GUIDELINES FOR RECALIBRATING PREDICTIVE MODELS OVER TIME BASED ON MODEL UNCERTAINTY

Bharadwaj Bommanayakanahalli, Corresponding Author
Traffic/ITS Engineer I
Arcadis U.S., Inc.
Houston, TX 77042, United States
Tel: (979) -739-6412 Email: bharadwaj472@rocketmail.com

Dominique Lord
Professor
Zachry Department of Civil Engineering, Texas A&M University
College Station, TX 77843, United States
Tel: (979) 458-3949 Email: d-lord@tamu.edu

Word Count: 3,947 words text + 8 tables * 250 words (each) = 5,947 words

Submission Date: August 1st, 2018

ABSTRACT

The Highway Safety Manual (HSM) recommends recalibrating Safety Performance Functions (SPFs) every two or three years. However, no formal research study has been conducted to support this recommendation. With no appropriate guidelines, transportation safety analysts could face the following two scenarios: 1) recalibrating the models when it is not needed; and, 2) not recalibrating them when it is, in fact, needed. This paper documents research that has been conducted to address this issue. Consequently, the objectives of this research are: 1) document a methodology to estimate the recalibration need based on the model uncertainty; and, 2) use the methodology to quantify a threshold to make a recalibration decision based on characteristics of data collected at the network level. Using this methodology, recalibrating SPFs is recommended when the total number observed crashes fall beyond its confidence intervals, which was set to the 95%-percentile. The methodology was also compared to another method known as the C-proxy calibration procedure for recalibrating models over time. The proposed methodology was evaluated using three intersection and segment datasets collected in Michigan and Toronto. The results show that the methodology provides good insights as to when recalibration is need, which also closely matches the C-proxy calibration procedure. However, the calibration methodology could be dependent on the selection of the confidence intervals (i.e., 90%, 95%, 99%). Hence, engineering judgment may therefore be needed for the recalibration procedure.

INTRODUCTION

The Highway Safety Manual (HSM) *(1)* is document that provides tools and methods to improve safety on streets and highways. The HSM prediction models known as Safety Performance Functions (SPFs) play an integral role in many safety evaluations and analyses documented in HSM *(2)*. While the current edition of the HSM documents SPFs for several facility types, these models should be calibrated when applied to a new jurisdiction (note: developing a jurisdiction-specific model instead is also an alternative to calibration. Given the scope of this study, however, we focus on documenting the calibration procedure). Similarly, even within the same jurisdiction, predictive models may need to be recalibrated (or a new model fitted) over time; recalibration is necessary since the frequency of crashes is expected to fluctuate as driver behavior (further affected by changes in the demographics of a region, driver awareness programs introduced by the DOTs), vehicle characteristics (e.g., advanced warning systems, better braking system), and roadway characteristics, among others, change over time.

Appendix A of part C of the HSM describes a method to calibrate (or recalibrate) the SPF models using a scalar calibration factor. As part of this method, the HSM recommends that the SPF models be recalibrated every 2 or 3 years. However, no formal studies have been performed for justifying this suggestion. In fact, these guidelines are not supported by previous research studies *(2-4)*. A transportation safety analyst adopting the HSM guidelines may experience one of the following unfavorable scenarios: 1) recalibrating the model and later realizing it was not necessary; 2) not recalibrating the model when such a need arises. Research is therefore needed to provide more reliable criteria to recognize the recalibration need. Consequently, this paper aims to develop better guidelines to recognize (or estimate) the recalibration need over time. To accomplish the study objective, this research employs the uncertainty associated with the predictive model to provide a strategy for recalibrating models over time.

LITERATURE REVIEW

The first edition of the HSM *(1)* documents a procedure to calibrate (or recalibrate) the SPF models to a new jurisdiction (or over time). In this procedure, after collecting the necessary data, a scalar calibration factor is calculated by taking the radio between the total observed and the total pre-calibration predicted number of crashes (as shown in Equation 1).

$$C = \frac{\sum N_{obs}}{\sum N_{pre}} \tag{1}$$

Where,

C = Calibration Factor;

N_{obs} = Number of observed crashes; and,

N_{pre} = Number of predicted crashes (pre-calibration); and,

Σ = Summation over all the sites that are part of the sample.

The calibrated model is then derived as:

$$N_{\mathrm{SPF}} = N_{\mathrm{pre}} \times \mathrm{C} \tag{2}$$

Where,

N_{SPF} = Predicted number of crashes after calibration.

Several research studies have applied the methodology described in the HSM to calibrate the SPF models to the local jurisdictions *(5-11)*. In these studies, several limitations have been identified. As such, it was noted that the calibration procedure can be a time-consuming task and demand significant resources. Several research studies have also noted issues regarding the required calibration sample size *(2, 12-15)*. Research studies as well noted the HSM limitations about appropriate guidelines to recalibrate the models over time *(2, 4, 15)*. Note that even within the same jurisdiction, predictive models may need to be updated because of continuous change in driver behavior, and vehicle and roadway characteristics over time, as discussed above. Updating predictive models can again be done through either applying a scalar calibration factor (as shown in equation 1) or refitting the model every few years (*16, 17)*.

Researchers have tried to provide revised guidelines on recalibration need over time. As such, Shirazi et al. *(3)* developed a methodology to determine when recalibration is warranted given a few data characteristics collected at the network level. The methodology involves calculating a simple network level factor referred to as C-proxy. The C-proxy for segment and intersection models is derived using Equations 3 and 4, respectively.

$$\tilde{C} = \frac{N_{obs}^{T}}{\left(e^{b_0+b_1*\ln(\bar{F})}\right)\times L^{T}} \quad (Segment\ Models) \tag{3}$$

$$\tilde{C} = \frac{N_{obs}^{T}}{\left(e^{b_0+b_1*\ln\left(\bar{F}_{major}\right)+b_2*\ln\left(\bar{F}_{minor}\right)}\right)\times N^{T}} \quad (Intersection\ Models) \tag{4}$$

Where,

$\tilde{C}$ = C-proxy;

N_{obs}^{T} = the total number of crashes in network;

$\bar{F}$ = the mean value of traffic flow in the network;

L^{T} = the total combined faculty length in the network;

$\bar{F}_{major}$ = the mean value of the traffic flow on major streets in the network;

$\bar{F}_{minor}$ = the mean value of the traffic flow on minor streets in the network; and,

N^{T}= total number of intersections in the network.

Shirazi et al. *(3)* recommended evaluating the C-proxy periodically and calculating its percentage change compared to the C-proxy that is evaluated in the reference year (note: the latest year in which the model was recalibrated is referred to as reference year). The authors suggested using a threshold of 10% to decide on recalibration. If the percentage change is greater than 10%, recalibration is recommended; else, no recalibration is warranted. Although not based on theoretical principles, the value of 10% was considered an appropriate balance between the frequency of recalibration and the effort needed for the calibration procedure.

In another study, Saha et al. *(4)* used the Bayesian estimation technique to establish guidelines for the frequency of recalibrating models. The authors' primary hypothesis was to evaluate the variation in the calibration factors for different facility types computed once every year, once every two years, and once every three years to determine the frequency of updating. The results of their study indicated that when the variation between C-factors (evaluated considering the total crashes) is less than or equal to 0.01, the model for 4-legged signalized intersections should be recalibrated every year, while the models for other facilities should be recalibrated every 2 years. Their results further suggest that when the variation (evaluated considering the total crashes) is greater than 0.01, but less than 0.05, the C-factors should be updated every year or every 2 years for 4-legged intersections, and every 3 years for other facility types. According to Saha et al. *(14)*, the limitations of their study included the lack of transferability to other jurisdictions and the fact that the data used only comes from arterial urban and suburban roads in Florida.

METHODOLOGY

This section describes the methodology for determining when predictive models need to be recalibrated over time.

Calculating the Variance and Confidence Intervals

The general nature of crash data is that it displays a high degree of randomness and usually exhibits over-dispersion *(18)*. The most commonly used or popular model that can handle over-dispersion is the Poisson-gamma or negative binomial (NB) model. This is the model used for the predictive methodology described in the HSM. Hence, this was also the model used in the present study.

The probability mass function (PMF) of the Poisson-gamma structure described above is given by the following equation:

$$f\left(y;\phi,\mu\right)=\binom{y+\phi-1}{\phi-1}\left(\frac{\phi}{\mu+\phi}\right)^{\phi}\left(\frac{\mu}{\mu+\phi}\right)^{y} \tag{5}$$

Where,

y = observed number of crashes;

μ = estimated number of crashes; and,

ϕ= inverse dispersion parameter of the Poisson-gamma distribution.

The term ϕ is usually defined as the "inverse dispersion parameter" of the Poisson-gamma distribution.

Wood *(19)* developed a methodology to calculate the variance and confidence intervals for generalized linear models (GLMs) with a Poisson and NB error structures. The methodology was developed for three cases: 1) the Poisson mean μ, 2) gamma mean m (the mean of a site within a population of sites that follows the same gamma distribution), and 3) predicted response z (i.e., applying the model to a new dataset). Table 1 shows the variances for these estimates. This table shows that the variances used to estimate the uncertainty of the gamma mean and the predicted response for the Poisson-gamma model both incorporate the inverse dispersion parameter, ϕ.

Table 1. Variance Estimation for Poisson and Poisson-gamma models (*19*)

Parameter	**Variance**		
		Poisson	**Poisson-gamma**
μ	v_{μ}	$\mu^2 Var(\eta)$	$\mu^2 Var(\eta)$
m	v_m	----	$\left\{\mu^2 \operatorname{var}(\eta) + \frac{\mu^2 \operatorname{var}(\eta) + \mu^2}{\phi}\right\}$
z	v_z	$\left\{\mu^2 Var(\eta) + \mu\right\}$	$\left\{\mu^2 Var(\eta) + \frac{\mu^2 Var(\eta) + \mu^2}{\phi} + \mu\right\}$

Note:

$Var(\eta) = \mathbf{x}_0' (\mathbf{XWX}')^{-1} \mathbf{x}_0$ (taken from the model's variance-covariance matrix)

μ = mean estimate of the Poisson or Poisson-gamma model

m = gamma mean

z = predicted mean

ϕ = inverse dispersion parameter of the Poisson-gamma model

v_{μ} = variance of the Poisson mean

v_m = variance of the gamma mean

v_z = variance of the predicted response

The methodology considers the entire network and evaluates the total number of observed crashes (y) and estimated number of crashes (μ) from the model. Similarly, it also evaluates the variance associated with the gamma mean (v_m). The variance can be directly added since the values of the parameters at each site are independent of other sites. The researchers also evaluated the variance of the predicted response (v_z), but the results were the same since the difference in 95% CIs was very close.

Under the central limit theorem (CLT), which states that the mean and the sum of a random sample from an arbitrary distribution have an approximately normal distribution when the sample size is sufficiently large, the distribution of the sum of the parameters (y, μ, ν_m) can be independently added and assumed to be normally distributed. The CLT is described in Equation 6.

$$\sum \text{mean}(x_i) \sim \text{Normal}\left(\sum(x_i), \text{Var}(x_i)\right) \tag{6}$$

The methodology is accomplished with the following four steps:

- Develop the SPF (flow-only model) using data from the initial period (*t*) and obtain the variance-covariance matrix of this SPF.

- Apply the SPF to the entire network in the subsequent period (*t+n*) and evaluate the variance of the gamma mean (ν_m) using the variance-covariance matrix and the estimated number of crashes for each of the individual sites.

- Compute the summation of the observed number of crashes for the entire network, the estimated number of crashes (using the latest calibrated model), and the variance of the gamma mean.

- Evaluate the 95% CIs of the gamma mean ($\sum \mu \pm 1.96 \times \sqrt{\sum \nu_m}$) (or 1.65 for 90% or 2.58 for 99%). The model should be recalibrated when the observed number of crashes lies outside the CIs. If it is not, use the same model in the subsequent analysis period.

- Repeat the same procedure for the subsequent time periods (using the latest recalibrated model) to investigate the need for recalibration.

Functional Form

The functional forms used for this study are the same ones documented in the HSM. For intersections, the functional form is described in Equation 7. The functional form contains two variables, which consist of the entering intersection flows.

$$\mu = \beta_0 * F_{Major}^{\beta_1} * F_{Minor}^{\beta_2} \tag{7}$$

Where, F_{Major} and F_{Minor} are the entering flows (AADT) for the major and minor streets, respectively.

In the present study, general AADT models were developed. These models only include traffic flow as variables, without specifically capturing the effects of other operational and geometric design variables via covariates. Their effects are indirectly captured by the traffic flow variables since not all the sites meet the baseline conditions.

For segments, the functional form is described in Equation 8. For this form, the segment length (L) is considered an exposure variable, which means that the coefficient is fixed to 1. The relationship between crash risk and traffic flow is assumed to be non-linear.

$$\mu = L * \beta_0 * F^{\beta_1} \tag{8}$$

Where, F is the traffic flow (AADT) traveling through the segment. The NB models were estimated using the software program R *(22).*

CHARACTERISTICS OF THE DATA

The methodology developed in the present study was evaluated using the following datasets:

- Toronto 4-legged intersections (6-year dataset): Crash data and entering traffic flows were collected at 868 4-legged signalized intersections in Toronto for a period of 6 years. This dataset has been used extensively by other researchers *(2, 20, 21).*
- Michigan 4-legged intersections (5-year dataset) (*13*): crash data along with the flows have been collected at 4-legged signalized intersections in Michigan for a period of five years. Entering traffic flow and crash data were collected at 349 intersections for each of the five years.
- Michigan 4-lane divided roads (5-year dataset) (*13*): The crash data along with the flows, and segment length have been collected on 4-lane divided segments in Michigan for a period of five years.

The characteristics of the datasets are shown in Table 2.

TABLE 2 Characteristics of Datasets Used in the Analysis

Year	**Toronto**			**Michigan**					
	4-Legged Intersections (n=868)*			**4-Legged Intersections (n=349)**			**4-Lane Divided Highways (n~370)**		
	Mean (std)			**Mean (std)**			**Mean (std)**		
	Major AADT	**Minor AADT**	**Crash**	**Major AADT**	**Minor AADT**	**Crash**	**AADT**	**Length (miles)**	**Crash**
1	27033 (10189)	10581 (8158)	9.53 (7.82)	20889 (15243)	8781 (7915)	8.38 (8.37)	10071 (5683)	0.99 (0.75)	3.74 (6.00)
2	27014 (10188)	10579 (8173)	9.38 (7.68)	20997 (15014)	8832 (7905)	8.23 (8.09)	9898 (5671)	0.99 (0.75)	3.64 (5.93)
3	27291 (10304)	10694 (8281)	10.04 (7.95)	21445 (15223)	9034 (8104)	8.56 (8.25)	9989 (5603)	0.99 (0.75)	3.28 (4.85)
4	27460 (10385)	10767 (8358)	11.31 (9.67)	21078 (15027)	8870 (7979)	8.49 (8.49)	10043 (5602)	0.96 (0.73)	3.84 (6.25)
5	27983 (10604)	10979 (8545)	11.53 (10.09)	21380 (15249)	8992 (8087)	8.35 (8.33)	10185 (5576)	0.96 (0.73)	4.16 (6.99)
6	28045 (10654)	11010 (8594)	11.56 (10.01)	-	-	-	-	-	-

**The sample size denoted by "n"*

Table 3 shows the model coefficients for the three datasets. The inverse dispersion parameter varies between 2.7 to 6.8.

TABLE 3 Model Output for Datasets Used in the Analysis

Coefficients	Mean Value (std.)		
	Toronto	Michigan	
	4-legged Intersections	4-legged Intersections	4-Lane Divided Highways
Intercept (ln β_0)	-8.84 (0.48)	-7.42 (0.66)	- 6.10 (0.80)
F_{Major} (β_1)	0.51 (0.05)	0.75 (0.07)	-
F_{Minor} (β_2)	0.64 (0.02)	0.23 (0.04)	-
F (β_1)	-	-	0.797 (0.09)
Inverse dispersion parameter (ϕ)	6.77 (0.62)	2.74 (0.295)	2.82 (0.46)
Model Performance			
Akaike information criterion (AIC)	4881.9	2036.62	1525.677
Bayesian information criterion (BIC)	4901.033	2052.04	1537.442
Median absolute deviation (MAD)	1.007	0.976	1.042
Mean square prediction error (MSPE)	1.091	1.103	1.133

Table 4 shows the variance-covariance matrices $(\mathbf{XWX'})^{-1}$ for the three datasets. Those were used for calculating $Var(\eta) = \mathbf{x}_0'(\mathbf{XWX'})^{-1}\mathbf{x}_0$.

Table 4. Variance-Covariance Matrix for the Three Datasets

Toronto – 4-legged Intersections			
	Intercept	F_{major}	F_{minor}
Intercept	0.2346591000	-0.0210500408	-0.0021762592
F_{major}	-0.0210500410	0.0022846600	-0.0002422538
F_{minor}	-0.0021762590	-0.0002422538	0.0005067881
Michigan – 4-legged Intersections			
	Intercept	F_{major}	F_{minor}
Intercept	0.436101710	-0.042030110	-0.002381334
F_{major}	-0.042030110	0.005249850	-0.001103465
F_{minor}	-0.002381334	-0.001103465	0.001513798
Michigan – 4-Lane Divided Segments			
	Intercept	F	
Intercept	0.63957695	-0.06921271	
F	-0.06921271	0.007514085	

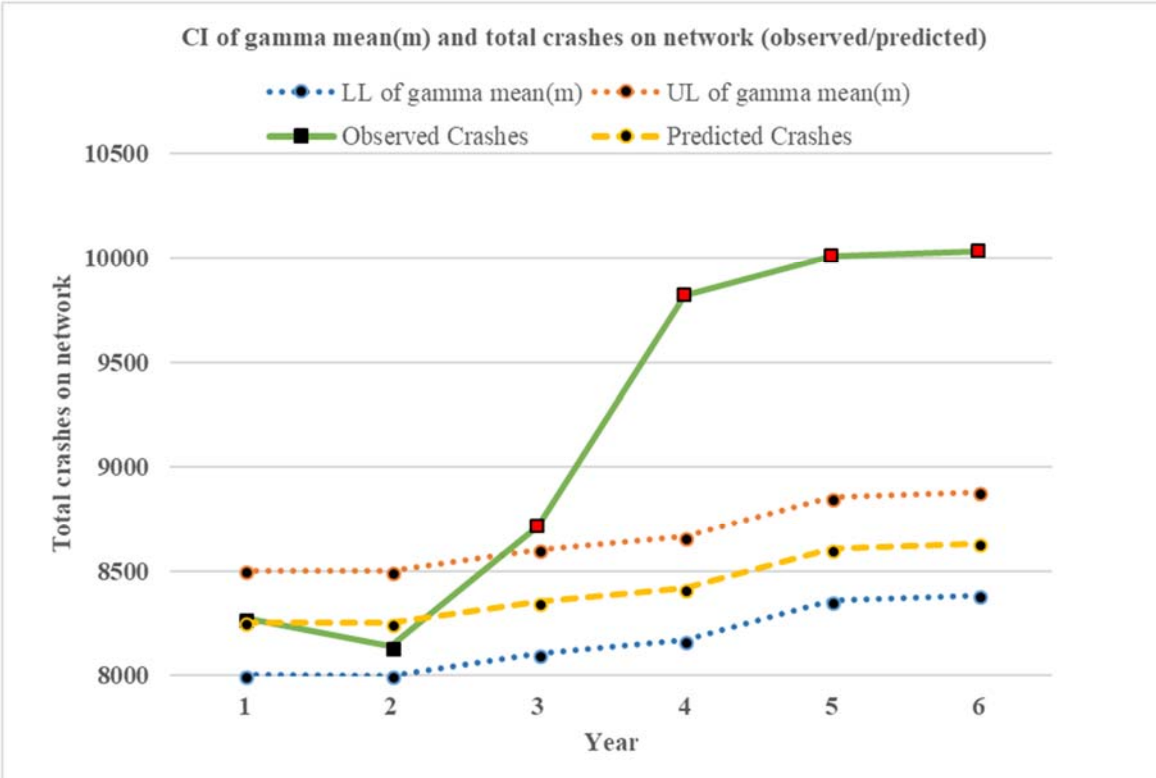
ANALYSIS RESULTS

This section describes the results of the proposed methodology. The section is divided into three parts, one for each dataset described above.

Toronto 4-Legged Intersections

Table 5 summarizes the application of the methodology for the Toronto 4-legged intersections dataset. This table shows the variance and the CIs for the summation of the gamma mean (m) over the period of 6 years using the method proposed in this research. Table 5 shows the CIs for the uncalibrated model. In other words, the CIs are calculated based on the base model applied in Year 1 for all the years. Figure 1 shows the observed number of crashes for all the sites and the 95% CI of the predicted values. Similar to Table 5, the predicted values shown in Figure 1 are evaluated using the base model which has not been recalibrated in subsequent years.

The results for the Toronto data indicate that the model needs to be recalibrated in both Year 3 to Year 6, respectively. In contrast, the results using Shirazi et al.'s (*3*) methodology indicate that the model needs to be calibrated in Years 4, 5 and 6. As shown in Figure 1, the model shows very little year-to-year variation. The observed crash counts, on the other hand, show a huge increase between Year 3 and Year 4 and remained stable between Year 4 and Year 6. Note that although the method suggested calibration in Year 3, the major deviance from the predicted CIs happens in year 4 (see Figure 1). These results are compatible with the C-proxy decision. The change in C-proxy in Year 4 as well increases substantially. Even though, in this study, the researchers used the 95% CIs, the results show that using a wider CIs (say 99%) would lead to no recalibrated model despite the characteristics shown in Figure 1. In the end, the selection of the CIs (90%, 95% or 99%) will need to be decided by the safety analyst given the cost and the effort needed to recalibrate models.



(Note: LL=Lower Limit; UL=Upper Limit)

FIGURE 1 Illustration of proposed method for Toronto 4-legged intersections

TABLE 5 Summary Results for the Toronto 4-Legged Intersections Based on the Proposed Methodology (Uncalibrated Model) and Shirazi et al. *(3)*.

Year	Total Crashes		Calibration Factor		Methodology				
					Prediction Confidence Interval		C-Proxy (*3*)		
	Observed	Predicted [b]	C-Factor	Change in C-Factor	Confidence Intervals	Calibration? (Yes/No)	C-proxy	Change in C-proxy	Calibration? (Yes/No)
1[a]	8276	8254	1.00	NA	NA	NA	0.94	NA	NO
2	8141	8251	0.99	1.00%	(8001.8, 8500.1)	NO	0.93	1.06%	NO
3	8714	8353	1.04	4.00%	(8103.9, 8602.3)	YES**	0.98	4.26%	NO
4	9818	8418	1.17	17.00%	(8168.3, 8666.6)	YES	1.10	17.02%	YES[c]
5	10010	8607	1.16	16.00%	(8357.9, 8856.2)	YES	1.09	15.96%	YES
6	10030	8634	1.16	16.00%	(8384.7, 8883.0)	YES	1.09	15.96%	YES

[a] The year where the model was initially developed
[b] Uncalibrated predicted crashes
[c] The first-time calibration need warranted was underlined.

Table 6 shows the results when the model is recalibrated and redefined as the base model for the year that it has been recalibrated. In this case, the proposed methodology shows that model needs to be recalibrated in Year 3 and Year 4, but does not need to be recalibrated in Year 5 and Year 6.

TABLE 6 Summary Results for the Toronto 4-Legged Intersections Based on the Proposed Methodology (Recalibrated Model)

Year	Total Observed	Total Predicted (Calibrated)	Confidence Intervals (CIs)	Calibration? (Yes/No)
1[a]	8276	8254	NA	NA
2	8141	8251	(8001.8, 8500.1)	NO
3	8714	8353	(8103.9, 8602.3)	YES[b]
4	9818	8781	(8521.2, 9041.0)	YES
5	10010	10039	(9748.5, 10329.7)	NO
6	10030	10070	(9779.7, 10360.9)	NO

[a] The year where the model was initially developed
[b] The first-time calibration need warranted was underlined.

Michigan 4-legged Intersections

Table 7 summarizes the application of the methodology for the Michigan 4-legged intersections dataset. This table shows the variance and the CIs for the summation of the gamma mean (m) over the period of 5 years using the proposed methodology. Figure 2 shows observed number of crashes for all sites and the 95% CIs of the predicted values. The predicted values shown in Figure 2 are evaluated using the base model which has not been recalibrated in subsequent years.

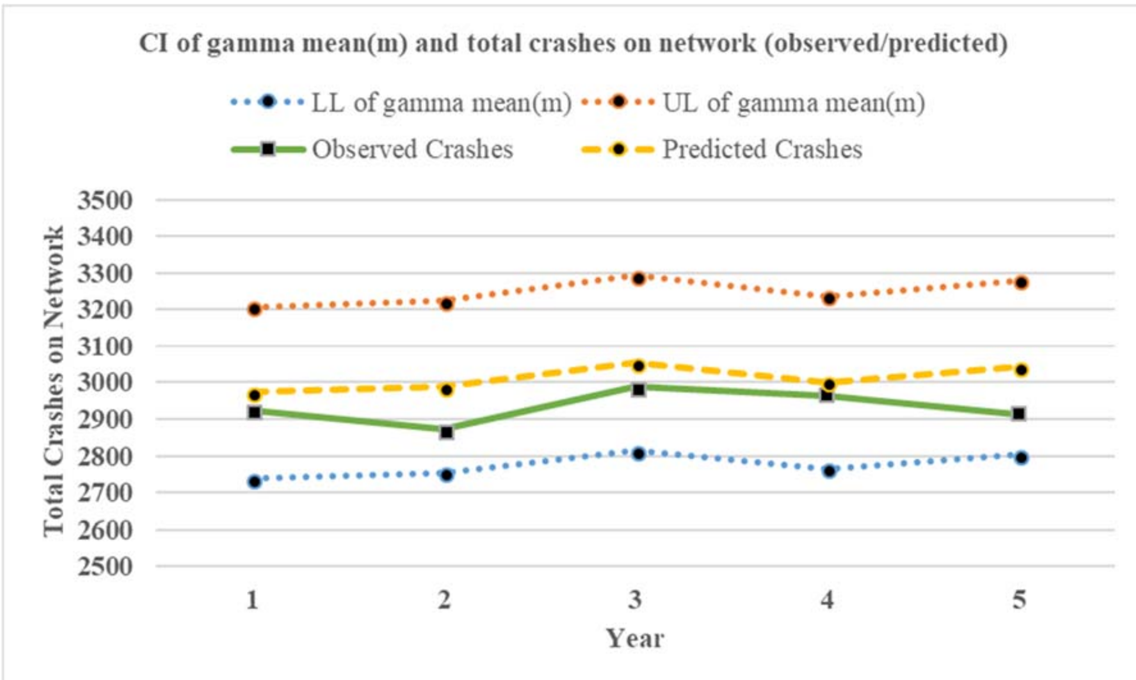


FIGURE 2 Illustration of proposed method for Michigan 4-legged intersections

As shown in both the table and figure, both methodologies indicate that there is no need to recalibrate the model within the 5-year period. For this dataset, recalibrating the model under the HSM guidelines would have been unnecessary. The model would have been recalibrating at least twice.

TABLE 7 Summary of the Results for Michigan 4-Legged Intersections Based on the Proposed Methodology (Uncalibrated Model) and Shirazi et al. (*3*)

Year	Total Crashes		Calibration Factor		Methodology				
					Prediction Confidence Interval		C-Proxy (*3*)		
	Observed	Predicted [b]	C-Factor	Change in C-Factor	Confidence Intervals	Calibration? (Yes/No)	C-proxy	Change in C-proxy	Calibration? (Yes/No)
1[a]	2925	2973	0.98	NA	NA	NA	0.93	NA	NA
2	2872	2988	0.96	2.04%	(2753.7, 3223.1)	NO	0.91	1.84%	NO
3	2989	3054	0.98	0.00%	(2814.3, 3293.3)	NO	0.92	0.75%	NO
4	2965	3000	0.99	1.02%	(2764.6, 3235.7)	NO	0.93	0.32%	NO
5	2914	3042	0.96	2.04%	(2802.9, 3280.8)	NO	0.90	2.91%	NO

[a] The year where the model was initially developed
[b] Uncalibrated predicted crashes

Michigan 4-lane divided segments

Table 8 summarizes the application of the methodology for the Michigan 4-lane divided highway segments dataset. This table shows the variance and the CIs for the summation of the gamma mean (m) over the period of 5 years using the proposed methodology. Figure 3 shows observed number of crashes for all the sites and the 95% CIs of the predicted values. The predicted values shown in Figure 3 are evaluated using the base model which has not been recalibrated in subsequent years.

For this dataset, the proposed methodology recommends recalibrating only in Year 5, while the methodology proposed in Shirazi et al. (*3*) recommends recalibrating the model both in Year 3 and Year 5. However, looking at the results more closely, if we select a confidence interval equal to 90%, the model would only need to be recalibrated in Year 3, since the observed value lies at the lower boundary of the 95% CIs. If the model is calibrated at Year 3 (calibrated down), it would be expected that the model would need to be recalibrated again in Year 4, but not year 5. This again shows that the selection of the CIs could play a role into when the model should be recalibrated.

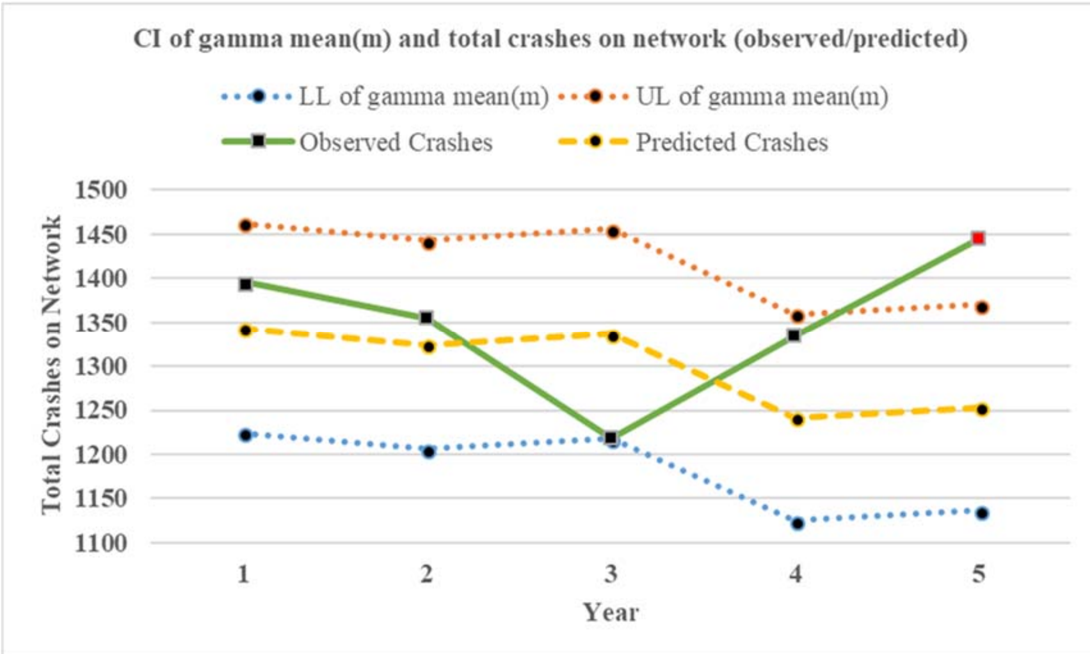


FIGURE 3 Illustration of proposed methodology for Michigan 4-lane divided segments

TABLE 8 Summary of the Results for Michigan 4-Lane Divided Segments Based on Proposed Methodology (Uncalibrated Model) and Shirazi et al. (*3*)

Year	Total Crashes		Calibration Factor		Methodology				
					Prediction Confidence Interval		C-Proxy (*3*)		
	Observed	Predicted [b]	C-Factor	Change in C-Factor	Confidence Intervals	Calibration? (Yes/No)	C-proxy	Change in C-proxy	Calibration? (Yes/No)
1[a]	1396	1343	1.04	NA	NA	NA	1.08	NA	NA
2	1354	1325	1.02	1.69%	(1206.5, 1442.6)	NO	1.07	0.80%	NO
3	1218	1337	0.91	12.36%	(1217.4, 1456.0)	NO	0.95	11.92%	YES[c]
4	1334	1242	1.07	3.33%	(1125.0, 1358.9)	NO	1.14	5.68%	NO
5	1444	1253	1.15	10.87%	(1136.8, 1369.9)	YES[c]	1.22	13.10	YES

[a] The year where the model was initially developed
[b] Uncalibrated predicted crashes
[c] The first-time calibration need warranted was underlined.

SUMMARY AND CONCLUSIONS

This paper has described the development of a methodology for recalibrating SPFs over time. The methodology uses variance associated with the SPFs to develop the guidelines for recalibration. The variance of the SPFs was evaluated using the methodology developed by Wood *(19)*. The methodology developed in this research suggests recalibration of the SPF when the observed number of crashes on the entire network (for a facility type) lies outside the 95% CIs of the gamma mean (*m*) of the entire network of that facility type, although other percentiles (say 90% or 99%) could be used for the decision of recalibrating models. The results of this study were compared to those using the methodology developed by Shirazi et al. *(3)*, which suggested to use a 10% threshold for recalibrating models.

The methodology proposed in this paper and the one from Shirazi et al. (*3*) were applied to three intersection and segment datasets from Toronto and Michigan. The results indicate that these methodologies can give safety analysts and practitioners better guidance regarding the frequency of recalibration than the one proposed in the HSM. Both provided similar recalibration guidelines, but the one proposed in this paper could be dependent on the selection of the CIs. Hence, engineering judgment may therefore be needed for the recalibration procedure.

The following are some of the limitations of the proposed methodology:

- The proposed methodology requires the variance-covariance matrix output from the models. Unfortunately, this matrix is not available for the models documented in the HSM (and most publications). The researchers are therefore recommending that the matrices be made available for the future editions of the MSH.

- Advanced statistical knowledge is needed to apply this methodology, since it involves the multiplication of multiple matrices. To overcome this, a function written in an Excel spreadsheet could be developed to automatize the recalibration process.

Future research should focus on determining the minimum number (and potentially the nature) of observations required to apply the proposed methodology to obtain the same output regarding the recalibration compared to using the full dataset. Another potential area of research is to simulate crash data for certain scenarios and identify the sample size requirements. Furthermore, the use of simulation could allow to properly identify false positives and false negatives with regards to selecting the appropriate CIs.

ACKNOWLEDGEMENTS

The authors would like to thank Dr. Mohammadali Shirazi for his input and comments provided for this paper.

CONTRIBUTION

The authors confirm contribution to the paper as follows: study conception and design: Bharadwaj Bommanayakanahalli, Dominique Lord; data collection: Dominique Lord; analysis and interpretation of results: Bharadwaj Bommanayakanahalli, Dominique Lord,; draft manuscript preparation: Bharadwaj Bommanayakanahalli, Dominique Lord. All authors reviewed the results and approved the final version of the manuscript.

REFERENCES

1. *Highway safety manual*. American Association of State Highway and Transportation Officials, Washington, D.C., 2010.

2. Lord, D., S.R. Geedipally, and M. Shirazi. Improved Guidelines for Estimating the Highway Safety Manual Calibration Factors. Final Report ATLAS-2015-10. Texas A&M Transportation Institute, College Station, TX, 2016.

3. Shirazi, M., S.R. Geedipally, and D. Lord. "A procedure to determine when safety performance functions should be recalibrated." *Journal of Transportation Safety & Security*, Vol. 9, No. 4, 2017, pp.457-469.

4. Saha, D., P. Alluri and A. Gan. "A Bayesian procedure for evaluating the frequency of calibration factor updates in highway safety manual (HSM) applications." *Accident Analysis & Prevention*, *98*, 2017, pp.74-86.

5. Xie, F., K. Gladhill, K. K. Dixon, and C. M. Monsere. "Calibration of Highway Safety Manual Predictive Models for Oregon State Highways." *Transportation Research Record 2241: Journal of the Transportation Research Board*, 2011, pp. 19–28.

6. Brimley, B. K., M. Saito, and G. G. Schultz. "Calibration of Highway Safety Manual Safety Performance Function." *Transportation Research Record 2279: Journal of the Transportation Research Board*, 2012, pp. 82–89.

7. Mehta, G., and Y. Lou (2013). Calibration and Development of Safety Performance Functions for Alabama. *Transportation Research Record 2398: Journal of the Transportation Research Board*, 2013, pp.75–82.

8. Brown, H., C. Sun, and P. Edara. Nuts and Bolts of Statewide HSM Calibration. *In Transportation Research Board 93rd Annual Meeting* (No. 14-2587), Washington, D.C., 2014.

9. Martinelli, F., F. La Torre and P. Vadi. “Calibration of the Highway Safety Manual's Accident Prediction Model for Italian Secondary Road Network.” *Transportation Research Record 2103: Journal of the Transportation Research Board*, 2009, pp. 1–9.

10. Williamson, M., and H. Zhou. Develop Calibration Factors for Crash Prediction Models for Rural Two-Lane Roadways in Illinois. *Procedia - Social and Behavioral Sciences*, Vol. 43, 2012, pp. 330–338.

11. Shin, H., Y. J. Lee, and S. Dadvar. The Development of Local Calibration Factors for Implementing the Highway Safety Manual in Maryland (No. MD-14-SP209B4J). Maryland State Highway Administration, Hanover, MD, 2014.

12. Banihashemi, M. “Sensitivity Analysis of Data Set Sizes for Highway Safety Manual Calibration Factors.” *Transportation Research Record 2279: Journal of the Transportation Research Board*, 2012, pp. 75–81.

13. Alluri, P., D. Saha, and A. Gan. Minimum sample sizes for estimating reliableHighway Safety Manual (HSM)calibration factors. *Journal of Transportation Safety & Security*, Vol. 8, No. 1, 2014, pp. 56–74.

14. Shirazi, M, D. Lord, and S. R. Geedipally. “Sample-Size Guidelines for Recalibrating Crash Prediction Models: Recommendations for the Highway Safety Manual.” *Accident Analysis & Prevention*, Vol. 93, 2016, pp. 160–168.

15. Bahar, G., and E. Hauer. User’s guide to develop highway safety manual safety performance function calibration factors. National Cooperative Highway Research Program, Transportation Research Board, Washington, D.C., 2014.

16. Connors, R. D., M. Maher, A. Wood, L. Mountain, and K. Ropkins. “Methodology for Fitting and Updating Predictive Accident Models with Trend.” *Accident Analysis & Prevention*, Vol. 56, 2013, pp. 82–94.

17. Wood, A. G., L. J. Mountain, R. D. Connors, M. J. Maher and K. Ropkins. “Updating Outdated Predictive Accident Models.” *Accident Analysis & Prevention*, Vol. 55, 2013, pp: 54–66.

18. Lord, D, and F. Mannering. “The Statistical Analysis of Crash-Frequency Data: A Review and Assessment of Methodological Alternatives.” *Transportation Research Part A: Policy and Practice*, vol. 44, no. 5, 2010, pp. 291–305.

19. Wood, G.R. “Confidence and prediction intervals for generalised linear accident models.” *Accident Analysis & Prevention*, Vol. 37, No. 2, 2005, pp.267-273.

20. Lord, D., and B. Persaud. "Accident Prediction Models with and Without Trend: Application of the Generalized Estimating Equations Procedure." *Transportation Research Record 1717: Journal of the Transportation Research Board*, 2000, pp. 102–108.

21. Lord, D., and B. Persaud. "Estimating the safety performance of urban road transportation networks." *Accident Analysis & Prevention*, Vol. 36, No. 4, 2004, pp. 609-620.

22. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2014. URL http://www.r-project.org/