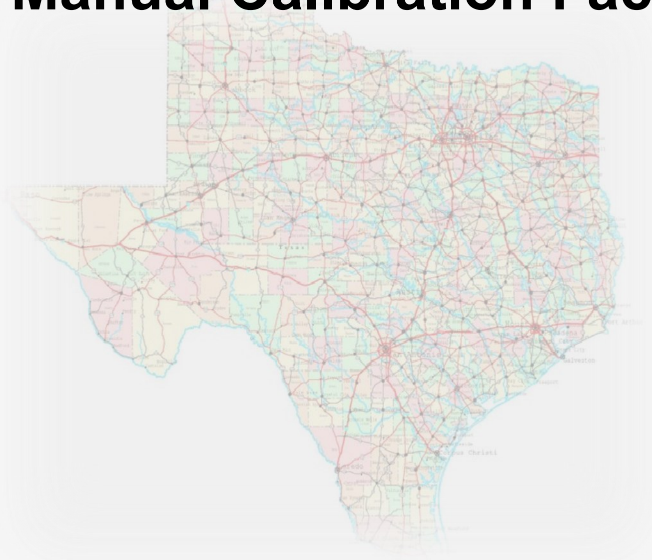
ATLAS CENTER

Advancing Transportation Leadership and Safety

RESEARCH

Improved Guidelines for Estimating the Highway Safety Manual Calibration Factors



www.ATLAS-Center.org

Improved Guidelines for Estimating the Highway Safety Manual Calibration Factors

Final Report
ATLAS-2015-10

by

Dominique Lord

Srinivas Geedipally

and

Mohammadali Shirazi

University of Michigan
2901 Baxter Rd., Room 124
Ann Arbor, MI 48109-2150

Texas A&M Transportation Institute
College Station, TX 77843-3135

January 2016

DISCLAIMER

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated under the sponsorship of the U.S. Department of Transportation's University Transportation Centers Program, in the interest of information exchange. The U.S. Government assumes no liability for the contents or use thereof.

ACKNOWLEDGMENTS

This research project was supported by the Center for Advancing Transportation Leadership and Safety (ATLAS Center). The ATLAS Center is supported by a grant from the U.S. Department of Transportation, Office of the Assistant Secretary for Research and Transportation, University Transportation Centers Program (DTRT13-G-UTC54). The ATLAS Center is a collaboration between the University of Michigan Transportation Research Institute (UMTRI) and the Texas A&M Transportation Institute (TTI).

The authors would like to thank Ms. Barbara Lorenz from TTI, Ms. Beth Jakubowski from UMTRI, and Mr. Robert C. Wunderlich, Director, Center for Transportation Safety at TTI, for their assistance for the successful completion of this research.

Technical Report Documentation Page

1. Report No. ATLAS-2015-10	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Improved Guidelines for Estimating the Highway Safety Manual Calibration Factors		5. Report Date 6. Performing Organization Code
7. Author(s) Dominique Lord; Srinivas Geedipally; Mohammadali Shirazi		8. Performing Organization Report No.
9. Performing Organization Name and Address Texas A&M Transportation Institute		10. Work Unit no. (TRAIS) 11. Contract or Grant No. DTRT13-G-UTC54
12. Sponsoring Agency Name and Address Advancing Transportation Leadership and Safety (ATLAS) Center 2901 Baxter Rd., Room 124, Ann Arbor, MI 48109-2150 U.S.A		13. Type of Report and Period Covered 14. Sponsoring Agency Code

15. Supplementary Notes

Supported by a grant from the U.S. Department of Transportation, OST-R, University Transportation Centers Program

16. Abstract

Crash prediction models can be used to predict the number of crashes and evaluate roadway safety. Part C of the first edition of the *Highway Safety Manual* (HSM) provides safety performance functions (SPFs). The HSM addendum that includes freeway and ramp chapters consist of severity distribution functions (SDFs) to estimate the crash severity as a function of geometric and traffic characteristics. In order to account for the differences in factors that were not considered or cannot be considered in the development of SPFs and SDFs, it is essential to calibrate them when they are applied to a new jurisdiction. The HSM recommends a one-size-fits-all sample size for calibration procedures that require crash data collected from randomly selected sites. However, the recommended sample size is not fully supported by documented studies, and several agencies have initiated SPF calibration efforts. In addition, there are no clear guidelines on when an agency should update their calibration factors (C-factors) and how they should make a decision on the need of region-specific calibration factors. The objectives of this research are to (1) review and document issues with the existing calibrating method in the HSM, (2) identify factors that influence the selection of the sample size for the SPFs calibration (or recalibration), (3) determine how frequently or when an agency should update their calibration factors, (4) determine whether or not having region-specific C-factors are justified and when they are needed, and (5) identify factors that influence the selection of the sample size for the SDFs calibration (or recalibration). The study objectives were accomplished using simulated and observed data. The guidelines included a discussion on (1) the sample size that is required to calibrate SPFs; (2) when the models should be recalibrated; (3) when the region-specific C-factors are recommended; and (4) the sample size that is required to calibrate SDFs.

17. Key Words Highway Safety Manual; Safety Performance Function; Severity Distribution Function; Recalibration; Calibration Factor; Sample Size		18. Distribution Statement Unlimited	
19. Security Classification (of this report) Unclassified	20. Security Classification (of this page) Unclassified	21. No. of Pages 72	22. Price

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: INTRODUCTION

PROBLEM STATEMENT

Crash prediction models can be used to predict the number of crashes and evaluate roadway safety. Part C of the first edition of the *Highway Safety Manual* (HSM) (AASHTO, 2010) provides crash prediction models, or what is often referred to as safety performance functions (SPFs), for roadway segments and intersections of four facility types: rural two-lane roads, rural multilane highways, urban and suburban arterials, and more recently freeways and interchanges. HSM prediction models were fitted and validated with data collected from a few selected numbers of states. Consequently, since crash frequency and its dispersion vary substantially from one jurisdiction to next, it is essential to calibrate SPFs when they are applied to a new jurisdiction. In other words, calibration is a tool to account for the differences in factors that were not considered or cannot be considered in the development of SPFs, such as weather, driver behavior, and reportability criteria between jurisdictions into predictive models. The characteristics of the SPF calibration procedure are presented in Appendix A of Part C of HSM. In this procedure, the calibration factor (C-factor) is eventually calculated as a ratio of the total number of observed crashes (N_{obs}) to the total number of predicted crashes (N_{pre}) (Equation 1), and is applied to the facility SPF as a scalar term:

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

The first version of the HSM recommends a one-size-fits-all sample size for calibration procedures that require crash data collected from randomly selected sites with a minimum number of approximately 100 crashes per year. However, this recommended sample size is not fully supported by documented studies and several agencies that have initiated SPF calibration efforts. Independent of the level of crash data history, the HSM still recommends using between 30 and 50 sites with at least 100 crashes. For sites with low crash history, this could be difficult to collect (Xie et al., 2011). On the other hand, for most facilities, this recommendation sounds to be too general and may not provide desirable results (Banihashemi, 2012; Alluri et al., 2016). The later issue, in addition to the fact that no documented study supported the initial HSM sample size recommendation, inspired researchers to investigate the quality of the C-factors estimated based on the HSM recommendation. Sensitivity analyses on C-factors derived from different sample sizes were documented in several studies to assess the HSM one-size-fits-all sample size recommendation. It has been reported that not only the HSM one-size-fits-all recommendation is inappropriate but is also insufficient to acquire the desirable accuracy in most cases.

Despite efforts that were put into proposing new sample size guidelines to recalibrate the SPFs, two shortcomings were identified in previous studies. First, it was assumed that the C-factor derived from the dataset in hand is the ideal (true) calibration factor. However, the ideal calibration factor is not known beforehand based on empirical data due to limitations with the data collection process. Consequently, the corresponding sample size guidelines could involve potential biases. This issue will be overcome in this study by conducting extensive simulation analyses. Using simulation, a true calibration factor can be determined, and then assess if the proposed sample size is large enough to achieve a calibration factor that is close to the true one. Second, sample size recommendations that were proposed in previous studies are based on a specific dataset usually collected at the state level. Therefore, given the fact that the characteristics of different roadways vary substantially, it is likely that these recommendations do not emerge to desirable results when applied to a new jurisdiction. In order to overcome this problem, the current study proposes its recommendations based on the crash data characteristics (i.e., the coefficient of variation [CV], which is the ratio of the standard deviation to the mean of the crashes). Therefore, agencies would be able to select a sample size that represents the characteristics of crash data for the type of facility analyzed.

Although easier than fitting a new model, calibrating crash prediction models is a time consuming task especially due to the energy or resources needed for collecting enough data (Lord and Bonneson, 2005; Xie et al., 2011; Brimley et al., 2012). Therefore, the agency may be interested to know when or how often predictive models should be recalibrated. This study addresses this issue by providing recalibration guidelines. The agency should examine the guidelines periodically. If the guidelines are met, the model should be considered for recalibration. The proposed procedure is based on the general characteristics of data at hand (i.e., (1) total number of crashes, (2) the mean value of average daily traffic [ADT] or the mean value of annual average daily traffic [AADT] in vehicles per day, and (3) the total segment length or the number of intersections).

States, or even large urban cities, may experience different numbers of crashes in different regions or parts of the city. This can be attributed to differences in terrain, population, weather, and other unobserved characteristics. Hence, it can impact the calibration procedure and consequently the C-factor when it is used for a very large area. This study first investigated whether or not having region-specific calibration factors are required and justified for large states, such as Texas. Next, region-specific guidelines are proposed to determine whether or not a region-specific C-factor is recommended for the type of facility under analysis. If the region-specific guidelines are met, the agency should derive a region-specific C-factor using region-specific data. Otherwise, the common statewide factor can be used. The proposed region-specific guidelines are similar to the one that will be proposed for when or how often recalibration of the models is recommended. The guidelines are also based on the general data at hand: (1) the total number of crashes, (2) the mean value of ADT/AADT, and (3) the total segment length (or the number of intersections).

Lastly, this research project evaluates the recent C-factor that was proposed to calibrate freeways and interchanges severity distribution functions (SDFs) in the HSM. Similar to SPFs, SDF models in HSM Chapters 18 and 19 were also fitted and validated using observed data from a few states. There is, therefore, a need to calibrate the models for local conditions. In this research, the validation of the C-factor that was proposed in the HSM is investigated using a simulation analysis. Then, different scenarios are examined to determine the calibration sample size for different conditions. The sample size guidelines are proposed based on the data that are used for calibration (i.e., the average CV of crash severities).

STUDY OBJECTIVES

The objectives of this research are to (1) review and document issues with the existing calibrating method in the HSM, (2) identify factors that influence the selection of the sample size for the SPFs calibration, (3) determine how frequently or when an agency should update their calibration factors, (4) determine whether or not having region-specific C-factors are justified and when they are needed, and (5) identify factors that influence the selection of the sample size for the SDFs calibration.

The study objectives will be accomplished using simulated and observed data. The guidelines will include a discussion on (1) the sample size that is required to calibrate SPFs; (2) when the models should be recalibrated; (3) when the region-specific C-factors are recommended; and (4) the sample size that is required to calibrate SDFs.

OUTLINE OF REPORT

The research conducted in this project is described in seven chapters. Chapter 2 provides a review of the most recent and important studies on the SPFs calibration in the literature. Chapter 3 examines the required sample size requirements for calibrating SPFs using a simulation protocol. Chapter 4 describes the guidelines for when the recalibration of predictive models is recommended. Chapter 5 documents guidelines on when region specific C-factors are needed. Chapter 6 describes the SDF calibration procedure and examines the required sample size to derive a reliable calibration factor using a simulation protocol. Chapter 7 summarizes the research conducted in this study and the proposed guidelines.

CHAPTER 2: LITERATURE REVIEW

This chapter provides a summary of the studies on the calibration of predictive models. Some of the most recent and important documents in the literature are described and reviewed in this chapter.

The chapter is divided into two sections. The first section describes the characteristics of the calibration method documented in Part C of the HSM (AASHTO, 2010). The second section provides a review of documented studies on the calibration procedure and the effects of sample size on the calibration process.

HIGHWAY SAFETY MANUAL CALIBRATION PROCEDURE

Crash prediction models are essential to predict the number of crashes and evaluate roadway safety. Part C of the first edition of the HSM provides crash prediction models or what is often referred to as SPFs for roadway segments and intersections for three facility types: rural two-lane roads, rural multilane highways, and urban and suburban arterials. HSM prediction models were fitted and validated with data collected from a few selected numbers of states. Consequently, since crash frequency varies substantially from one jurisdiction to another, it is essential to calibrate SPFs when they are applied to a new jurisdiction. In other words, calibration is a tool to account the differences in factors, such as climate, driver behavior, between jurisdictions into predictive models.

The SPF calibration procedure is presented in greater details in Appendix A of Part C of the HSM. The general steps of the procedure are as follows:

- **Step 1—Identifying the predictive model.** The SPF models are provided in Chapters 10 to 12 of Part C of the HSM. These chapters cover rural two-lane roads, rural multilane highways, and urban and suburban arterials, respectively. (Note: the same calibration procedure is used for the models documented in HSM Chapters 18 and 19.)
- **Step 2—Sampling the sites.** The HSM recommends deriving the calibration factors using a randomly selected sample that includes 30–50 sites with total of at least 100 crashes per year. For the cases where the required data are readily available for a larger number of sites, however, the larger set is recommended to be used to derive the calibration factor.
- **Step 3—Obtaining the required data.** The data collection consists of two components: (1) the total number of observed crashes obtained from randomly selected sites, and (2) the site characteristics data required to predict the number of crashes using the predictive model. The site characteristics data are classified into two groups: (1) the required data and (2) the desired data. While the required data are essential to predict the crashes, the desired data can enhance the prediction accuracy.

- **Step 4—Finding the predicted number of crashes using the predictive model.** Once the roadway characteristics data are collected and compiled, they are applied to the particular SPF to calculate the number of predicted crashes for the facility type.
- **Step 5—Calculating the calibration factor.** The calibration factor is calculated by the ratio of the number of observed crashes to the number of predicted crashes as follows:

$$C = \frac{\sum N_{obs}}{\sum N_{pre}} \quad (2)$$

where:

- N_{obs}= observed number of crashes.
- N_{pre}= predicted number of crashes.

The calibration factor is then multiplied to the facility SPF as a scalar term.

PREVIOUS STUDIES

This section describes some of the most recent studies on the calibration of predictive models. This section is divided into two subsections. First, studies that provided a review and critique of the HSM one-size-fits-all sample size recommendation are summarized. Next, several case studies are reviewed, and some of the issues and challenges researchers have encountered with the HSM calibration process are discussed. In addition, recent studies that have attempted to propose, improve, or compare alternative calibration procedures are briefly covered.

Studies on Calibration Sample Size

As a general guideline, the HSM recommends estimating the calibration factors using a randomly selected sample that includes 30 to 50 sites with a total of at least 100 crashes per year. However, the one-size-fits-all sample size recommendation should be reviewed given the fact that different roadway types have different levels of homogeneity and the minimum sample size is a function of the population homogeneity (Alluri et al., 2016). Taking this fact into account, several researchers have tried to evaluate the HSM recommendation and propose new guidelines.

Banihashemi (2012) reviewed the HSM sample size requirements by performing a sensitivity analysis on C-factors derived from samples with different sizes. The author used a dataset collected in Washington State and performed a sensitivity analysis for three types of facilities: rural two lane roads, rural multilane highways, and urban and suburban arterials. This study first found the calibration factor that is derived from the dataset in hand and referred to it as the ideal (true) calibration factor. Then, for each given sample size, 10 samples were generated randomly and their corresponding C-factors calculated. Next, assuming that the estimated measures follow the normal distribution, the quality of each sample size was quantified by measuring the probability that the calibration factor lied within 5 percent or 10 percent (depending on the desired accuracy) of the ideal calibration factor. The sample size that ensures the estimated calibration factor lies within 10 percent of the ideal calibration factor with a reasonable probability was recommended in new guidelines. The study showed that the HSM 30–50 sites criterion is too small to derive a reliable C-factor for most roadway types.

Alluri et al. (2016) used data collected in Florida to determine the minimum sample size that results in a reliable calibration factor for the same three types of facilities described above. A similar procedure as the one proposed by Banhashemi (2012) was used here as well to assess the accuracy of the C-factors and estimate the minimum sample size. In this study, for each given sample size, 30 subsets of data were generated and the corresponding C-factors calculated. Assuming a normal distribution, the minimum sample size is selected when, with a high probability, the estimated C-factors lie within 10 percent of the ideal C-factor. The analysis showed that not only the HSM generalized one-size-fits-all sample size is not appropriate, but also this criterion is insufficient to acquire the desired accuracy. The recommendations provided in the paper are based on the criterion that, with a high probability, the calibration factors lie within 10 percent of the ideal factor. However, for cases where sufficient data are available and a higher accuracy is sought, the recommendations based on 5 percent of the ideal factor were provided as well. The recommended minimum sample size for reaching the 5 percent accuracy almost doubles compared to the recommendations of achieving the 10 percent accuracy.

Trieu et al. (2014) performed a sensitivity analysis on the calibration sample size requirement to evaluate and critique the accuracy of the HSM sample size guideline for two-lane two-way undivided urban arterial roadways. Given different percentages of a complete dataset, the samples were generated from the complete dataset for 500 iterations (Note: albeit this paper referred its method as a Monte Carlo simulation, it seems that the samples were obtained directly from the original dataset. Monte Carlo simulation, however, is referred to as parametric sampling methods which samples are generated from parametric distributions). Then, C-factors for each size-group were classified based on their errors from the ideal C-factor in 5 percent increments. As the sample size increased, C-factor observations with high error range decreased. For samples generated from 50 percent (or more) of the complete dataset, all C-factors lied within 10 percent of the ideal C-factor. The paper concluded that the current HSM sample size criterion may not yield a reliable C-factor. The authors analyzed the AADT distribution for a group of C-factors that were generated with a sample size of 37 sites (the sample size that satisfies the HSM criterion). The results showed that the AADT distribution could influence the C-factor reliability.

Bahar (2014) recently introduced a procedure to evaluate the required sample size based on data at hand and the desired accuracy for the C-factor variance (note: the work done in this research was initially performed in parallel and independently from the work of Bahar). The procedure is anticipated to be incorporated into the second edition of the HSM. The proposed procedure is based on Dr. Ezra Hauer’s work that was documented in the appendices of the report. First, it was assumed that the source of the dispersion for the C-factor comes only from the crash data.

Based on this assumption, the variance of the C-factor was derived and presented in different alternative equations. Next, five different approaches were recommended to assess the required sample size. In fact, these alternative approaches are only different based on the accuracy they provide or the information that is required to calculate the C-factor variance. Alternative 1 assumes no over-dispersion in the data, and can be used only as an initial estimate for the sample size. This alternative requires a decent guess about the C-factor. Alternative 5 provides a rough approximation about the sample size based on the average AADT, average segment length, a decent guess about the C-factor and the desired accuracy for the C-factor variance. Alternatives 2, 3, and 4, on the other hand, can be used when a better estimate for the required sample size is sought. Alternative 2 can be used when the observed crash data, AADT, and the segment length are available or known. On the other hand, alternatives 3 and 4 can be used once either the AADT or crash data are missing but instead the analyst can have a good guess about the C-factor. To use alternatives 2, 3, or 4, the safety analyst needs to begin with a sample size (for example 50 sites), then calculate the C-factor and its variance. If the desired variance for the C-factor is not achieved, the analyst should increase the sample size and collect more data. This trial and error procedure continues until the desired accuracy is fulfilled.

Bahar (2014) method can be reviewed as follows:

- First, to calculate the variance of the C-factor, it was assumed that the over-dispersion parameter is fixed, based on the models documented in the HSM, and does not change from one jurisdiction to another; however, this may not be the case as the characteristics of the crash data vary significantly between different jurisdictions, including the level of the dispersion (Lord and Bonneson, 2005). This can lead to a significant bias.
- Second, the method requires the analyst to make an assumption about the desired C-factor variance. The author suggests using $0.1 \times C$ as the desired variance for the C-factor. However, this recommendation can be problematic since the desired accuracy itself can change from one round of trial and error to the next. In other words, the C-factor will change as more sites are added; hence, the estimated variance will change too.
- Third, the analyst needs at least three out of the four sources of information (i.e., AADT, crash data, segment length, or a decent guess about the C-factor) to use the methodology. In our viewpoint, however, once it is assumed the crash data are the only source of the C-factor dispersion (as was the case in Hauer's derivation of the equations), crash data are the only information that is needed to estimate the required sample size. The issue is explained in greater details in Chapter 3.
- Fourth, there are no clear guidelines on how many more sites the analyst needs to collect in the next round of trial and error if the desired accuracy for the variance was not achieved in the current trial.
- Fifth, the procedure described in the report can be a tedious and time consuming task, since several of the alternatives require a trial and error approach and the next trial depends on calibration results itself.

The proposed method in this current study overcomes the issues described above. First, we account for the effect of the over-dispersion parameter that is based on the data to be used for the calibration process. Second, the desired accuracy and data collection process are not related to calibration results itself. Third, the only information required to collect data can be derived from crash data (even for segments, the segment length is not needed). Fourth, we provide clear guidelines on how many sites the analyst needs to collect in each scenario for the calibration process. Fifth, the procedure is straightforward. Once the analyst derived the information from the crash data, he or she can use the guidelines to select the required sample size.

Studies on Calibration Procedure

Upon the release of the HSM, several states have tried to develop state-specific calibration factors. Oregon (Xie et al., 2011) was one of the pioneering states that developed state-specific calibration factors. In recent years, calibration factors were generated for other states, such as Utah (Brimley et al., 2012), Illinois (Williamson and Zhou, 2012), Alabama (Mehta and Lou, 2013), Missouri (Brown et al., 2014), and Maryland (Shin et al., 2014). This section briefly reviews some of the documents that use the HSM calibration method to calibrate SPFs to local conditions and covers some of the issues and challenges researchers have encountered. In addition, some of the studies that tried to improve, propose, or compare alternative calibration procedures are addressed.

Although the methodology described in HSM is straightforward, Xie et al. (2011) indicated several issues and limitations in calibrating the SPFs for Oregon roadway facilities. The researchers noted that the methodology necessitates detailed data and is a time-consuming task for collecting and compiling all the necessary data. In particular, they reported that pedestrian volume at urban intersections and the traffic volume (vehicles/day) of minor roads at rural locations were the most difficult data to collect. Moreover, meeting the HSM sample size guideline was not applicable for some roadway types with low level of historical crash data. In this case, the authors decided to use all possible and available data instead. Due to a small crash rate in Oregon, the calibration factor for most facilities in this state was less than one.

Brimley et al. (2012) used three years of crash data that occurred on 157 rural two-lane two-way roadway segments in Utah to calibrate the HSM SPF for this facility type. The calibration factor was estimated to be 1.16, which indicated that the HSM SPF underestimates the number of crashes. Then, the researchers tried to develop a jurisdiction-specific model. For this purpose, a negative binomial regression model was applied to the same dataset, but considered additional variables that were assumed to be associated with crashes. Four jurisdiction-specific models, two conventional and two models with traffic flow to a power (the natural log of the AADT) were developed and compared based on the Bayesian information criterion (BIC). The model with traffic flow to a power had the lowest BIC and was selected as the preferred model among the four proposed models. The modeling results showed that AADT, segment length, speed limit, and the percentage of multiunit trucks were seen to be significantly associated with the number of crashes.

Mehta and Lou (2013) used the HSM recommended methodology to calibrate the SPF for two facility types: two-lane two-way rural roads and 4-lane divided highways in Alabama. Then, they proposed a new approach to find or calculate the calibration factor. This approach treated the calibration factor as a constant in the negative binomial regression model. Next, for each facility type, four state-specific models were developed with the negative binomial regression error structure. The authors used several performance measures and a validation dataset to compare different approaches. While the new calibration method did not emerge as a better method compared to the one recommended by HSM, one of the state-specific models outperformed all approaches including the HSM calibration procedure.

Brown et al. (2014) documented the calibration of the HSM SPFs for Missouri and addressed some practical solutions to some challenges encountered in the process. Addressing the HSM sample size guidelines and data requirement, balancing the minimum length and homogeneity of segments, and inconsistency with the crash data were some of the challenges the researchers encountered.

Martinelli et al. (2009) analyzed the transferability of the HSM calibration procedure to data collected outside North America. They calibrated the HSM SPF for the rural two-lane highways in the Italian province of Arezzo, a region with different roadway characteristics than those built in the US. Four different models and three calibration strategies were assumed to develop 12 district C-factors. The applied calibration strategies were as follows:

- The ratio of the number of observed crashes to the number of predicted crashes (the HSM calibration procedure).
- The ratio of the densities of the observed crashes to predicted crashes.
- The ratio of the weighted average over the length of the observed crashes to the predicted crashes.

The SPF model that incorporated Crash Modification Factors (CMFs) and was applied to the stratified classes defined by the HSM procedure and calibrated by the weighted average over the length C-factor provided the best results.

Since roadway and vehicle characteristics, and the driver behavior continuously change over time, crash prediction models can quickly become outdated. Because fitting a new model requires significant data and is a time-consuming and expensive task, it is essential to find an efficient approach for updating outdated models. Similar to calibrating predictive models to local conditions, calibration can be used to update the predictive models as well. Connors et al. (2013) documented several methodological issues that arise from updating predictive models initially developed in England through both scalar calibration and the re-fitting of models. One issue that was documented was related to selecting the scalar factor based on goodness-of-fit (GOF) criteria. The researchers noted that the selection of scalar factors was dependent on the GOF criteria. In the end, they nonetheless suggested to use the HSM recalibration procedure.

Wood et al. (2013) analyzed two issues regarding the updating of the same crash prediction models addressed in Connors et al. (2013) study. First, they looked at the temporal transferability of the model as a function of its complexity. The researchers concluded that the more complex the model is, the better its temporal transferability. Then, the authors investigated two general approaches in updating the predictive models: (1) refitting the old model considering the same variables but with new data sources, and (2) calibration through a scalar factor. Both methods are more practical and more efficient compared to fitting a new crash prediction model and both emerged to desired results in their study. Moreover, they analyzed the original model, which had a term for capturing time trend. The authors stated that since the pattern may not remain stable over time, the model with a trend term can lead to a significant bias in estimations. Therefore, simpler calibration procedures, such as refitting or scalar calibration, were more reliable.

The C-factor might be different within a large region because attributes within that region are not uniform across the entire area. For example, the HSM recommends finding separate C-factors for large jurisdictions that are characterized by different topographical or weather conditions (AASHTO, 2010). Unfortunately, the HSM does not provide guidelines for determining the detailed conditions when separate C-factors are warranted or justified. Bahar (2014) studied this issue and suggested two approaches:

- First, Bahar studied how much bias can affect the prediction results of the C-factor. The hypothesis is that there is no need to be more precise with the C-factor than for the base model or the product of CMFs. Based on this hypothesis, a conservative guideline was provided, "…the coefficient variation of the C-factor does not need to be less than, say, half of the coefficient of variation of the product of the CMFs." However, the document does not provide clear guidelines on what the typical CV of the product of the CMFs should be. Hence, calculation of the CV of the product of CMFs was left to the user.
- Second, it was suggested to group data based on different variables and conditions, such as AADT, segment length or crash severities. If a major difference in C-factors was observed, a separate C-factor is suggested. This method is not necessarily for a region or terrain but can be used to consider the effect of the different variables more accurately. For example, different C-factors can be recommended for different ranges of AADTs.

The method proposed by Bahar (2014) is based on the availability of detailed data that are used for the calibration process. More specifically, it is assumed that the analyst first collects all data that are required for the calibration, and then the analyst goes through grouping the variables to determine whether or not a separate C-factor is needed, for example, for different AADT ranges or a region. This method may not be efficient since the analyst may need to know if a separate C-factor is desired (specifically for a region) in advance before the calibration procedure begins in order to collect enough data for the required sample size. In this research, we address this issue in Chapter 5.

Several studies were also conducted to improve the accuracy of SPFs. Kim and Lee (2013) proposed an iterative four steps procedure to develop SPFs that reflect the categorical impact of exposure variables varying with freeway segments. First, freeway segments were classified into three similar groups where in each group the dispersion of exposure variables is minimized. In the second step, several distributions (Poisson and negative binomial, geometric, and discrete uniform) were assumed and tested using the Kolmogorov-Smirnov GOF test. All categories showed a good fit with the negative binomial distribution. In the third step, several SPF models were estimated using the negative binomial regression model. The model with log transformation of AADT and segment length provided the best results (although the later may not be theoretically sound, see Lord et al. [2005]). In step four, the validity of differences among the clustered groups was tested. This four-step procedure produced more accurate results.

In summary, several studies have noted that the calibration of predictive models is a time-consuming task in addition to problems associated with the collection, readiness, and completeness of the data. Moreover, independent of the level of crash data history for different types of facilities, the HSM still recommends using between 30 and 50 sites with at least 100 crashes. The small sample size proposed by HSM inspired researchers to investigate the quality of C-factors. Sensitivity analyses on C-factors derived from different sample sizes were conducted by several researchers to assess the HSM one-size-fits-all sample size recommendation. Not only is the HSM one-size-fits-all recommendation inappropriate, but it is also insufficient to acquire the desirable accuracy in most cases. In these studies, it was assumed that the C-factor that is derived from the full dataset is the true C-factor. Then, the quality of each given sample size was quantified by comparing the C-factor obtained from that sample size with the true C-factor. For instance, if the C-factor lied within 10 percent of the true C-factor with a high probability, the sample size would be classified as a reliable option.

CHAPTER 3: SAFETY PERFORMANCE FUNCTION SAMPLE SIZE REQUIREMENTS

This chapter documents the analyses used for determining the required sample size needed to calibrate predictive models. A simulation protocol is proposed to simulate a wide range of scenarios and evaluate the required sample size for each case. Then, the sample size guidelines are provided and validated. The guidelines are based on the characteristics of crash data used for calibration.

This chapter is divided into four sections. The first section describes the simulation protocol used in this research. In the second section, the simulation results are presented and discussed for a range of scenarios. Next, in the third section, the sample size guidelines are provided and discussed. The fourth section presents the results related to the evaluation and validation of the sample size guidelines using two observed datasets.

SIMULATION PROTOCOL

This section presents the Monte Carlo simulation protocol that was used in the research. Before describing the simulation protocol, recall that the calibration factor (C-factor) for each facility can be estimated as follows:

$$C = \frac{\sum N_{obs}}{\sum N_{pre}} \tag{3}$$

where:

- C= calibration factor.
- N_{obs}= the observed number of crashes.
- N_{pre}= the predicted number of crashes.

A simulation scenario, in this study, was specified by the specified mean for the predicted number of crashes, a calibration factor, and an inverse dispersion parameter (a measure of dispersion). The first step of the simulation began with generating the AADT variable from a lognormal distribution. To simplify the simulation process, a flow-only crash prediction model was selected and its intercept was modified until the specified mean of the predicted number of crashes in the selected scenario was achieved. Next, given the modified model, for each site, the predicted number of crashes was calculated. The observed number of crashes was then generated from a negative binomial distribution with the identified mean and inverse dispersion parameter. Last, for each given sample size (n), n sites were randomly selected and the sample's calibration factor (C_n) was calculated. This step was repeated for 1,000 iterations. The quality of the each given sample size was quantified with a method similar to the one that was proposed by Banihashemi (2012), which is described as follows:

1. Calculate the mean ($\mathrm{Avg}(C_n)$) and standard deviation ($\mathrm{sd}(C_n)$) of the generated calibration factors.
2. Assume the calibration factors, which are generated from 1,000 simulated iterations, follow a normal distribution and calculate following two statistics:

$$Z_{min} = \frac{0.9 \times C_N - \mathrm{Avg}(C_n)}{\mathrm{sd}(C_n)} \quad (4)$$

$$Z_{max} = \frac{1.1 \times C_N - \mathrm{Avg}(C_n)}{\mathrm{sd}(C_n)} \quad (5)$$

   where C_N is the calibration factor that is derived from the simulated dataset (Note: the size of the dataset was set to 5,000 observations or sites) and is referred to as the true calibration factor.

3. Find the probability that the calibration factor lies within 10 percent of the true calibration factor as follows:

$$P = \Phi(Z_{max}) - \Phi(Z_{min}) \quad (6)$$

   where Φ(.) indicates the cumulative density function (CDF) of the normal distribution, and the probability P indicates the probability that the sample size calibration factor lies within 10 percent of the true calibration factor.

The simulation procedure is summarized in the following steps.

Step 1 – Initialization

1.1 Set the scenario by specifying the desired mean for the predicted number of crashes (μ_{pre}), a calibration factor (C) and an inverse dispersion parameter (φ).

1.2 Set the size of the simulation dataset (N).

1.3 For each site, generate the AADT variable from a log-normal distribution with a given mean and standard deviation.

Step 2 – Simulating the Dataset

2.1 Take a crash prediction model and modify its intercept in a way that the mean of the predicted number of crashes is matched to the scenario.

2.2 Generate the crash prediction mean (N_{pre}) at each site using the modified functional form.

2.3 Generate the observed number of crashes (N_{obs}) in each site from a negative binomial distribution with a mean equal to $N_{Pre} \times C$ and the given inverse dispersion parameter (φ).

Find the calibration factor for the simulated dataset as follows and referred to it as the true calibration factor:

$$C_N = \frac{\sum_N N_{obs}}{\sum_N N_{pre}} \quad (7)$$

Note that C_N shall be close to the assumed calibration factor (C).

Step 3 – Test the Quality of Each Given Sample Size

3.1 Repeat the following steps for 1,000 iterations:

3.1.1 For a given sample size (n), randomly select (n) sites.

3.1.2 Calculate the sample's calibration factor as follows (Eq. 7):

$$C_n = \frac{\sum_n N_{obs}}{\sum_n N_{pre}}$$

3.2 Measure the quality of each given sample size using Equations 4 to 6.

SIMULATION RESULTS

For all simulation runs, the size of the simulation dataset was set to 5,000 observations. The simulation scenarios were generated considering a range for the mean of the predicted number of crashes, calibration factors, and inverse dispersion parameters. The range of these factors varied as follows:

Predicted mean (μ_{pre}) = {0.5, 2.5, 5}

Inverse dispersion parameter (φ) = {0.5, 1, 5}

Calibration factor (C) = {0.5, 1.0, 1.5, 2.0}

The inverse dispersion parameters of 0.5, 1, and 5, respectively, represent a high, medium, and small dispersion values. The calibration factors that were evaluated include C < 1, C = 1, and C > 1. In order to test the quality of each sample size (n), the sample size range varied from 50 to 500 in 25 increments, from 500 to 1,000 in 50 increments, and from 1,000 to 1,500 in 100 increments.

The simulation process was performed for a segment-type crash prediction model that was fitted and validated for six-lane divided rural roadways using Texas and California data. The data were collected for an on-going national research project. The model is shown as follows:

$$N_{pre} = b_0 \times L \times e^{1.2359 \times \ln(\text{AADT})} \quad (8)$$

The variable L indicates the segment length and is set to 0.2 mile for all sites in the simulation runs. The AADT variable was simulated from a lognormal distribution with a mean and standard deviation of 38,329 veh/day and 15,510 veh/day, respectively, based on the characteristics of the data that were used to develop the model. For each simulation run, the intercept (b_0) was manipulated until the mean of the predicted number of crashes identified by the specified scenario is achieved.

For each scenario, a dataset with a size of 5,000 sites was simulated. Next, for each given sample size, 1,000 C-factors were randomly generated based on the simulation protocol. The quality of the given sample size is quantified using Equations 4 to 6. As an example, Table 1 shows the simulation results for a scenario in which the predicted mean, calibration factor, and inverse dispersion parameter are equal to 2.5, 1.5, and 1, respectively. For each given sample size, Table 1 shows the probability that the calibration factor lies within 10 percent of the true calibration factor. As expected, the confidence level is increased as the sample size increases.

Table 1. Simulation Results for a Scenario with a Predicted Mean of 2.5, Calibration Factor of 1.5, and Inverse Dispersion Parameter of 1.

Sample Size	Confidence Probability	Sample Size	Confidence Probability
50	0.4433	**475**	0.9454
75	0.5184	**500**	0.9476
100	0.5931	**550**	0.9535
125	0.6353	**600**	0.9634
150	0.6987	**650**	0.9761
175	0.7105	**700**	0.9755
200	0.7403	**750**	0.9847
225	0.8064	**800**	0.9886
250	0.8232	**850**	0.9916
275	0.8501	**900**	0.9917
300	0.8568	**950**	0.9958
325	0.8710	**1,000**	0.9956
350	0.8874	**1,100**	0.9979
375	0.9022	**1,200**	0.9991
400	0.9043	**1,300**	0.9993
425	0.9162	**1,400**	0.9997
450	0.9422	**1,500**	0.9997

For a range of predicted means, calibration factors, and inverse dispersion parameters, Table 2 shows the recommended sample size to fulfill a 90 percent, 80 percent, and 70 percent confidence level where the calibration factor lies within 10 percent of the actual calibration factor. Note that, for instance, with a 80 percent confidence interval, there is a 20 percent chance that the calibration factor does not lie within 10 percent of the true calibration factor. As indicated in this table, generally, as the observed mean of crashes increases, the required sample size to attain a certain level of confidence (such as 90 percent confidence) is decreased. Likewise, as the inverse dispersion parameter increases (which represents smaller dispersion or variation in the dataset), a smaller sample size is needed to reach a certain level of confidence. Furthermore, as indicated by the simulation results shown in Table 2, not only the HSM one-size-fits-all recommended sample size of 30–50 sites is not appropriate, but is also insufficient to attain a 90 percent or 80 percent of confidence in all surveyed scenarios. The simulation protocol was also applied to an intersection model and similar sample size requirements were required for comparable scenarios. Appendix A shows the results for the intersection model.

Table 2. Sample Size Requirements to Fulfill 90 Percent, 80 Percent, and 70 Percent Confidence Level.

Predicted Crash Mean	Calibration Factor	Observed Crash Mean*	Inverse Dispersion Parameter		
			0.5	1	5
0.5	0.5	0.25	1,300 (90%)** 900 (80%) 650 (70%)	1,100 (90%) 700 (80%) 500 (70%)	950 (90%) 650 (80%) 400 (70%)
	1.0	0.50	950 (90%) 650 (80%) 450 (70%)	750 (90%) 475 (80%) 325 (70%)	550 (90%) 375 (80%) 250 (70%)
	1.5	0.75	900 (90%) 600 (80%) 400 (70%)	650 (90%) 400 (80%) 250 (70%)	375 (90%) 250 (80%) 175 (70%)
	2.0	1.00	800 (90%) 500 (80%) 350 (70%)	550 (90%) 325 (80%) 225 (70%)	325 (90%) 200 (80%) 150 (70%)
2.5	0.5	1.25	750 (90%) 475 (80%) 350 (70%)	550 (90%) 300 (80%) 225 (70%)	300 (90%) 175 (80%) 125 (70%)
	1.0	2.50	750 (90%) 475 (80%) 300 (70%)	450 (90%) 300 (80%) 200 (70%)	175 (90%) 125 (80%) 75 (70%)
	1.5	3.75	650 (90%) 400 (80%) 300 (70%)	375 (90%) 225 (80%) 175 (70%)	150 (90%) 100 (80%) 75 (70%)
	2.0	5.00	650 (90%) 400 (80%) 300 (70%)	350 (90%) 225 (80%) 175 (70%)	125 (90%) 75 (80%) 50 (70%)
5.0	0.5	2.50	750 (90%) 475 (80%) 300 (70%)	450 (90%) 300 (80%) 200 (70%)	175 (90%) 125 (80%) 75 (70%)
	1.0	5.00	650 (90%) 425 (80%) 300 (70%)	350 (90%) 225 (80%) 175 (70%)	125 (90%) 75 (80%) 50 (70%)
	1.5	7.50	650 (90%) 400 (80%) 250 (70%)	350 (90%) 225 (80%) 150 (70%)	125 (90%) 75 (80%) 50 (70%)
	2.0	10.00	650 (90%) 400 (80%) 300 (70%)	350 (90%) 225 (80%) 150 (70%)	100 (90%) 75 (80%) 50 (70%)

*The observed crash mean might be slightly different for different runs of simulations due to randomness. However, this table shows only the rounded values.
**The numbers in parenthesis show the confidence level.

In order to further investigate the reason behind the two characteristics described in the previous paragraph (i.e., the required sample size to fulfill a certain confidence level is decreased by increasing the mean of the observed number of crashes and increasing the inverse dispersion parameter), it is advised to approximate Equation 3 with Equation 9 as a ratio of the mean of the observed number of crashes ($\overline{\mathrm{N}}_{\mathrm{obs}}$) to the mean of the predicted number of crashes ($\overline{\mathrm{N}}_{\mathrm{pre}}$):

$$C_N = \frac{\bar{N}_{obs}}{\bar{N}_{pre}} \quad (9)$$

A reliable sample size can be approximated with a sample that can sufficiently estimate the mean of the observed number of crashes. (Note that in this analysis, the effect of the variation of $\bar{N}_{pre}$ is considered to be negligible compared to $\bar{N}_{obs}$.) As the deviation of data around its mean decreases, smaller sample size is required to estimate its mean.

Once the inverse dispersion parameter increases, data will be less dispersed. Therefore, the corresponding observed mean of crash data can adequately be estimated with a smaller sample size. Likewise, once the observed mean of crash data increases, even though the standard deviation of dataset increases, the deviation of data around the mean will be reduced. Therefore, the observed mean can be estimated with a smaller sample size. The theoretical derivation of these factors, which supports the simulation study and how they affect the sample size calculations, can be found in Appendix B.

Taking these observations into account, the simulation results were sorted based on the ratio of the standard deviation to the mean of the observed number of crashes that is referred to as the CV. CV of the observed crash data is defined as follows:

$$CV = \frac{sd(N_{obs})}{\bar{N}_{obs}} \quad (10)$$

where $\bar{N}_{obs}$ and $sd(N_{obs})$, respectively, denote the mean and the standard deviation of the observed number of crashes. Table 3 shows the relationship of CV of the crash data versus the sample size that is required to fulfill certain levels of accuracies. As indicated in this table, the required sample size to attain certain levels of confidences (such as 90 percent) increases as the crash CV increases. For scenarios in which the CVs are approximately the same, the simulation results show that almost the same sample size is needed to fulfill certain levels of accuracies. For instance, Table 3 shows that a scenario with an observed mean of 10.36 and standard deviation of 17.83 has the same CV as the one with an observed mean of 3.91 and standard deviation of 6.72. For both scenarios, the CV is equal to 1.72, so a same sample size is required.

Table 3. CV of Crash Data vs. the Calibration Sample Size.

Crash CV.	Crash Mean	Crash Sd.	Sample Size*		
			90% Confidence	80% Confidence	70% Confidence
2.62	0.26	0.67	1,300 (0.904)	900 (0.810)	650 (0.714)
2.27	0.25	0.57	1,100 (0.918)	700 (0.811)	500 (0.702)
2.16	0.51	1.10	950 (0.902)	650 (0.811)	450 (0.719)
2.09	0.26	0.53	950 (0.919)	650 (0.805)	400 (0.703)
2.04	0.77	1.58	900 (0.907)	600 (0.807)	400 (0.710)
1.89	1.01	1.91	800 (0.900)	500 (0.808)	350 (0.723)
1.88	1.28	2.41	750 (0.907)	475 (0.818)	350 (0.734)
1.84	0.49	0.91	750 (0.900)	475 (0.803)	325 (0.704)
1.80	2.63	4.73	750 (0.921)	475 (0.817)	300 (0.712)
1.72	10.36	17.83	650 (0.913)	400 (0.800)	300 (0.730)
1.72	3.91	6.72	650 (0.903)	400 (0.807)	300 (0.700)
1.70	5.15	8.75	650 (0.902)	400 (0.800)	300 (0.716)
1.68	7.35	12.34	650 (0.911)	400 (0.823)	250 (0.703)
1.66	0.75	1.24	650 (0.915)	400 (0.829)	250 (0.704)
1.59	0.49	0.77	550 (0.905)	375 (0.834)	250 (0.716)
1.56	1.00	1.55	550 (0.900)	325 (0.801)	225 (0.714)
1.55	1.24	1.93	550 (0.924)	300 (0.808)	225 (0.709)
1.41	2.52	3.55	450 (0.915)	300 (0.831)	200 (0.737)
1.36	0.74	1.00	375 (0.907)	250 (0.803)	175 (0.711)
1.32	3.73	4.93	375 (0.902)	225 (0.806)	175 (0.710)
1.30	4.99	6.47	350 (0.909)	225 (0.800)	175 (0.735)
1.27	10.21	13.00	350 (0.906)	225 (0.823)	150 (0.727)
1.26	7.59	9.60	350 (0.906)	225 (0.800)	150 (0.713)
1.21	0.99	1.20	325 (0.914)	200 (0.809)	150 (0.740)
1.13	1.25	1.42	300 (0.928)	175 (0.811)	125 (0.760)
0.96	2.51	2.42	175 (0.917)	125 (0.827)	75 (0.732)
0.90	3.77	3.38	150 (0.913)	100 (0.842)	75 (0.759)
0.86	5.01	4.29	125 (0.909)	75 (0.810)	50 (0.702)
0.81	7.54	6.09	125 (0.928)	75 (0.853)	50 (0.759)
0.80	10.09	8.03	100 (0.910)	75 (0.848)	50 (0.744)

*The numbers in parenthesis show the confidence probability with each sample size.

Our methodology can be compared with Bahar (2014) by looking at the Equation B.7 in Appendix B of the Bahar (2014) manuscript to derive the variance of the C-factor. It is described as follows:

$$Var(\hat{C}) = \frac{C^2}{\sum N_{obs}} + \frac{C^2 \sum K N_{obs}^2}{(\sum N_{obs})^2} \quad (11)$$

where K denote the over-dispersion parameter (i.e., 1 /φ).

If we divide both sides of the above equation by C^2, then we have:

$$CV^2(\hat{C}) = \frac{Var(\hat{C})}{C^2} = \frac{1}{\sum N_{obs}} + \frac{\sum KN_{obs}^2}{(\sum N_{obs})^2} \tag{12}$$

Now, instead of calculating the variance of the C-factor, the goal can be focused on minimizing the variation of the estimated C-factor around its true value (the left hand side of the equation) to the sought level of accuracy. This is similar to the goal that we seek in step 3 of our simulation protocol. In addition, all information we need to minimize the CV of the C-factor to the required accuracy level is on the right hand side of the equation and this only depends on crash data. As stated in the background section (the third issue with Bahar [2014] methodology), once the crash data were assumed to be the only source of the C-factor dispersion or variation, the only information needed to select a reliable sample size can be derived from the crash dataset itself.

SAMPLE SIZE GUIDELINES

Based on simulation results, Table 4 shows the sample size guidelines to fulfill a confidence level of 90 percent, 80 percent, and 70 percent for a range of CVs. The guidelines can be used for all types of facilities and for both segment and intersection models. In order to use the sample size guidelines, the agency is required to secure the facility crash mean and standard deviation to calculate the CV of the crash data. Then, given the CV, a sample size that fulfills the desired level of accuracy can be selected from the table. The sample size guidelines show the minimum sample size to meet a given level of accuracy. In cases when more data are readily available, the agency is advised to use the full dataset. On the other hand, for cases when the agency cannot meet the minimum sample size guidelines, the agency is advised to develop a state-specific crash prediction model.

Table 4. SPF Sample Size Guidelines.

CV.	Confidence Level		
	90%	**80%**	**70%**
3.0†	1,500	1,100	700
2.8	1,400	1,000	650
2.6	1,300	900	600
2.4	1,200	800	550
2.2	1,000	650	450
2.0	900	550	400
1.8	750	450	300
1.6	600	350	250
1.4	450	300	200
1.2	300	200	150
1.0	200	125	75
0.8	100	75	50
≤ 0.6	50	30	30

†Values larger than 3.0 are would require even larger sample sizes, which may not be practical to collect, so they are not presented in this table. Agencies may need to revise the datasets for recalibrating the models.

VALIDATION OF THE SAMPLE SIZE GUIDELINES WITH OBSERVED DATA

In this section, the sample size guidelines are validated using two observed datasets, each characterized by a different CV. For the purpose of the analysis and since the dataset is large enough, the full dataset calibration factor is considered to be the ideal (true) calibration factor (C_N). Then, the recommended sample size is validated by generating 1,000 C-factors and measuring the confidence level using Equations 4 to 6.

The sample size guidelines are first validated for the 4-lane divided urban arterial facility with multivehicle non-driveway crashes collected in Texas from 2012 to 2014 at 4,265 locations. The facility SPF is provided in Chapter 12 of the HSM. Except for the median width, all sites in the dataset meet the base conditions. For the median width, the CMF was calculated and applied to the model before calculating the calibration factor. The CV of crash data is 2.86 (with a mean and standard deviation of 2.36 and 6.75, respectively). For this case, the guidelines approximately recommend collecting 1,400, 1,000, and 650 sites to achieve a confidence level of 90 percent, 80 percent, and 70 percent, respectively. The results show that the sample size of 1,400, 750, and 650 can attain a 94 percent, 88 percent, and 76 percent of accuracy, respectively, which even though a little bit conservative, fulfill the desired levels of accuracies. Figure 1 shows a normal distribution given by generated C-factors (1,000 observations) based on a sample size of 1,400 (solid line), 1,000 (dashed line), and 650 (dotted line). The 10 percent error interval around the true calibration is marked with dotted lines. As shown in this figure, the normal distribution becomes more concentrated around the true calibration factor as the sample size increases and can better estimate the true C-factor.

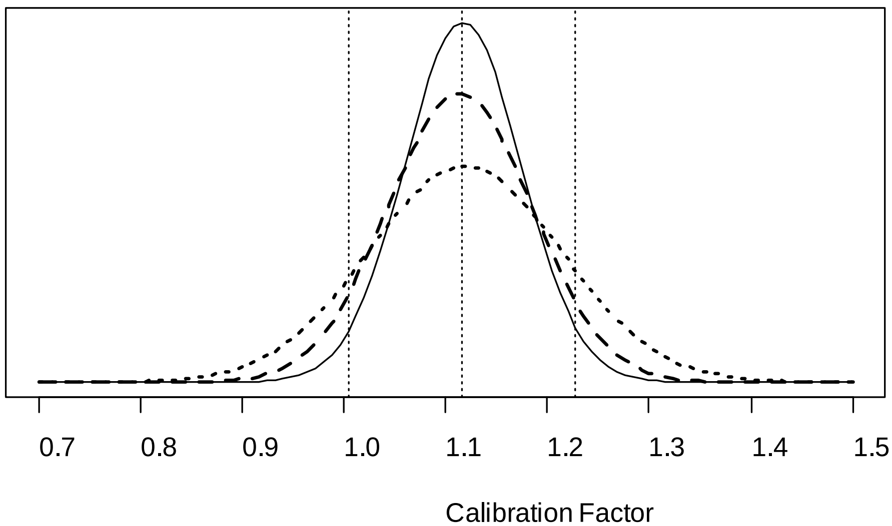


Figure 1. Accuracy of the Estimation of Calibration Factor by Sample Size of 1,400 (Solid Line), 1,000 (Dashed Line), and 650 (Dotted Line) for Texas 4-Lane Divided Urban Arterial Facility.

For this facility type, based on Florida data, Alluri et al. (2016) recommended collecting 500 sites to fulfill a 90 percent confidence that the calibration factor lies within 10 percent of the true factor. However, for the Texas dataset, which is characterized by a different crash frequency and dispersion than the one in Florida, the sample size of 500 only satisfies 68 percent of the confidence level. As previously stated, guidelines that are prepared with data from a specific state can be criticized for two reasons. First, while preparing the guidelines, it is assumed that the C-factor derived from the available dataset is the ideal (true) calibration factor. However, the true calibration factor is not known beforehand based on empirical data due to limitations with the data collection process. Second, since the characteristics of different roadways vary substantially, it is likely that these recommendations do not emerge to desirable results when applied to a new jurisdiction.

The simulation results are also verified with the crash data collected in 1995 from 868 4-legged signalized intersections located in Toronto, Ontario; the dataset has been used extensively by others (Miaou and Lord, 2003; Lord et al., 2008; Miranda-Moreno and Fu, 2007). The CV of crash data for this dataset is 0.87 (with a crash mean and standard deviation of 11.56 and 10.01, respectively). For this case, the guidelines approximately recommend collecting 100, 75, and 50 sites to fulfill a confidence level of 90 percent, 80 percent, and 70 percent, respectively. The results show that the sample size of 100, 75, and 50, respectively, can attain a 91 percent, 85 percent, and 74 percent levels of confidence, which are comparable to each given desired level of accuracy. Similar to the previous figure, Figure 2 shows a normal distribution given by generated C-factors (1,000 observations) based on a sample size of 100 (solid line), 75 (dashed line), and 50 (dotted line). As the sample size increases, the normal distribution becomes more concentrated around the true calibration factor and can have a better estimate.

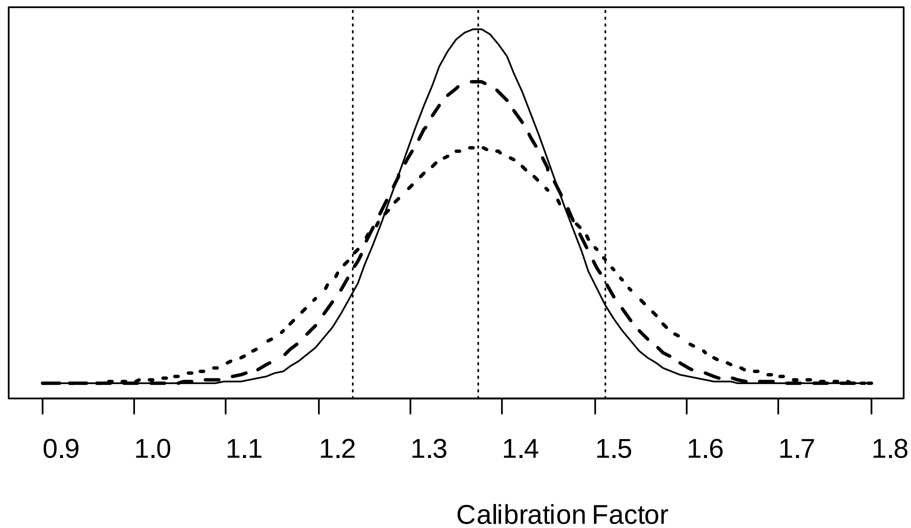


Figure 2. Accuracy of the Estimation of the Calibration Factor by a Sample Size of 100 (Solid Line), 75 (Dashed Line), and 50 (Dotted Line) for 4-Legged Signalized Intersections in Toronto, Ontario.

CHAPTER 4: SAFETY PERFORMANCE FUNCTION RECALIBRATION GUIDELINES

This chapter documents the proposed guidelines on when the predictive models are recommended for recalibration. The proposed guidelines are evaluated for different observed datasets in Texas and are based on the general characteristics of the data at hand: (1) the total numbers of crashes, (2) the ADT/AADT average, and (3) the total segment length (or the total number of intersections).

The chapter is divided into two sections. First, the recalibration guidelines are developed and discussed. Next, the guidelines are validated using two observed datasets in Texas.

SAFETY PERFORMANCE FUNCTION RECALIBRATION

As stated in Chapter 1, crash prediction models are crucial to predict the number of crashes and evaluate roadway safety. However, developing a new model demands a great deal of time, energy, and money. Although more efficient than fitting a new model, the recalibration of prediction models could still be a time-consuming and expensive task due to limitations with the data collection and completeness. Taking this issue into account, the agency may need to know when or how often SPFs should be recalibrated.

Recalibration Guidelines

In this section, recalibration guidelines are developed and discussed. The proposed guidelines are based on the general characteristics of data at hand: (1) the total number of crashes, (2) the ADT/AADT average, and (3) the total segment length. It is assumed that if the fluctuation in number of crashes is due to a comparable change in ADT/AADT, the agency may not need to conduct the recalibration. Otherwise, recalibration is advised.

Recall that the calibration factor can be calculated for a facility as the following:

$$C = \frac{\sum_i N_{i,obs}}{\sum_i N_{i,pre}} \tag{13}$$

where:

- C= Calibration factor (C-factor).
- $N_{i,obs}$= the observed number of crashes at site i.
- $N_{i,pre}$=the predicted number of crashes at site i.

The typical HSM base model to predict the number of crashes on roadway segments is as follows:

$$N_{SPF}^{i} = e^{b_0+b_1\ln(ADT_i)+\ln(L_i)} \quad (14)$$

where ADT_i and L_i, respectively, denote the ADT and the segment length at site i. Moreover, the coefficients b_0 and b_1, respectively, denote the SPF intercept and the ADT coefficient. Given Equation 14, Equation 13 can be written as:

$$C = \frac{\sum_i N_{i,obs}}{\sum_i \prod_j e^{b_0+b_1\ln(ADT_i)+\ln(L_i)} \times CMF_{ij}} \quad (15)$$

where CMF_{ij} denote the jth CMF at site i.

Now, let us assume the functional form in Equation 14 remains the same, all the sites meet base conditions (thus CMFs can be ignored) and the average ADT is used instead of the ADT at each site; consequently, the parameter $\tilde{C}$ (referred to as the C-factor proxy in this study) can be calculated as follows:

$$\tilde{C} \approx \frac{N_{obs}^{T}}{e^{b_0+b_1\ln(\overline{ADT})} \times L^{T}} \quad (16)$$

where:

- N_{obs}^{T} = the total number of crashes in the sample.
- $\overline{ADT}$ = the average ADT of all sites.
- L^{T} = the total combined length of all sites.

The only variables that are required to calculate the $\tilde{C}$ are (1) the total number of crashes, (2) the average AADT/ADT, and (3) total segment length. Now, it can be argued that if the number of crashes increases or decreases in line with ADT fluctuations (i.e., same direction and magnitude), the transportation agency may not need to go out of its way to recalibrate the predictive model (i.e., find a new calibration factor).

In order to use the guidelines, the agency is required to calculate the $\tilde{C}$ periodically and compare it with the $\tilde{C}$ that was estimated in the reference year (Note: the latest year that the predictive model was recalibrated is referred to as the reference year. In addition, the parameter $\tilde{C}$, which was calculated in the reference year, is denoted as $\tilde{C}_{REF}$.) If the relative difference between the current $\tilde{C}$ and $\tilde{C}_{REF}$ is less than a certain threshold (say 10 percent), the agency can ignore the recalibration and keep the current model; otherwise, the transportation agency is recommended to recalibrate the predictive model.

Procedure to Use the Guidelines

This section summarizes the procedure on how to calculate the $\tilde{C}$ and its use in the recalibration guidelines for roadway segment and intersection models. Before starting the steps, it is important to note that the proposed procedure serves only as guidelines. There might be other factors that affect the recalibration decision. For example, if significant systematic safety improvements happened in the jurisdiction analyzed, the agency would be required to update the models regardless of what the recalibration guidelines state. In addition, the agency is advised to recalibrate the model whenever possible and adequate resources are available.

Segment Models

The general steps to calculate the $\tilde{C}$ and its use in the recalibration guidelines for roadway segment models are as follows:

Step 1. Find the total number of crashes (N_{obs}^{T}) and the total segment length (L^{T}) on the network facility.

Step 2. Find the average ADT ($\overline{ADT}$) (or AADT) on the facility.

Note that if the ADT is not available to the agency on all sites, randomly collect ADT for a limited number of sites that provide the overall representation of the network.

Step 3. Consider the base SPF model (i.e., the model without CMFs) from the HSM. Let b_0 and b_1 denote the intercept and the coefficient of ADT, respectively. Estimate the approximate average predicted number of crashes ($\widetilde{N}_{pre}$) using the following functional form:

$$\widetilde{N}_{pre} = e^{b_0 + b_1 \times \ln(\overline{ADT})} \quad (17)$$

Step 4. Find the $\tilde{C}$ using the following equation:

$$\tilde{C} = \frac{N_{obs}^{T}}{\widetilde{N}_{pre} \times L^{T}} \quad (18)$$

Step 5. Find the variable $\tilde{e}$ as follows:

$$\tilde{e} = \frac{|\tilde{C} - \tilde{C}_{REF}|}{\tilde{C}_{REF}} \times 100 \quad (19)$$

where $\tilde{C}_{REF}$ denote the $\tilde{C}$ that was calculated in the reference year.

Step 6. If $\tilde{e} > 10\%$, the model needs to be recalibrated; calibrate the model and set the current $\tilde{C}$ as the new $\tilde{C}_{REF}$. Otherwise, keep the current $\tilde{C}_{REF}$ and use the calibration factor that was estimated in the reference year.

Intersection Models

A similar procedure with small modifications can be proposed to investigate whether or not the intersection models are required to be recalibrated. The general steps to calculate the $\tilde{C}$ and its use in the recalibration guideline for intersection models are as follows:

Step 1. Find the total number of crashes (N_{obs}^{T}) and the total number of intersections (N) on the network facility.

Step 2. Find the average traffic flow on major street ($\bar{F}_{major}$) and minor street ($\bar{F}_{minor}$).

Note: If the traffic flows are not available to the agency at all intersections, randomly collect ADT for a limited number of intersections that provide the overall representation of the intersection type the agency is interested in.

Step 3. Consider a base SPF model from the HSM. Let b_0, b_1, and b_2 denote the intercept, the coefficient of the traffic flow on major street, and the coefficient of the traffic flow on minor street, respectively. Find the approximate average predicted number of crashes ($\widetilde{N}_{pre}$) as:

$$\widetilde{N}_{pre} = e^{b_0 + b_1 \times \ln(\bar{F}_{major}) + b_1 \times \ln(\bar{F}_{minor})} \quad (20)$$

Step 4. Find the $\tilde{C}$ using the Equation 18. Note: for this case, in Equation 18, the total length variable (L^{T}) should be replaced with the total number of intersections (N).

Step 5. Find $\tilde{e}$ using the Equation 19.

Step 6. If $\tilde{e} > 10\%$, the model needs to be recalibrated; calibrate the model and set the current $\tilde{C}$ as the new $\tilde{C}_{REF}$. Otherwise, keep the current $\tilde{C}_{REF}$ and use the calibration factor that was estimated in the reference year.

VALIDATING THE GUIDELINES WITH OBSERVED DATASETS

In this section, the guidelines are evaluated for two different types of facilities. First, the guidelines are applied to the Texas urban 4-lane divided arterials. The data for this facility were collected in one- and three-year frequencies from 2007 to 2014. All variables in the dataset met the base conditions except the median width. Therefore, the model includes one CMF for the median width. Next, the same guidelines are examined for Texas multilane divided rural segments. The dataset was divided in three-year frequency from 2007 to 2014. For this dataset, the SPF included three CMFs: lane width, right shoulder width, and the median width. Table 5 shows the results for the Texas urban 4-lane divided arterials using three-year frequency data. First, it is assumed that the agency calibrated the model using data collected from 2007 to 2009. Although the calibration will be conducted after year 2009 is completed, for convenience purpose, the calibration year is referred to as 2009. Also, we refer this year to as the reference year and set the $\tilde{C}$ that was estimated in 2009 ($\tilde{C}_{2009}$) as $\tilde{C}_{REF}$ (i.e., $\tilde{C}_{REF}$=1.055). Next, in 2010, the data from 2008 to 2010 are used to calculate the $\tilde{C}$ ($\tilde{C}_{2010}$). The relative difference between $\tilde{C}_{2010}$ and $\tilde{C}_{REF}$ is equal to 4.95 percent. Since the change in $\tilde{C}$ is less than 10 percent, the recalibration is not needed. The relative difference between the actual calibration factor in 2010 from the reference is equal to 3.48 percent and is relatively small; so the actual results adequately confirm the decision of not to conduct the recalibration. In 2011, the agency could use the three years data from 2009 to 2011 to calculate the $\tilde{C}_{2011}$. The $\tilde{C}_{2011}$ is compared to the $\tilde{C}_{REF}$. (Note: since we did not calibrate the model in 2010, the reference year is still 2009 and $\tilde{C}_{REF} = 1.055$.) Since the $\tilde{C}$ is changed by more than 10 percent, the model should be recalibrated. After the recalibration, the reference year and $\tilde{C}_{REF}$ are modified accordingly. In 2012, the $\tilde{C}$ is changed by almost 10 percent; so recalibration is advised. The actual calibration factor is also increased by about 10 percent, so the decision based on the guidelines developed is justified. The model is recalibrated and the reference year is set to 2011 and $\tilde{C}_{2011}$ is set as the $\tilde{C}_{REF}$. The same procedure is repeated for subsequent years. As indicated in Table 5, the model, again, is needed for recalibration in 2014 once the analyst observes almost a 18 percent change in $\tilde{C}$.

Table 5. Recalibration for Urban 4-Lane Divided Arterials (Three-Year Frequency).

Year	2007–2009	2008–2010	2009–2011	2010–2012	2011–2013	2012–2014
Total Crashes	8573	8140	7613	8613	9114	10072
Avg. Predicted Crashes	7.74	7.66	7.65	7.71	7.61	7.64
Total Length (Mile)	1050.3	1060.2	1083.2	1105.9	1105.9	1105.2
$\tilde{C}$	1.055	1.002	0.919	1.011	1.083	1.192
$\tilde{C}_{REF}$*	-	**1.055 (2009)**	1.055 (2009)	**0.919 (2011)**	**1.011 (2012)**	1.011 (2012)
Change in $\tilde{C}$ (%)	-	4.95	**12.87**†	**10.0**	7.20	**17.99**
C	1.011	0.976	0.887	0.978	1.039	1.116
Change in C (%)	-	3.45	**12.28**	**10.21**	6.301	**14.12**

* The number in parenthesis indicates the reference year (the time that the model was recalibrated). In addition, the time that the reference year was changed for the first time is marked in bold. † Underlined values: $\tilde{C} \geq 10\%$.

Table 6 shows the recalibration results for the same facility described above (Texas urban 4-lane divided arterials); however, this time, one-year crash frequency data are used. It is assumed that the model was recalibrated using data collected in 2007. We set the $\tilde{C}_{2007}$ as $\tilde{C}_{REF}$. Later on, the $\tilde{C}_{2008}$ is calculated using the 2008 data. The change in $\tilde{C}$ is less than 10 percent, so recalibration is not recommended. Next, using 2009 data, the $\tilde{C}_{2009}$ is calculated and compared with $\tilde{C}_{REF}$. Since the change is more than 10 percent, the model is recommended for recalibration. The model is recalibrated, and the reference year and the $\tilde{C}_{REF}$ are modified accordingly. Later, for two years (in 2010 and 2011), the change in $\tilde{C}$ remain less than 10 percent. Therefore, for two years, no recalibration is advised. With 2012 data, however, the $\tilde{C}$ is increased by 23.21 percent. Since it is more than 10 percent, the model is recommended for recalibration. With 2012 data, the actual C-factor is also increased by 21.37 percent, which validates the recalibration decision made by the C-factor proxy.

Table 6. Recalibration for the Texas Urban 4-Lane Divided Arterials (One-Year Frequency).

Year	2007	2008	2009	2010	2011	2012	2013
Total Crashes	2714	3023	2591	2580	2560	3307	3304
Avg. Predicted Crashes	2.82	2.58	2.55	2.55	2.57	2.54	2.55
Total Length (Mile)	881.8	1050.3	1060.1	1083.2	1105.9	1105.9	1105.2
$\tilde{C}$	1.09	1.12	0.96	0.93	0.90	1.18	1.17
$\tilde{C}_{REF}$*	-	**1.09 (2007)**	1.09 (2007)	0.96 (2009)	0.96 (2009)	0.96 (2009)	**1.18 (2012)**
Change in $\tilde{C}$ (%)	-	2.31	**12.22†**	2.40	5.85	**23.21**	0.49
C	1.012	1.07	0.93	0.90	0.87	1.13	1.10
Change in C (%)	-	5.75	**7.86**	3.24	6.48	**21.37**	2.95

* The number in parenthesis show the reference year (the time that the model was recalibrated). In addition, the time that the reference year was changed for the first time is marked in bold. † Underlined values: $\tilde{C} \geq 10\%$.

In order to evaluate the guideline for datasets with more CMFs, the guidelines are applied to Texas multilane divided rural segments data collected from 2007 to 2014 in three-year intervals. The SPF included three CMFs: the lane width, right shoulder width, and the median width. Table 7 shows the recalibration results for this dataset. Let us first assume the model was calibrated in 2009. Then, as it is indicated in Table 7, until 2013 all corresponding changes in $\tilde{C}$ remain below than 10 percent, so the model is not advised for recalibration for four years. The small change in actual calibration factor also validates the decision made by the C-factor proxy. In 2013, however, the $\tilde{C}$ is increased by 10.04 percent compared to the $\tilde{C}_{REF}$, so the model is recommended for recalibration. The significant increase (10.07 percent) in actual calibration factor also validates the decision made by the recalibration guidelines. In 2014, since the difference between the C-factor proxy in 2014 and the reference year is more than 10 percent, the model is recommended for recalibration again.

Table 7. Recalibration for the Texas Multilane Divided Rural Segments (Three-Year Frequency).

Year	2007–2009	2008–2010	2009–2011	2010–2012	2011–2013	2012–2014
Total Crashes	1974	1993	1988	2175	2118	2511.000
Avg. Predicted Crashes	4.53	4.35	4.32	4.42	4.20	4.23
Total Length (Mile)	489.6	511.1	518.5	539.2	514.1	535.8
$\tilde{C}$	0.891	0.897	0.888	0.912	0.980	1.109
$\tilde{C}_{REF}$*	-	**0.891 (2009)**	0.891 (2009)	0.891 (2009)	0.891 (2009)	**0.980 (2013)**
Change in $\tilde{C}$ (%)	-	0.67	0.27	2.41	**10.04†**	**13.15**
C	0.961	0.961	0.946	0.994	1.058	1.120
Change in C (%)	-	0.02	1.59	3.41	**10.07**	**5.91**

* The number in parenthesis show the reference year (the time that the model was recalibrated). In addition, the time that the reference year was changed for the first time is marked in bold. † Underlined values: $\tilde{C} \geq 10\%$.

CHAPTER 5: REGION SPECIFIC CALIBRATION GUIDELINES

This chapter documents the application of region-specific calibration factors for large areas and when they are justified. More specifically, guidelines are proposed on when region-specific C-factors are recommended. These guidelines are similar to the ones that were proposed to recalibrate the predictive models in Chapter 4, but with some minor modifications. The proposed guidelines are validated using two observed datasets. A region can be defined by administrative boundaries, topography, or weather among others.

This chapter is divided into three sections. The first section describes when having region specific calibration factors is justified. The second section presents the guidelines for using region-specific C-factors. The last section covers how the guidelines were validated using two observed datasets.

DEVELOPING REGION SPECIFIC CALIBRATION FACTORS

This section documents when having region-specific calibration factors for large states, such as Texas is justified. The empirical data from two facility types, one in Texas and the other one in Michigan, are used to accomplish the task.

First, the region-specific C-factors are estimated for Texas urban 4-lane divided arterials (the same dataset was used in previous chapter to validate the recalibration guidelines). The dataset is collected in both one- and three-year frequencies. Texas is divided into four regions: north, south, east, and west. The division of the state is based on administrative boundaries used by the Texas Department of Transportation. Table 8 and

Table 9 summarize the calibration factors calculated for each region using the one- and three-year frequency data, respectively. As indicated in these tables, regardless of which frequency of data (one year or three year) are used, the difference between the C-factors calculated in different regions are significant. For instance,

Table 9 indicates that, in 2007, the difference between the C-factor in north region and the one in west region is about 35 percent. This observation validates the need for region-specific C-factors for large states, such as Texas.

Table 8. Region Specific Calibration Factors for Texas Urban 4-Lane Divided Arterials (One-Year Frequency).

Region	District Numbers	Calibration Factor						
		2007	2008	2009	2010	2011	2012	2013
North	1,2 ,3,9,10,18,19,23	0.890	1.070	0.820	0.855	0.713	0.928	0.922
South	13,14,15,16,21,22	1.056	0.978	0.848	0.817	0.945	1.234	1.091
East	11,12,17,20	1.048	1.081	1.137	0.980	0.997	1.268	1.378
West	4 ,5,6,7, 8, 24, 25	1.258	1.339	1.185	1.197	0.981	1.286	1.156
Total	All	1.012	1.070	0.932	0.902	0.872	1.131	1.098

Table 9. Region Specific Calibration Factors for Texas Urban 4-Lane Divided Arterials (Three-Year Frequency).

Region	District Numbers	Calibration Factor					
		07-09	08-10	09-11	10-12	11 -13	12-14
North	1,2 ,3,9,10,18,19,23	0.947	0.894	0.814	0.809	0.854	0.928
South	13,14,15,16,21,22	0.950	0.904	0.826	1.038	1.120	1.089
East	11,12,17,20	1.081	1.102	1.001	1.103	1.208	1.407
West	4 ,5,6,7, 8, 24, 25	1.284	1.248	1.129	1.165	1.114	1.247
Total	All	1.011	0.976	0.887	0.978	1.039	1.116

Figure 3 shows the region-specific C-factors for the urban 4-lane divided arterials in different regions of Texas in 2014 using the three years frequency data from 2012 to 2014.

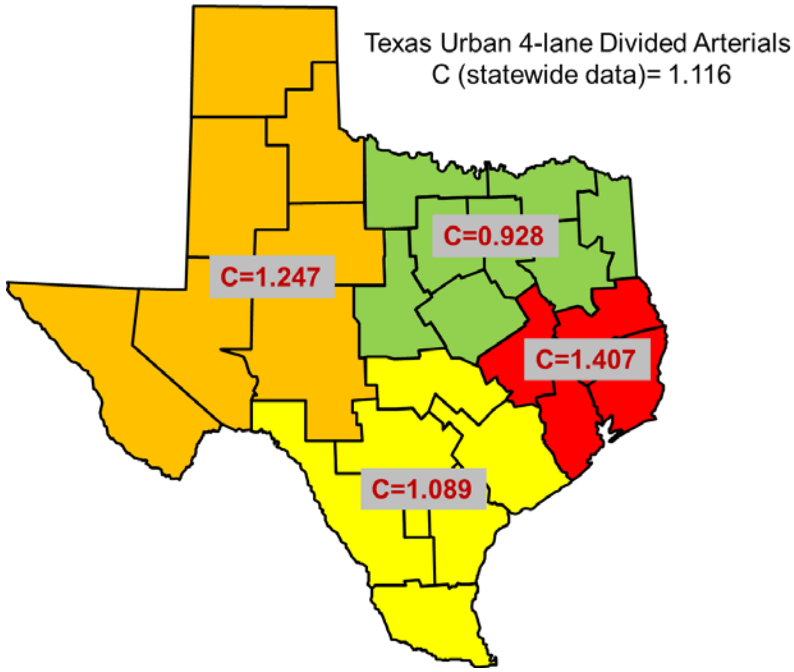


Figure 3. Region Specific Calibration Factors for Texas Urban 4-Lane Divided Arterials (Three-Year Frequency) in 2014.

In order to further investigate and validate the observation documented above for Texas urban 4-lane divided arterials, the region-specific C-factors were developed for Michigan 4-legged signalized intersections as well. This dataset consisted of one-year crash frequency. The state of Michigan is divided into seven regions, which is based on administrative boundaries. The regions were indicated by numbers from one to seven in Table 10. As shown in Table 10, the calibration factors are significantly dissimilar for different regions. For instance, in 2008, the difference between the C-factor in region 1 is almost 2.5 times larger than the C-factor in region 7. Having region-specific C-factors for Michigan 4-legged signalized intersections is also highly recommended.

It is worth pointing out that, for each year, there are about 50 sites in each region that can be used to find the region-specific C-factors. The CV of crash data for different regions in Michigan 4-legged signalized intersections data varies between 0.8 and 1.0. Based on sample size guidelines in Chapter 3, using 50 sites provide only 70 percent confidence that the C-factor lies within the 10 percent of the true factor; that is, the C-factor derived for different regions might still provide a biased or erroneous estimate. However, the effect of the sample size bias on the observed trend is rather negligible (considering the huge difference in C-factors derived), especially since the same trend repeated for all five years analyzed. In fact, to minimize the effect of the sample size bias, Table 10 included the five years average of the calibration factors too. This table shows that the five year average C-factors for different regions are also significantly different, so the region-specific C-factors are advised.

Table 10. Region Specific Calibration Factors for Michigan 4-Legged Signalized Intersections (One-Year Frequency).

Region	Region Name	Calibration Factor					
		2008	2009	2010	2011	2012	Avg.
1	Superior	2.62	2.43	2.44	2.19	2.49	2.43
2	North	2.44	2.35	2.44	2.36	2.21	2.36
3	Grand	1.66	1.67	1.75	2.07	1.78	1.79
4	Bay	2.09	2.25	2.13	2.01	2.01	2.10
5	Southwest	2.63	2.73	2.70	2.71	2.48	2.65
6	University	1.99	1.79	1.83	1.86	1.98	1.89
7	Metro	1.14	1.07	1.10	1.07	1.08	1.09
Total		1.83	1.80	1.82	1.84	1.78	1.81

In summary, based on these observations, it is justified to find and use region-specific C-factors instead of a state-wide C-factor for large states, such as Texas or Michigan. However, important questions should first be asked before segmenting a state or large region into sub-regions. When are the region-factors required? Can we get any intuition about its requirement before doing the recalibration process? For instance, the analyst may want to know if the sample should be collected into different regions or statewide level. The next section addresses this issue by providing the region-specific guidelines. The guidelines are based on the general characteristics of data at hand and can be used to inform the agency (or the analyst) whether or not having the region-specific C-factor is required.

REGION-SPECIFIC GUIDELINES

This section documents the proposed guidelines for determining when region-specific calibration factors are recommended. Similar guidelines to the ones proposed for recalibration of predictive models in Chapter 4 are used here as well to investigate the need for region-specific factors. First, the C-factor proxy similar to the one shown in Equation 16 is estimated for each region ($\tilde{C}_r$). Next, the $\tilde{C}_r$ is compared with the statewide $\tilde{C}$ ($\tilde{C}_s$). If the relative difference ($\tilde{e}_r$) is more than a certain threshold, the agency is recommended to calculate and use the region-specific C-factor. Otherwise, the statewide C-factor can be used. It is recommended to set the threshold to 10–20 percent. (We used a 10 percent threshold in this study.) The proposed procedure can be used for both segment and intersection models. The procedure for each roadway type is described below.

Segment Models

The general steps to calculate the $\tilde{C}_r$ and $\tilde{C}_s$, and how to use the region-specific guidelines for segment roadway models are as follows:

Step 1. Find the total number of crashes and the total segment length in the state ($N^T_{s,obs}$ and L^T_s) and in the region ($N^T_{r,obs}$ and L^T_r).

Step 2. Find the average ADT in the state ($\overline{ADT_s}$) and in the region ($\overline{ADT_r}$).

Step 3. Take the base SPF (i.e., the model without CMFs) from the HSM. Let b_0 and b_1, respectively, denote the intercept and the coefficient of ADT. Find the approximate average predicted number of crashes in the state ($\widetilde{N}_{s,pre}$) and in the region ($\widetilde{N}_{r,pre}$) as follows:

$$\widetilde{N}_{s,pre} = e^{b_0+b_1\times\ln(\overline{ADT_s})} \tag{21}$$

$$\widetilde{N}_{r,pre} = e^{b_0+b_1\times\ln(\overline{ADT_r})} \tag{22}$$

Step 4. Find the parameter $\tilde{C}$ in the state ($\tilde{C}_s$) and in the region ($\tilde{C}_r$) as follows:

$$\tilde{C}_s = \frac{N^T_{s,obs}}{\widetilde{N}_{s,pre}\times L^T_s} \tag{23}$$

$$\tilde{C}_r = \frac{N^T_{r,obs}}{\widetilde{N}_{r,pre}\times L^T_r} \tag{24}$$

Step 5. Find $\tilde{e}_r$ as follows:

$$\tilde{e}_r = \frac{|\tilde{C}_r-\tilde{C}_s|}{\tilde{C}_s} \times 100 \tag{25}$$

Step 6. If $\tilde{e}_r > 10\%$, calculate and use the region-specific calibration factor. Otherwise, use the statewide calibration factor.

Intersection Models

A similar procedure with some modifications can be proposed for intersection models. The general steps to calculate the $\tilde{C}_r$ and $\tilde{C}_s$, and how to use the region-specific guidelines for intersection models are as follows:

Step 1. Find the total number of crashes and the total number of intersections in the state ($N^T_{s,obs}$ and N_s) and in the region ($N^T_{r,obs}$ and N_r).

Step 2. Find the average traffic flow on the state major ($\bar{F}_{s,major}$) and minor ($\bar{F}_{s,minor}$) streets and on the region major ($\bar{F}_{r,major}$) and minor ($\bar{F}_{r,minor}$) streets.

Step 3. Take the SPF (only the base model without CMFs) from the HSM. Let b_0, b_1, and b_2, respectively, denote the intercept, the coefficient of traffic flow on the major street, and the coefficient of the traffic flow on the minor street. Find the approximate average predicted number of crashes in the state ($\widetilde{N}_{s,pre}$) and in the region ($\widetilde{N}_{r,pre}$) as:

$$\widetilde{N}_{s,pre} = e^{b_0+b_1\times\ln(\bar{F}_{s,major})+b_2\times\ln(\bar{F}_{s,minor})} \tag{26}$$

$$\widetilde{N}_{r,pre} = e^{b_0+b_1\times\ln(\bar{F}_{r,major})+b_2\times\ln(\bar{F}_{r,minor})} \tag{27}$$

Step 4. Find the parameter $\tilde{C}$ in the state ($\tilde{C}_s$) and in the region ($\tilde{C}_r$) using Equations 23 and 24, respectively. Note: the variables L_s^T and L_r^T should be replaced with N_s and N_r, respectively.

Step 5. Find $\tilde{e}_r$ using Equation 25.

Step 6. If $\tilde{e}_r > 10\%$, calculate and use the region-specific C-factor. Otherwise, use the statewide calibration factor.

VALIDATING THE REGION-SPECIFIC GUIDELINES

In this section, the region specific guidelines are validated using two observed datasets. First, the guidelines are examined with Texas urban 4-lane divided arterials data from 2012 to 2014. Next, the guidelines are evaluated with Michigan 4-legged signalized intersections dataset collected in 2008.

Table 11 shows the application of the region-specific guidelines to the Texas urban 4-lane arterials dataset collected from 2012 to 2014. As indicated in the table, for the north region, the difference between $\tilde{C}_{North}$ and $\tilde{C}_s$ is almost 15 percent, so a region-specific calibration factor for that region is recommended. The difference between the actual north and statewide C-factors is 16.85 percent, which validates the decision based on the guidelines. The guidelines, however, does not recommend considering a region-specific C-factor for the south region ($\tilde{e}_{South}$=2.94 percent). The difference between south region and statewide C-factors is also negligible (e_{South}= 2.42 percent). Next, the region-specific C-factor is also recommended for the east region (e_{East}=23.63 percent). Here too the actual difference between C-factors is significant (e_{East}= 26.08 percent), which confirms the decision based on the guidelines. For the west region, the $\tilde{e}_{West}$ is equal to 13.75 percent, which is greater than 10 percent, so the region-specific factor is also recommended.

Table 11. Region-Specific Guidelines – Texas Urban 4-Lane Divided Arterials (Three-Year Frequency Data from 2012 to 2014).

Region	North	South	East	West	Statewide
Total Crashes	3214	2851	2902	1105	10072
Avg. Predicted Crashes	8.40	8.50	8.16	4.13	7.64
Total Length (Mile)	376.7	289.8	241.3	197.4	1105.2
$\tilde{C}$	1.015	1.157	1.474	1.356	1.192
Change in $\tilde{C}$ (%)	**14.85**†	2.94	**23.63**	**13.75**	-
C	0.928	1.089	1.407	1.247	1.116
Change in C (%)	**16.85**	2.42	**26.08**	**11.74**	-

† Underlined values: $\tilde{C} \geq 10\%$.

Table 12 indicates the results for application of the region-specific guidelines for Michigan 4-legged signalized intersections dataset collected in 2008. As shown in this table, the region-specific factor was recommended for all regions, except for region 6. For all these regions, the difference between the actual region and statewide C-factors is also very large. For region 6, the region-specific factor is not recommended. The actual C-factor also affirms the decision based on the guidelines.

Table 12. Region Specific Guidelines - Michigan 4-Legged Signalized Intersections (One-Year Frequency Data in 2008).

Region	1	2	3	4	5	6	7	State-wide
Total Crashes	250	372	454	344	483	374	578	2855
Avg. Predicted Crashes	3.02	4.28	7.61	4.65	4.95	5.16	13.06	6.08
No. of Intersections	46	50	51	50	52	50	50	349
$\tilde{C}$	1.80	1.74	1.17	1.48	1.88	1.45	0.89	1.35
Change in $\tilde{C}$ (%)	**33.85**†	**29.19**	**13.15**	**9.97**	**39.44**	7.71	**34.25**	-
C	2.62	2.44	1.66	2.09	2.63	1.99	1.14	1.83
Change in C (%)	**43.17**	**33.33**	**9.29**	**14.21**	**43.72**	8.74	**37.70**	-

† Underlined values: $\tilde{C} \geq 10\%$.

CHAPTER 6: SEVERITY DISTRIBUTION FUNCTION CALIBRATION FACTORS

This chapter documents the development of the methodology for calibrating the HSM SDFs and sample size guidelines to calculate the calibration factor. The chapter is divided into five sections. The first section describes the characteristics of the SDFs documented in Chapters 18 and 19 of the HSM. The second section provides a description of the calibration methodology and investigates the principles behind the equation that was proposed to derive the C-factor. The third section evaluates the proposed equation for a range of simulated scenarios. The fourth section presents the simulation results for estimating the required sample size to calibrate the models based on the characteristics of data at hand. The fifth section documents the sample size guidelines.

SEVERITY DISTRIBUTION FUNCTIONS

The total number of crashes at each site (s) can be classified into two severity categories: fatal-injury (FI) and property damage only (PDO) crashes. The FI crashes can further be classified into four severity level categories: fatal (K), incapacitating injury (A), non-incapacitating injury (B), and possible injury (C). An SDF is a discrete choice model to predict the likelihood of each severity level described above (i.e., K, A, B, or C). The SDF usually includes explanatory variables, such as the geometric design of the sites, traffic control features, or traffic characteristics. Since the SDF accounts for all severity levels together, a single change in variables such as roadway characteristics could result in simply shifting the number of crashes between different severity level alternatives. The SDFs for freeways and ramps currently presented in the HSM Chapters 18 and 19, respectively, were developed using the multinomial logit (MNL) model.

The theoretical framework of the SDF can be presented as follows. Assume an MNL model with the severity level C as the base scenario in the model. Let $p^s_{pre,KAB}$ and $p^s_{pre,C}$ be the pre-calibration likelihoods for the severity level KAB (K+A+B) and the severity level C at site s respectively. Let u_K, u_A, and u_B, respectively, denote the deterministic components (the utility function in the context of the MNL) of the severity level K, A, and B. Then, the pre-calibration predicted likelihoods for the severity levels KAB (K+A+B) and C can be determined as follows:

$$p^s_{pre,KAB} = \frac{e^{u^s_k}+e^{u^s_A}+e^{u^s_B}}{1+e^{u^s_k}+e^{u^s_A}+e^{u^s_B}} \qquad (28)$$

$$p^s_{pre,C} = \frac{1}{1+e^{u^s_k}+e^{u^s_A}+e^{u^s_B}} \qquad (29)$$

In addition, the likelihoods for the severity level K, A, and B are:

$$p_{pre,K}^{s} = \frac{e^{u_k^s}}{1+e^{u_k^s}+e^{u_A^s}+e^{u_B^s}} \text{ ; } p_{pre,A}^{s} = \frac{e^{u_A^s}}{1+e^{u_k^s}+e^{u_A^s}+e^{u_B^s}} \text{ ; } p_{pre,B}^{s} = \frac{e^{u_B^s}}{1+e^{u_k^s}+e^{u_A^s}+e^{u_B^s}}$$

where $p_{pre,K}^{s}$, $p_{pre,A}^{s}$, and $p_{pre,B}^{s}$ denote the pre-calibration likelihoods for the severity levels K, A, and B at site s, respectively.

SEVERITY DISTRIBUTION FUNCTION CALIBRATION FACTOR

Similar to the SPFs, the SDFs are also fitted and validated using data from a few selected numbers of states in the United States; since the roadway characteristics vary substantially from one jurisdiction to another, the SDF models are required to be calibrated to the local conditions. Recall that in Equations 28 and 29, the utility function of the severity level C was assumed as the base utility (i.e., $e^{u_C} = 1$); to calibrate the models to the local condition, the base utility can be modified with a scalar C-factor. Consequently, we have,

$$p_{post,KAB}^{s} = \frac{e^{u_k^s}+e^{u_A^s}+e^{u_B^s}}{\frac{1}{C}+e^{u_k^s}+e^{u_A^s}+e^{u_B^s}} \tag{30}$$

where C denote the calibration factor and $p_{post,KAB}^{s}$ and $p_{post,C}^{s}$ are the post-calibration likelihoods for the severity levels KAB and C. In addition, we have,

$$p_{post,K}^{s} = \frac{e^{u_k^s}}{\frac{1}{C}+e^{u_k^s}+e^{u_A^s}+e^{u_B^s}} \text{; } p_{post,A}^{s} = \frac{e^{u_A^s}}{\frac{1}{C}+e^{u_k^s}+e^{u_A^s}+e^{u_B^s}} \text{; } p_{post,B}^{s} = \frac{e^{u_B^s}}{\frac{1}{C}+e^{u_k^s}+e^{u_A^s}+e^{u_B^s}} \tag{31}$$

Taking Equations 28 to 31 into account, we have:

$$\frac{p_{post,KAB}^{s}}{p_{post,C}^{s}} = C \times \frac{p_{pre,KAB}^{s}}{p_{pre,C}^{s}} \tag{32}$$

As a result, the scalar C-factor modifies the ratio of the predicted KAB to the predicted C crashes.

In order to estimate the SDF C-factor for freeways and ramps, the HSM proposed the following procedure:

(1) Find the average observed probability of KAB crashes ($p_{obs,KAB}$) as follows:

$$p_{obs,KAB} = \frac{\sum_s N_{obs,KAB}^{s}}{\sum_s N_{obs,KABC}^{s}} \tag{33}$$

(2) Find the average pre-calibration (C=1) predicted probability of KAB crashes ($p_{pre,KAB}$) as:

$$p_{pre,KAB} = \frac{\sum_s N_{pre,KAB}^{s}}{\sum_s N_{pre,KABC}^{s}} \tag{34}$$

(3) Find the calibration factor using the Equation (35):

$$C = \frac{p_{obs,KAB}}{1-p_{obs,KAB}} \times \frac{1-P_{pre,KAB}}{P_{pre,KAB}} \tag{35}$$

Alternatively, one can show that the Equation 35 is equivalent to Equation 36 as follows:

$$C = \frac{\sum_s N_{obs,KAB}^s / \sum_s N_{pre,KAB}^s}{\sum_s N_{obs,C}^s / \sum_s N_{pre,C}^s} \tag{36}$$

Let us further investigate on how Equation 36 can be interpreted. As described earlier, since roadway characteristics vary from one jurisdiction to another, the SDFs need to be calibrated. For this purpose, one can modify the probability that is assigned to the KAB and the one that is assigned to the C severity levels. One way to look at this problem is to minimize the bias between the total observed and predicted KAB, and the total observed and predicted C crashes; the ratio of the KAB and C crashes can be modified by $\sum_s N_{obs,KAB}^s \,/\, \sum_s N_{pre,KAB}^s$ and $\sum_s N_{obs,C}^s \,/\, \sum_s N_{pre,C}^s$, respectively. Therefore,

$$p_{post,KAB}^s = \frac{\frac{\sum_s N_{obs,KAB}^s}{\sum_s N_{pre,KAB}^s} \times (e^{u_k^s} + e^{u_A^s} + e^{u_B^s})}{\frac{\sum_s N_{obs,C}^s}{\sum_s N_{pre,C}^s} \times 1 + \frac{\sum_s N_{obs,KAB}^s}{\sum_s N_{pre,KAB}^s} \times (e^{u_k^s} + e^{u_A^s} + e^{u_B^s})} \tag{37}$$

$$p_{post,C}^s = \frac{\frac{\sum_s N_{obs,C}^s}{\sum_s N_{pre,C}^s} \times 1}{\frac{\sum_s N_{obs,C}^s}{\sum_s N_{pre,C}^s} \times 1 + \frac{\sum_s N_{obs,KAB}^s}{\sum_s N_{pre,KAB}^s} \times (e^{u_k^s} + e^{u_A^s} + e^{u_B^s})} \tag{38}$$

This can transformed as follows:

$$p_{post,KAB}^s = \frac{(e^{u_K} + e^{u_A} + e^{u_B})}{\left(\frac{\sum_s N_{obs,C}^s / \sum_s N_{pre,C}^s}{\sum_s N_{obs,KAB}^s / \sum_s N_{pre,KAB}^s}\right) \times 1 + (e^{u_k^s} + e^{u_A^s} + e^{u_B^s})} \tag{39}$$

$$p_{post,C}^s = \frac{\frac{\sum_s N_{obs,C}^s / \sum_s N_{pre,C}^s}{\sum_s N_{obs,KAB}^s / \sum_s N_{pre,KAB}^s} \times 1}{\left(\frac{\sum_s N_{obs,C}^s / \sum_s N_{pre,C}^s}{\sum_s N_{obs,KAB}^s / \sum_s N_{pre,KAB}^s}\right) \times 1 + (e^{u_k^s} + e^{u_A^s} + e^{u_B^s})} \tag{40}$$

Taking the Equation 36 into account, we have,

$$p_{post,KAB}^s = \frac{e^{u_k^s} + e^{u_A^s} + e^{u_B^s}}{\frac{1}{C} + e^{u_k^s} + e^{u_A^s} + e^{u_B^s}}$$

$$p_{post,C}^{s} = \frac{\frac{1}{C}}{\frac{1}{C}+e^{u_k^s}+e^{u_A^s}+e^{u_B^s}}$$

which are equivalent to Equations 30 and 31 (i.e., the post-calibration likelihoods).

VALIDATION OF THE C-FACTOR

This section documents the validation of the calibration factor derived from Equation 36. The validation is performed using the following steps. First, a calibration factor (C) is assumed. In the next step, given a large sample size, for each site s, KAB and C crash data are randomly generated from a negative binomial distribution. The corresponding C-factor is then calculated (C_m). The later step is repeated for 1,000 iterations. In the end, the mean of the generated C-factors is calculated ($\bar{C}_m$) (the average C_m) and the bias between the $\bar{C}_m$ and C (i.e., $|\bar{C}_m - C|$) is measured and evaluated with the desired ε-accuracy.

This section is divided into two parts. First, the simulation protocol is presented to generate the calibration factors for different scenarios with different characteristics, followed by a process to evaluate them. In the second part, the simulation protocol is applied to different scenarios and results are presented.

Simulation Protocol

The following simulation protocol is proposed to validate the C-factor derived from Equation 36. In this simulation protocol, we evaluate the bias between $\bar{C}_m$ and C.

Step 1—Initialization

1.1 Set a scenario by specifying:

- Average predicted ratio for severity level KAB ($\bar{P}_{pre,KAB}$). Note: $\bar{P}_{pre,C} = 1 - \bar{P}_{pre,KAB}$.
- An SDF C-factor (C).
- An inverse dispersion parameter for severity level KAB (φ_1) and C (φ_2).

1.2 Take an SPF model for KABC crashes and modify its intercept to adopt the desired mean for total KABC crashes. Then, calculate the predicted number of KABC crashes for each site ($N_{pre,KABC}^{s}$).

1.3 Use a beta distribution with a mean of $\bar{P}_{pre,KAB}$ and standard deviation of 0.1 to generate the predicted KAB probability ratio at each site ($P_{pre,KAB}^{s}$). Note: $P_{pre,C}^{s} = 1 - P_{pre,KAB}^{s}$.

1.4 Assign the predicted number of crashes to the KAB and C severity levels based on $P_{pre,KAB}^{s}$ and $P_{pre,C}^{s}$, respectively.

Step 2—Generating the Observed Dataset

Repeat the following steps for 1,000 iterations.

2.1 Find the post-calibration probability ratios ($P^s_{post,KAB}$ and $P^s_{post,C}$) by solving the following two equations:

$$\frac{P^s_{post,KAB}}{P^s_{post,C}} = C \times \frac{P^s_{pre,KAB}}{P^s_{pre,C}} \tag{41}$$

$$P^s_{post,KAB} + P^s_{post,C} = 1 \tag{42}$$

2.2 Generate the observed number of crashes for the severity level KAB ($N^s_{obs,KAB}$) from a negative binomial distribution with a mean equal to $N^s_{pre,KABC} \times P^s_{post,KAB}$ and the inverse dispersion parameter of φ_1.

2.3 Generate the observed number crashes for the severity level C ($N^s_{obs,C}$) from a negative binomial distribution with a mean equal to $N^s_{pre,KABC} \times P^s_{post,C}$ and the inverse dispersion parameter of φ_2.

2.4 Calculate the calibration factor using the following equation:

$$C_m = \frac{\sum_s N^s_{obs,KAB} / \sum_s N^s_{pre,KAB}}{\sum_s N^s_{obs,C} / \sum_s N^s_{pre,C}} \tag{43}$$

where C_m indicates the calibration factor in iteration m.

Step 3—Evaluation

3.1 Find the average of the generated C-factors ($\bar{C}_m$).

3.2 Find the bias $|\bar{C}_m - C|$ and the percentage of relative error as $\frac{|\bar{C}_m - C|}{C} \times 100$ and compare it with the desired accuracy.

Simulation Results

This subsection reports the simulation results for different scenarios. We set the predicted mean for total KABC crashes to 10 and modify the intercept of the KABC predictive model (the same model that was used in SPF sample size evaluation) to attain the desired mean. For simplicity, it was assumed that $\varphi = \varphi_1 = \varphi_2$ (i.e., KAB crashes have the same inverse dispersion parameter as C crashes). In total, 36 scenarios were generated and repeated for 1,000 iterations. The range of the average predicted ratio for the severity level KAB ($\bar{P}_{pre,KAB}$), the C-factor (C), and the inverse dispersion parameter (φ) were varied as follows:

$\bar{P}_{pre,KAB} = \{0.2,0.4,0.6,0.8\}$

$C = \{0.5, 1, 1.5, 2\}$

$\varphi = \{0.5,1,5\}$

The range for the C-factor includes C<1, C=1, and C>1. In addition, the range for the inverse dispersion parameter covers the high, medium, and low dispersion. Table 13 shows the relative bias error for a range of scenarios described above. As it is shown in Table 13, the relative bias error is minimal for different scenarios. If a threshold of 5 percent is assumed for the accuracy, then for almost all scenarios, the C-factor bias lied below the accuracy threshold. The minimum bias is for the scenarios where the C-factor is equal to 1 (i.e., C=1). Based on this result, the equation presented in the HSM for calibrating the SDF is valid and appropriate.

Table 13. The C-Factor Relative Bias (%) for Different Scenarios.

Avg. Predicted KAB Ratio ($\bar{P}_{pre,KAB}$)	C-Factor (C)	Inverse Dispersion Parameter (φ)		
		0.5	**1.0**	**5.0**
0.2	**0.5**	3.67%	3.82%	3.76%
	1.0	0.01%	0.09%	0.02%
	1.5	2.91%	2.67%	2.81%
	2.0	4.97%	4.79%	5.02%
0.4	**0.5**	2.56%	2.84%	2.63%
	1.0	0.24%	0.07%	0.05%
	1.5	1.65%	1.84%	1.77%
	2.0	2.84%	2.88%	2.84%
0.6	**0.5**	3.01%	2.98%	3.03%
	1.0	0.04%	0.10%	0.05%
	1.5	1.51%	1.53%	1.55%
	2.0	2.37%	2.56%	2.60%
0.8	**0.5**	5.14%	5.39%	5.25%
	1.0	0.09%	0.08%	0.02%
	1.5	2.45%	2.17%	2.24%
	2.0	3.70%	3.66%	3.49%

SAMPLE SIZE EVALUATION

This section presents the results of the analysis to determine the required sample size to estimate the SDF calibration factor. The section is divided into two parts. First, a simulation protocol is presented to simulate and evaluate the sample size for different scenarios. Second, considering different scenarios, the results of the simulation are presented.

Simulation Protocol

The following protocol is proposed to simulate a wide range of scenarios and evaluate the required sample size for each case. In this protocol, the quality of each sample size is measured with the same method used in Chapter 3 to evaluate the quality of the sample size to calibrate SPFs. With this method, it is assumed that the calibration factors that were generated for 1,000 iterations, given a sample size, follow a normal distribution. Then, the sample size that fulfils the desired confidence levels that the calibration factor lies within 10 percent of the true factor is determined.

Step 1—Initialization

1.1 Set a scenario by specifying:

- Average predicted ratio for severity level KAB ($\bar{P}_{pre,KAB}$). Note: $\bar{P}_{pre,C} = 1 - \bar{P}_{pre,KAB}$.
- An SDF C-factor (C).
- An inverse dispersion parameter for severity level KAB (φ_1) and C (φ_2).

1.2 Consider an SPF model for KABC crashes and modify its intercept to adopt the desired mean for total KABC crashes. Then calculate the predicted number of KABC crashes for each site ($N^s_{pre,KABC}$).

1.3 Use a beta distribution with a mean of $\bar{P}_{pre,KAB}$ and standard deviation of 0.1 to generate the predicted KAB probability ratio at each site ($P^s_{pre,KAB}$). Note: $P^s_{pre,C} = 1 - P^s_{pre,KAB}$.

1.4 Assign the predicted number of crashes to the severity levels KAB and C based on $P^s_{pre,KAB}$ and $P^s_{pre,C}$, respectively.

Step 2—Generating the Observed Dataset

2.1 Find the post-calibration probability ratios ($P^s_{post,KAB}$ and $P^s_{post,C}$) by solving Equations 41 and 42.

2.2 Generate the observed number of crashes for the severity level KAB ($N^s_{obs,KAB}$) from a negative binomial distribution with a mean equal to $N^s_{pre,KABC} \times P^s_{post,KAB}$ and the inverse dispersion parameter of φ_1.

2.3 Generate the observed number crashes for the severity level C ($N^s_{obs,C}$) from a negative binomial distribution with a mean equal to $N^s_{pre,KABC} \times P^s_{post,C}$ and the inverse dispersion parameter of φ_2.

2.4 Calculate the calibration factor as follows where C_N indicates the calibration factor for the full dataset with a size N:

$$C_N = \frac{\sum_{s=1}^{N} N^s_{obs,KAB} / \sum_{s=1}^{N} N^s_{pre,KAB}}{\sum_{s=1}^{N} N^s_{obs,C} / \sum_{s=1}^{N} N^s_{pre,C}} \quad (44)$$

Step 3—Evaluation of the Sample Size

3.1 Repeat the following steps for 1,000 iterations:

3.1.1 For a given sample size (n), randomly select (n) sites.

3.1.2 Calculate the sample's calibration factor as follows:

$$C_n = \frac{\sum_{s=1}^{n} N_{obs,KAB}^{s} / \sum_{s=1}^{n} N_{pre,KAB}^{s}}{\sum_{s=1}^{n} N_{obs,C}^{s} / \sum_{s=1}^{n} N_{pre,C}^{s}} \quad (45)$$

3.2 Measure the quality of each given sample size as follows:

3.2.1 Calculate the mean and standard deviation of the generated calibration factors and denote them as $\mathrm{Avg}(C_n)$ and $\mathrm{sd}(C_n)$, respectively.

3.2.2 Assume calibration factors that are generated from 1,000 iterates of simulation follow a normal distribution and then calculate following two statistics:

$$Z_{min} = \frac{0.9 \times C_N - \mathrm{Avg}(C_n)}{\mathrm{sd}(C_n)} \quad (46)$$

$$Z_{max} = \frac{1.1 \times C_N - \mathrm{Avg}(C_n)}{\mathrm{sd}(C_n)} \quad (47)$$

3.2.3 Find the probability that the calibration factor lies within 10 percent of the true calibration factor (p) as:

$$P = \Phi(Z_{max}) - \Phi(Z_{min}) \quad (48)$$

where $\Phi(.)$ indicates the CDF of the normal distribution.

Simulation Results

For simplicity, it was assumed that $\varphi = \varphi_1 = \varphi_2$ (i.e., KAB crashes have the same inverse dispersion parameter as C crashes). In addition, it was assumed that sites with zero KABC crashes do not add much information in calculating the SDF C-factor, so they were not considered in the simulated dataset (they were excluded from the dataset). In total, 5,000 sites with at least 1 KABC crash were generated and used for the sample size evaluation. In total, 36 different scenarios were evaluated using the simulation protocol described above. The range of the characteristics was varied (similar to those that were considered for the SDF C-factor validation) as follows:

$$\bar{P}_{pre,KAB} = \{0.2, 0.4, 0.6, 0.8\}$$

$$C = \{0.5, 1, 1.5, 2\}$$

$$\varphi = \{0.5, 1, 5\}$$

Note that $\bar{P}_{pre,C} = 1 - \bar{P}_{pre,KAB}$; the range of the $\bar{P}_{pre,C}$ is varied as {0.8, 0.6, 0.4, 0.2}. (Technically, a $\bar{P}_{pre,C}$ below 0.6 is not realistic in practice. However, the full range was used for completeness purposes.) As described in the previous section, the range for the C-factor includes C<1, C=1, and C>1. In addition, the range for the inverse dispersion parameter covers high, medium, and low dispersion. Furthermore, the range for $\bar{P}_{pre,KAB}$ combined with the given C-factors covers a wide range of scenarios for the $\bar{P}_{obs,KAB}$ (and $\bar{P}_{obs,C}$). The sample size range was increased in increments of 25 for a sample size varying from 50 to 200, in increments of 50 for a sample varying from 200 to 1,000, and in increments of 100 for a sample size varying from 1,000 to 1,500. Table 14 shows the simulation results and the sample size requirements for different scenarios.

Table 14. Sample Size Requirement.

$\frac{\bar{P}_{pre,KAB}}{\bar{P}_{pre,C}}$	C-factor (C)*	$\frac{\bar{P}_{obs,KAB}}{\bar{P}_{obs,C}}$*	Inverse dispersion parameter (φ)		
			0.5	1.0	5.0
0.20/0.80	0.5	0.11/0.89	1,200 (90%)** 800 (80%) 550 (70%)	900 (90%) 600 (80%) 400 (70%)	400 (90%) 250 (80%) 175 (70%)
	1.0	0.20/0.80	1,100 (90%) 750 (80%) 550 (70%)	800 (90%) 600 (80%) 400 (70%)	300 (90%) 200(80%) 125 (70%)
	1.5	0.27/0.73	1,200 (90%) 800 (80%) 550 (70%)	750 (90%) 500 (80%) 350 (70%)	300 (90%) 175 (80%) 125 (70%)
	2.0	0.33/0.67	1,200 (90%) 750 (80%) 550 (70%)	800 (90%) 450 (80%) 350 (70%)	250 (90%) 175 (80%) 100 (70%)
0.40/0.60	0.5	0.25/0.75	1,100 (90%) 750 (80%) 500 (70%)	750 (90%) 500 (80%) 350 (70%)	300 (90%) 175 (80%) 125 (70%)
	1.0	0.40/0.60	1,100 (90%) 650 (80%) 550 (70%)	750 (90%) 500 (80%) 350 (70%)	300 (90%) 150 (80%) 100 (70%)
	1.5	0.50/0.50	1,100 (90%) 700 (80%) 500 (70%)	700 (90%) 450 (80%) 350 (70%)	250 (90%) 150 (80%) 100 (70%)
	2.0	0.57/0.43	1,200 (90%) 800 (80%) 550 (70%)	700 (90%) 450 (80%) 300 (70%)	250 (90%) 150 (80%) 125 (70%)
0.60/0.40	0.5	0.43/0.57	1,200 (90%) 800 (80%) 600 (70%)	700 (90%) 500 (80%) 350 (70%)	250 (90%) 175 (80%) 100 (70%)
	1.0	0.60/0.40	1,200 (90%) 800 (80%) 550 (70%)	700 (90%) 450 (80%) 300 (70%)	250 (90%) 150 (80%) 125 (70%)
	1.5	0.69/0.31	1,100 (90%) 750 (80%) 500 (70%)	800 (90%) 500 (80%) 350 (70%)	300 (90%) 150 (80%) 125 (70%)
	2.0	0.75/0.25	1,100 (90%) 750 (80%) 550 (70%)	700 (90%) 500 (80%) 350 (70%)	300 (90%) 150 (80%) 125 (70%)
0.80/0.20	0.5	0.67/0.33	1,200 (90%) 750 (80%) 550 (70%)	750 (90%) 450 80%) 350 (70%)	300 (90%) 175(80%) 125 (70%)
	1.0	0.80/0.20	1,100 (90%) 750 (80%) 550 (70%)	800 (90%) 500 (80%) 350 (70%)	350 (90%) 200 (80%) 150 (70%)
	1.5	0.86/0.14	1,200 (90%) 800 (80%) 600 (70%)	750 (90%) 550 (80%) 350 (70%)	350 (90%) 250 (80%) 150 (70%)
	2.0	0.89/0.11	1,200 (90%) 800 (80%) 600 (70%)	800 (90%) 550 (80%) 350 (70%)	400 (90%) 250 (80%) 175 (70%)

*The generated dataset might have slightly different C-factor or average observed severity ratios due to randomness but here we reported only the rounded values.
**The numbers in parenthesis shows the confidence level that the sample size provides.

As indicated in Table 14, for a same inverse dispersion parameter φ, the effect of the average observed KAB and C probability ratio on the required sample size is rather small (i.e., the range of the required sample size approximately remains the same). For example, in spite of what the average observed KAB to C ratio is, the required sample size to fulfill 90 percent level of accuracy, is 1,100–1,200 sites for $\varphi = 0.5$, 700–900 sites for $\varphi = 1$, and 250–400 sites for $\varphi = 5$. On the other hand, the level of dispersion can have a significant effect on the required sample size. For instance, if the average observed KAB/C probability is 0.20/0.80, to fulfill 90 percent of confidence, we need 1,100 sites for $\varphi = 0.5$, 800 sites for $\varphi = 1$, and 300 sites for $\varphi = 5$. Thus, it can be hypothesized that the variation in KAB or C crashes is more critical than their ratio in estimating the required sample size.

Let CV_{KAB} and CV_C denote the CV of KAB and C severity levels, respectively:

$$CV_{KAB} = \frac{sd(N_{obs,KAB})}{\bar{N}_{obs,KAB}} \tag{49}$$

$$CV_C = \frac{sd(N_{obs,C})}{\bar{N}_{obs,C}} \tag{50}$$

Now, define the average of the CVs of KAB and C crashes as the following:

$$CV_{Avg.} = \frac{(CV_{KAB} + CV_C)}{2} \tag{51}$$

Then, to account the effect of the level of dispersion, one can sort the simulation results based on $CV_{Avg.}$ Table 15 shows the required sample size given the $CV_{Avg.}$ The $CV_{Avg.}$ ranged from 0.87 to 1.74. For other values, an interpolation can be used to estimate required sample size. As shown in Table 15, as $CV_{Avg.}$ increases, a smaller sample size is required to find a reliable calibration factor. We use this observation to propose SDF sample size guidelines in the next section.

Table 15. Average CV of KAB and C Levels vs. Required Sample Size.

$CV_{Avg.}$	CV_{KAB}	CV_C	Required Sample Size		
			90%*	80%	70%
1.74	1.75	1.73	1,200 (0.90)**	850 (0.83)	550 (0.72)
1.74	1.50	1.98	1,200 (0.90)	850 (0.82)	600 (0.74)
1.73	1.49	1.97	1,200 (0.90)	800 (0.81)	600 (0.73)
1.73	1.83	1.63	1,200 (0.90)	800 (0.81)	600 (0.73)
1.73	1.91	1.55	1,200 (0.92)	800 (0.82)	550 (0.71)
1.71	1.63	1.79	1,200 (0.90)	750 (0.81)	550 (0.72)
1.69	1.77	1.60	1,100 (0.90)	750 (0.82)	550 (0.71)
1.69	1.75	1.62	1,200 (0.91)	800 (0.83)	550 (0.72)
1.68	1.66	1.71	1,200 (0.90)	800 (0.82)	550 (0.72)
1.68	1.77	1.59	1,200 (0.92)	750 (0.81)	550 (0.72)
1.67	1.56	1.78	1,100 (0.90)	750 (0.80)	550 (0.72)
1.65	1.59	1.71	1,100 (0.91)	750 (0.82)	500 (0.71)
1.65	1.54	1.76	1,100 (0.90)	750 (0.81)	550 (0.72)
1.64	1.57	1.71	1,100 (0.92)	700 (0.80)	500 (0.71)
1.63	1.68	1.57	1,100 (0.91)	750 (0.82)	500 (0.72)
1.62	1.57	1.67	1,100 (0.90)	650 (0.80)	550 (0.76)
1.50	1.67	1.34	800 (0.90)	600 (0.83)	400 (0.72)
1.47	1.73	1.21	900 (0.91)	600 (0.83)	400 (0.72)
1.43	1.20	1.66	800 (0.91)	550 (0.82)	350 (0.71)
1.43	1.53	1.33	750 (0.90)	500 (0.81)	350 (0.71)
1.42	1.19	1.64	750 (0.90)	550 (0.82)	350 (0.71)
1.41	1.22	1.61	800 (0.91)	500 (0.81)	350 (0.71)
1.39	1.45	1.34	800 (0.91)	450 (0.81)	350 (0.72)
1.34	1.24	1.44	750 (0.92)	450 (0.80)	350 (0.73)
1.34	1.33	1.35	800 (0.91)	500 (0.81)	350 (0.72)
1.33	1.24	1.42	700 (0.91)	500 (0.84)	350 (0.74)
1.33	1.38	1.27	750 (0.92)	500 (0.83)	350 (0.75)
1.32	1.32	1.32	650 (0.90)	450 (0.81)	350 (0.74)
1.32	1.32	1.31	700 (0.91)	450 (0.80)	350 (0.73)
1.31	1.35	1.27	750 (0.91)	500 (0.83)	350 (0.73)
1.30	1.29	1.31	700 (0.91)	450 (0.81)	300 (0.72)
1.29	1.27	1.32	700 (0.92)	450 (0.80)	300 (0.70)
1.08	0.79	1.36	400 (0.90)	250 (0.82)	175 (0.73)
1.07	1.34	0.79	400 (0.90)	250 (0.81)	175 (0.72)
1.04	0.80	1.27	350 (0.90)	250 (0.83)	150 (0.71)
0.99	0.79	1.19	350 (0.94)	200 (0.82)	150 (0.75)
0.97	1.15	0.79	300 (0.91)	200 (0.81)	125 (0.71)
0.95	1.08	0.82	300 (0.90)	175 (0.82)	125 (0.74)
0.95	0.88	1.01	300 (0.92)	175 (0.82)	125 (0.74)
0.94	1.03	0.85	250 (0.90)	175 (0.83)	100 (0.71)
0.92	1.02	0.83	300 (0.93)	175 (0.82)	125 (0.73)
0.92	0.83	1.01	300 (0.92)	175 (0.82)	125 (0.71)
0.90	0.86	0.95	250 (0.90)	150 (0.81)	125 (0.77)
0.90	0.83	0.97	300 (0.93)	150 (0.80)	125 (0.75)
0.89	0.90	0.88	250 (0.91)	150 (0.80)	100 (0.71)
0.89	0.91	0.87	300 (0.93)	150 (0.80)	100 (0.72)
0.88	0.84	0.91	250 (0.92)	150 (0.81)	125 (0.77)
0.87	0.88	0.87	250 (0.91)	150 (0.80)	100 (0.72)

* Confidence level accuracy

** The numbers in parenthesis show the actual confidence level by the sample size.

In order to further investigate the potential reason behind the later observation (the relationship between $CV_{Avg.}$ and the required sample size), we have rewritten Equation 36 as the following alternative:

$$C = \frac{\frac{\bar{N}_{obs,KAB}}{\bar{N}_{obs,C}}}{\frac{\bar{N}_{pre,KAB}}{\bar{N}_{pre,C}}} \tag{52}$$

Let us assume that the only source of the C-factor dispersion is the observed KAB/C ratio (the dispersion in predicted ratio was assumed to be relatively negligible to the observed ratio). Next, it can be argued that a better estimate for both $\bar{N}_{obs,KAB}$ and $\bar{N}_{obs,C}$ would result in a better estimate for the ratio as well. As crash data become more dispersed around either $\bar{N}_{obs,KAB}$ or $\bar{N}_{obs,C}$, a larger sample size is required to have accurate estimates for $\bar{N}_{obs,KAB}$ or $\bar{N}_{obs,C}$. On the other hand, once data are less dispersed around both $\bar{N}_{obs,KAB}$ and $\bar{N}_{obs,C}$, a smaller sample can be needed to attain reliable estimates. Consequently, if either CV_{KAB} or CV_C is large, we need a larger sample size; conversely, if both are small, we need a smaller sample size. The average of CV_{KAB} and CV_C was used to account both effects simultaneously.

SAMPLE SIZE GUIDELINES

Given the simulation results, Table 16 provides the sample size guidelines to fulfill a confidence level of 90 percent, 80 percent, and 70 percent for a range of $CV_{Avg.}$. Similar to the SPFs sample size guidelines, these guidelines can also be used for any SDF models, either an intersection or a segment model. In order to use the sample size guidelines, the agency needs to secure or estimate the KAB and C crash mean and standard deviation for sites with at least 1 KABC crash separately. Once done, the average of the CV_{KAB} and CV_C to find the $CV_{Avg.}$ is calculated. Then, given the $CV_{Avg.}$, a sample size that fulfills the desired level of accuracy can be selected from Table 16. The sample size guidelines show the minimum sample size needed to meet a given level of accuracy. In cases when more data are readily available, the agency is advised to use the full dataset. On the other hand, for cases when the agency cannot meet the minimum sample size guidelines, the agency is advised to consider developing a state-specific SDF. The recommended sample size is smaller than the guidelines proposed for developing new SDFs (Ye and Lord, 2014).

Table 16. SDF Sample Size Guidelines.

$CV_{Avg.}$	Confidence Level		
	90%	**80%**	**70%**
2.0†	1,400	900	650
1.8	1,200	800	600
1.6	1,000	650	550
1.4	800	450	350
1.2	600	350	200
1.0	400	200	150
0.8	200	125	100
≤ 0.6	100	75	50

†Values larger than 2.0 are would require even larger sample sizes, which may not be practical to collect. Hence, they are not presented in this table. Agencies may need to revise the datasets for recalibrating the models.

CHAPTER 7: CONCLUSIONS AND RECOMMENDATIONS

This chapter summarizes the results and analyses from earlier chapters to formulate conclusions and guidelines.

CONCLUSIONS

The research was conducted to (1) identify factors that influence the selection of the sample size for the SPFs calibration (or recalibration), (2) determine how frequently or when an agency should update their calibration factors, (3) determine whether or not having region-specific C-factors are justified and when they are needed, and (4) identify factors that influence the selection of the sample size for the SDFs calibration (or recalibration). The study objectives were accomplished using simulated and observed data.

The calibration of predictive models is a time-consuming task in addition to problems associated with the collection, readiness, and completeness of the data. Independent of the level of crash data history for different types of facilities, the HSM still recommends using between 30 and 50 sites with at least 100 crashes for the calibration of SPFs. It was also reported in the literature that not only the HSM one-size-fits all recommendation is inappropriate but is also insufficient to acquire the desirable accuracy in most cases. In this research, an extensive simulation analysis was performed for a range of predicted means, calibration factors, and inverse dispersion parameters. As the mean of observed number of crashes and the given inverse dispersion parameter increases (a measure of variation in the data), a smaller sample size is needed to achieve desirable levels of accuracies. These observations were used to propose sample size guidelines based on the ratio of the standard deviation to the mean of the crash data (i.e., the CV of the crash data that will be used for calibrating the model). The study results showed that as the CV increases, the required sample size to attain a certain confidence level increases as well. The proposed sample size guidelines can be used for all facility types and prediction models either for intersections or segments (independent of the length of the sections).

Since the recalibration of prediction models is a time-consuming and expensive task, the agency may need to know when or how often SPFs should be recalibrated. The HSM recommends deriving calibration factors at least every two to three years. However, there is no research to support this recommendation. In this research, we have developed guidelines about when the predictive models are advised to be recalibrated. The guidelines were successfully validated by applying them to two observed datasets in Texas. The guidelines are straightforward and can be used for any predictive model, irrespective of the type of facility, such as an intersection model or a roadway segment model. The only variables that are needed to use the guidelines are: (1) the total number of crashes, (2) the average ADT or AADT (or the average traffic flow on major and minor roads), and (3) the total segment length (or number of intersections for intersection models). The agency is required to secure these variables periodically and follow the steps stated in the guidelines.

Currently, there exists no clear guidance on whether or not region-specific calibration factors are needed in a particular jurisdiction. Using two observed crash datasets, one collected in Texas and the other one in Michigan, we have developed region-specific calibration factors and compared them between regions. The derived factors revealed that the region-specific factors were advised for both datasets, at least for some regions. Guidelines similar to the one for the recalibration were proposed to identify when developing a region-specific factor is justified. The region-specific guidelines are also based on the general characteristics of data at hand. It requires (1) the total number of crashes, (2) the average ADT or AADT (or the average traffic flow on major and minor roads), and (3) the total segment length (or number of intersections).

Similar to the SPFs, the HSM SDFs were also fitted and validated with data obtained from a few selected numbers of states. Therefore, calibration is needed when they are applied to a new jurisdiction. The HSM provided a method to calculate the calibration factor for SDFs. In this study, the calibration factor that was proposed to calibrate SDFs in HSM Chapters 18 and 19 was investigated and validated using simulation. Then the required sample size to derive a reliable SDF C-factor was investigated for a wide range of scenarios using simulation. Based on the simulation results, sample size guidelines were provided based on the average CVs of the KAB and C crashes. The proposed sample size guidelines can be used for all facility types and prediction models either for intersections or segments (independent of the length of the sections).

RECOMMENDATIONS

This section summarizes the most important guidelines developed in this research.

What Is the Required Sample Size to Calibrate the SPF Models?

The sample size guidelines show the minimum sample size needed to meet a given level of accuracy. In cases when more data are readily available, the agency is advised to use the full dataset. When the agency cannot meet the minimum sample size guidelines, it is advised to develop a state-specific crash prediction model. The following steps are used in finding the required sample size:

Step 1. Calculate the sample mean of observed crashes ($\bar{\mathrm{N}}_{\mathrm{obs}}$).

Step 2. Calculate the sample standard deviation of observed crashes ($sd(N_{obs})$).

Step 3. Estimate the coefficient of variation $\left(CV = \frac{sd(N_{obs})}{\bar{\mathrm{N}}_{\mathrm{obs}}}\right)$.

Step 4. Using Table 4, find the required sample size based on the CV.

The recommended minimum sample size was presented in Table 4, which is reproduced below:

CV.	Confidence Level		
	90%	**80%**	**70%**
3.0†	1,500	1,100	700
2.8	1,400	1,000	650
2.6	1,300	900	600
2.4	1,200	800	550
2.2	1,000	650	450
2.0	900	550	400
1.8	750	450	300
1.6	600	350	250
1.4	450	300	200
1.2	300	200	150
1.0	200	125	75
0.8	100	75	50
≤ 0.6	50	30	30

†Values larger than 3.0 are would require even larger sample sizes, which may not be practical to collect, so they are not presented in this table. Agencies may need to revise the datasets for recalibrating the models.

When Recalibration Is Needed?

In order to identify when calibration is needed, the agency should secure these three variables periodically: (1) total number of crashes, (2) the average ADT or AADT (or the average traffic flow on major and minor streets in case of intersections), and (3) total segment length (or the total number of intersections). The following steps are used for segment models to decide when to calibrate (similar guidelines for intersection models can be found in Chapter 4).

Step 1. Find the total number of crashes (N_{obs}^{T}) and the total segment length (L^{T}) on the network facility.

Step 2. Find the average ADT ($\overline{ADT}$) (or AADT) on the facility.

Note that if the average ADT is not available to the agency on all sites, it is advised to randomly collect ADT for a limited number of sites that provide the overall representation of the network to find the mean value of the ADT.

Step 3. Consider the base SPF model (i.e., the model without CMFs) from the HSM. Let b_0 and b_1 denote the intercept and the coefficient of ADT, respectively. Estimate the approximate average predicted number of crashes ($\widetilde{N}_{pre}$) using the following functional form:

$$\widetilde{N}_{pre} = e^{b_0+b_1\times\ln(\overline{ADT})}$$

Step 4. Find the $\tilde{C}$ using the following equation:

$$\tilde{C} = \frac{N_{obs}^T}{\widetilde{N}_{pre}\times L^T}$$

Step 5. Find the variable $\tilde{e}$ as follows:

$$\tilde{e} = \frac{|\tilde{C}-\tilde{C}_{REF}|}{\tilde{C}_{REF}} \times 100$$

where $\tilde{C}_{REF}$ denote the $\tilde{C}$ that was calculated in the reference year. The reference year is the latest or most recent year that the model was calibrated.

Step 6. If $\tilde{e} > 10\%$, the model needs to be recalibrated; calibrate the model and set the current $\tilde{C}$ as the new $\tilde{C}_{REF}$. Otherwise, keep the current $\tilde{C}_{REF}$ and use the calibration factor that was estimated in the reference year.

When Region-Specific Calibration Factors Are Needed?

To determine the need of region-specific calibration, the agency needs to secure (1) the total number of crashes, (2) the average ADT or AADT (or the average traffic flow on major and minor streets in case of intersections), and (3) total segment length for the region of interest and whole state (or the total number of intersections). The following steps are used for segment models to decide the need of region-specific calibration (similar guidelines for intersection models can be found in Chapter 5):

Step 1. Find the total number of crashes and the total segment length in the state ($N_{s,obs}^T$ and L_s^T) and in the region of interest ($N_{r,obs}^T$ and L_r^T).

Step 2. Find the average ADT in the state ($\overline{ADT_s}$) and in the region ($\overline{ADT_r}$).

Step 3. Take the base model SPF (i.e., the model without CMFs) from the HSM. Let b_0 and b_1, respectively, denote the intercept and the coefficient of ADT. Find the approximate average predicted number of crashes in the state ($\widetilde{N}_{s,pre}$) and in the region ($\widetilde{N}_{r,pre}$) as:

$$\widetilde{N}_{s,pre} = e^{b_0+b_1\times\ln(\overline{ADT_s})}$$

$$\widetilde{N}_{r,pre} = e^{b_0+b_1\times\ln(\overline{ADT_r})}$$

Step 4. Find the parameter $\tilde{C}$ in the state ($\tilde{C}_s$) and in the region ($\tilde{C}_r$) as follows:

$$\tilde{C}_s = \frac{N^T_{s,obs}}{\tilde{N}_{s,pre} \times L^T_s}$$

$$\tilde{C}_r = \frac{N^T_{r,obs}}{\tilde{N}_{r,pre} \times L^T_r}$$

Step 5. Find $\tilde{e}_r$ as follows:

$$\tilde{e}_r = \frac{|\tilde{C}_r - \tilde{C}_s|}{\tilde{C}_s} \times 100$$

Step 6: If $\tilde{e}_r > 10\%$, calculate and use the region-specific C-factor. Otherwise, use the statewide calibration factor.

What Is the Required Sample Size to Calibrate the SDF Models?

The sample size guidelines show the minimum sample size needed to meet a given level of accuracy (Note: only the sites with at least 1 KABC crash are considered for calibration). In cases when more data are readily available, the agency is advised to use the full dataset. When the agency cannot meet the minimum sample size guidelines, it is advised to develop a state-specific SDF. The following steps are used in finding the required sample size:

Step 1. Calculate the sample mean of observed KAB crashes ($\bar{N}_{obs,KAB}$) and C crashes ($\bar{N}_{obs,C}$)

Step 2. Calculate the sample standard deviation of observed KAB crashes $\left(sd\left(N_{obs,KAB}\right)\right)$ and C crashes $\left(sd\left(N_{obs,C}\right)\right)$.

Step 3. Estimate the coefficient of variation for KAB crashes $\left(CV_{KAB} = \frac{sd(N_{obs,KAB})}{\bar{N}_{obs,KAB}}\right)$ and C crashes $\left(CV_C = \frac{sd(N_{obs,C})}{\bar{N}_{obs,C}}\right)$.

Step 4. Calculate the average CV $\left(CV_{Avg.} = \frac{CV_{KAB} + CV_C}{2}\right)$.

Step 5. Using Table 16, find the required sample size based on the $CV_{Avg.}$.

The recommended minimize sample size was presented in Table 16, which is reproduced below:

$CV_{Avg.}$	Confidence Level		
	90%	**80%**	**70%**
2.0†	1,400	900	650
1.8	1,200	800	600
1.6	1,000	650	550
1.4	800	450	350
1.2	600	350	200
1.0	400	200	150
0.8	200	125	100
≤ 0.6	100	75	50

†Values larger than 2.0 are would require even larger sample sizes, which may not be practical to collect. Hence, they are not presented in this table. Agencies may need to revise the datasets for recalibrating the models.

REFERENCES

Alluri, P., D. Saha, and A. Gan (2016). “Minimum Sample Sizes for Estimating Reliable Highway Safety Manual (HSM) Calibration Factors.” *Journal of Transportation Safety & Security*, *8*(1), 56-74.

American Association of State Highways and Transportation Officials (AASHTO) (2010), *Highway Safety Manual (1st Edition).* Transportation Research Board of the National Academies, Washington, D.C.

Bahar, G. (2014). *User’s Guide to Develop Highway Safety Manual Safety Performance Function Calibration Factors.* National Cooperative Highway Research Program, Transportation Research Board, Washington, D.C.

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

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

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

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

Kim, D. K., and C. Lee (2013). “Safety Performance Functions Reflecting Categorical Impact of Exposure Variables for Freeways.” *Transportation Research Record: Journal of the Transportation Research Board*, 2398(1), 67–74.

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

Lord, D., and J.A. Bonneson (2005). “Calibration of Predictive Models for Estimating the Safety of Ramp Design Configurations.” *Transportation Research Record*, pp. 88–95.

Lord, D., S.D. Guikema, and S.R. Geedipally (2008). “Application of the Conway–Maxwell–Poisson Generalized Linear Model for Analyzing Motor Vehicle Crashes.” A*ccident Analysis & Prevention*, 40 (3), pp. 1123–1134.

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

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*, pp.75–82.

Miaou, S. P., and D. Lord (2003). “Modeling Traffic Crash-Flow Relationships for Intersections: Dispersion Parameter, Functional Form, and Bayes versus Empirical Bayes Methods.” *Transportation Research Record 1840: Journal of the Transportation Research Board*, pp. 31–40.

Miranda-Moreno, L. F., and L. Fu (2007). “Traffic Safety Study: Empirical Bayes or Full Bayes?” Presented at the *86th Annual Meeting of the Transportation Research Board* (No. 07-1680), Washington, D.C.

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

Trieu, V., S. Park, and J. McFadden (2014). “Use of Monte Carlo Simulation for a Sensitivity Analysis of Highway Safety Manual Calibration Factors.” *Transportation Research Record 2435: Journal of the Transportation Research Board*, pp. 1–10.

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

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

Ye, F., and D. Lord (2014). “Comparing Three Commonly Used Crash Severity Models on Sample Size Requirements: Multinomial Logit, Ordered Probit and Mixed Logit Models.” *Analytic Methods in Accident Research*, Vol. 1, pp. 72–85.

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

APPENDIX A: SIMULATION RESULTS FOR AN INTERSECTION MODEL

The same simulation protocol described in Chapter 3 was also applied to an intersection predictive model. The model was fitted and validated for 4-legged signalized intersections using the Texas and California crash data. The data were collected for an on-going national research project. The model is shown in Equation A.1:

$$N_{pre} = b_0 \times {F_1}^{0.175} \times {F_2}^{0.325} \qquad \text{(A.1)}$$

The variables F_1 and F_2, respectively, denote the traffic flow on the major and minor streets. The traffic flow on the major streets was simulated from a lognormal distribution with a mean and standard deviation of 45,003 veh/day and 17,066 veh/day, respectively. The traffic flows on the minor streets were generated from a lognormal distribution with a mean and standard deviation of 13,931 veh/day and 11,454 veh/day, respectively. Both flow characteristics are based on the characteristics of the data that were used to develop the original model. For each simulation run, the intercept (b_0) was manipulated until the mean of the predicted number of crashes was equal to value of the desired scenario. The simulation results for a range of scenarios (the same scenarios that were described in Chapter 3) are shown in Table A.1. As described in this table, similar and compatible sample size requirements to the ones proposed for the segment model are shown. The small variations are due to the simulation randomness.

Table A.1 Simulation Results for the Intersection Model*.

Predicted Crash Mean	Calibration Factor	Observed Crash Mean**	Inverse Dispersion Parameter		
			0.5	1	5
0.5	0.5	0.25	1300 (90%)*** 850 (80%) 600 (70%)	1100 (90%) 700 (80%) 550 (70%)	900 (90%) 700 (80%) 400 (70%)
	1.0	0.50	900 (90%) 600 (80%) 400 (70%)	700 (90%) 450 (80%) 325 (70%)	550 (90%) 350 (80%) 225 (70%)
	1.5	0.75	800 (90%) 500 (80%) 325 (70%)	550 (90%) 375 (80%) 250 (70%)	400 (90%) 250 (80%) 175 (70%)
	2.0	1.00	800 (90%) 450 (80%) 325 (70%)	550 (90%) 325 (80%) 225 (70%)	325 (90%) 200 (80%) 125 (70%)
2.5	0.5	1.25	700 (90%) 450 (80%) 300 (70%)	450 (90%) 300 (80%) 200 (70%)	300 (90%) 175 (80%) 125 (70%)
	1.0	2.50	700 (90%) 375 (80%) 300 (70%)	375 (90%) 225 (80%) 175 (70%)	175 (90%) 125 (80%) 75 (70%)
	1.5	3.75	600 (90%) 375 (80%) 250 (70%)	325 (90%) 225 (80%) 150 (70%)	150 (90%) 100 (80%) 75 (70%)
	2.0	5.00	600 (90%) 375 (80%) 250 (70%)	325 (90%) 200 (80%) 150 (70%)	125 (90%) 75 (80%) 50 (70%)
5.0	0.5	2.50	700 (90%) 375 (80%) 300 (70%)	375 (90%) 225 (80%) 175 (70%)	175 (90%) 125 (80%) 75 (70%)
	1.0	5.00	600 (90%) 375 (80%) 250 (70%)	325 (90%) 200 (80%) 150 (70%)	125 (90%) 75 (80%) 50 (70%)
	1.5	7.50	550 (90%) 375 (80%) 250 (70%)	300 (90%) 200 (80%) 125 (70%)	100 (90%) 75 (80%) 50 (70%)
	2.0	10.00	550 (90%) 325 (80%) 225 (70%)	300 (90%) 200 (80%) 150 (70%)	100 (90%) 75 (80%) 50 (70%)

* The small difference in sample size requirement in this table and Tables 2 in Chapter 3 is due to the randomness in simulation.

**The observed crash mean might be slightly different for different runs of simulations due to randomness. However, this table shows only the rounded values.

***The number in parenthesis shows the confidence level for the sample size.

APPENDIX B: DERIVATION OF THE SAMPLE SIZE REQUIREMENT

First, let us define key variables and parameters. Let C be the calibration factor. Let $N_{i\,,obs}$ and $N_{i\,,pre}$, respectively, denote the observed and predicted number of crashes at site i. Let $\bar{N}_{obs}$ and $\bar{N}_{pre}$ denote the true observed and predicted mean of crashes, respectively. In addition, let φ be the inverse dispersion parameter of the NB distribution or model.

The true calibration factor can be calculated as follows (with a significantly large sample size):

$$C_{true} = \frac{\sum_{i=1}^{n\to\infty} N_{i\,,obs}}{\sum_{i=1}^{n\to\infty} N_{i\,,pre}} = \frac{\bar{N}_{obs}}{\bar{N}_{pre}} \tag{B.1}$$

The calibration factor, however, can be estimated as follows with a sufficient number of sites (n):

$$\hat{C} = \frac{\sum_{i=1}^{n} N_{i\,,obs}}{\sum_{i=1}^{n} N_{i\,,pre}} \tag{B.2}$$

Let us assume the observed number of crashes is the only source of dispersion in $\hat{C}$. Therefore, the variance of the C-factor ($\hat{C}$) can be derived as follows:

$$Var(\hat{C}) = Var\left(\frac{\sum_{i=1}^{n} N_{i\,,obs}}{\sum_{i=1}^{n} N_{i\,,pre}}\right) = \frac{Var(\sum_{i=1}^{n} N_{i\,,obs})}{(\sum_{i=1}^{n} N_{i\,,pre})^2} = \frac{\sum_{i=1}^{n}\left(N_{i,obs}+\frac{N_{i,obs}^2}{\varphi}\right)}{(\sum_{i=1}^{n} N_{i\,,pre})^2} \tag{B.3}$$

Since the variation of the predicted number of crashes were assumed to be negligible compared to the observed number of crashes, we have:

$$Var(\hat{C}) = \frac{\sum_{i=1}^{n}\left(N_{i,obs}+\frac{N_{i,obs}^2}{\varphi}\right)}{\bar{N}_{pre}^2 \times n^2} \tag{B.4}$$

The ultimate goal is to minimize the variation of $\hat{C}$ around its true value. Therefore, in order to achieve the goal, one can minimize the CV (the ratio of the standard deviation to the mean) of $\hat{C}$. Hence, the following function can be minimized:

$$CV(\hat{C}) = \frac{s.d(\hat{C})}{C_{true}} = \frac{\sqrt{\frac{\sum_{i=1}^{n}\left(N_{i,obs}+\frac{N_{i,obs}^2}{\varphi}\right)}{\bar{N}_{pre}^2 \times n^2}}}{C_{true}} = \frac{\frac{\sqrt{\sum_{i=1}^{n}\left(N_{i,obs}+\frac{N_{i,obs}^2}{\varphi}\right)}}{\bar{N}_{pre} \times n}}{\frac{\bar{N}_{obs}}{\bar{N}_{pre}}} = \frac{\sqrt{\sum_{i=1}^{n}\left(N_{i,obs}+\frac{N_{i,obs}^2}{\varphi}\right)}}{n \times \bar{N}_{obs}} \tag{B.5}$$

For a sufficiently large n, we have:

$$CV(\hat{C}) = \frac{\sqrt{n \times \bar{N}_{obs} + \frac{1}{\varphi} \sum_{i=1}^{n} N_{i,obs}^2}}{n \times \bar{N}_{obs}} \tag{B.6}$$

Equation (B.6) can be written as:

$$CV(\hat{C}) = \sqrt{\frac{1}{n \times \bar{N}_{obs}} + \frac{1}{\varphi} \times \frac{\sum_{i=1}^{n} N_{i,obs}^2}{n^2 \times (\bar{N}_{obs})^2}} \tag{B.7}$$

Note that the following inequality is always true:

$$\sum_{i=1}^{n} N_{i,obs}^2 \leq n^2 \times (\bar{N}_{obs})^2.$$

Furthermore, as $\bar{N}_{obs}$ increases, the inequality can become even larger.

If either of the variables $\bar{N}_{obs}$, φ or n increases, the coefficient variation of C will be decreased. The analyst goal can then be focused on providing a balance between these three variables to minimize the $CV(\hat{C})$.