**NCHRP 17-93: Updating Safety Performance Functions for Data-Driven Safety Analysis**

**Working White Paper: Recommended Refinements to Calibration Function Guidance**

**Prepared for:**

**National Cooperative Highway Research Program**

**Transportation Research Board**

**The National Academies**

**TRANSPORTATION RESEARCH BOARD OF THE NATIONAL ACADEMIES OF SCIENCES, ENGINEERING AND MEDICINE**

**PRIVILEGED DOCUMENT**

This document, not released for publication, is furnished only for review to members of or participants in the work of NCHRP. This document is to be regard as fully privileged, and dissemination of the information included herein must be approved by NCHRP.

Submitted by:

Raghavan Srinivasan and Taha Saleem

The University of North Carolina at Chapel Hill

Highway Safety Research Center

James Bonneson

Kittelson and Associates

Vikash Gayah and Kristen Kersavage

VHB, Inc.

October 2023

**ACKNOWLEDGMENT OF SPONSORSHIP**

This work was sponsored by the American Association of State Highway and Transportation Officials, in cooperation with the Federal Highway Administration, and was conducted in the National Cooperative Highway Research Program, which is administered by the Transportation Research Board of the National Research Council.

**DISCLAIMER**

This is an uncorrected draft as submitted by the contractor. The opinions and conclusions expressed or implied herein are those of the contractor. They are not necessarily those of the Transportation Research Board, the Academies, or the program sponsors.

TABLE OF CONTENTS

[Introduction 1](#_Toc158130100)

[Topic 7a. Selecting the Form of the Calibration Function 1](#_Toc158130101)

[Background 1](#_Toc158130102)

[Issues 2](#_Toc158130103)

[Research Approach 2](#_Toc158130104)

[Topic 7b. Overdispersion Parameter for Original vs. Calibrated CPM 3](#_Toc158130105)

[Background 3](#_Toc158130106)

[Issues 4](#_Toc158130107)

[Research Approach 4](#_Toc158130108)

[Topic 8. Sample Size, Number of Coefficients, and Model Fit 5](#_Toc158130109)

[Background 5](#_Toc158130110)

[Issues 5](#_Toc158130111)

[Research Approach 5](#_Toc158130112)

[References 6](#_Toc158130113)

[Appendix – Terminology 6](#_Toc158130114)

Introduction

The Subtask 3.5 Working Paper (Prioritized List of Candidate Guidelines for Management of Crash Prediction Models) was submitted to NCHRP for review on July 10, 2020. This paper presented a prioritized list of the topics for which guidance could be developed based on the funding available in Phase I of the project. After discussion with the project panel, it was decided that the topics identified in Table 1 could be addressed in Phase I with the understanding that the associated guidelines would be refined in Task 7.

Table . Prioritized List of Research Topics.

|  |  |  |  |
| --- | --- | --- | --- |
| Subtask Title | Subtask Number | Topic Number | Topic |
| When to calibrate or update | 3.1 | 1 | Calibrate if there is a change in location |
| 2 | Update if there is a change in safety over time |
| How to calibrate or update | 3.2 | 1 | Combined calibration data for similar site types |
| 2 | Reliability of extrapolating calibration factor across crash type and severity categories |
| 3 | Maintain standardized database |
| 4 | Prioritizing input data elements |
| 5 | Minimum site sample size |
| 6 | Calibrated model fit statistics |
| 7 | Selecting the calibration function |
| 8 | Calibrated model fit statistics (calibration function) |

Working Paper 4 (Guidelines Describing How to Calibrate or Update a Crash Prediction Model) was submitted to NCHRP for review in late 2020. It addresses the topics listed in Table 1 for Subtask 3.2. The depth of the research into each topic varied, where the amount of research undertaken for a given topic was based on the available Phase I funding and the topic’s relative priority.

This paper describes proposed refinements to Topics 7 and 8 that could be undertaken in Task 7. These refinements represent additional research intended to provide more depth or breadth to the guidance provided in Working Paper 4. The proposed refinements are identified in following list:

* Topic 7a. Guidance for selecting the functional form of the calibration function.
* Topic 7b. Guidance for when to use the overdispersion parameter for the original model or that for the calibrated model.
* Topic 8. Considerations of minimum sample size, number of coefficients, and model fit.

The remainder of this paper provides an overview of each topic and describes the additional research envisioned for it. Each topic is presented in a separate section.

Topic 7a. Selecting the Form of the Calibration Function

Background

Srinivasan et al. (2016) investigated the use of alternative functional forms to calibrate a crash prediction model (CPM) to a region of interest. The form they preferred is described in Working Paper 4 (i.e., Equation 25). However, they also acknowledged that other calibration functions may fit better, depending on facility types and the region of interest. The functions considered by Srinivasan et al. (2016) raised various terms of the CPM to a power, where the value of the power was determined by regression analysis.

In contrast, Bahar and Hauer (2014; Appendix D) suggest that analysts consider the development of a calibration function that is applied as a multiplicative adjustment to the uncalibrated CPM, much like the calibration factor *C* is used with the HCM Part C CPMs. The function they described includes segment length and annual average daily traffic (AADT) volume as variables that are associated with coefficients determined by regression analysis. Neither Srinivasan et al. nor Bahar and Hauer evaluated the merit of this calibration function using data. Research is needed is to develop guidelines for determining whether the calibration factor or calibration function is most appropriate and, if the latter is chosen, for selecting the most appropriate form for the calibration function.

Issues

There are several issues that should be addressed in the development of calibration function selection guidelines. One issue applies to calibration functions applied to road segment CPMs that include segment length as a multiplicative factor raised to the power 1.0. This type of CPM is included in HSM Part C for project-level evaluations. It is also used in Safety Analyst for segment network screening applications. If the calibration function includes segment length as a factor or otherwise alters the power to which segment length is raised (from 1.0), then the predicted number of crashes in a 2-mile section will *not* be double the predicted number of crashes in a similar 1-mile section. This characteristic can questionably influence the results when comparing alternative designs or identifying sites for safety improvement.

A second issue relates to the use of calibration functions that raise one or more CMFs to a power that is other than 1.0. This approach effectively alters the value of each CMF. To illustrate this issue, consider a treatment associated with a CMF of 0.5. The analyst will expect that the treatment will reduce crashes by 50 percent if applied to the site. However, if by calibration the CMF value is raised to the power of 0.4, the predicted crash frequency will be reduced by only 24 percent (= 100 – 100 × 0.50.4). This outcome is counterintuitive and may not be acceptable to analysts.

A third issue is whether the use of a calibration function should be accompanied by an increase in the minimum site sample size. The guidelines described in Working Paper 4 indicate that at least X sites (i.e., observations) are needed to estimate a single-coefficient calibration function (i.e., the constant *C* as described in HSM Part C). Statistical references indicate that the minimum number of observations should increase with an increase in the number of coefficients estimated from the data. Thus, it is logical that the value of X should increase as a function of the number of coefficients in the calibration function. A consequence of having too few observations (relative to the number of coefficients) is that the calibration function could over-fit the data. When a calibrated CPM over-fits the calibration data, its predicted values are less reliable when the CPM is applied to sites outside of the calibration database.

Research Approach

The objective of this research is to develop guidelines for selecting the most appropriate functional form of the calibration function (including the simplest form, which is the constant *C*). The guidelines should identify all potential application issues that arise in the use of a calibration function; including those issues identified in the previous section. The guidelines should be sensitive to CPM application (i.e., is it for use in safety management applications or for use in project-level analysis). They should consider the increase in site sample size necessitated by the additional coefficients in the function. They should consider the practical implications of requiring the use of regression analysis to estimate the function coefficients (e.g., increase in analyst training requirements related to statistical regression analysis).

The research should include the following tasks:

* Identify a wide range of alternative calibration function forms (including the multiplicative constant *C*). Identify the issues associated with each alternative form. As a minimum, these issues should include those noted previously as well as adherence to logical boundary conditions and intuitiveness of the predicted results.
* Develop guidance that assists analysts in subjectively assessing the various issues associated with each calibration function form and selecting one or two forms that can be estimated and further evaluated.
* Using the results from the research conducted for Topic 8 (described in a subsequent section), develop guidance for determining the minimum site sample size based on the number of estimated coefficients in the calibration function.
* Describe a procedure for using regression analysis to estimate the calibration function coefficients and their standard errors. Include guidelines for interpreting the standard errors and whether they indicate that the coefficient improves the fit of the calibrated CPM.
* Provide guidelines for using alternative measures for assessing the goodness-of-fit (e.g., AIC, modified R2) provided by the calibrated CPM and for assessing whether the model may be over-fitting the calibration data.
* Demonstrate the aforementioned guidelines in a case study.

Topic 7b. Overdispersion Parameter for Original vs. Calibrated CPM

Background

A CPM is typically estimated using data for a specified region (e.g., state or group of states). For this discussion, this CPM is referred to as the “original CPM” and the associated data are referred to as the “original data.” The CPMs in HSM Part C are considered original CPMs. In many instances, an original CPM is used to evaluate sites in a region that is not represented in the original data. In this situation, the HSM recommends that the analyst calibrate the original CPM using data representing the region of interest. Thus, the analyst assembles “calibration data” from the region of interest and uses it to produce a “calibrated CPM.”

The premise behind CPM calibration is that the original CPM is likely to provide biased predictions when it is used to evaluate sites in a region not represented in the original data. In contrast, a calibrated CPM should ensure unbiased predictions for the subject region.

Bahar and Hauer (2014) developed a procedure for estimating the minimum sample size needed for the calibration data when using the HSM CPM calibration procedure. Their procedure incorporates the overdispersion parameter *k* associated with the original data and original CPM (i.e., *koriginal*) Their procedure assumes that *koriginal* also describes the site-to-site variation in average crash frequency for the region to which the CPM is calibrated. Sawalha and Sayed (2006) have noted that this assumption may be weak in some instances. They recommend computing the overdispersion parameter associated with the calibration data and calibrated CPM (i.e., *kcalibrated*).

The procedure described in Working Paper 4 does not incorporate *koriginal* in the calculation of minimum site sample size for the calibration data. This approach was taken [instead of the procedure by Bahar and Hauer (2014)] because it is not known whether *koriginal* adequately describes the calibration data. Rather, the procedure in Working Paper 4 bases the sample size calculation on the computed variability in the calibration data (but it does not describe a procedure for computing *kcalibrated*).

The HSM is silent on this topic. Current practice is to use *koriginal* with a calibrated CPM to compute the expected average crash frequency for a site in the region of interest.

Issues

The guidelines documented in Working Paper 4 do not identify a recommended overdispersion parameter because it is unknown whether *kcalibrated* is more (or less) reliable than *koriginal*. The most appropriate overdispersion parameter to use with a calibrated CPM is that which reliably describes the site-to-site variation in average crash frequency for the region to which the CPM is applied. Logically, the overdispersion parameter obtained from the calibration data *kcalibrated* describes this variation. However, there is no research available to indicate if, or when, it is more reliable than *koriginal*. Notably, *kcalibrated* may have a large standard error given the inherently small sample sizes used for calibration data (relative to sample sizes used for original data). Moreover, it is not known whether the reliability of *kcalibrated* is changed when a calibration function is used.

Research Approach

The objective of this research is to develop guidelines for determining whether to use the overdispersion parameter from the original model *koriginal* or that from the calibrated model *kcalibrated*. If it is determined that latter parameter is appropriate, the guidelines should consider the practical implications of requiring analysts to use regression analysis to estimate *kcalibrated* (e.g., increase in analyst training requirements related to statistical regression analysis).

The research should include the following tasks:

* Acquire data for one site type (e.g., rural two-lane highway segments) from several states. The number of sites in this database will need to be of the size used for model estimation (not calibration). For a given site type and using a common model form, find the best-fit model to each database and its associated overdispersion parameter. Determine whether this parameter varies from state to state by a statistically significant amount.
* Repeat the previous task for one or more additional site types. Confirm whether the overdispersion parameter varies from state-to-state and whether the amount of variation is influenced by site type or database sample size. If the state-to-state variation is practically significant, conclude that *kcalibrated* should be used with a calibrated CPM and continue with the next task. If it is not, then conclude that it is acceptable to use *koriginal* with a calibrated CPM.
* Conduct a sensitivity analysis to determine the implications of uncertainty in the estimate of *kcalibrated* when calculating expected average crash frequency. Determine whether this uncertainty introduces bias in the calculated crash frequency. Determine the extent to which this uncertainty increases the uncertainty of the calculated crash frequency. Determine if the findings suggest a threshold level of uncertainty (or coefficient of variation) that, if exceeded, results in an “unreliable” estimate of *kcalibrated*.
* Develop a series of simulated databases to quantify the effect of the following independent variables on the standard error of the estimated *kcalibrated*: site sample size, average crash frequency per site, number of CPM variables, and true overdispersion parameter value *ktrue*. The simulated databases should reflect a factorial design for a range of representative values for each independent variable.
* Estimate a common model form for all databases. Determine if the relationship between the independent variables and the standard error of the estimated *kcalibrated* can be derived. If such a relationship can be defined, prepare guidelines describing their use to determine the minimum sample size to produce a reliable estimate of *kcalibrated*. If such a relationship cannot be defined, determine if an iterative procedure can be developed for assessing the sample size adequacy of a calibration database and whether more sites need to be added to this database. An example iterative procedure of this type is described in Appendix B of the report by Bahar and Hauer (2014).
* Describe a procedure for using regression analysis to estimate *kcalibrated* and its standard error.
* Based on the findings from previous tasks, prepare guidelines that can be used to determine whether and when to use the overdispersion parameter from the original model *koriginal* or that from the calibrated model *kcalibrated*.
* Demonstrate the aforementioned guidelines in a case study.

Topic 8. Sample Size, Number of Coefficients, and Model Fit

Background

As described in the section associated with Topic 7a, one issue associated with the use of a calibration function is whether the calibration database has sufficient sample size to support the additional coefficients associated with the function. One rule-of-thumb in the statistical literature is that the sample size should include at least 30 sites; with this number increased by 10 sites for each additional coefficient that is estimated (beyond the first coefficient).

Srinivasan et al. (2016) described several alternative functional forms for the calibration function. The number of coefficients varied from form to form. The coefficients ranged in number from 2 to 4 depending on the form used.

Issues

A consequence of having too few sites in the calibration database (relative to the number of estimated coefficients) is that the calibration function may over-fit the data. If the calibrated CPM is over-fit to the data, its predictions may be less reliable when applied to sites not represented in the calibration database. The presence of multicollinearity among the variables associated with two or more coefficients can also affect the calibrated CPM’s goodness-of-fit and whether it is over-fit to the data.

Research Approach

The objective of this research is to develop guidelines for determining whether a calibration function over-fits the CPM to the calibration data. The guidelines should consider the number of coefficients in the calibration function. It should describe statistical tests and the use of graphical tools (e.g., CURE plot) to facilitate the detection of over-fit CPMs.

The research should include the following tasks:

* Identify statistical techniques (e.g., cross validation), statistical tests (e.g., likelihood ratio test [Venkataraman et al. 2016], modified *R2* [Fridstrom et al., 1995]), and graphical tools (e.g., CURE plot [Hauer, 2015]) that can be used to detect over-fit CPMs. Document how they are used for this purpose.
* Acquire a real-world database previously used to estimate a CPM for a specific site type and which includes data from two or more states. The data for one state will be called the “original” data. It will be used to estimate a CPM (this model will be called the “original CPM”). The data for the other state (or states) will be called the “calibration” data. It will be used to calibrate the original CPM. Prior to calibrating the original CPM, the calibration data should be disaggregated into subsets of 30, 50, 100, and 200 sites. Resampling can be an option, if needed, to produce the desired number of subset databases.
* Estimate the original CPM coefficients using the original database.
* Specify the calibration functional forms to be evaluated. One form should have one coefficient (i.e., a multiplicative constant *C*), the second form should have two coefficients, the third form should have three coefficients, etc.
* Using a calibration function with one coefficient and the original CPM, estimate the function coefficients for a specific calibration functional form using one calibration data subset. Repeat this estimation for each calibration data subset.
* Repeat the previous task for each of the specified calibration functional forms.
* Plot the GOF measures as a function of the number of the number of estimated coefficients. Use the results to determine the number of coefficients associated with an over-fit CPM.
* Repeat the previous tasks for a different site type and real-world database.
* Assemble a database of the results from the previous tasks (one observation for each subset database). The data base should include the following variables: site type, site sample size of the calibration data subset, number of coefficients, average crash frequency per site, overdispersion parameter, various GOF measures, and an indicator variable for over-fit status (1 if over-fit, 0 if not over-fit).
* Determine if a relationship can be derived to estimate the minimum site sample size (i.e., smallest number of observations that provide a statistically valid fit but do not exhibit over-fit characteristics). If such a relationship can be defined, prepare guidelines describing their use to determine the minimum sample size to produce a reliable calibration function with *n* coefficients. If such a relationship cannot be defined, determine if an iterative procedure can be developed for assessing the sample size adequacy of a calibration database and whether more sites need to be added to this database when the calibration function has *n* coefficients.
* Based on the findings from the previous tasks, develop guidelines for determining whether a calibration function over-fits the CPM to the calibration data.
* Demonstrate the aforementioned guidelines in a case study.

References

Bahar, G., and E. Hauer. (2014). *User’s Guide to Develop Highway Safety Manual Safety Performance Function Calibration Factors. Final Report*. NCHRP Project HR 20-7(332). Transportation Research Board, Washington, D.C.

Fridstrom, L., Ifver, J., Ingebrigtsen, S., Kulmala, R., and Thomsen, L. K. (1995). “Measuring the contribution of randomness, exposure, weather, and daylight to the variation in road accident counts.” *Accident Analysis and Prevention*, Vol. 27, pp. 1-20.

Hauer, E. (2015). *The Art of Regression Modeling in Road Safety*. Springer International, New York.

Venkataraman, N., V. Shankar, J. Blum, B. Hariharan, and J. Hong. (2016). “Transferability Analysis of Heterogeneous Overdispersion Parameter Negative Binomial Safety Performance Functions: A Case Study from California.” Paper 16-5969, Presented at the 95th Annual meeting of the Transportation Research Board, Washington, D.C.

Sawalha, Z., and T. Sayed, “Transferability of Accident Prediction Models.” *Safety Science*, Vol. 44, 2006, pp. 209-219.

Srinivasan, R., Colety, M., Bahar, G., Crowther, B., and Farmen, M. (2016). “Estimation of Calibration Functions for Predicting Crashes on Rural Two-Lane Roads in Arizona.” *Transportation Research Record 2583*. Transportation Research Board, Washington, D.C., pp. 17–24.

Appendix – Terminology

**Predictive Method.** Each chapter in HSM Part C describes the predictive method for a type of highway facility—the process to quantify the safety performance of a road. Each predictive method includes several crash prediction models. Models are developed for specific facility types (e.g., urban three-leg signalized intersection).

**Crash Prediction Model (CPM).** For safety management applications (e.g., HSM Part B, SafetyAnalyst), a crash prediction model consists of a safety performance function (SPF). For design applications (e.g., HSM Part C), a crash prediction model consists of an SPF and some combination of SPF adjustment factors (AFs), crash modification factors (CMFs), calibration factor, crash type distribution proportions, and crash severity distribution proportions. Some HSM Part C prediction models include a severity distribution function (SDF) instead of severity distribution proportions. The SDF is used to predict the crash severity distribution proportions as a function of site characteristics.

**Predictive Model Equation.** The SPF, CMFs, and calibration factor components of a CPM. These components are used to compute the predicted average crash frequency for a specified combination of crash type and severity category.

**Statistical Model.** A statistical model represents an empirically-derived predictive relationship that is based on statistical analysis of data. The following are statistical models: SPF, AF, CMF, and SDF.

**Model Re-estimation.** When an existing statistical model is re-estimated, its empirical coefficients are replaced by new estimates that are quantified through statistical analysis using (1) data that is different from that for which it was originally estimated, or (2) statistical assumptions or analysis techniques that are different from those used for initial model estimation. In some instances, knowledge gained from recent research can suggest the need for a new model variable—which may also trigger the need for re-estimation.

**Model Calibration.** Model calibration is a process that produces an adjustment factor (or factors) to be used with a CPM to account for spatial differences between the location used for model estimation and the location the model is being used to evaluate. Models are developed for one or more jurisdictions based on a sample data set. Model calibration allows analysts to transfer models between jurisdictions to provide more reliable estimates of crash frequency. Calibration accounts for differences in safety between regions not addressed through model variables. If the CPM being calibrated was originally estimated using data for a different time period (and location) than that being evaluated, then the calibration process also updates the CPM (see Model Updating).

**Model Updating.** Model updating is a process of maintaining CPMs through re-estimation, calibration, and other methods for the purpose of accounting for temporal changes between the time period used to calibrate the model and the time period the model is being used to evaluate. The objective of the updating process is to increase prediction reliability by better reflecting conditions for the time period of interest at a given location. When the model is updated through calibration, the process is sometimes called “recalibration” in the literature.

The HSM advises analysts to calibrate the HSM CPMs to the region of interest before using them for the evaluation of sites in that region. It also advises them to update the calibration factor every two or three years.

**Homogeneous Segment.** A homogeneous segment is defined to be a portion of roadway whose geometric design elements and traffic characteristics are very similar along the road’s length. The elements and characteristics used to assess segment homogeneity are those that are known to have some influence on safety.