mentioning that for the proposed CM-Function in this study the same curves were used for development and validation. This could improve the prediction accuracy of the proposed CM-Function, since some confounding information might have been included in it. Another question arises: is the CM-Function developed in one area or during one period applicable to another area or period? From this study, the answer seems to be no. In addition, it should not be concluded that the HSM and Gooch CM-Functions are similar, although their prediction results of bins are very close to each other in this study. The cumulative residual (known as CURE) plots were analyzed (not documented here due to space limitation), and results showed that they differed significantly. For more details about CURE plots, readers are referred to Hauer (9).

This study developed the CM-Function for horizontal curves using the cross-sectional method. Compared to the regression model approach, the cross-sectional method used in this study is more straightforward. No statistical models or complicated estimating computations were needed (except when fitting the ratios against radii). Thus, most of the issues linked to regression models are avoided. However, there are still a few limitations with this approach. First, the curve splitting (i.e., how to divide the curves) affects the resulted ratios and hence the developed CM-Functions. Usually, if curves are divided into more bins, much noise will be included in ratios. On the other hand, if too few bins are used, the data will be over-deducted and some useful information may be lost. This study divided the curves into eight bins based on the number of curves, and used the curve mean to present the feature of each bin. This is slightly different than how the data were manipulated in the previous study (12). It is recommended to compare the splitting methods and see how to split it to obtain optimal results in the future. Second, when fitting the ratio against curve radius to develop CM-Function, the functional form was selected by comparing a couple of common ones. In this study, the power function looked good and was used. There is no theoretical ground for selecting the power form, and likely there exists another form which could fit the plotted observations better. This is similar to the question in regression modeling, where modelers frequently choose the additive linear form to link the mean (or logarithmic mean) with independent variables. Comparatively, the cross-sectional method used in this study is more flexible. The final CM-Function can have various forms. And finally, the comparison method may suffer from the confounding variable problem, which is quite similar to the omitted variable bias in the regression models. In this study, a bin of curves was compared to a bin of “dummy” tangents, and the difference in safety performance was attributed to the horizontal curves, or specifically the radii. However, there might be some other factors influencing crash risk, e.g., terrain. It is possible that sharper curves are more common in mountainous areas. This problem exists not only in this study but in almost all kinds of cross-sectional studies. In addition, most of the researchers in the previous studies focused on radius (or degree of curve) only when analyzing curve safety. However, preliminary analysis showed that other factors, e.g., deflection angle, also influenced curve safety significantly and simultaneously. With the cross-sectional analysis method used in this study, safety analysts are able to investigate the combined safety effects of multiple features, e.g., combinations of radius and deflection angle (this is ongoing and will be documented in the future manuscripts).
REFERENCES


