Developing Analytical Procedures for Calibrating the *Highway Safety Manual* Predictive Methods

Xiao Qin, M. Razaur Rahman Shaon, and Zhi Chen

Performance-based safety goals and objectives are more attainable with the use of the *Highway Safety Manual* (HSM). However, the safety performance functions (SPFs) in the HSM may not be accurate when used with local jurisdictions. Each SPF and crash modification factor (CMF) assume a set of base site conditions that might not be realistic or representative of local highways. The calibration procedures provided in Part C, Appendix A, of the HSM should therefore be modified to accommodate local data availability as well as roadway, traffic, and crash characteristics. Furthermore, a set of base conditions applicable to local highways should be determined. Results show that the HSM models underestimate the total average number of crashes on highway segments for all rural highway facility types in South Dakota. To quantify highway safety performance better, this study analyzed the underlying factors contributing to the underestimation and proposed procedures to improve crash prediction accuracy. The HSM calibration was performed with crash data from rural two-lane, two-way highways and rural multilane divided and undivided highways during a 5-year period (2008–2012). The procedures included establishing new base conditions, developing jurisdiction-specific SPFs, and converting CMFs to new base conditions. The comparison results show that the customized models outperform the HSM models in predicting sites with crashes.

The past several decades have witnessed a remarkable improvement in highway safety because of better design, better roads, intense law enforcement, educational programs, and advanced vehicle safety technologies. Despite these advancements, NHTSA reported that 9,902,000 vehicles were involved in crashes in the United States in 2012, in which 22,912 people died and more than 2 million people were injured. Almost 41% of the motor vehicle crashes were related to nonjunction or roadway segments (*1*).

Transportation agencies need to set safety targets for reducing fatalities and serious injuries to continue meeting the economic needs of local communities while keeping motorists safe. The 2010 publication of the *Highway Safety Manual* (HSM) by AASHTO has made it easier to define safety performance expectations (*2*). The HSM provides guidance for safety analyses that use scientific and statistically sound methods developed from decades of highway safety research (*2*). The predictive method in the HSM estimates expected average crash frequencies and injury severity levels for a specific roadway segment or intersection.

The HSM predictive models help forecast how many crashes will happen for specific roadway segments or intersections; however, safety conditions change over time, so these models should be calibrated to avoid compromising safety estimates, producing unrealistic results, and undermining agency accountability. Furthermore, the models should be customized if the results are to be compared with an agency's estimates based on historical crash data, as the HSM models were developed from safety data collected in selected states. It is equally important to define, establish, and convert base conditions that are appropriate for local highway segments, as each safety performance function (SPF) and crash modification factor (CMF) recommended in the HSM assumes a set of base conditions that may not be realistic or representative for local roadways. The need for jurisdiction-specific SPFs should be evaluated if the jurisdiction or state did not contribute to the development of SPFs in the HSM.

This study sought to answer the following questions:

- How well does the HSM predictive model correspond to local safety data such as that for South Dakota?
- If the model results do not adequately fit local data, what type of analysis should be performed to identify why the model did not fit?
- How can the analysis be used to better the HSM calibration for a more accurate crash prediction performance?

This paper presents the development and modification of the HSM predictive method for various highway facility types in rural areas, focusing on bridging the gap between the HSM prediction and actual observation. The modification of the HSM predictive method included the establishment of jurisdiction-specific base conditions, the development of SPFs, and the conversion of CMFs for the facility types. The results of the HSM predictive models and jurisdiction-specific predictive models were compared. It was anticipated that the calibrated methods would be more accurate and reliable for predicting crash frequencies and should therefore be used to make more informed decisions on safety improvements.

LITERATURE REVIEW

The HSM provides a three-step process for predicting crash frequencies for various facility types (*2*). The SPF, CMFs, and calibration factors are given a set of values for base conditions. The SPF predicts crash frequency as a function of annual average daily traffic (AADT) for roadway segments with basic geometric and traffic conditions. CMFs can be multiplied to the crash frequency for sites possessing

Department of Civil and Environmental Engineering, University of Wisconsin–Milwaukee, P.O. Box 784, Milwaukee, WI 53201-0784. Corresponding author: X. Qin, qinx@uwm.edu.

Transportation Research Record: Journal of the Transportation Research Board, No. 2583, Transportation Research Board, Washington, D.C., 2016, pp. 91–98.
DOI: 10.3141/2583-12

characteristics that differ from the base conditions. Each CMF represents one type of change. Once all available CMFs have been considered, a calibration factor, *C,* serves as the ultimate adjustment for all other known or unknown, measurable or immeasurable differences, such as climate, driver populations, animal populations, crash reporting thresholds, and crash reporting system procedures. Each of the three steps yields the opportunity for calibration if more accurate results are desired.

A distinct homogeneous segment is key to the success of developing and implementing SPFs. In the HSM, a roadway segment is defined as "a part of a roadway which is not interrupted by an intersection and consists of homogeneous geometric and traffic control features" (*2*). The segmentation of a total roadway network based on multiple variables can lead to very short homogeneous roadway segments (*3*). The presence of short segments can result in many segments with zero crashes, which can become problematic for proper statistical inference. Predictive models have no prescribed minimum segment length, but the HSM suggests a length of no less than 0.10 mi. Qin and Wellner studied the relationship between segmentation and safety screening analysis by using different lengths of sliding windows to identify the crash hot spot (*4*). The authors concluded that short segments as well as those that are too long create a bias in the identification of sites with safety problem. Ogle et al. demonstrated that segment lengths of less than 160 m cause uncertain results in crash analyses (*5*).

Since the SPF carries most of the weight in predicting crashes, calibrating the SPF may be more critical and effective than making other modifications. Brimley et al. hypothesized that some new variables such as speed limit, the presence or absence of a shoulder rumble strip, passing ability, and percentage of single-unit and multiple-unit trucks may have a significant correlation with the number of crashes that were not used in the HSM SPF for rural two-lane, two-way roads (*6*). The authors found that speed limit and percentage of multiple-unit trucks have a significant correlation with crash frequency. Following the HSM procedure, Abdel-Aty et al. developed statewide SPFs for various subtypes of multilane roadways and freeway segments in Florida (*7*). The authors found that crash frequency was either overestimated or underestimated at low or high AADT levels. District-specific SPFs were developed to account for variations across Florida, and a population group-level calibration factor was used in place of a state-level calibration factor. In general, the base conditions should be the most representative segments, as they guarantee a large-enough sample for development of statistically robust models. However, the most representative roadway type may vary from state to state or region to region. SPF calibration may not be representative of the larger sample and may not be rigorous enough when used with a small sample size, and the resulting estimation deviance may be further amplified after applying CMFs.

After the SPF is calibrated, CMFs are multiplied by the SPF. CMFs can be in the format of a factor or a function. For instance, a CMF for centerline rumble strips is a factor, whereas a CMF for a horizontal curve is a function of curve length and degree of the curve. As both the SPF and CMFs account for the safety effects of measured variables, the unmeasured factors can be estimated via an overall calibration factor. The HSM recommends use of calibration factors to adjust for regional differences (*8, 9*). Banihashemi evaluated the quality of calibration factors generated from data sets of different percentages from the complete data set (*10*). The author evaluated a different data size percentage that falls within a 5% to 10% limit of the ideal calibration factor. Mehta and Lou treated the calibration factor as a special case of the negative binomial (NB) regression model, but the results showed that the HSM-recommended method outperformed the proposed method in estimating the calibration factor (*11*).

DATA SOURCE AND PROCESSING

The required and desirable data items for applying the HSM predictive method are provided in Appendix A of the HSM. Roadway inventory data and historical crash information are required for calibrating the HSM predictive models. All data used in this study were provided by the South Dakota Department of Transportation (DOT). Roadway segment data were obtained from the roadway inventory system (RIS) maintained by the South Dakota DOT. Crash data from 2008 to 2012 were collected from the South Dakota Accident Records System. The scope of this study for calibration purposes was limited to rural state highway segments. The facilities included rural two-lane, two-way highways (RT), rural multilane undivided highways (RM4U), and rural multilane divided highways (RM4D).

The roadway geometric features and traffic information were stored in eight separate event tables in the RIS. So that homogeneous highway segments that meet the HSM requirements could be generated, data event tables containing various highway and traffic information were processed, integrated, and reduced to a format that is appropriate for calibration. Processing highway segment data was complex because the roadway geometric attributes were not the same in all event tables. The event tables were merged by using information from the mileage category, the common attribute, with a type of dynamic segmentation method that allows multiple sets of attributes to be associated with a linear (mileage) feature. The merging sequence went from coarse to fine to minimize the number of disaggregate segments and avoid creating extremely short segments at the beginning of a merging sequence. After all event tables from RIS were merged, a roadway shapefile was generated with the South Dakota DOT linear referencing system.

ArcMap assigned crashes to each segment according to the spatial distance. All crashes occurring on state roads within rural boundaries were joined with road data. The linkage distance between the crash and the segment was used to explore the crash count distribution. It is apparent that the linking count does not significantly increase after the distance reaches 50 ft; therefore, a 50-ft buffer distance was used for each crash when the linkage between the crash data set and the roadway data set was finalized. All crashes (including animal collisions) within a 50-ft buffer distance of a roadway segment were associated with that segment.

APPLICATION OF HSM PREDICTIVE MODELS

For any facility type, the HSM provides the following predictive model equation:

$$N_{\text{predicted}} = N_{\text{spf}} \times C \times (\text{CMF}_1 \times \cdots \times \text{CMF}_n) \quad (1)$$

where

$N_{\text{predicted}}$ = predicted average crash frequency for a site,
N_{spf} = predicted average crash frequency for the base conditions for a site called SPF,

TABLE 1 Calibration Results by Segment Type

Facility Type	Sample Size	Sample Length (mi)	Crashes: Observed (2008–2012)	Crashes: Predicted (5 years)	Calibration Factor, C_r
RT	16,828	6,362	10,418	8,861	1.18
RM4U	1,210	152	940	822	1.14
RM4D	1,619	634	1,791	1,139	1.57
Grand Total	19,657	7,149	13,149	10,822	1.22

NOTE: $C_r = \frac{\sum_{\text{all sites}} \text{Observed Crashes}}{\sum_{\text{all sites}} \text{Predicted Crashes}}$.

C = calibration factor, where C_r indicates a roadway segment and C_i an intersection,

and series of CMFs account for changes in the number of crashes related to specific site characteristics or safety treatments.

The HSM recommends a minimum sample size of 30 to 50 sites for calibrating the HSM predictive method for a specific type of highway facility. The entire group of sites for a facility type should have at least 100 crashes per year. All available sites should be used if the facility type has fewer than 30 sites (*2*). To eliminate possible bias caused by drawing samples from all sites, this calibration used all available sites in the facility type. Table 1 shows the calculation of calibration factors following the HSM predictive method.

Table 1 indicates that the HSM predictive methods underestimate the crash counts in all facility types. On average, the volume of crashes occurring on South Dakota rural state highway segments is 22% higher than HSM predictions. The calibration factors vary between 1.14 and 1.57 by facility type. The estimation difference can be caused by crash trends in various parts of South Dakota, the influence of variables included, or unutilized or unobserved variables.

ANALYSIS

Although the calibration factor has been estimated for each highway facility type, individual site performance is unknown. It is possible for the calibration factor to vary considerably among sites within the same facility type; if so, new information should be sought to explain the within-facility variability.

A logical starting point is to review the calibration factor for each site over the entire state highway system and identify any spatial patterns that may support a regional stratification. Next, RIS roadway attributes that are not included in the HSM predictive method (e.g., posted speed limit, the degree of curve, and grade) should be reviewed as these factors may contribute to crash occurrence and may explain variability occurring within the same facility. Finally, the effect of the variables already included in the HSM predictive method (e.g., AADT, lane width, and shoulder width) should be evaluated because the HSM SPFs and CMFs were not developed with local data. Furthermore, the quantitative relationship presented as the coefficients of the SPF does not necessarily hold in South Dakota. The following analysis provides factual information and statistical evidence to support any modification to the current HSM predictive models.

Geographic Distribution of Ratio

The regional factor is often considered as the surrogate measure for conditions other than those used in the HSM predictive method (e.g., weather, animal population, terrain, crash reporting threshold, criteria). Ratio is introduced in this section. The ratio is more similar to site-specific calibration factor. Figure 1 is a map of the ratio for each highway segment. The ratio value is calculated with following equation:

$$\text{ratio}_i = \frac{\text{observed}_i}{\text{predicted}_i} \qquad (2)$$

where

ratio_i = value of the ratio for each site,
observed_i = observed crash count for a site, and
predicted_i = predicted crash count for the same site using SPF and CMFs.

The spatial distribution of the ratio suggests there may be two distinctive zones for segments in South Dakota. In western South Dakota, the predicted number of segment crashes is overestimated, whereas it is underestimated in eastern South Dakota. The dense highway network in the east may contribute to the average calibration factor of 1.22. Despite noticeable patterns, the mixed results present a challenge in drawing boundaries between regions.

The four regions in South Dakota defined by the South Dakota DOT are Rapid City, Pierre, Aberdeen, and Mitchell. The Rapid City region is in the western part of South Dakota, Pierre is in the center of the state, and Aberdeen and Mitchell are on the eastern side of South Dakota (one in the northeast and one in the southeast). It is practical to use these four regions when developing any jurisdiction-specific SPFs or estimating any region-specific calibration factors. These regional factors were expected to represent the crash variations associated with any geographical or spatial effects among facility types.

Residual Analysis

A spatial analysis indicates whether the regional factor should be included to better represent geographic disparities. The difference between observation and prediction reveals whether the relationship between the predicted crash frequency and the explanatory variables should be adjusted for higher accuracy (namely, adding new

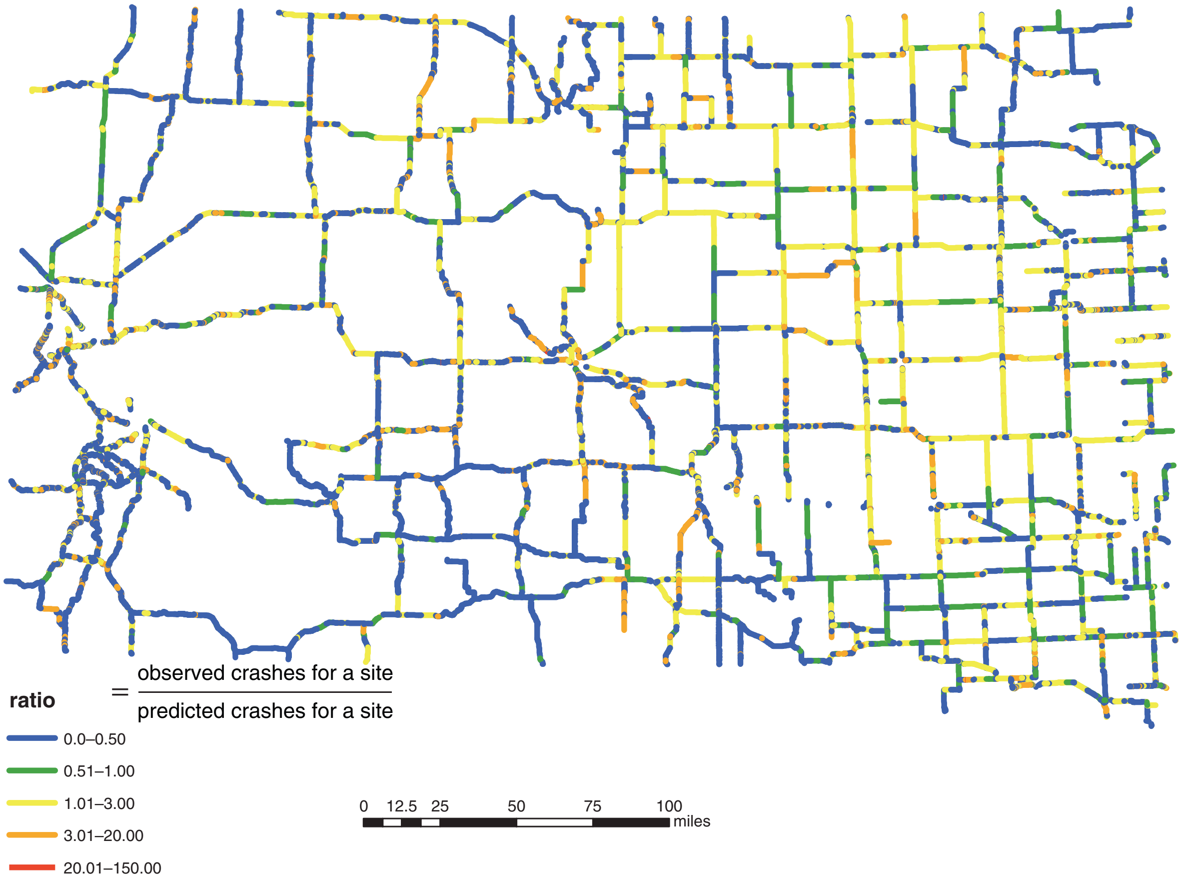

FIGURE 1 Distribution of segment ratio.

variables to the current SPFs or calibrating SPFs with local data). In this exploration, the residual was used to examine which factors contribute to the difference between observation and prediction. The residual was formulated as the observation minus the prediction multiplied by C_r:

$$\text{residual}_i = N_{\text{observed}_i} - N_{\text{predicted}_i} \tag{3}$$

where

$$N_{\text{predicted}_i} = N_{\text{SPF}} \times (\text{CMF}_1 \times \cdots \times \text{CMF}_n) \times C_r \text{ and}$$

$$\sum_{\text{all sites within a facility}} \text{residual}_i = 0$$

In this context, the residual measures the variation in prediction within each facility type. A residual analysis was conducted for all facility types. First, the correlation matrix between residuals and all available segment characteristics was calculated and plotted. None of the Pearson correlation coefficients were found to be larger than 0.3, indicating a weak correlation between the residual and segment characteristics. The locally weighted scatter plot smoothing (LOESS) curve was fitted for each of the four variables with the largest correlation coefficients to examine the nonlinear dependency between residuals and explanatory variables. Figure 2 displays the residual plot for the RT facility. The residual plots for other highway facilities are omitted.

A closer review of the figure suggests the LOESS curves led to no apparent dependence between residuals and other explanatory variables. So, no distinct trends were found between residuals and variables used in this study.

Base Conditions

The HSM suggests that the jurisdiction-specific base conditions be designed to represent the most common characteristics of facilities. For segments, base conditions should be defined according to the most representative geometric features. The length distribution for various combinations of lane and shoulder width was used to identify the conditions with the largest share of mileage. Geometric features representing the maximum segment length in South Dakota are as follows:

- For RT, the only difference found in the base condition was a 4-ft shoulder width; the HSM recommends a 6-ft shoulder width.
- For RM4U, a 2-ft shoulder was established as a base condition.
- Base conditions for both lane width and right shoulder width in RM4D also differed from the HSM base conditions. A 13-ft lane width and 4-ft right shoulder width were established as the new base condition for segments in RM4D.

The sample sizes conforming to these conditions were reviewed. The proportion of observed crashes was 9.11%, 3.3%, and 6.14%

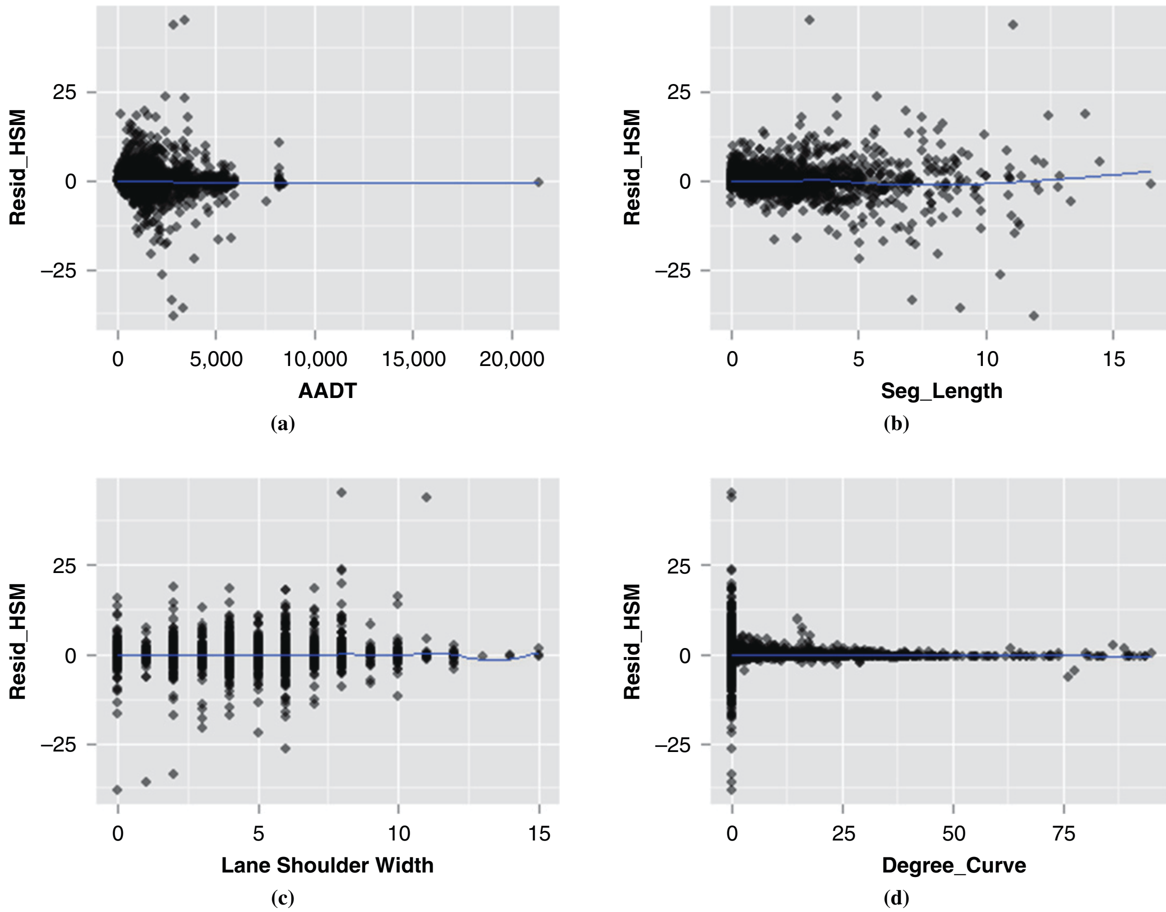

FIGURE 2 Residual plot with correlated variables for RT.

of the total data set for RT, RM4U, and RM4D facility types, respectively. The length proportions were 11.8%, 7.8%, and 15.6%, respectively. The numbers of observed crashes for both RM4U and RM4D were very low. Developing jurisdiction-specific SPFs with a very small sample size may yield impractical results and undermine safety effects. To increase the representation of the jurisdiction-specific base conditions, the CMFs for all facility types were reinvestigated along with the length distribution for a different combination of geometric features. The establishment of jurisdiction-specific base conditions is as follows:

- In the HSM, the CMF for lane width is the same for all sites with a lane width greater than or equal to 12 ft. A jurisdiction-specific base condition for lane width was created for all facility types by bundling together all sites with a lane width greater than or equal to 12 ft.
- The HSM provides CMFs for all sites with 0 ft, 2 ft, 4 ft, 6 ft, and 8 ft or more of shoulder width. All sites were bundled in 2-ft increments, that is, 0 to 2 ft, 2 to 4 ft, 4 to 6 ft, 6 to 8 ft, and 8 ft or more. The maximum share of roadway length was found in the 2- to 4-ft shoulder width for the RT and RM4D facilities and 4- to 6-ft shoulder width for the RM4U facility. All sites with a 2- to 4-ft shoulder width were considered as base sites for RT and RM4D facilities, and a 4- to 6-ft shoulder width was considered as a base site for RM4U facilities.
- South Dakota DOT data reported that sites with a centerline rumble strip prevailed on RT highways and that 34% of highway miles did not have a centerline rumble strip. The CMF provided in the HSM for a centerline rumble strip is a fixed value (0.94), so it is possible to consider the safety performance of a centerline rumble strip as an explanatory variable in the SPF. The centerline rumble strip was not considered as a base condition but as an explanatory variable for developing the SPF.
- The HSM recommends a 0% grade level as a base condition. The CMF value for grade is divided into three regions: level, moderate, and steep. All segments with a grade of less than or equal to 3% were considered to be level and were treated as base sites for the RT facility.
- The site length distribution was explored and the results showed that the major share of length has a 30-ft median width. It was observed that the site lengths with a median width less than 30 ft were quite high compared with sites that had more than 30 ft of median width. Weighted by segment length, the CMF for a median width less than or equal to 30 ft accounts for a 1.01% variation in predicted crashes. Relaxing the 30-ft median width in the base conditions will not compromise the prediction accuracy; thus, a median width of less than or equal to 30 ft was chosen as a base condition for the RM4D facility.
- All other variables were the same as base condition criteria described in the HSM.

After the jurisdiction-specific base conditions were established, the sample size conforming to these base conditions was compared with the sample size conforming to the HSM base conditions. Table 2 provides the sample size comparison for all facilities.

TABLE 2 Comparison of Base Conditions

Base Condition	HSM	Jurisdiction Specific
RT		
Lane width	12 ft	12 ft or more
Shoulder width	6 ft	2–4 ft
Centerline rumble strips	None	Used as variable in SPF
Grade level	0%	Less than or equal to 3%
RM4U		
Lane width	13 ft	12 ft or more
Shoulder width	6 ft	4–6 ft
RM4D		
Lane width	13 ft	12 ft or more
Right shoulder width	8 ft	2–4 ft
Median width	30 ft	30 ft or less
Sample Size		
RT		
Length: mile (% of whole data set)	230.93 (3.60)	1,832.96 (28.80)
Observed crashes: count (% of whole data set)	449 (4.30)	2,437 (23.40)
RM4U		
Length: mile (% of whole data set)	4.12 (2.70)	49.02 (32.20)
Observed crashes: count (% of whole data set)	18 (1.91)	296 (31.50)
RM4D		
Length: mile (% of whole data set)	22.205 (3.50)	206.1 (32.50)
Observed crashes: count (% of whole data set)	56 (3.10)	518 (28.92)

NOTE: Base conditions found to be different from the HSM recommended base conditions are provided in this table. Other base condition variables not presented here were considered to be the same as the HSM recommended base conditions.

The jurisdiction-specific base conditions were established by relaxing some of the base conditions proposed in the HSM; no new CMFs needed to be developed. The CMFs provided in the HSM can be used directly or adjusted for select geometric features.

Crash Modification Factor–Function Conversion

The HSM describes the CMF as follows: "CMF represents the relative change in estimated average crash frequency due to a change in one specific condition (when all other conditions and site characteristics remain constant)" (*2*). The specific conditions represent the base conditions for each facility type. If the base conditions are different, CMFs must be converted accordingly.

The jurisdiction-specific base lane width for all facility types is 12 ft or more. In the HSM, the CMF is the same for a lane width of 12 ft or larger; no CMF adjustment was needed for lane width. The base shoulder width for both RT and RM4D facilities in the jurisdiction-specific model was a 2- to 4-ft shoulder width. The CMF for shoulder width in the RT facility was divided into three domains according to AADT (*2*). The CMF value for an AADT of fewer than 400 vehicles and more than 2,000 vehicles was a scale factor from the study conducted by Zegeer et al. (*12*). The CMF for an AADT of between 400 and 2,000 vehicles a day was formulated as a linear equation where the variable is (AADT – 400), the intercept is the CMF value for AADT less than 400, and the slope is $CMF_{>2,000} - CMF_{<400}/2{,}000 - 400$. The conversion of CMFs according to the jurisdiction-specific base conditions was conducted with the CMF values provided in the HSM. When the CMF value is a scale factor, the new base condition will be reset to 1 when the CMF value is a scale factor. The other base conditions can be adjusted by a corresponding multiplier.

The base condition for shoulder width in the RM4U facility was 4 to 6 ft. This base condition can be considered the same as the HSM base condition because it used the CMF for maximum width in each band. No adjustment was required for the shoulder-width CMFs for jurisdiction-specific models. The CMF for right shoulder width at jurisdiction-specific base conditions for RM4D was calculated the same as the RT facility. For right shoulder width, all CMF values were scale values that were adjusted with corresponding multipliers by resetting the CMF value for the new base conditions to 1. The original CMF value for 2 to 4 ft was reset to 1, and other CMF values were adjusted by multiplying 1/1.09.

JURISDICTION-SPECIFIC PERFORMANCE FUNCTIONS

A review of the results obtained from the HSM predictive model application shows that jurisdiction-specific predictive models are more appropriate for South Dakota. It was expected that the SPFs developed with South Dakota data would be more accurate and reliable for predicting crash frequencies. The question of whether a calibration factor is necessary remains.

Similar to the SPFs in the HSM, the NB regression analysis was used to develop these SPFs. The centerline rumble strip was used as an explanatory variable to develop a jurisdiction-specific SPF for the RT facility type. The region categorical variable was considered in the jurisdiction-specific SPFs to check whether a geographical effect exists. The results obtained from the NB models for all facilities were reviewed to check the statistical significance of each variable used. The region variable was found to be statistically significant only for RM4U. The NB model was then rebuilt after

TABLE 3 Comparison of Jurisdiction-Specific SPFs with HSM SPFs

Facility Type	HSM SPF	Jurisdiction-Specific SPF
RT	$AADT \times L \times 365 \times 10^{-6} \times e^{(-0.312)}$	$AADT \times L \times 365 \times 10^{-6} \times e^{(0.10212-0.21439 \times \text{Rumble})}$
RM4U	$e^{(-9.653+1.176 \times \ln(AADT)+\ln(L))}$	$e^{(-3.1966+0.2482 \times \ln(AADT)+\ln(L)+1.3816 \times \text{Regi(Mitchell)}+1.0515 \times \text{Regi(Pierre)}+1.4866 \times \text{Regi(Rap\_City)})}$
RM4D	$e^{(-9.025+1.049 \times \ln(AADT)+\ln(L))}$	$e^{(-14.4774+1.8191 \times \ln(AADT)+\ln(L))}$

excluding the region variable from the RT and RM4D facilities. Updated model results show that all explanatory variables were statistically significant. A comparison of the SPFs for all facility types and the HSM SPFs is shown in Table 3.

The predictive models were applied and then compared with prediction accuracy statistics so the better-fitting model could be chosen. Some individual site predictions may overestimate crash frequency, while others may underestimate the frequency. It is important to access the prediction variation within each facility to assess the overall prediction accuracy of the model. The sum of absolute errors (SAE) and symmetric mean absolute percentage error (SMAPE) were used as the accuracy measurement to examine the model's goodness of fit within each facility type.

The SAE estimates the summation of absolute difference between the observed and predicted crashes, whereas SMAPE compares the relative error between models. Two versions of SMAPE were used to compare model performance, as the value of SMAPE is not as symmetric as it sounds. The over- and underestimation are not treated equally in the first version of SMAPE. The second version of SMAPE was used to measure the direction of data bias generated on each site. The formulas used to estimate the prediction accuracy can be formulated as follows:

$$\text{SAE} = \sum_{i=1}^{N} \left| Y_i - \hat{Y}_i \right| \tag{4}$$

$$\text{SMAPE\_1} = \frac{1}{N} \sum_{t=1}^{N} \frac{\left| \hat{Y}_i - Y_i \right|}{Y_i + \hat{Y}_i} \tag{5}$$

$$\text{SMAPE\_2} = \frac{\sum_{t=1}^{N} \left| \hat{Y}_i - Y_i \right|}{\sum_{t=1}^{N} \left(Y_i + \hat{Y}_i \right)} \tag{6}$$

where Y_i is the observation and $\hat{Y}_i$ is the prediction.

Table 4 presents the SAE and SMAPE values calculated for both jurisdiction-specific and HSM methods. In Table 4, the calibration factors calculated with jurisdiction-specific predictive models were closer to 1. The values of calibration factors obtained from jurisdiction-specific models show that it is not necessary to use a calibration factor for calibrated models, as the error rate is only 1% to 3% without the calibration factor.

The HSM models appear to perform better compared with jurisdiction-specific models because the SAEs were comparatively smaller in the HSM models for all facility types. However, in a comparison of SMAPE values, both versions of SMAPE show that the jurisdiction-specific models have a lower percentage of error. The mixed results may be related to the large percentage of sites without crashes; almost 78%, 64%, and 64% of sites for RT, RM4U, and RM4D, respectively, did not have crashes. It may be possible that the sites without crashes dominate the SAE estimate. Safety engineers value a more accurate estimate for sites with crashes over better predictions for sites without crashes. The statistics were calculated for 11 sites both with and without crashes, and the results are shown in Table 5.

The jurisdiction-specific predictive models provide more meaningful results and clearly outperform the HSM models at sites with crashes, as supported by both SAE and SMAPE statistics.

TABLE 4 Comparison of Prediction Accuracy

	HSM				Jurisdiction Specific			
Type	SAE	SMAPE_1	SMAPE_2	C_r	SAE	SMAPE_1	SMAPE_2	C_r
RT	8,410.01	0.89	0.43	1.18	8,835.55	0.87	0.42	0.99
RM4U	856.37	0.80	0.49	1.14	886.06	0.77	0.47	1.01
RM4D	1,404.74	0.82	0.48	1.57	1,505.32	0.78	0.42	1.03

TABLE 5 Prediction Accuracy Between Sites with Crashes and Sites Without Crashes

	Sites Without Crashes		Sites With Crashes					
	SAE		HSM			Jurisdiction Specific		
Type	HSM	Jurisdiction Specific	SAE	SMAPE_1	SMAPE_2	SAE	SMAPE_1	SMAPE_2
RT	2,286.6	2,787.79	6,123.41	0.49	0.36	6,047.1	0.45	0.33
RM4U	275.88	373.64	580.49	0.45	0.39	512.42	0.37	0.34
RM4D	291.29	447.5	1,113.45	0.48	0.42	1,057.82	0.40	0.34

CONCLUSIONS

Quantifying highway safety performance and predicting the number of crashes on a roadway segment are important for identifying effective safety countermeasures. The HSM crash predictive method has been used by transportation agencies to screen highway networks and find problem areas for further safety review, to compare safety design alternatives, evaluate site-specific safety issues, and plan future safety projects. Although calibration procedures are available in Appendix A of the HSM, they need to be refined or modified to accommodate local data availability and include qualities of local roadways, traffic trends, and crash characteristics. Proper calibration procedures and guidance for future calibration activities are necessary. In this study, the calibration was guided by rigorous analytical procedures that thoroughly investigated the causes of substantial underestimation by the HSM predictive method.

The presence of large deviations prompted the review of the spatial distribution of the calibration factor, the residual analysis for all available variables, and the identification of local base conditions. The findings led to the development of jurisdiction-specific SPFs, the inclusion of regional effects in the SPF, and the conversion of CMFs to base conditions if they differ from those in the HSM. The prediction accuracy measured by SAE and SMAPE indicates lower relative error percentage by the jurisdiction-specific models for all facility types. The results were further divided into a group of sites without crashes and a group of sites with crashes. The SAE and SMAPE results show that jurisdiction-specific models better fit the crash data for sites with crashes. The calibration factor was found to no longer be necessary when used with jurisdiction-specific models because the SPFs and CMFs can account for most variations within the data, and the differences between predicted and observed values are not significant (between 1% and 3%).

ACKNOWLEDGMENTS

The authors thank the South Dakota Department of Transportation and the Mountain–Plains Consortium at North Dakota State University for funding this research.

REFERENCES

1. *Traffic Safety Facts: 2012 Data.* NHTSA, Washington, D.C., 2014.
2. *Highway Safety Manual.* AASHTO, Washington, D.C., 2010.
3. Resende, P. T. V., and R. F. Benekohal. Effects of Roadway Section Length on Accident Modeling. In *Traffic Congestion and Traffic Safety in the 21st Century: Challenges, Innovations, and Opportunities,* ASCE, Reston, Va., 1997.
4. Qin, X., and A. Wellner. Segment Length Impact on Highway Safety Screening Analysis. Presented at 91st Annual Meeting of the Transportation Research Board, Washington, D.C., 2012.
5. Ogle, J. H., P. Alluri, and W. Sarasua. MMUCC and MIRE: The Role of Segmentation in Safety Analysis. Presented at 90th Annual Meeting of the Transportation Research Board, Washington, D.C., 2011.
6. Brimley, B. K., M. Saito, and G. G. Schultz. Calibration of Highway Safety Manual Safety Performance Function. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2279,* Transportation Research Board of the National Academies, Washington, D.C., 2012, pp. 82–89.
7. Abdel-Aty, M. A., P. C. Lee, P. Eng, J. Park, J.-H. Wang, M. Abuzwidah, and S. Al-Arifi. *Validation and Application of Highway Safety Manual (Part D) in Florida.* Florida Department of Transportation, Tallahassee, 2014.
8. Martinelli, F., F. La Torre, and P. Vadi. Calibration of the Highway Safety Manual's Accident Prediction Model for Italian Secondary Road Network. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2103,* Transportation Research Board of the National Academies, Washington, D.C., 2009, pp. 1–9.
9. Xie, F., K. Gladhill, K. K. Dixon, and C. M. Monsere. Calibration of Highway Safety Manual Predictive Models for Oregon State Highways. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2241,* Transportation Research Board of the National Academies, Washington, D.C., 2011, pp. 19–28.
10. Banihashemi, M. Highway Safety Manual, New Model Parameters vs. Calibration of Crash Prediction Models. Presented at 90th Annual Meeting of the Transportation Research Board, Washington, D.C., 2011.
11. Mehta, G., and Y. Lou. Calibration and Development of Safety Performance Functions for Alabama. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2398,* Transportation Research Board of the National Academies, Washington, D.C., 2013, pp. 75–82.
12. Zegeer, C. V., R. C. Deen, and J. G. Mayes. *The Effect of Lane and Shoulder Widths on Accident Reductions on Rural, Two-Lane Roads.* Kentucky Department of Transportation, Frankfort, 1980.

The Standing Committee on Highway Safety Performance peer-reviewed this paper.