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

**Working White Paper: Describing How to Calibrate or Update a Crash Prediction Model**

**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](#_Toc158130115)

[Guidelines for Calibrating or Updating a CPM 1](#_Toc158130116)

[Guideline Application 2](#_Toc158130117)

[Typical Case 2](#_Toc158130118)

[Special Case 2](#_Toc158130119)

[Project Planning Procedure 2](#_Toc158130120)

[Scoping Process 2](#_Toc158130121)

[Minimum Sample Size Assessment Process 4](#_Toc158130122)

[Site Identification Process 4](#_Toc158130123)

[Data Assembly Process 5](#_Toc158130124)

[Calibration Procedure 7](#_Toc158130125)

[Step 1 – Compute Site Sample Statistics and Assess Sample Adequacy 8](#_Toc158130126)

[Step 2 – Compute the Predicted Average Crash Frequency 10](#_Toc158130127)

[Step 3 – Compute the Calibration Factor 10](#_Toc158130128)

[Step 4 – Outlier Detection and Resolution 12](#_Toc158130129)

[Step 5 – Assess Model Fit Based on the Calibration Factor 15](#_Toc158130130)

[Step 6 – If Applicable, Compute the Calibration Function 17](#_Toc158130131)

[References 18](#_Toc158130132)

[Appendix – Terminology 18](#_Toc158130133)

Introduction

This paper provides guidelines for calibrating or updating the predictive equation in a crash prediction model (CPM). The guidelines are applicable to the CPMs documented in Part C of the *Highway Safety Manual* (HSM) (HSM, 2010). They are also applicable to CPMs developed for a specific jurisdiction (provided that these jurisdiction-specific CPMs are developed using techniques similar to those used to develop the HSM CPMs). These guidelines are intended to provide a reasonable balance between the effort required to implement the calibration process and the predictive reliability of the calibrated CPM.

A CPM consists of a *predictive model equation*, a crash type distribution, and a crash severity distribution. The predictive model equation includes a safety performance function (SPF), crash modification factors (CMFs; also known as SPF adjustment factors), and a calibration factor (or factors). Additional terms related to CPMs are defined in the Appendix.

As noted previously, the guidelines described herein are intended to support the cost-efficient calibration and updating of predictive model equations. Hence, consideration was given to the inclusion of techniques having a lower implementation “cost” as long as they were shown to be capable of providing reliable results.

The guidelines generally follow the calibration guidance provided in the following three documents:

* *Highway Safety Manual* (HSM, 2010)
* *User’s Guide to Develop Highway Safety Manual Safety Performance Function Calibration Factors* (Bahar and Hauer, 2014)
* *Improved Prediction Models for Crash Types and Crash Severities* (Ivan et al., 2018).

The predictive model equations provided in the HSM include a single calibration factor. However, research has found that in some situations, more reliable predictions can be obtained when a calibration *function* is used with the CPM (e.g., Srinivasan et al., 2016). Henceforth in this paper, the phrase “calibration factor (or factors)” is used to acknowledge that the calibration process may produce one or more calibration factors. The process used to determine whether the calibration process should produce one factor or multiple factors is discussed in a subsequent section of this paper.

Guidelines for Calibrating or Updating a CPM

This section describes guidelines for calibrating or updating the predictive model equation in a CPM. The guidelines consist of procedures, techniques, and statistical tests.

The calibration process produces a calibration factor (or factors) that can be used with a predictive model equation to obtain reliable estimates of average crash frequency for a specific facility type in a given geographic or administrative region. It is intended to adjust a predictive model equation that was developed with data from one jurisdiction so that it can provide reliable results when used in another jurisdiction. Calibration provides a method to account for the collective set of differences between these jurisdictions that result in differences in the average crash frequency (beyond that which can be explained by model components). This collective set of differences may include climate, topography, driver population, wild animal population, design practice, crash reporting threshold, prevailing weather patterns, enforcement levels, and crash reporting system procedures.

The objective of the guidelines is to provide practitioners with procedures for achieving a cost-efficient balance between model predictive reliability and the level of effort required to calibrate and update the model. The guidelines are focused on the calibration and updating of the predictive model equation. Guidelines for updating the crash type and severity distribution components of a CPM are not addressed.

The guidelines are intended for calibrating or updating predictive model equations used for design evaluation, such as those documented in Part C of the HSM (HSM, 2010). The guidelines can also be used to calibrate or update predictive models used for network screening. The guidelines are applicable to CPMs that are developed using techniques similar to those used to develop the HSM Part C CPMs.

Guideline Application

The application of these guidelines produces a calibration database and a calibration factor (or factors) for one CPM. Each observation in the database represents the data for one site. The data for a site includes its crash count, traffic volume, and site characteristics that are represented in the subject CPM’s variables. The site characteristics typically describe geometric design elements and traffic control devices.

Typical Case

The calibration process is described herein as consisting of two procedures. The first procedure is the project planning procedure. It is followed by the calibration procedure. Except for the situation described in the next section, each procedure is completed in sequence to calibrate one predictive model equation. The calibration process is repeated for each predictive model equation of interest.

Special Case

In some cases, the agency may desire to calibrate several CPMs that differ only in terms of their target crashes (i.e., each CPM has the same facility type, region, and site type). In this situation, the calibration process can be efficiently undertaken by assembling one calibration database that is inclusive of all the desired crash type and severity categories. The level of efficiency gained with this approach increases with the number of the non-crash data elements that are common to the CPMs of interest.

The project planning procedure should be repeated for each CPM of interest prior to implementing the calibration procedure for any one CPM. This approach will identify a minimum sample size estimate for each CPM of interest. The largest of these estimates will then define the minimum sample size needed for the one calibration database that is then used to calibrate all CPMs.

Project Planning Procedure

This section provides an overview of the preparatory activities that are undertaken at the start of a calibration project. The first subsection provides an overview of the project scoping process. The second subsection describes the minimum sample size needed for calibration. The third subsection provides an overview of the site identification process. The last subsection describes the data assembly process.

Scoping Process

The scoping process entails the specification of the facility type, region, site type, crash type, severity category, and calibration period that describe each CPM for which calibration factors are being developed. At the conclusion of this process, the CPM (or CPMs) of interest are identified.

The presentation to follow is based on the assumption that the agency desires to either (1) calibrate some or all of a set of existing CPMs that were developed using data from other jurisdictions, or (2) update some of all of the CPMs that it has previously developed or calibrated using data from their jurisdiction. In this manner, the discussion to follow guides the agency in selecting the CPMs of interest for calibration (or updating) based on the facility types, regions, site types, crash types, and severity categories of interest.

Facility Type of Interest

The facility type designation is used to assign roadways to categories with distinctly different speed environments, access functions, and design criteria. These characteristics tend to have a significant influence on traveler safety. It is rationalized that CPM reliability is improved when the CPM is developed for a specific facility type.

Facility type descriptors typically include area type (i.e., urban or rural) and road class (e.g., freeway, highway, arterial street). The facility types recognized in Part C of the HSM include: rural, two-lane roads; rural multilane highways; urban and suburban arterial streets; rural and urban freeways; and rural and urban interchange ramps.

The facility type of interest is used to identify the CPMs to be calibrated. Documentation describing the development of each candidate CPM should be consulted to determine whether it is a match to the facility type of interest.

Region of Interest

It is rationalized that CPM reliability is improved when the CPM is calibrated for a region that has the following conditions uniformly represented throughout its borders: climate, topography, driver population, wild animal population, design practice, crash reporting threshold, prevailing weather patterns, enforcement levels, and crash reporting system procedures. For large jurisdictions where these conditions may vary widely from border to border, it may be desirable to identify two or more regions within the jurisdiction such that each region has uniform conditions within their borders. In this manner, the CPM is uniquely calibrated for each region. For example, a state with one region described as a geographic plain and a second region described as mountainous might choose to develop a separate calibration factor for each region.

Site Types of Interest

Each CPM is developed for application to a specific site type. There are two main site type categories: segment and intersection. Within the segment category, a site type can be designated by its area type (i.e., urban or rural), number of through lanes, and functional classification. Within the intersection category, a site type can be designated by its area type, number of legs, traffic control type (e.g., signal), and design configuration (e.g., conventional intersection, roundabout).

There is some overlap in the facility-type and site-type designations such that some site-type categories are predetermined once the facility type of interest is identified. For example, “area type” is both a facility-type category and a site-type category. If the facility type of interest is designated to include only urban facilities, then the site type of interest is also designated as urban.

The site type of interest is used to identify the CPM to be calibrated. Documentation describing the development of the candidate CPM should be consulted to determine whether it is a match to the site type of interest. This documentation should be reviewed to identify the specific characteristics and criteria used to define the site type represented by the candidate CPM. The HSM Part C chapters describe the characteristics and segmentation criteria used to identify each site type category associated with an HSM CPM.

Crash Type and Severity Category of Interest (Target Crashes)

Each CPM is developed to predict a specific crash type and severity category. Some of the crash types represented by CPMs in the HSM include: all crash types combined, single-vehicle crashes, multiple-vehicle crashes, vehicle-pedestrian crashes and vehicle-bicycle crashes. Similarly, several different crash severity categories are represented by CPMs in the HSM. Some of these categories include: all severity categories combined, fatal-and-injury combined (i.e., K, A, B, or C severity designation), and “KAB” combined (i.e., K, A, or B severity designation). If the CPM is developed to predict crashes of all types and severity categories, then it is described as predicting “total” crashes.

The crash type and severity category of interest is referred to hereafter as the “target” crash type. The target crashes are used to identify the CPM to be calibrated. Documentation describing the development of the candidate CPM should be consulted to determine whether it predicts the average frequency of target crashes. The HSM Part C chapters describe the target crashes associated with each HSM CPM.

Calibration Period

The calibration period is defined as the number of consecutive years for which observed crash data are acquired and used to estimate the calibration factor (or factors). The calibration period should be at least one year in duration. The period may be increased to two or three years if needed to reach the minimum sample size (as described in the next section). *There is only one calibration period for a given CPM and it spans the same time period for all sites in the associated calibration database.*

The average crash frequency of a site is likely to change over time due to changes in the design practice, vehicle crashworthiness, driver behavior, and so on. The magnitude of the change can be significant after several years have passed. For this reason, the calibration period should not exceed three years.

The calibration period must have a duration that is a multiple of 12 months to avoid seasonal effects. For ease of application, it is suggested that the calibration periods consist of “full” calendar years (i.e., January through December).

The calibration period is specified by the consecutive dates that define it. For example, the calibration period could be specified for January 1, 2010 to December 31, 2011, in which case its duration is two years. Similarly, the calibration period could be specified for just January 1, 2010 to December 31, 2010, in which case its duration is one year. The calibration period duration has units of “years.”

The same calibration period should be used for all CPMs that are calibrated for a specified region, facility type, and site type. However, exceptions to this guidance may be made where necessary such that some regions use one calibration period and other regions use another calibration period.

Minimum Sample Size Assessment Process

The sample size assessment process is used to determine the minimum number of sites needed to produce a calibration factor (or factors) that enables the CPM to produce reliable estimates of the predicted average crash frequency for sites in the region of interest.

Based on the HSM, the required minimum site sample size for the calibration database for one CPM is 30 sites. If a jurisdiction has fewer than 30 sites for a particular facility type, then it is desirable to use all of the available sites for calibration.

The sample of sites selected for inclusion in the calibration database must collectively represent at least 100 observed target crashes during the calibration period. The required minimum number of sites may need to be increased beyond 30 sites to ensure satisfaction of this criterion. This criterion should be assessed during the data assembly process, after the crash data have been assigned to the sites in the calibration database. The data assembly process is described in a subsequent section.

During the calibration procedure (described in a subsequent section), the minimum site sample size needed to obtain statistically valid calibration factor (or factors) will be calculated. At that time, additional sites may need to be added to the calibration database if indicated by these calculations.

Site Identification Process

Some agencies have an existing database that has been developed for road inventory and management purposes. This database is often based on segments that describe the characteristics of short sections of roadway. It may also identify segments that intersect to form intersections. The road inventory database is likely to provide useful information about the geometric design elements, traffic characteristics, traffic control devices, and crash records for the sites included in the calibration database.

Desirably, the segments in the agency’s road inventory database can be used directly as sites in the calibration database. However, the analyst should confirm that the sites in the agency database are consistent with the site type of interest and its defined characteristics. Documentation describing the development of the candidate CPM should be reviewed to identify the specific characteristics and criteria used to define the site type represented by the CPM. The HSM Part C chapters describe segmentation criteria that are used to identify each site type associated with a HSM CPM.

For segments, the site-type-of-interest’s defining characteristics should be consistent for the length of the segment (i.e., homogeneous). If a few of the segments in the agency database are not homogeneous, then these segments should be subdivided to produce segments that are homogeneous, or they should be removed from the calibration database.

If most of the segments in the agency’s road inventory database are not homogeneous, then the calibration sites should be established manually using aerial photographs or road design plans. The site type characteristics and criteria associated with the CPM are used to disaggregate the facilities of interest into a set of homogeneous sites. This process typically requires a manual review of photographs or plans to establish appropriate site boundaries. Milepost (or mile marker) designations are then associated with these boundaries to facilitate the site’s linkage to agency databases that contain geometric design elements, traffic characteristics, traffic control devices, and crash records.

For segments, each site should be between 0.1 and 1.0 mi in length. Lengths in this range should be long enough to have statistical validity and short enough to be realistically homogeneous.

It is desirable that the selected calibration sites be reasonably representative of the range of site characteristics to which the predictive model will be applied. However, no formal stratification by traffic volume or other site characteristics is needed in selecting the calibration sites. As such, the sites can be selected in a manner to make the data assembly process as efficient as practical.

Each site in the calibration database should be selected without regard to its associated number of crashes reported during the calibration period. In other words, calibration sites should not be selected on the basis of their reported crash counts. Where practical, this may be accomplished by selecting calibration sites randomly from a larger set of candidate sites.

The site identification process only need to be performed the first time that calibration is undertaken for a given region, facility type, and site type. The same sites may be used again for model updating in subsequent years—provided that their crash count, traffic volume, and site characteristics data are updated to the current time period. More generally, when updating a model, there is no need to develop a new database if an existing database with sites suitable for calibration is already available. If some new sites need to be chosen to supplement an existing database, the new sites should be selected randomly from a larger set of candidate sites.

Data Assembly Process

The data assembly process entails the identification, acquisition, reduction, and organization of the data needed for calibration of the CPM of interest. The analyst should conduct a jurisdiction-specific assessment of available existing data. A key outcome of this assessment is the identification of data that the analyst may need to acquire from supplemental sources (e.g., aerial photography). A work plan for data assembly that is developed through this assessment can be a useful basis for estimating the resources required for the calibration project.

The steps associated with this process are described in the following list.

1. Identify the data elements needed for calibrating the CPM of interest.
2. Identify the jurisdiction-specific databases that contain the desired data elements.
3. Identify the missing data elements and establish a process to acquire them from supplemental sources.
4. Assemble all data elements and combine them in a calibration database where each site represents one observation.

Guidance associated with these steps is provided in the following subsections. A complete data assembly process should only need to be performed the first time that calibration is performed for a given region, facility type, and site type. For model updating in subsequent years, the same sites may be used again with the crash count, traffic volume, and site characteristics data updated to reflect the new calibration period. Data describing the geometric design elements and traffic control devices will only need to be updated for those sites that have had a corresponding change in design or devices in the intervening years.

At the conclusion of the data assembly process, the analyst should determine the number of observed target crashes represented in the calibration database. If there are less than 100 observed target crashes, the analyst should return to the Site Identification Process section and add more sites to the database.

Data Elements Needed for Calibration

The calibration database needs two main categories of data. One category includes the data elements that serve as input values to the CPM. The second category is the observed crash data. Data for both categories are needed for each site represented in the calibration database.

**Input Data Elements.** The input data elements needed in the calibration database include all site characteristics that are used to apply the CPM of interest. The number of input data elements for a given model is dependent on the number of variables in the SPF and CMFs that comprise the CPM’s predictive model equation. These data elements typically include traffic characteristics, geometric design elements, and traffic control features.

Documentation describing the development of the candidate CPM should be consulted to determine the specific definition of each input variable to ensure that it is counted, measured, or computed in a manner that is consistent with the variable’s use in the CPM selected for calibration. This information is provided in Section 5 of each HSM Part C chapter if the subject CPM is from the HSM.

For the CPMs in HSM Part C, the input data elements have been categorized as “required” and “desirable” for inclusion in the calibration database. Default values are provided for calibration purposes when site-specific data are not available.

The analyst should check the data for each site to confirm that all geometric design elements and traffic control features are unchanged for the duration of the calibration period. Two options are available if any element or feature at a site changes during this period. One option is to exclude the site from the calibration database. Another option is to reduce the duration of the calibration period (such that all elements and features are unchanged at all sites during the reduced period).

The analyst should check the data for each site to confirm that a long-term work zone was not present during the calibration period. Two options are available if a work zone was present at a site during this period. One option is to exclude the site from the calibration database. Another option is to reduce the duration of the calibration period (such that no long-term work zone was present at any site during the reduced period).

The guidance in the section titled Calibration Period still applies if the calibration period is reduced following the guidance in the two previous paragraphs. Specifically, there is only one calibration period for a given CPM and it spans the same time period for *all* sites in the associated calibration database.

If any traffic flow characteristic (e.g., AADT) changes during the calibration period, the average value of the characteristic during the calibration period should be used in the calibration database.

**Crash Data.** The calibration database must include all target crashes that are observed (i.e., reported) at a site of interest during the calibration period. If the database is being used to calibrate two or more CPMs for a common region, facility, and site type, then the database will need to include the target crashes associated with each CPM of interest.

The crash attributes needed from each crash report include crash location, date and time, intersection-relationship, severity, and crash type. These data are needed for each crash that occurs near a site of interest during the calibration period. The crash location and intersection-relationship data are used to assign crashes to the correct site. The date of the crash is used to verify that the crash occurred during the calibration period.

Each crash must be assigned to the appropriate site. Documentation describing the development of the candidate CPM should be consulted to determine the specific criteria used to assign a crash to a site. This information is provided in Section 5 of each HSM Part C chapter if the subject CPM is from the HSM.

Data Sources

The development of a calibration factor (or factors) for a CPM requires the integration of data describing the crash history, traffic volume, geometric design elements, and traffic control features. Some of these data (e.g., crash history) will be obtained from agency databases. Other data may need to be assembled (i.e., collected, measured, or estimated) from supplemental sources. These sources may include internet imagery archives (e.g., Google Maps, Google Street View), mobile data collection logs (e.g., LiDAR data, videotapes), as-built design plans, and field data collection (e.g., site visits, traffic counts). Once these data are integrated into a calibration database, the agency can update the data when recalibrating the associated CPM at a later time.

The calibration database assembly process entails identifying and consolidating the sources of data within the jurisdiction. The data source must have sufficient detail to allow the analyst to determine the facility type and site type. The use of an agency’s road inventory data as a data source is discussed in the section titled Site Identification Process.

Missing Data

Most agencies will need to collect data to supplement their existing databases. Possible sources of supplemental geometric data include: internet imagery archives, mobile data collection logs, as-built design plans, and field data collection. The assembly of data from these sources can require measurable time. For each data element of interest, a process should be established (and documented) that describes how the data will be measured, processed, and checked. The process will vary based on the source of the data. Particular care should be taken to ensure that the data will be collected in a consistent manner for all sites and that it will be representative of conditions present during the calibration period.

If the CPM of interest is updated in subsequent years, the aforementioned document describing the procedures for collecting supplemental data should be used to guide the collection of missing data for any new sites that are added to the calibration database.

**Missing Volume Data.** Guidance in HSM Part C addresses the case where AADT volume data are available for only some of the years in the calibration period. The following rules may be applied to estimate the AADT volumes for years in which no data are available. If these rules are applied, the fact that some AADT volumes are estimated should be documented with the analysis results.

* If AADT volume is available for only a single year, that same volume is assumed to apply to all years of the evaluation period.
* If two or more years of AADT data are available, the AADT volumes for intervening years are computed by interpolation.
* The AADT volumes for years before the first year for which data are available are assumed to be equal to the AADT volume for that first year.
* The AADT volumes for years after the last year for which data are available are assumed to be equal to the AADT volume for that last year.

Calibration Procedure

The procedure described in this section is used to compute the value of the calibration factor (or factors) for the predictive model equation in the CPM of interest. The procedure is developed for calibrating or updating one predictive model equation. The procedure is repeated for each predictive model equation of interest.

It is possible that a predictive model equation will include one or more variables for which a default value is being used in the calibration database. It is also possible that a predictive model equation will include a constant representing a distribution value (e.g., proportion of total nighttime crashes). If any default values or crash distributions are included in this manner, then they should be confirmed to be representative of the region of interest. If they are not representative, then jurisdiction-specific values and distributions should be estimated and inserted into the calibration database before the predictive model equation is calibrated.

The calibration procedure involves six steps. These steps are described in the following subsections.

Step 1 – Compute Site Sample Statistics and Assess Sample Adequacy

During this step, the sites in the calibration database are used to compute the minimum site sample size needed to produce a specified level of reliability in the calibration factor (or factors). These computations are described in the following three tasks.

Task A. Compute Site Sample Statistics

The crash data for the calibration sites is used to compute the average and standard deviation of the observed crash counts. Initially, these statistics are computed for the calibration database assembled as part of the project planning procedure. However, as explained later in this step, data for additional sites may need to be added to the calibration database. If data are added, then these statistics will need to be calculated again and the results used to determine a new estimate of the minimum site sample size. The following sequence of equations is used to compute the statistics of interest:

Equation

Equation

Equation

Equation

where

= average number of observed crashes per site during the calibration period (crashes/site/period);

= average segment length (mi);

*so* = standard deviation of the number of crashes observed during the calibration period;

*cv,No* = coefficient of variation of the observed crashes;

*No,i,j* = observed annual crash count for site *i* and year *j* (crashes/yr);

*Ls,i* = segment length for site *i* (mi) (= 1.0 if the database represents intersections);

*nsites* = number of sites in the calibration database (sites); and

*nca* = number of years in the calibration period (yr).

Task B. Determine Minimum Site Sample Size

Using the statistics computed in the previous section, Table 1 is consulted to identify the minimum number of sites for calibrating the predictive model equation for the CPM of interest. This minimum value is based on the coefficient of variation of the observed crashes *cv,No* (from Equation 4) and a specified threshold value for the calibration factor coefficient of variation *cv,c*. Bahar and Hauer (2014) recommend the use of a threshold value between 0.10 and 0.15 for sample size calculations. A smaller value for this coefficient corresponds to a more reliable estimate of the calibration factor. However, smaller values are also associated with a larger minimum sample size.

Table . Minimum number of sites for calibration.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Computed Coefficient of Variation of the Observed Crashes *cv,No* | Minimum Number of Sites by Target Calibration Factor Coefficient of Variation *cv,C*a | | | | | |
| 0.10 | 0.11 | 0.12 | 0.13 | 0.14 | 0.15 |
| 2.85 | 812 | 671 | 564 | 481 | 414 | 361 |
| 2.60 | 676 | 559 | 469 | 400 | 345 | 300 |
| 2.35 | 552 | 456 | 384 | 327 | 282 | 245 |
| 2.15 | 462 | 382 | 321 | 274 | 236 | 205 |
| 1.95 | 380 | 314 | 264 | 225 | 194 | 169 |
| 1.75 | 306 | 253 | 213 | 181 | 156 | 136 |
| 1.60 | 256 | 212 | 178 | 151 | 131 | 114 |
| 1.45 | 210 | 174 | 146 | 124 | 107 | 93 |
| 1.35 | 182 | 151 | 127 | 108 | 93 | 81 |
| 1.20 | 144 | 119 | 100 | 85 | 73 | 64 |
| 1.10 | 121 | 100 | 84 | 72 | 62 | 54 |
| 1.00 | 100 | 83 | 69 | 59 | 51 | 44 |
| 0.90 | 81 | 67 | 56 | 48 | 41 | 36 |
| 0.85 | 72 | 60 | 50 | 43 | 37 | 32 |
| 0.75 | 56 | 46 | 39 | 33 | 30 | 30 |
| 0.70 | 49 | 40 | 34 | 30 | 30 | 30 |
| 0.60 | 36 | 30 | 30 | 30 | 30 | 30 |
| < 0.60 | 30 | 30 | 30 | 30 | 30 | 30 |

a – Number of sites for calibration (*nsites,min*) computed as where, *cv,No* = coefficient of variation of the observed crashes and *cv,c* = calibration factor coefficient of variation.

Table 1 is used by first choosing the column of interest. This choice is based on the selected threshold calibration factor coefficient of variation *cv,c*. Then, the first column is consulted to determine the row that best corresponds to the value of the coefficient of variation of the observed crashes *cv,No*. Finally, the minimum site sample size *nsites,min* is found at the intersection of the row and column of interest (interpolation can be used if needed). Alternatively, the equation in the table footnote can be used to estimate the minimum site sample size, if desired.

The equivalent minimum crash sample size *No,min* can be computed as the product of the average number of observed crashes per site during the calibration period and the minimum number of sites *nsites,min* (i.e., *No,min* = × *nsites,min*).

Task C. Adjust Database If Necessary

If the number of sites in the calibration database is less than the minimum site sample size determined in Task B, then it is an indication that the differences among sites (in terms of their safety characteristics) is sufficiently large that it is unlikely that the computed calibration factor will have the desired reliability (as indicated by the specified calibration factor coefficient of variation). Several options are available to mitigate this issue. These options are described in the following list.

1. Add sites to the database (following the guidance provided in the section titled, Site Identification Process).
2. Increase the duration of the calibration period (up to three years).
3. If sites were initially identified from a large region (e.g., entire state), subdivide the jurisdiction into geographic regions that are individually more consistent in terms of climate, topography, driver population, wild animal population, design practice, crash reporting threshold, prevailing weather patterns, enforcement levels, and crash reporting system procedures (such that the computed value of the coefficient of variation of the observed crashes is reduced).
4. Increase specified threshold value for the calibration factor coefficient of variation (up to 0.15).

Any combination of these options can be used. If an option is selected, then the minimum site sample size needs to be recalculated using the process outlined in this step. This process of “(a) compute statistics, (b) determine minimum sample size, and (c) if needed, adjust database” is repeated until the actual number of sites in the database equals or exceeds the computed minimum site sample size.

Step 2 – Compute the Predicted Average Crash Frequency

The predictive model equation for the CPM of interest is used to compute the predicted average number of crashes during the calibration period for each site in the calibration database. The predicted value is specific to the target crashes of interest that were identified in the planning process. The following equation is used for this purpose:

Equation

where

*Np,u,i =* predicted average number of crashes for site *i* during the calibration period and unadjusted by the calibration factor (crashes/period);

*Np,u,i,avg =* predicted average crash frequency for site *i* using average AADT for calibration period and unadjusted by the calibration factor (crashes/yr); and

*nca* = number of years in the calibration period (yr).

For this step, the predictive model equation must be applied without using the EB Method and without employing a calibration factor (or factors). In other words, applying the predictive model equation entails the calculation of the associated SPF and all CMFs, and then using them to compute the predicted average crash frequency without adjustment by the calibration factor (or factors).

Step 3 – Compute the Calibration Factor

The calibration factor (or factors) is computed in this step. The following sequence of tasks is used to compute the calibration factor for the subject predictive model equation.

Task A. Sum the Predicted Average Number of Crashes for All Sites Combined

The sum of the predicted average number of crashes is computed using the following equation:

Equation

where *Np,u,aS* is the total predicted average number of crashes for all sites during the calibration period and unadjusted by the calibration factor (crashes/period); and all other variables are previously defined.

Task B. Sum the Observed Crash Counts for All Sites Combined

As a first activity, the observed number of target crashes is computed for each site. The following equation is used for this purpose:

Equation

where *No,i* is the observed crash count for site *i* during the calibration period (crashes/period); *No,i,j* is the observed annual crash count for site *i* and year *j* (crashes/yr); and all other variables are previously defined.

Then, the sum of the predicted average number of crashes is computed using the following equation:

Equation

where *No,aS* is the total observed crash count for all sites during the calibration period (crashes/period); and all other variables are previously defined.

Task C. Compute the Calibration Factor.

The calibration factor is computed using the following equation:

Equation

where *C*is the calibration factor to adjust the CPM for local conditions; and all other variables are previously defined.

When all steps of the calibration procedure are completed, the computed calibration factor is rounded to two decimal places for all subsequent applications.

Task D. Compute the Coefficient of Variation

The calibration factor coefficient of variation is computed using the following sequence of equations:

Equation

Equation

Equation

where

*sop* = standard deviation of the residual number of crashes (i.e., difference between the observed crash count and the predicted average crash frequency) during the calibration period, (crashes/site/period);

*cv,c* = calibration factor coefficient of variation;

*sc* = standard error of the calibration factor;

and all other variables are previously defined.

If the computed calibration factor coefficient of variation is larger than the desired threshold value (as specified in Step 1), then the analyst should consider increasing the site sample size, calibration period duration (up to three years), or desired threshold value (up to 0.15). If any of these actions is taken, then return to Step 2 of the calibration process and repeat Steps 2 through 3.

Step 4 – Outlier Detection and Resolution

The data in the calibration database should be checked to determine whether there are any outlier sites present. In this regard, an *outlier site* is defined as a site with a count that is exceptionally smaller or larger than that for other sites with similar volume, length, and design characteristics.

Prior to determining whether a site is an outlier, it must first be identified as an extremely rare case. A site is considered to be an *extremely rare case* when its predicted average crash frequency is significantly different from its observed annual crash count, relative to all other sites in the calibration database. If this difference (i.e., “residual error”) is a result of an inaccurate prediction or incorrect observed crash count, then the site is considered an *outlier*. If one or more outliers are included in the data used to compute the calibration factor, the value of this factor will likely be biased.

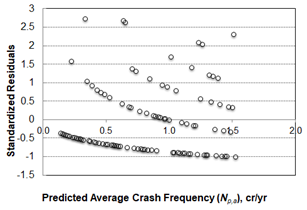
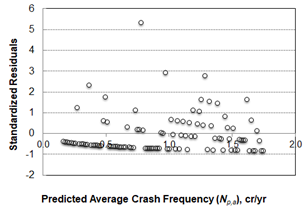
Several techniques are available for identifying extremely rare cases. A technique based on the examination of standardized residuals is described in this section. It provides a reliable indication of the presence of extremely rare cases in databases used to develop or calibrate CPMs.

The standardized residual technique is based on an assessment of the standardized residuals for each site in the database. The standardized residual is computed by dividing the residual error associated with a site by its standard deviation. The standardized residuals tend to be uniformly distributed over the range of the predicted average number of crashes represented in the database. The standardized residual is asymptotic to the normal distribution for databases having a larger overall average crash frequency and a smaller overdispersion parameter. Extremely rare cases are evidenced by having a standardized residual that is several standard deviations away from 0.0.

An example plot of standardized residuals is shown in Figure 1a. Each data point shown in this figure corresponds to one site. The calibration period is one year. The data points tend to lie along one of several bands that extend from the upper left side of the figure to the lower right side. Each band corresponds to an integer crash count. For example, the lowest band in the figure corresponds to sites for which there were no observed crashes during the year. None of the data points in the figure have an unusually large (or small) standardized residual value so no extremely rare cases appear to be present in the data.

Figure 1b has one data point associated a standardized residual value of 5.2. This value is much larger than that of the other data points. Hence, it is considered an extremely rare case. The data associated with this site should be further examined to determine if the site is an outlier.

The use of standardized residuals for detecting outliers is described in this section. The following sequence of tasks is used to implement this technique and resolve outlier-related issues.



***a. Example data with no extremely rare cases. b. Example data with an extremely rare case.***

Figure . Example standardized residual plots.

Task A. Identify Extremely Rare Cases

The standardized residual error is computed for each site using the following calculations.

*Calculation 1.* *Compute the adjusted predicted average number of crashes*. The following equation is used to compute the adjusted predicted number of crashes during the calibration period at each site:

Equation

where *Np,a,i* is the predicted average number of crashes for site *i* during the calibration period and adjusted by the calibration factor (crashes/period); and all other variables are previously defined.

The overdispersion parameter for the calibration data is estimated in the next three tasks. The calculations are based on least-squares regression analysis of the database. This parameter can be computed using other procedures (e.g., maximum likelihood regression, method of moments). If the analyst desires to estimate this parameter using another procedure, they can skip Calculations 2, 3, and 4.

*Calculation 2*. *Compute the variable Y*. This variable is computed for each site using the following equation:

Equation

where *Yi* is an intermediate variable used in a subsequent calculation; and all other variables are previously defined.

*Calculation 3. Compute the dispersion coefficient*. This variable is computed using the following equation.

Equation

where *K* is the dispersion coefficient used in a subsequent calculation; and all other variables are previously defined.

*Calculation 4. Compute the overdispersion parameter*. This variable is computed using the following equation:

Equation

where *ki* = overdispersion parameter for site *i* and other variables are as previously defined.

*Calculation 5. Compute the crash frequency variance*. This variance is computed for each site using the following equation:

Equation

where *V*[*X*]*i* is the crash frequency variance for site *i* during the calibration period (crashes/period2); and all other variables are previously defined.

*Calculation 6. Compute the standardized residual.* The standardized residual is computed for each site using the following equation:

Equation

where *rp,i* is the standardized residual for site *i* and all other variables are previously defined.

*Calculation 7. Identify extremely rare cases*. Lower and upper boundary values are selected from Table 2 using the average number of observed crashes per site (computed previously using Equation 1) and the average overdispersion parameter value (computed using the equation in the table footnote).

Table . Standardized residual boundaries for identifying extremely rare cases.

|  |  |  |  |
| --- | --- | --- | --- |
| Average Number of Crashes () (cr/site/period) | Average Overdispersion Parameter () | Standardized Residual Boundaries | |
| Lower Boundary | Upper Boundary |
| ≥ 6 | 0.01 | -2.2 | 3.2 |
|  | 0.1 | -1.9 | 3.9 |
|  | 0.2 | -1.7 | 4.2 |
|  | 0.5 | -1.2 | 4.7 |
|  | 1.0 | -1.0 | 5.0 |
| < 6 | any | -1.1 | 4.9 |

Note: If the sites are segments with length *Ls,i* in miles, then for this table is computed as ; otherwise .

Once the boundary values are selected, a site with a standardized residual value that is smaller than the lower boundary or larger than the upper boundary is considered an extremely rare case.

This assessment can be most easily undertaken by creating a plot with the predicted average number of crashes on the *x*-axis and the scaled residual on the *y*-axis (see Figure 1). A horizontal line can be located on the plot at the upper boundary value and a second horizontal line can be located at the lower boundary value. Any data point above the upper line or below the lower line is considered an extremely rare case.

Task B. Identify Outliers and Resolve

All extremely rare cases should be individually evaluated to determine whether they are outliers. If the discrepancy is a result of an error (i.e., inaccurate prediction or incorrect observed crash count) then the site is considered an *outlier*. The source of the error should be identified and corrected if possible. For example, an inaccurate prediction could be the result of an AADT volume with one digit mistakenly left out or a lane width value with its digits transposed. If the error cannot be corrected, then the observation should be removed from the database and, if possible, replaced with a new (randomly selected) site.

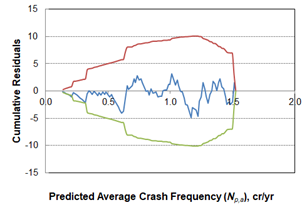
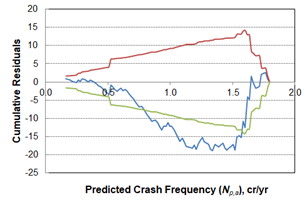
If there is no reason to believe that the extremely rare case is a result of error, then it should not be deleted without careful consideration. The analyst should investigate the site’s associated characteristics to determine the reason why the site is so different from the other sites. The insights obtained may reveal some weaknesses in the site selection process. It may confirm that the extreme site is truly not representative of the site type of interest—in which case it should be replaced with a new site.

If any sites are added to (or removed from) database, or if the data elements associated with one or more sites are corrected, then the analyst should return to Step 2 and complete Steps 2 through 4.

Step 5 – Assess Model Fit Based on the Calibration Factor

The fit of the calibrated predictive model equation to the data is assessed in this step. The CURE technique is used for this purpose. It was developed by Hauer and Bamfo (1997) for assessing how well a regression model fits the crash data used to estimate the model coefficients. This technique is equally applicable to the evaluation of a calibrated CPM’s fit to the calibration database.

Hauer (2015) indicates that a model that fits the data reasonably well produces a cumulative residual trend line that “…meanders around the horizontal axis in a manner consistent with a symmetric random walk.” This type of meandering is illustrated in Figure 2a (using a blue trend line).



***a. Example of reasonably good fit. b. Example of possible poor fit.***

Figure . Example CURE plots.

In contrast, a poor fit can be indicated when significant portions of the cumulative residual trend line increase and remain above the horizontal “0” line, or decrease and remain below the horizontal “0” trend line. If these extended portions are well above or below the “0” line, it could be a result of model bias-in-fit (i.e., a suboptimal model form that produces a systematic discrepancy between the observed and predicted values [Hauer, 2015]). This condition is suggested in Figure 2b for the range of residuals associated with a predicted average crash frequency of about 1.3 to 1.7 crashes/yr. In this instance, a calibration function may be needed to mitigate the CPM’s bias-in-fit and improve its predictive reliability.

Task A. Calculate Cumulative Residuals

This section summarizes the calculations required to compute the cumulative residuals and their associated 95th percentile confidence interval.

*Calculation 1. Prepare the database*. During this task, sites in the calibration database are sorted by the predicted average number of crashes adjusted by the calibration factor *Np,a,i*.

*Calculation 2. Compute the cumulative residuals*. The cumulative residual is computed for each of the sites in the database. The *m*th cumulative residual is computed from the residual error for all sites from 1 to *m*. This calculation is shown in the following equation:

Equation

where *rc,m* is the *m*th cumulative residual (where 1 ≤ *m* ≤ *nsites*) and all other variables are previously defined.

*Calculation 3. Compute the standard deviation of the cumulative residual.* The standard deviation of the cumulative residual is computed for each of the sites in the database. The standard deviation for the *m*th cumulative residual is computed using the following equation:

Equation

where *sr,m* is thestandard deviation of the *m*th cumulative residual (where 1 ≤ *m* ≤ *nsites*); and all other variables are previously defined.

*Calculation 4. Compute the zp value associated with the 95th percentile confidence interval*. The *zp* values used to define the confidence interval of the cumulative residuals are computed using the following equations:

Equation

Equation

where *zp* is the number of standard deviations associated with a cumulative probability *p*; and all other variables are previously defined.

*Calculation 5. Compute the upper and lower confidence interval limits*. The upper and lower confidence interval limits are computed for each of the sites in the database. The upper and lower confidence limits for the *m*th cumulative residual are computed using the following equations:

Equation

Equation

where

*rc,m,0.025* *=* lower boundary for cumulative residual trend line based on cumulative probability of 0.025;

*rc,m,0.975* *=* upper boundary for cumulative residual trend line based on cumulative probability of 0.975;

and all other variables are previously defined.

Task B. Assess Model Fit

When the underlying model fits the data, about 95 percent of all calibration databases will have 5 percent or less of the cumulative residuals that exceed the confidence interval limits. In other words, if no more than 5 percent of the cumulative residuals go beyond the upper or lower limits, the functional form of a CPM is unlikely to be creating bias-in-fit over the range of the predicted crash frequency.

On the other hand, if the percentage of cumulative residuals that exceed the upper or lower limits exceeds 5 percent, it is an indication of a potentially poor fit of the CPM to the data.

If the calibrated CPM is determined to provide a good fit to the data, then the calibration process is complete, and Step 6 can be omitted.

If the calibrated CPM is determined to provide a poor fit to the data, then a calibration function should be considered. The analyst proceeds to Step 6 to compute this function.

Step 6 – If Applicable, Compute the Calibration Function

If the findings from Step 5 indicate that a calibration function should be considered, then a regression analysis is undertaken to fit the following model to the observed crash data:

Equation

where

*No =* observed crash count during the calibration period (crashes/period);

*Np,u =* predicted average number of crashes during the calibration period unadjusted by calibration factor (crashes/period); and

*Ci =* calibration factor *i* to adjust the CPM for local conditions.

The regression analysis should incorporate a log-link function to ensure that the predicted values respect the non-negative average crash frequency associated with each site. The equation suitable for log-linear regression analysis is:

Equation

where

*ci* = regression coefficient i;

*C*3 = exp[*c3*];

*C*4 = *c4*;

and all other variables are previously defined.

The regression analysis should be based on a negative binomial distribution of the residuals given that this distribution is typically found in observed crash data. The two calibration factors in Equation 25 are represented as regression coefficients in Equation 26. The best-fit model coefficients should be based on the maximum-likelihood objective function where the overdispersion parameter is also considered an estimable coefficient. The regression analysis should be undertaken using a software tool that automates the model estimation process and produces a standard error for each regression coefficient. The tool developed by Lyon et al. (2018) can be used for this purpose.

The fit of the calibration function should be assessed to determine if the regression coefficient *c4* is significantly different from 1.0. The test statistic *t* is computed as follows *t* = (*c4* – 1.0)/*sc4*; where *sc4* is the standard error of *c4* (as obtained from the regression model output). The absolute value of *t* will need to exceed 1.645 to have 90 percent confidence that the coefficient value is truly different from 1.0.

If the absolute value of *t* is less than 1.645, then there is insufficient evidence that the calibration function improves the CPM’s fit to the calibration data. The calibration factor computed in Step 5 should be adopted for subsequent application of the CPM in the region of interest.

If the absolute value of *t* equals or exceeds 1.645, then the regression coefficients should be used with the calibration function in Equation 25 to calibrate the subject CPM. The analyst should return to Step 4 and use the technique described therein to confirm that the function does reveal any additional outliers. Then the analyst should return to Step 5 and use the technique described therein to confirm that the function does not create any bias-in-fit. If the results of Step 5 indicate that the calibration function provides an acceptable fit to the data, then the calibration function determined in this step should be adopted for subsequent application of the CPM in the region of interest. If the results indicate that the function does not provide an acceptable fit to the data, then the calibration factor computed in Step 5 should be adopted for subsequent application of the CPM in the region of interest.

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.

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

Hauer,E., and J. Bamfo. (1997). “Two Tools for Finding what Function Links the Dependent Variable to the Explanatory Variables.” *Proceedings of the ICTCT 97 Conference*. Lund, Sweden, pp. 1-19.

HSM (2010). *Highway Safety Manual*. American Association of State Highway and Transportation Officials. Washington, D.C.

Ivan, J., S. Al Mamun, N. Ravishanker, B. Persaud, C. Lyon, R. Srinivasan, B. Lan, S. Smith, T. Saleem, M. Abdel-Aty, J. Lee, A. Farid, and J-H. Wang. (2018). *Improved Prediction Models for Crash Types and Crash Severities. Final Report.* NCHRP Project 17-62. Transportation Research Board, Washington, D.C.

Lyon, C., B. Persaud, and F. Gross. (2018). *The Calibrator—An SPF Calibration and Assessment Tool, Updated User Guide*. Report No. FHWA-SA-17-016. Federal Highway Administration, Washington, D.C.

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.