Analysis of Data Selection and Inclusion for Highway Safety Manual Model Calibration

Bridget Scheyd Robicheaux

Louisiana State University and Agricultural and Mechanical College

Follow this and additional works at: https://digitalcommons.lsu.edu/gradschool_theses

Part of the Civil and Environmental Engineering Commons

Recommended Citation
https://digitalcommons.lsu.edu/gradschool_theses/385

This Thesis is brought to you for free and open access by the Graduate School at LSU Digital Commons. It has been accepted for inclusion in LSU Master's Theses by an authorized graduate school editor of LSU Digital Commons. For more information, please contact gradetd@lsu.edu.
ANALYSIS OF DATA SELECTION AND INCLUSION FOR HIGHWAY SAFETY MANUAL MODEL CALIBRATION

A Thesis

Submitted to the Graduate Faculty of the Louisiana State University and Agricultural and Mechanical College in partial fulfillment of the requirements for the degree of Master of Science

in

The Department of Civil and Environmental Engineering

by

Bridget S. Robicheaux
B.S., Louisiana State University, 2007 December 2014
ACKNOWLEDGMENTS

I would like to thank my advisor, Dr. Brian Wolshon, for his assistance with this project and for his support throughout my studies at LSU. From introducing transportation engineering to me in his CE 3600 class in my undergraduate studies to convincing me to return to graduate school, I would not be where I am today without him. Also, special thanks to Dr. Sherif Ishak and Dr. Chester Wilmot for their willingness to help serve on my advisory committee and for all of the knowledge they passed down to me through their classes.

I would also like to thank my supportive husband, who has taken care of our daughter when I was spending countless nights and weekends trying to get all of my schoolwork done. I cannot wait to have more time to spend as a family of FOUR soon! To my sweet Agnes Mae who was always so understanding when I had to miss out on time with her to get my work done and who has made all of my efforts worthwhile. I love you more than you know! To my mom for all of the love and support she has given me – I would not have survived graduate school without you!

This thesis was developed from a project which was completed under the direction of the Louisiana Transportation Research Center and the Louisiana Department of Transportation and Development. I would like to thank the Project Review Committee members and all other DOTD personnel involved in this project for the expertise that they provided and to Dr. Helmut Schneider, Jose Rodriguez, and Dr. David Sathiaraj for their technical contributions to this project. I would also like to thank the two undergraduate students who helped in the data collection efforts, Matt Benoit and Aidan Eymard. I would also like to thank the Louisiana Board of Regents for granting me a fellowship that allowed me to return to school.
# TABLE OF CONTENTS

ACKNOWLEDGMENTS .................................................................................................................. ii

LIST OF TABLES .......................................................................................................................... v

LIST OF FIGURES ....................................................................................................................... vi

ACYRONYMS, ABBREVIATIONS, AND SYMBOLS ............................................................... vii

ABSTRACT ................................................................................................................................... ix

CHAPTER 1. INTRODUCTION ....................................................................................................1
  1.1 Objectives ......................................................................................................................3
  1.2 Scope ..............................................................................................................................3

CHAPTER 2. LITERATURE REVIEW .........................................................................................5
  2.1 HSM Predictive Method ..............................................................................................5
  2.2 Data Collection ..............................................................................................................8
  2.3 Tailoring the HSM for Individual Application ............................................................10
  2.4 Dataset Size ..................................................................................................................12
  2.5 Calibration Factors .......................................................................................................14
  2.6 Development of Local SPFs ........................................................................................16
  2.7 Findings ........................................................................................................................19

CHAPTER 3. METHODOLOGY .................................................................................................20
  3.1 HSM Calibration Procedure .........................................................................................21
  3.2 Data Acquisition and Processing .................................................................................21
  3.3 Site Selection Methods ................................................................................................24
  3.4 Crash Data ....................................................................................................................28
  3.5 Data Compilation and Computation ............................................................................31
  3.6 SPFs and CMFs ............................................................................................................31
  3.7 Calculating Calibration Factors ...................................................................................35

CHAPTER 4. DISCUSSION OF RESULTS ................................................................................37
  4.1 Results of Calibration Factor Development .................................................................37
  4.2 Results of Calibration Factor Validation .....................................................................43
  4.3 Comparison of Calibration Factors ..............................................................................44

CHAPTER 5. SUMMARY AND CONCLUSIONS .....................................................................46
  5.1 General Findings ..........................................................................................................46
  5.2 Calibration Factor Recommendations .........................................................................48
  5.3 Future Work ..................................................................................................................49

REFERENCES ..............................................................................................................................51
LIST OF TABLES

Table 1: Segment Dataset Sizes in Prior Calibration Studies ........................................ 13
Table 2: Calibration Factors from Other States’ Calibrations ....................................... 14
Table 3: Property Damage Only Crash Reporting Threshold by State .............................. 15
Table 4: Maximum Allowable AADT for HSM SPF Application ................................. 23
Table 5: Number of Segments Used in Each Iteration of Calibration Factors ............... 26
Table 6: Summary of Data Inclusion by Facility Type ..................................................... 36
Table 7: Calibration Factors by Facility Type ............................................................... 38
Table 8: Percent Change between Iteration 1 and 2 Calibration Factors ..................... 39
Table 9: Percent Change between Iteration 2 and 3 Calibration Factors ..................... 40
Table 10: Difference and Percent Change between Iteration 2 and 4 Calibration Factors ... 41
Table 11: Percent Change between Iteration 4 Crash Definitions ................................. 42
Table 12: Percent Change between Predicted Crashes and Actual Crashes .................. 44
Table 13: Percent Change in Initial and 2012 Calibration Factors .............................. 45
Table 14: Calibration Factor by Facility Type ............................................................. 48
LIST OF FIGURES

Figure 1: Louisiana Rural Two-Lane Road Data ................................................................. 22

Figure 2: Sample of Driveway Type and Location ............................................................ 28

Figure A-1: Excel® Macro Used to Remove the Crashes Occurring Within 250 ft. of an Intersection .................................................................................................................. 53

Figure A-2: Excel® Macro Used to Remove the Crashes Occurring Within a Curve for Rural Two-Lane Roads ............................................................ 54

Figure A-3: Excel® Macro Used to Count Crashes for Rural Two-Lane Without Crashes on Curves or Within 150 ft. of Intersection ......................................................... 55
# ACRONYMS, ABBREVIATIONS, AND SYMBOLS

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AASHTO</td>
<td>American Association of State Highway and Transportation Officials</td>
</tr>
<tr>
<td>HSM</td>
<td>Highway Safety Manual</td>
</tr>
<tr>
<td>FHWA</td>
<td>Federal Highway Administration</td>
</tr>
<tr>
<td>LA DOTD</td>
<td>Louisiana Department of Transportation and Development</td>
</tr>
<tr>
<td>SPF</td>
<td>Safety Performance Function</td>
</tr>
<tr>
<td>CMF</td>
<td>Crash Modification Factor</td>
</tr>
<tr>
<td>C</td>
<td>Calibration Factor</td>
</tr>
<tr>
<td>$N_{\text{predicted}}$</td>
<td>Predicted Average Crash Frequency</td>
</tr>
<tr>
<td>TWLTL</td>
<td>Two-Way Left-Turn-Lane</td>
</tr>
<tr>
<td>AADT</td>
<td>Average Annual Daily Traffic Volume (vehicles per day)</td>
</tr>
<tr>
<td>DOT</td>
<td>Department of Transportation</td>
</tr>
<tr>
<td>PDO</td>
<td>Property Damage Only</td>
</tr>
<tr>
<td>$P_{RA}$</td>
<td>Proportion of Total Crashes Constituted by Related Crashes</td>
</tr>
<tr>
<td>2U</td>
<td>Urban/Suburban Two-Lane Road</td>
</tr>
<tr>
<td>3T</td>
<td>Urban/Suburban Three-Lane Highway with TWLTL</td>
</tr>
<tr>
<td>4U</td>
<td>Urban/Suburban Four-Lane Undivided Highway</td>
</tr>
<tr>
<td>4D</td>
<td>Urban/Suburban Four-Lane Divided Highway</td>
</tr>
<tr>
<td>5T</td>
<td>Urban/Suburban Five-Lane Highway with TWLTL</td>
</tr>
<tr>
<td>L</td>
<td>Length of Roadway Segment (miles)</td>
</tr>
<tr>
<td>a,b</td>
<td>Regression Coefficients</td>
</tr>
<tr>
<td>$C_r$</td>
<td>Calibration Factor for Roadway Segments of a Specific Type Developed for use for a Particular Geographical Area</td>
</tr>
<tr>
<td>$N_{\text{spf ru}}$</td>
<td>Base Total Expected Average Crash Frequency for a Roadway Segment</td>
</tr>
<tr>
<td>Symbol</td>
<td>Description</td>
</tr>
<tr>
<td>--------</td>
<td>-------------</td>
</tr>
<tr>
<td>N_{spf rs}</td>
<td>Predicted Total Average Crash Frequency for Roadway Segment Base Conditions</td>
</tr>
<tr>
<td>N_{spf rd}</td>
<td>Base Total Expected Average Crash Frequency for a Roadway Segment</td>
</tr>
<tr>
<td>N_{predicted rs}</td>
<td>Predicted Average Crash Frequency of an Individual Roadway Segment for the Selected Year</td>
</tr>
<tr>
<td>N_{hr}</td>
<td>Predicted Average Crash Frequency of an Individual Roadway Segment (excluding vehicle-pedestrian and vehicle-bicycle collisions)</td>
</tr>
<tr>
<td>N_{pedr}</td>
<td>Predicted Average Crash Frequency of Vehicle-Pedestrian Collisions for an Individual Roadway Segment</td>
</tr>
<tr>
<td>N_{biker}</td>
<td>Predicted Average Crash Frequency of Vehicle-Bicycle Collisions for an Individual Roadway Segment</td>
</tr>
<tr>
<td>N_{bmv}</td>
<td>Predicted Average Crash Frequency of Multiple-Vehicle Non-Driveway Collisions for Base Conditions</td>
</tr>
<tr>
<td>N_{bsv}</td>
<td>Predicted Average Crash Frequency of Single-Vehicle Crashes for Base Conditions</td>
</tr>
<tr>
<td>N_{bndwy}</td>
<td>Predicted Average Crash Frequency of Multiple-Vehicle Driveway Collisions for Base Conditions</td>
</tr>
</tbody>
</table>
ABSTRACT

The application of the American Association of State Highway and Transportation Officials’ (AASHTO) Highway Safety Manual (HSM) to Louisiana roads is a key component to the Louisiana Department of Transportation and Development’s (LA DOTD) plan to improve safety on state highways and reach the goal of “Destination Zero Deaths.” To apply the HSM the research sought to develop state-specific HSM calibration factors for eight facility types. During the development process, the data-intensive computational procedure followed to compute the calibration factors revealed numerous issues associated with the inclusion and definitions of various data elements required by the HSM. These included coding errors, missing required data elements in the Louisiana roadway and crash databases, and varying definitions of which specific crashes should or should not be included in the sample. Because of this, some of the resulting factors were unexpected, in particular, those for urban three-lane and urban five-lane highways which were lower than anticipated. To investigate the effect of including or excluding various data elements and varying crash definition assumptions, the HSM calibration factors were developed using a series of computational iterations in which the amount of data and assumed crash conditions were varied from one iteration to the next. The overall results of this thesis demonstrate the extent of the variability and sensitivity of HSM calibration factors to the inclusion of data that may or may not be included in roadway databases and how crashes that occur within various distances away from intersections are included or excluded.
CHAPTER 1. INTRODUCTION

The state of Louisiana consistently ranks near the bottom in national statistics regarding highway safety, particularly traffic crash related fatalities. To counter these conditions, the Louisiana Department of Transportation and Development (LA DOTD) has initiated several safety-related campaigns over the recent decade; among these is the Louisiana Strategic Highway Safety Plan. The goal of this program is to reach “Destination Zero Deaths” on Louisiana roadways by reducing the human and economic toll on Louisiana’s surface transportation system through various collaborative efforts and an integrated 4E approach (Education/Enforcement, Engineering and Operations, Emergency Services, Everyone Else!).

In conjunction with these Louisiana efforts, the past several decades have seen many new highway safety related innovations on a national level. One of the most promising recent developments has been the 2010 publication of the Highway Safety Manual (HSM) by the American Association of State Highway and Transportation Officials (AASHTO). The HSM includes analytical tools and techniques for quantifying the safety effects of planning, design alternatives and configurations, and operations and maintenance decisions. However, since the HSM has been developed based on national trends and statistics, it must be calibrated for local use. This calibration allows it to better represent local conditions. The HSM provides guidance on the calibration procedure in Volume 2 Appendix A. While the computational procedure follows a very prescriptive, straightforward process, it can become considerably more complicated for a variety of reasons. Among the most significant are the identification of sites that meet the required HSM criteria and selecting them in a statistically representative manner. Another issue is acquiring and including all data to meet the computational requirements of the HSM and maintaining their consistency with HSM procedures. Addressing these needs often requires users to apply less-than-
ideal methods and accept factors that, while imperfect, reflect the best estimates possible under the circumstances.

This thesis presents the results of a recent research effort to inform and support future users of the HSM by examining how data availability and assumption-making affects the HSM calibration factor development process. The research started out as a project for the Louisiana Transportation Research Center to calculate statewide calibration factors for the application of the HSM on various state highway classifications in Louisiana. This was accomplished by applying the HSM Part C predictive models to eight different rural, urban, and suburban roadway classifications. However, given limitations of certain data elements and collection/coding methods in the state, it was apparent that there was a limited understanding of how data availability and assumption-making affect the HSM calibration process. As a result, additional research to quantify and understand these effects was thought to be needed. Based on this, the focus of this thesis is on testing methods that can be used to address these issues and better understand their quantitative effects on the resulting factors.

In the following sections of this thesis, the objectives of the research and the scope of this effort are described. This is followed by a review of literature to summarize the current state of HSM-application practice, including its development, content, and prior applications that are relevant to this work. Later, details of the data and methods used to compute the Louisiana-specific HSM calibration factors for the various road types are described. This includes the collection and analysis of Louisiana crash data as well as DOTD roadway traffic, design and control data to support the creation of computational methods for calibration factor calculation. The factors were calculated in an iterative process which is described in greater detail in the Methodology chapter. This is accompanied by the quantitative results of these efforts. Finally, important conclusions
that were drawn from this work are discussed as well as suggestions for the application of this new knowledge and suggestions for future work are offered.

1.1 Objectives

The goal of this research was to not only calibrate the HSM for Louisiana roads but to do this in a way that would determine the effects of various methods of data inclusion and crash definitions. To accomplish this, the following objectives were completed:

1. Develop state-specific calibration factors for the application of the HSM in Louisiana,
2. Perform the calibration in an iterative process with varying inclusion of data elements and different definitions of “intersection” and “non-intersection” crashes,
3. Validate predicted crashes against observed 2012 crash data, and
4. Develop new calibration factors based on 2012 crash data.

1.2 Scope

The development of the HSM calibration factors for the various facility types in this study were based on a statewide database that included all parishes and city areas throughout Louisiana. These segment types included:

- Rural Two-Lane Roads,
- Rural Four-Lane Divided (and Undivided Highways), and
- Urban and Suburban Arterials (including Two-Lane, Three-Lane with Center Two-Way Left-Turn-Lane (TWLTL), Four-Lane Divided and Undivided, and Five-Lane with Center TWLTL facilities).

Four iterations of these eight calibration factors were developed with increasing data inclusion. In the fourth iteration, crashes were classified as “intersection-related” or “non-intersection-related” using three different methods. This resulted in six sets of eight calibration
factors. After these were compared to the actual 2012 crash data, suggestions were made for the application of these values and potential future work.
CHAPTER 2. LITERATURE REVIEW

Prior to initiating the research, a review of literature relevant to the current state of HSM-application practice was conducted. In addition to a summarization of the HSM predictive method, the review also focused on two significant issues, data collection and segment selection, that have arisen when other states have used the HSM predictive method as well as outcomes of the other states’ calibration efforts.

This literature review summarizes the strategies, difficulties, and outcomes of the calibration efforts in various states, with the goal of determining common trends with the calibration process or potential “best practices” for performing calibration. Most importantly, this literature review focuses on how the inclusion or exclusion of data and the assumptions made regarding the definition and inclusion of crashes near intersections affect the resulting calibration factors.

2.1 HSM Predictive Method

One of the key components of the HSM is the predictive method, a computational procedure used to predict the safety of a roadway segment or intersection in terms of its expected crash frequency and severity based on its geometric and roadway characteristics and traffic volumes. This method can be implemented to “reduce the vulnerability of historical crash-based methods to random variations of crash data” (AASHTO, 2010).

To apply the predictive method, safety performance functions (SPFs) found in Part C of the HSM are used to calculate the predicted average crash frequency of a site under base conditions. Chapters 10, 11, and 12 of the HSM list the base conditions assumed for each SPF, including twelve foot lane widths, five driveways per mile for rural two-lane roads, thirty foot median widths for rural multilane divided roads, and no automated speed enforcement (AASHTO,
An SPF is a regression equation that relates crash frequency with exposure, generally average annual daily traffic (AADT). The HSM has separate SPFs for roadway segments and intersections, and the segment SPFs, the focus of this study, predict crashes based on AADT and segment length. Each facility type – rural two-lane, rural multilane, and urban and suburban arterials – has its own set of SPFs specific to the segment type.

While SPFs predict the crashes along a segment under base conditions, crash modification factors (CMFs) are then multiplied by the SPF to account for local variations from the base conditions. A CMF is the ratio of the expected number of crashes with a particular treatment or condition to the expected number of crashes without that treatment or condition. Examples of CMFs include lane and shoulder width adjustments, presence of lighting and rumble strips, and the presence of automated speed enforcement. A CMF value of less than one corresponds with a reduction in crashes (i.e. a road segment with centerline rumble strips), while a CMF value greater than one represents a condition that would increase crashes from the base conditions (i.e. lack of shoulder).

One final step to predict crashes along the segment remains after multiplying the SPF with the appropriate CMFs. The Part C SPFs were developed using national data, which can “vary substantially from one jurisdiction to another for a variety of reasons including climate, driver populations, animal populations, crash reporting thresholds, and crash reporting system procedures” (AASHTO, 2010). Therefore, for the Part C SPFs to provide reliable results for a jurisdiction, they must be calibrated for local application. This can be accomplished through the use of a calibration factor (C) or the development of a unique set of jurisdiction-specific SPFs based on local traffic and road characteristics.
The calibration factor is the ratio of the total number of observed crashes, determined by historical crash data, to the total number of predicted crashes, which is estimated using the predictive method. This represents the difference between the jurisdiction for which the base models were developed and those to which they are being applied. These variations can be caused by various reasons including driver characteristics and crash reporting thresholds. A calibration factor of less than one suggests that the HSM overestimates the average crash frequency of a jurisdiction, while a calibration factor of greater than one suggests that the HSM underestimates the average crash frequency of a jurisdiction. When the calibration factor is used, the predictive method takes the form of the following equation, which can be found in the HSM:

\[ N_{\text{predicted}} = N_{\text{spf,x}} \times (\text{CMF}_{1x} \times \text{CMF}_{2x} \times \ldots \times \text{CMF}_{yx}) \times C_x \]  

where,

- \( N_{\text{predicted}} \) = predicted average crash frequency for a specific year on site type \( x \),
- \( N_{\text{spf,x}} \) = predicted average crash frequency determined for base conditions of the SPF developed for site type \( x \),
- \( \text{CMF}_{yx} \) = crash modification factors specific to site type \( x \) and specific geometric design and traffic control features \( y \), and
- \( C_x \) = calibration factor to adjust SPF for local conditions for site type \( x \).

An alternative approach to making crash prediction results applicable for a particular jurisdiction involves developing jurisdiction-specific SPFs using statistically valid methods such as negative binomial regression. The HSM suggests that when feasible, calibration factors and local SPFs should be developed; this allows each state to determine its preferred method of calibration (AASHTO, 2010). This report focused on the development of calibration factors and not local SPFs; however, both methods are discussed in the following sections.

One of the first and, arguably, most critical steps in the predictive method is the selection of road segments (i.e., “sites”) that are representative of the total population of all roads of similar type in a study location. The HSM recommends a minimum of 30 to 50 sites upon which at least
100 crashes occur annually for each road type. The HSM defines a roadway segment as “a section of continuous traveled way that provides two-way operation of traffic, that is not interrupted by an intersection, and consists of homogeneous geometric and traffic control features” (AASHTO, 2010). Once the segments are selected, the next challenge that arises is collecting the extensive amount of data elements, which can be found in Table A-2 in Volume 2 Appendix A of the HSM. Some of the required data elements include segment length, AADT, lengths and radii of horizontal curves, lane and shoulder width, presence of lighting and on-street parking, and driveway density. Results of prior HSM work showed that a common issue with this process, however, is that state departments of transportation (DOTs) generally do not collect all data elements required for calibration, and they tend to focus on state routes only. Another issue involves the use (or exclusion) of segments with horizontal curves. The HSM requires two additional data elements for horizontal curves on rural two-lane roads, including curve length and radii. Because the inclusion of this information often significantly increases the effort required for data collection, curved segments are often not included in the calibration, further limiting the predictive capability of the HSM predictive method. The challenges of data collection are discussed below in more detail.

2.2 Data Collection

A select number of states have already performed HSM calibration, and the results of their studies are discussed in greater detail in the following sections. One issue that was common among the HSM calibration reports was the difficulty in collecting the required and recommended data elements. These include some data elements that are standard in state roadway databases but also many that needed to be collected using various methods on a segment-by-segment basis.

The University of Florida Transportation Research Center calibrated SPF s to reflect Florida roadway conditions (Srinivasan, Haas, Dhakar, Hormel, Torbic, & Harwood, 2011), and a Utah
Department of Transportation (UDOT) study calibrated the SPF for rural two-lane roadway segments throughout the state of Utah (Saito, Brimley, & Schultz, 2011). Both of these studies included tangent sections along state routes but excluded curved road segments and local roads from the calibration due to data limitations and time restrictions. In Maryland, Morgan State University calibrated all segment facility types for the Maryland State Highway Administration along state routes only (Shin, Lee, & Dadvar, 2014). Curved sections were included despite the additional data collection that was required to do so. An even more comprehensive study was conducted in Oregon with Oregon State University and Portland State University for the Oregon Department of Transportation and the Oregon Transportation Research and Education Consortium. This calibration included curved sections of roadway, and although crashes were initially assigned “intersection-related” or “non-intersection related” based on distance from the intersection (250 ft.), they were then individually evaluated to confirm the classification. This study, like the others, was also limited to state highway system segments due to lack of data on local roads (Dixon, Monsere, Xie, & Gladhill, 2012).

The Utah research team made use of one of their local programs, Roadview Explorer, along with Google Earth™, UDOT traffic tables, crash data, and construction data for data collection (Saito et al., 2011). The Oregon calibration also required multiple sources, including video logs and Google Earth™ to collect the necessary data. Video logs were used to collect driveway density, centerline rumble strips, presence of TWLTL, lighting information, Roadside Hazard Rating (RHR), and sideslope. Google Earth™ was utilized to count driveways and determine the driveway type. The researchers suggested the addition of these missing elements to the Oregon database (Dixon et al., 2012). Because the HSM is still in its infancy, most states have not adjusted their roadway databases to include the required elements for HSM calibration; however, future
inclusion of these data elements would save a great deal of time and effort to perform or update the calibration.

The University of North Carolina Highway Safety Research Center calibrated segment SPFs for the North Carolina Department of Transportation and also had to utilize multiple sources for data collection. These sources included aerial photographs, GIS files, roadway inventory files, and the Traffic Engineering Accident Analysis System database (Srinivasan & Carter, 2011). The Missouri study, which was performed by the University of Missouri for the Missouri Department of Transportation (MoDOT), made use of the MoDOT Transportation Management Systems database, Automated Road Analyzer videos and images, and Google Earth™ to collect the necessary data elements (Sun, Brown, Edara, Claros, & Nam, 2013).

The Maryland study (Shin et al., 2014), a Clemson University study that calibrated rural two-lane roads in Georgia (Alluri, 2010), and an initial Louisiana study performed by the University of Louisiana at Lafayette to calibrate Louisiana’s rural multilane divided and undivided segments (Sun, Magri, Shirazi, Gillella, & Li, 2011) also required the use of multiple data sources. In these previously mentioned studies, Utah, Oregon, North Carolina, Missouri, Maryland, Georgia, and Louisiana, data availability and collection were cited as the most difficult and time-consuming aspects of their projects.

While as many data elements as possible should be collected, this time-consuming task can become quite expensive. Knowing when to stop data collection and how large of a sample size is large enough is vital to having a reliable calibration.

2.3 Tailoring the HSM for Individual Application

It has been mentioned previously that one of the most significant challenges of applying the predictive method is selecting a sample that satisfies the HSM minimum recommended size.
The HSM does allow for individual interpretation of the guidelines and for users to adjust their calibration process to suit their state’s unique needs and characteristics, which can be seen in the studies below.

The Utah DOT study tailored the HSM predictive method to suit their unique road characteristics by removing roads with excessively high AADT or low speeds, stating that these are not representative of Utah’s highways. Although this affects the randomness of the dataset, they justify this by stating that those segments’ “extreme characteristics undermine the predicting capability of the SPFs” (Saito et al., 2011). The North Carolina research team also individualized the HSM predictive method by excluding segments with AADT less than 500 because these segments were deemed unreliable in their coding (Srinivasan & Carter, 2011).

The Oregon study evaluated fatal and injury crashes only because Property Damage Only (PDO) crashes are self-reported in Oregon, meaning a great deal of non-injury crashes remain unreported (Dixon et al., 2012). The Florida study also analyzed fatal and injury crashes alone, since PDO crashes are not included in the Florida crash database (Srinivasan et al., 2011). Due to the differences in crash reporting by different states, in cases such as Oregon and Florida, it was more fitting to predict fatal and injury crashes rather than total crashes.

In the Maryland study, the City of Baltimore was excluded from the calibration. While this limited the applicability of the calibration factors, this showed the importance of recognizing differences in geographical locations throughout the state and how some areas may be excluded for representative results (Shin et al., 2014). All of these studies show how HSM calibration can be customized to fit the needs and characteristics of a particular state or jurisdiction.
2.4 Dataset Size

Another issue with the HSM predictive method is the seemingly random selection for number of target sites (30-50) and crashes (at least 100 annually) occurring on these sites. Due to the varying exposure of each facility type, many states found that one single target number of sites seemed overly simplistic and not practical. GENEX Systems performed a study for the state of Washington which compared various dataset sizes to further analyze this (Banihashemi, 2012). The researchers at GENEX stated that there is “not much research available to provide guidance to the States on how the “goodness” of the calibration factor relates to the size of the calibration dataset” (Banihashemi, 2012).

The Oregon research team also had an issue with the seemingly arbitrary HSM minimum dataset requirements. Because most urban and suburban crashes occur at intersections and not along roadway segments, a large number of segments was required to meet the 100 crash threshold. The project team selected urban and suburban sites until the threshold was reached (Dixon et al., 2012). From the Oregon and Washington research, it is evident that the “one size fits all” approach is not appropriate for different facility types’ calibration dataset sizes. Table 1 below shows the dataset size used by various states, when this information was made available.

The table shows that the majority of states’ dataset sizes remained reasonably close to the HSM minimum requirement; however, the Florida and Alabama studies were able to use thousands of segments in their calibration dataset. In the Florida study, a Python Script was written to automatically divide the state-maintained roads into homogeneous roadway segments. This segmenting procedure created thousands of sites for calibration – about 4,800 for rural two-lane roads, 1,400 for rural multilane highways, and 17,000 urban and suburban segments. This procedure would not have been feasible if done manually; with the Python script, the HSM’s
minimum site requirements were far exceeded (Srinivasan et al., 2011). The University of Alabama completed a study for the Alabama Department of Transportation predicting crashes along Alabama’s rural two-lane roads and four-lane divided segments (Mehta & Lou, 2012). The Critical Analysis Reporting Environment database was used to gather roadway and crash data to perform the calibration. For the rural two-lane roads 5,991 segments were selected out of a total of 64,736 homogeneous segments, and for the four-lane divided highways 4,000 segments were selected out of 10,576 homogeneous segments in the system (Mehta & Lou, 2012). Because additional data collection using multiple sources was not necessary for the Florida and Alabama studies, it was feasible to use thousands of sites for the calibration.

Table 1: Segment Dataset Sizes in Prior Calibration Studies

<table>
<thead>
<tr>
<th>Facility Type</th>
<th>Number of Segments Used in Calibration</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Utah</td>
</tr>
<tr>
<td>Rural Two Lane</td>
<td>157</td>
</tr>
<tr>
<td>Rural Multilane Undivided</td>
<td>-</td>
</tr>
<tr>
<td>Rural Multilane Divided</td>
<td>-</td>
</tr>
<tr>
<td>Urban Two Lane</td>
<td>-</td>
</tr>
<tr>
<td>Urban Three Lane with TWLTL</td>
<td>-</td>
</tr>
<tr>
<td>Urban Four Lane Undivided</td>
<td>-</td>
</tr>
<tr>
<td>Urban Four Lane Divided</td>
<td>-</td>
</tr>
<tr>
<td>Urban Five Lane with TWLTL</td>
<td>-</td>
</tr>
</tbody>
</table>

Facility Type | Georgia | Missouri | Maryland |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Rural Two Lane</td>
<td>399</td>
<td>196</td>
<td>251</td>
</tr>
<tr>
<td>Rural Multilane Undivided</td>
<td>-</td>
<td>-</td>
<td>19*</td>
</tr>
<tr>
<td>Rural Multilane Divided</td>
<td>-</td>
<td>37</td>
<td>160</td>
</tr>
<tr>
<td>Urban Two Lane</td>
<td>-</td>
<td>73</td>
<td>252</td>
</tr>
<tr>
<td>Urban Three Lane with TWLTL</td>
<td>-</td>
<td>-</td>
<td>138</td>
</tr>
<tr>
<td>Urban Four Lane Undivided</td>
<td>-</td>
<td>-</td>
<td>145</td>
</tr>
<tr>
<td>Urban Four Lane Divided</td>
<td>-</td>
<td>66</td>
<td>244</td>
</tr>
<tr>
<td>Urban Five Lane with TWLTL</td>
<td>-</td>
<td>59</td>
<td>115</td>
</tr>
</tbody>
</table>

* Only 19 sites available; did not meet HSM minimum
** Total number of segments for all urban facility types
2.5 Calibration Factors

The outcome of most of the previously mentioned reports was a set of calibration factors for implementation along state routes. These varied from state to state and can be found in Table 2.

Table 2: Calibration Factors from Other States’ Calibrations

<table>
<thead>
<tr>
<th>Facility Type</th>
<th>Utah</th>
<th>Oregon</th>
<th>Florida*</th>
<th>North Carolina</th>
<th>Louisiana**</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rural Two Lane</td>
<td>1.16</td>
<td>0.74</td>
<td>1.005</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Rural Multilane Undivided</td>
<td>-</td>
<td>0.36**</td>
<td>-</td>
<td>-</td>
<td>0.98</td>
</tr>
<tr>
<td>Rural Multilane Divided</td>
<td>-</td>
<td>0.78</td>
<td>0.683</td>
<td>0.97</td>
<td>1.25</td>
</tr>
<tr>
<td>Urban Two Lane</td>
<td>-</td>
<td>0.63</td>
<td>1.025</td>
<td>1.54</td>
<td>-</td>
</tr>
<tr>
<td>Urban Three Lane with TWLTG</td>
<td>-</td>
<td>0.82</td>
<td>1.038</td>
<td>3.62</td>
<td>-</td>
</tr>
<tr>
<td>Urban Four Lane Undivided</td>
<td>-</td>
<td>0.65</td>
<td>0.729</td>
<td>4.04</td>
<td>-</td>
</tr>
<tr>
<td>Urban Four Lane Divided</td>
<td>-</td>
<td>1.43</td>
<td>1.628</td>
<td>3.87</td>
<td>-</td>
</tr>
<tr>
<td>Urban Five Lane with TWLTG</td>
<td>-</td>
<td>0.64</td>
<td>0.669</td>
<td>1.72</td>
<td>-</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Facility Type</th>
<th>Georgia</th>
<th>Missouri</th>
<th>Maryland</th>
<th>Alabama</th>
<th>Washington</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rural Two Lane</td>
<td>0.937</td>
<td>0.82</td>
<td>0.6960</td>
<td>1.392</td>
<td>1.472</td>
</tr>
<tr>
<td>Rural Multilane Undivided</td>
<td>-</td>
<td>-</td>
<td>2.2632***</td>
<td>-</td>
<td>2.035</td>
</tr>
<tr>
<td>Rural Multilane Divided</td>
<td>-</td>
<td>0.98</td>
<td>0.5833</td>
<td>1.103</td>
<td>1.253</td>
</tr>
<tr>
<td>Urban Two Lane</td>
<td>-</td>
<td>0.84</td>
<td>0.6818</td>
<td>-</td>
<td>1.452</td>
</tr>
<tr>
<td>Urban Three Lane with TWLTG</td>
<td>-</td>
<td>-</td>
<td>1.0784</td>
<td>-</td>
<td>1.545</td>
</tr>
<tr>
<td>Urban Four Lane Undivided</td>
<td>-</td>
<td>-</td>
<td>0.8783</td>
<td>-</td>
<td>2.624</td>
</tr>
<tr>
<td>Urban Four Lane Divided</td>
<td>-</td>
<td>0.98</td>
<td>0.8268</td>
<td>-</td>
<td>1.995</td>
</tr>
<tr>
<td>Urban Five Lane with TWLTG</td>
<td>-</td>
<td>0.73</td>
<td>1.1892</td>
<td>-</td>
<td>0.985</td>
</tr>
</tbody>
</table>

* Fatal + Injury Crashes
** Initial Louisiana study performed by ULL
*** Only 19 sites available; did not meet HSM minimum

If only the facility types that met the minimum HSM requirements are considered (disregarding the facility types with less than 30-50 segments), the calibration factors ranged from 0.63 (rural two-lane roads in Oregon) to 4.04 (urban four-lane undivided roads in North Carolina),
with the majority of the calibration factors relatively close to 1.0. While the values of calibration factors for urban roads in North Carolina is surprisingly high, the research team was unable to determine a reason for these extremely high values (Srinivasan & Carter, 2011).

One possible explanation of the differences among the calibration factors is the difference in crash reporting thresholds by state. Each state has its own requirements for filing traffic crash reports. In Oregon, non-injury crashes are self-reported, and $1500 in property damage is required to file a report, with many minor crashes remaining unreported (Dixon et al., 2012). In Florida, PDO crashes are not stored in their crash database, which is why they only performed the calibration using fatal and injury crashes. Table 3 below shows the Property Damage Only crash reporting threshold for each of the states mentioned in this report (NHTSA, 2014). From this table, the vast differences among crash reporting thresholds by state is obvious.

<table>
<thead>
<tr>
<th>PDO Crash Reporting Threshold by State</th>
</tr>
</thead>
<tbody>
<tr>
<td>Utah</td>
</tr>
<tr>
<td>Oregon $1,500</td>
</tr>
<tr>
<td>Florida PDO not stored in system</td>
</tr>
<tr>
<td>North Carolina</td>
</tr>
<tr>
<td>Alabama No Threshold</td>
</tr>
<tr>
<td>Georgia $500</td>
</tr>
<tr>
<td>Missouri $500</td>
</tr>
<tr>
<td>Maryland &quot;Immobilizing&quot; Property Damage</td>
</tr>
<tr>
<td>Louisiana $500</td>
</tr>
<tr>
<td>Washington $700</td>
</tr>
<tr>
<td>Virginia $1,000</td>
</tr>
</tbody>
</table>

Other possible explanations as to the differences between calibration factors by state include driver behavior, road conditions, animal populations, and the method in which each research team classified “intersection” and “non-intersection” crashes.
2.6 Development of Local SPFs

While the focus of this research has been on developing calibration factors to apply to the HSM Part C SPFs, the other option for calibration is developing local SPFs using statistically valid methods. While this method is even more data-intensive than developing a calibration factor, the results are expected to be more reliable, since calibration factors assume a linear relationship between the jurisdiction’s crash frequency and the predicted crash frequency using the nationally-developed SPFs. This section explains the results of studies that have compared both methods of calibration and which method was more representative of local conditions.

In North Carolina, SPFs were developed with AADT as the only independent variable. Because these SPFs are so basic, with only one independent variable, they are most useful for network screening, or identifying locations that would benefit the most from a safety improvement. More complex SPFs were developed for rural two-lane roads with AADT and additional site characteristics including shoulder width and type and terrain (Srinivasan & Carter, 2011). While this study found that North Carolina-specific SPFs would provide more reliable results than calibrating the HSM Part C SPFs, this was not possible for some facility types due to a lack of sufficient data. The research team suggested that when deciding which method to select, the future application of the SPF must be taken into consideration. If the purpose of the SPFs is to help prioritize locations for safety improvements, developing a simple SPF based on AADT alone would be sufficient. However, if the SPFs are to be used to predict crashes for a specific project, it would likely be more straightforward to calibrate the HSM SPFs due to the extensive statistical knowledge required to develop more reliable SPFs with more independent variables (Srinivasan & Carter, 2011). The Georgia study also explained that the future application of the SPFs must be
taken into consideration when selecting a preferred method of calibration. Their study ultimately recommended that local SPFs be implemented along rural two-lane roads (Alluri, 2010).

State-specific SPFs were also developed in Florida for rural two-lane roads and urban and suburban four-lane divided highways, and their accuracy was compared to that of the calibrated SPFs to determine the best method for predicting average crash frequency in Florida (Srinivasan et al., 2011). After the calibration factors and local SPFs were developed, it was determined that crash prediction was not improved overall by using the Florida-specific SPFs. Statewide calibration factors were preferred because they reduced the number of segments with extreme under- or over-predictions. Further future analysis was recommended to determine if the reliability of the state-specific SPFs improved with time (Srinivasan et al., 2011).

State-specific SPFs were developed in Utah for rural two-lane segments using negative binomial and hierarchical Bayesian modeling techniques and compared to calibration factors, with the goal of determining the best method for predicting crashes along Utah rural two-lane highways (Saito et al., 2011). Both techniques were deemed suitable for predicting crashes along Utah’s rural two-lane segments; however, using the HSM SPFs requires applying 12 CMFs, and using the newly-developed negative binomial model only requires four data elements – AADT, segment length, combo-unit truck percentage, and speed limit. While the calibrated HSM model predicted crashes more accurately than the state-specific SPF, it did not provide a substantial enough benefit to make it worth the additional time and effort for data collection, and the local SPF was recommended for statewide use (Saito et al., 2011).

Calibration factors and state-specific SPFs were also developed in Alabama for rural two-lane roads and four-lane divided segments to determine the best method for statewide implementation (Mehta & Lou, 2012). Ultimately this study finds that using a state-specific SPF
with lane width, speed limit, and year as the independent variables provides the most accurate crash prediction results; however, the calibration factor method provides satisfactory results and requires much less effort to carry out (Mehta & Lou, 2012).

The Virginia Transportation Research Council performed a study to predict crashes along rural and urban two-lane segments in Virginia to assist in the prioritization of highway safety improvements (Garber, Haas, & Gosse, 2010). In this study, calibration factors and state-specific SPFs for rural and urban two-lane highways were developed, and the state-specific SPFs provided much more accurate crash prediction results than the calibrated HSM SPFs and are therefore recommended for predicting safety along state routes (Garber et al., 2010).

The Washington study compared calibrated SPFs with newly developed SPFs for rural two-lane roads in Washington. The results of their study showed that the calibrated Part C SPFs work sufficiently well and require less data than developing new models, so new models are not necessary (Banihashemi, 2012).

Overall, there are mixed preferences on calibration methods, with North Carolina (Srinivasan & Carter, 2011) and Georgia (Alluri, 2010) both preferring local SPFs but recommending that users take the future use of the calibration into consideration when selecting a method, and Florida (Srinivasan et al., 2011) preferring the calibration factors to local SPFs. Virginia (Garber et al., 2010) recommended the use of local SPFs because their calibration factors were not well-representative of local conditions, and Alabama (Mehta & Lou, 2012) preferred local SPFs also, although they did find that the calibration factors performed satisfactorily and would be easier for users to implement. Utah (Saito et al., 2011) and Washington (Banihashemi, 2012) both found that the calibration factors performed sufficiently well and were much simpler to carry out than developing new models. Determining the best method of predicting average crash
frequency depends on a variety of factors including the future application of the predictions as well as data availability within each state. Ideally, each state will develop calibration factors and state- or region-specific SPFs to determine which method provides the more reliable results.

2.7 Findings

The site selection and data collection methodology employed in previous state DOT efforts provides a basis upon which the Louisiana calibration factors were calculated. The review of prior work is also useful to address some of the potential issues that may arise during the calibration process, including data limitations, selecting enough segments to meet the HSM requirements, and difficulty in defining crashes as “intersection-” or “non-intersection-related.” The review of prior studies demonstrate that, while there have been numerous applications of the HSM predictive method throughout the country, specific aspects of data inclusion and assumption-making remains highly variable from state-to-state. Unfortunately, the practical quantitative implications of these decisions remain relatively unknown. To address this lack of information, this study will take a systematic approach to evaluate the effects of including and excluding various data elements in Louisiana, as shown in the following chapter.
CHAPTER 3. METHODOLOGY

As the application of the HSM predictive method was new to Louisiana it was recognized that many aspects of the project would require developmental work despite the fact that the computational processes for computing calibration factors have been well-established. The most significant area of need in Louisiana was data availability. The LA DOTD statewide roadway database, while quite comprehensive in some areas (pavement characteristics, lane width, location reference, etc.) was also lacking in others (shoulder surface type, driveway density, lighting, posted speed, embankment slope steepness, etc.). This fact limited the application of the HSM models to their full predictive potential. As a result, the research methodology described in this chapter was developed to address this need and examine the effects of this and other limitations of the manual in its current form.

In this section, the calibration procedure documented in Appendix A of Volume 2 and the predictive method summarized in Chapters 10, 11, and 12 of the HSM as they related to this research are highlighted. This section also describes the issues that arose during the calibration procedure and how they were resolved, including the development of an iterative series of calibration factors to show how the inclusion and exclusion of data affected the results. Ultimately, this became one of the more useful and telling findings of the research and, perhaps, the most valuable contribution of this work to both the LA DOTD and to other agencies seeking to develop their own calibration for their local jurisdictions.

It should be noted that only road segments, not intersections, were included in the calibration development of this research, and only total crashes were evaluated, with no separation of fatal, injury, or PDO crashes. All SPFs and CMFs used in the computational process were taken from Part C of the HSM.
3.1 HSM Calibration Procedure

The five primary steps of the calibration process, as explained in Appendix A of the HSM, are given below:

- identification of facility types for which the applicable Part C predictive model is to be calibrated,
- selection of sites for calibration of the predictive model for each facility type,
- obtaining data for each facility type applicable to a specific calibration period,
- application of the appropriate model to predict total crash frequency for each site during the calibration period as a whole, and
- the calculation of the individual calibration factors for use in the Part C predictive models.

The selection of the eight different types of roadways for calibration was established by the LA DOTD. These included the following roadway classifications:

- Rural Two-Lane Roads,
- Rural Four-Lane Undivided Highways,
- Rural Four-Lane Divided Highways, and
- Urban and Suburban Arterials, including
  - Two-Lane,
  - Three-Lane with Center TWLTL facilities,
  - Four-Lane Undivided,
  - Four-Lane Divided, and
  - Five-Lane with Center TWLTL facilities.

3.2 Data Acquisition and Processing

Once the facility types had been selected, the data collection process began. This step of the project involved the acquisition and organization of roadway design and attribute data files and historical traffic crash data from 2009 through 2012. While the calibration factors were developed using data from the study period, 2009 through 2011, 2012 data was also collected to later evaluate the reliability of the calibration factors. All of these data elements were obtained from the LA DOTD. This data was accessed in digital form using various data files, including a roadway data file, a crash file, an intersection file, and a curve file. Using the LA DOTD roadway data file BM_STL_CONTROLS (last updated on March 14, 2013), segments were organized into eight
separate spreadsheets by facility type using the Highway Class (HIGHWAY_CLASS) identifier.

An illustrative segment of the data file spreadsheet for rural two-lane roads is shown in Figure 1.

![Figure 1: Louisiana Rural Two-Lane Road Data](image)

In Figure 1, the way in which the Louisiana DOTD roadway data file was organized as well as several of the key data variables required for the calculation of the calibration factors can
be seen. These include segment length (SECTION_LENGTH), average daily traffic (ADT), number of lanes (NUM_LANES), lane and shoulder widths (PAVEMENT_WIDTH_PRI and SHOULDER_WIDTH_PRI), and many others. Using this data, disaggregated by facility type, the spreadsheets were then sorted by their average annual daily traffic (AADT). The HSM Part C SPFs are only applicable for specified AADT values. Those segments whose AADT values were not within the range specified by the HSM were excluded from the calibration segments. The maximum allowable AADT values can be found in Table 4, and the minimum values for all facility types is zero vehicles per day.

Table 4: Maximum Allowable AADT for HSM SPF Application

<table>
<thead>
<tr>
<th>Facility Type</th>
<th>Maximum AADT Value (veh/day)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rural Two Lane</td>
<td>17,800</td>
</tr>
<tr>
<td>Rural Multilane Undivided</td>
<td>33,200</td>
</tr>
<tr>
<td>Rural Multilane Divided</td>
<td>89,300</td>
</tr>
<tr>
<td>Urban Two Lane</td>
<td>32,600</td>
</tr>
<tr>
<td>Urban Three Lane with TWLTL</td>
<td>32,900</td>
</tr>
<tr>
<td>Urban Four Lane Undivided</td>
<td>40,100</td>
</tr>
<tr>
<td>Urban Four Lane Divided</td>
<td>66,000</td>
</tr>
<tr>
<td>Urban Five Lane with TWLTL</td>
<td>53,800</td>
</tr>
</tbody>
</table>

All rural two-lane divided, rural five-lane with TWLTL, urban two-lane divided, and urban six- and eight-lane segments from the roadway data file were excluded from the calibration segments because the HSM Part C SPFs did not apply to segments of this type. Next, the Census Category (CENSUS_CATEGORY), Federal Aid Urban Area (FED_AID_URBAN_AREA), and Population Group (POP_GROUP) identifiers were used to further determine whether a segment was located in a rural or urban area. While the Highway Class identifier was initially used to determine the facility type, sometimes these other three categories did not correspond with one another or with the Highway Class, making it difficult to distinguish between rural and urban.
roadway segments. Because it was not feasible to review every segment individually to determine this, when two of the three categories classified a segment as “rural,” the segment was moved to the rural spreadsheet, and likewise for urban segments.

Another concern when determining calibration factors was selecting homogeneous segments. Because the segments in the DOTD roadway database were already split into homogeneous sections with a corresponding AADT, the original segment length was used, rather than dividing or combining any segments.

3.3 Site Selection Methods

The next step of the calibration process was to select segments for calibration. In this research, an iterative series of four calibration factors were computed for each facility type. The purpose of the iterations was to illustrate the effect of making changes to the data collection effort and assumptions with regard to what crashes, segments, and/or data elements to include or exclude. In the first set, the HSM process was followed to the letter. However, it was the least labor-intensive because it used only readily-available and easily-accessible data. Results of later iterations reflected incrementally increasing amounts of data (and corresponding levels of effort) to incorporate greater details into the computational process to reveal the relative changes in the calibration factor values. A description of each iteration of calibration factors is included in the following sections.

3.3.1 Iteration 1

The first iteration of calibration factors reflected a minimal level of effort in terms of data collection and processing. The primary assumption here was to use only the data available in the DOTD roadway and crash databases, and no effort was made to identify or remove crashes which may have been coded “improperly” based on the differences in definitions of “non-intersection-
related” crashes by the police and the HSM. Readily available data included segment length, AADT, shoulder width, and median width. Fifty sample segments were randomly selected for each roadway type. The lone exception were urban three-lane roadways with TWLTLs because there were only 32 such segments in the state database. Random selection was done within the Microsoft® Excel® spreadsheets using a random number generator.

3.3.2 Iteration 2

In the second iteration, only existing information from the LA DOTD roadway database was included in the calculation. However, rural two-lane curved road segments and their associated crashes were excluded due to curve data limitations, and crashes occurring within 250 ft. of the center of the intersection were excluded from the analysis to account for the likely inconsistencies between how police defined “intersection-related” crashes and how it was assumed for the application of the HSM. In Louisiana, neither of these was a trivial matter and a considerable computational effort was required to make both of these changes. Roadway curve data is included in a separate data file, and both that information and crash locations and intersection centroids had to be geospatially located within the LA DOTD roadway files which use a different referencing system.

The 250 foot distance was selected somewhat arbitrarily, but was intended to be consistent with prior HSM calibration efforts (Dixon et al., 2012) (Garber et al., 2010) (Srinivasan & Carter, 2011) and reflect the potential for intersections to influence crashes 250 ft. or eight to ten car lengths beyond an intersection center point due to queueing of vehicles or improper signal timing. Interestingly, though not surprisingly, in many cases the exclusion of intersection-related crashes required the inclusion of a much greater number of segments to reach the HSM-specified 100 annual crash minimum, as shown in Table 5. Because fewer crashes were being assigned as
segment-related, the original fifty segments in three cases (rural two-lane, rural multilane undivided, and urban five-lane) did not include 100 crashes annually; therefore, more segments were selected until the HSM minimum recommended value was reached.

Table 5: Number of Segments Used in Each Iteration of Calibration Factors

<table>
<thead>
<tr>
<th>Facility Type</th>
<th>Iteration 1</th>
<th>Iteration 2</th>
<th>Iteration 3</th>
<th>Iteration 4 (50’, 150’, &amp; 250’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rural Two Lane</td>
<td>50</td>
<td>100</td>
<td>6,188</td>
<td>99</td>
</tr>
<tr>
<td>Rural Multilane Undivided</td>
<td>50</td>
<td>150</td>
<td>219</td>
<td>80</td>
</tr>
<tr>
<td>Rural Multilane Divided</td>
<td>50</td>
<td>50</td>
<td>521</td>
<td>50</td>
</tr>
<tr>
<td>Urban Two Lane</td>
<td>50</td>
<td>50</td>
<td>1,403</td>
<td>30</td>
</tr>
<tr>
<td>Urban Three Lane with TWLTL</td>
<td>32</td>
<td>32</td>
<td>32</td>
<td>32</td>
</tr>
<tr>
<td>Urban Four Lane Undivided</td>
<td>50</td>
<td>50</td>
<td>469</td>
<td>49</td>
</tr>
<tr>
<td>Urban Four Lane Divided</td>
<td>50</td>
<td>50</td>
<td>553</td>
<td>49</td>
</tr>
<tr>
<td>Urban Five Lane with TWLTL</td>
<td>50</td>
<td>226</td>
<td>226</td>
<td>145</td>
</tr>
</tbody>
</table>

3.3.3 Iteration 3

A third iteration of calibration factors was calculated using the same methods of the second iteration, but it included every applicable segment in the LA DOTD database. This was undertaken for several reasons. The most important of these was to reduce the potential skewing effect of the results due to outliers. For example, it was observed that some of the randomly selected segments included significantly more or less crashes than others and, when the sites were selected randomly, it was thought that each sample could vary up or down based on the pure chance inherent in a random selection process. It was assumed that the analyses of the complete dataset would tend to reduce this potential variation from one sample set to another. The second reason was that it was very easy to do with a quick coding change. Rather than stopping the selection process after the minimum number of 100 observed crashes was observed, the selection process was allowed to run until the full applicable database was included in the computation.
3.3.4 Iteration 4

The results of the fourth and final iteration reflect the results of an effort to include as many accessible sources of data as could be identified. Similar to the preceding iterations, sample segments were selected randomly using a random number generator, and several data elements that were not included in the LA DOTD database were collected using Google Earth™ and Google StreetView™. These included key design characteristics like driveway type (major or minor residential, major or minor commercial, or major or minor industrial/institutional) and density, roadway lighting, on-street parking, posted speed, offset and spacing location of fixed objects four inches or great in diameter in the roadside, and approximate embankment slide slope steepness.

The collection of this information was extremely time-consuming and required the conversion of 534 sample segments of varying length from the LA DOTD location referencing system to latitude and longitude coordinates then making visual observations and measurements of each of these key data elements. An example illustration of the process used to classify driveway types and locations is shown in Figure 2. After geo-locating the segment it was necessary to first identify and exclude minor intersection entrances, then identify each driveway by type, then count them and convert them into a segment density. This process was also repeated for each of the other design features and elements for each of the 534 segments in the sample. The amount of segments and level of detailed observation resulted in a labor effort of about three person-months.

Within Iteration 4, the definition of crashes was also varied using three different distances. Under each assumption, observed crashes on the segment were excluded from analysis at distances of 50 ft., 150 ft., and 250 ft. from the center point of the intersection.
The fourth iteration was also used to compare the predicted crash frequency with the actual 2012 crash frequency once the calibration factors had been developed.

3.4 Crash Data

Once all of the data had been collected, the next step was to use the crash data to determine how many crashes occurred on each of the segments. Two sets of crash statistics, one during the study period (2009-2011) and one during the evaluation period (2012) were used. Microsoft® Excel® Macros were written to link the segments in the roadway data file with the corresponding traffic crashes in the crash file based on the control section (CONTROL_SECTION) and logmile from/to (LOGMILE_FROM, LOGMILE_TO) identifiers. Three separate Excel® files were required to complete this process of linking the segments and crashes. The first file (CRASHES) contained the list of all the traffic crashes that occurred in Louisiana from 2009 through 2012. This
file also included relevant information associated with the crashes, such as control section and logmile locations. The second (CURVES) and third (INTERSECTIONS) files contained the list of all curve and intersection locations along state routes, including the control section and logmile to allow the files to be merged with the crash data.

Before adding the crashes along each segment, intersection-related crashes were removed from the CRASHES file because this calibration focused on roadway segments, and the HSM has intersection-related crash SPFs that require a separate calibration for each intersection type. While the crashes in the crash database were coded as “intersection” or “non-intersection”, this was entered by the police officer who recorded the crash report. Because the officer’s definition of intersection crashes only included crashes occurring within the physical threshold boundaries of the intersection created by the radii of return on each of the approach legs, many crashes caused by or associated with the intersection were systematically excluded from this group. An example of such an excluded crash could be a crash 100 ft. before the intersection caused by queueing or a sudden stop made by a car due to improper yellow time at the intersection. After the first iteration of calibration factors were developed based on the crash database coding of intersection or non-intersection crashes, removal of these crashes for the remaining iterations was completed without regard to how the individual crash was coded. Because the coordinates of the crashes were known (CRASHES) and the coordinates of the centers of all intersections were also known (INTERSECTIONS), crashes occurring within a specified distance from the center of the intersection were assumed to have occurred due to factors associated with the intersection and were not included in the segment crash count. For Iterations 2 and 3, crashes within 250 ft. of the center of the intersection were removed, and for Iteration 4, crashes were removed using three different distances, 50 ft., 150 ft., and 250 ft.
In addition, and only for the rural two-lane road type, crashes that occurred within a curve were also excluded from the CRASHES file before running the corresponding crash count. This eliminated the need for additional data collection as the rural two-lane facility type required additional data for horizontal curves. The curve lengths were removed from the overall segment length for the rural two-lane segments; however, the distance for intersection-related crashes (50 ft., 150 ft. or 250 ft.) was not removed from the segment length for this project. Excel® Macros were used to exclude the crashes related to intersections and/or within a curve by comparing the control section and the logmile from the CRASHES file to the control section and the logmile-from and logmile-to of the INTERSECTIONS and/or the CURVES files as shown in Appendix A.

For the fourth iteration, three different crash assessments were performed for each road type category: one without the crashes that occurred within 50 ft. from the intersection, one without those within 150 ft., and one without those within 250 ft. For rural two-lane, however, the crashes that occurred within a curve (and the curve length) were excluded for all of the three crash counts in addition to those that occurred within the intersection influence area. A very important point of note is that individual crash reports were not manually reviewed to examine the cause and exact location of the crashes. Rather, the coordinates of the crashes as entered into the Louisiana database were simply compared to the locations of intersections to determine whether or not the crashes were “intersection-related.” This is a key point because it is recognized that the crash locations labeled in the database were often different from the true “exact” location of the crash. However, a manual review of every crash report would have required reviews of over 2,500 specific crash events representing a time commitment that was beyond the scope of this research.

To count the total number of crashes that occurred within a road type, the control section and the logmile of the CRASHES file were compared to the control section and the logmile-from
and logmile-to of the roadway database and added in a new column. This was done for both 2009-2011 and 2012 separately. A sample segment of the Excel® Macro used to count the crashes for a particular road type can be found in Appendix A.

3.5 Data Compilation and Computation

After organizing the Louisiana road segments by facility type and assigning crashes to the corresponding segments, the next step of the project was to calculate the HSM SPFs and CMFs. Despite its complexity, the highly systematic nature of this computational process lent itself to spreadsheet-oriented calculation. To this end, project-specific macro routines were also written in Microsoft® Excel® to expedite the calibration procedure. These macros can be found in Appendix B of this report. SPFs for base conditions were calculated for every segment within each spreadsheet using the equations found in Chapters 10, 11, and 12 of the HSM. The HSM Part C SPFs used in this process can also be found in the following sections.

3.6 SPFs and CMFs

3.6.1 Rural Two-Lane Segments

The SPF used to predict average crash frequency for rural two-lane road segments was

\[ N_{spf \text{rs}} = AADT \times L \times 365 \times 10^{-6} \times e^{-0.312} \]  

(2)

where,

- \( N_{spf \text{rs}} \) = predicted total average crash frequency for roadway segment base conditions,
- \( AADT \) = average annual daily traffic volume (vehicles per day), and
- \( L \) = length of roadway segment (miles).

This equation was used to calculate the predicted average crash frequency under base conditions. Initially, only the rural two-lane segment CMFs for lane width and shoulder width were applied, and base conditions were assumed for the remaining CMFs because of insufficient data. These CMFs included horizontal curves, grades, driveway density, shoulder type, centerline rumble strips, passing lanes, two-way left-turn lanes, roadside design, lighting, and automated
speed enforcement. Data for the remaining CMFs was gathered using Google Earth™ for each of the selected segments for Iteration 4.

For the rural two-lane lane width and shoulder width CMFs, the default value for “proportion of total crashes constituted by related crashes” (PRA) was used. This value indicated that run-off-the-road, head-on, and sideswipe crashes typically represent 57.4 percent of total crashes for rural two-lane segments. Local data can be used to update these numbers in the future to better reflect Louisiana conditions.

3.6.2 Rural Multilane Highways

The SPF used to predict average crash frequency for rural multilane undivided segments was

$$N_{spf\ ru} = e^{(a + b \cdot \ln (AADT) + \ln (L))}$$ (3)

where,
- \(N_{spf\ ru}\) = predicted total average crash frequency for roadway segment base conditions,
- AADT = average annual daily traffic volume (vehicles per day) on roadway segment,
- L = length of roadway segment (miles), and
- a,b = regression coefficients (-9.653 and 1.176 respectively for total crashes).

The SPF used to predict average crash frequency for rural multilane divided segments was

$$N_{spf\ rd} = e^{(a + b \cdot \ln (AADT) + \ln (L))}$$ (4)

where,
- \(N_{spf\ rd}\) = predicted total average crash frequency for roadway segment base conditions,
- AADT = average annual daily traffic volume (vehicles per day) on roadway segment,
- L = length of roadway segment (miles), and
- a,b = regression coefficients (-9.025 and 1.049 respectively for total crashes).

These SPFs were used to calculate the predicted average crash frequency under base conditions. The rural multilane undivided segment CMFs for lane width and shoulder width were then applied. Initially, base conditions were assumed for the CMFs for shoulder type, sideslopes,
lighting, and automated speed enforcement. In the fourth iteration, data for the shoulder type and lighting CMFs was gathered in Google Earth™. The sideslopes were roughly approximated based on reviews of the road segments in Google Earth™, and no automated speed enforcement was assumed. For the rural multilane divided segments, the lane width, right shoulder width, and median width CMFs were initially applied, but the lighting and automated speed enforcement CMFs could not be calculated due to insufficient data. For the fourth iteration, lighting data was gathered using Google Earth™, and no automated speed enforcement was assumed in all cases.

For the rural multilane lane width and shoulder width CMFs, the default values for “proportion of total crashes constituted by related crashes” (PRF) were used. These values indicate that run-off-the-road, head-on, and sideswipe crashes typically represent 27 percent and 50 percent of total crashes for rural multilane undivided and divided segments, respectively. Local data can be used to update these numbers in the future to better reflect Louisiana-specific conditions.

3.6.3 Urban and Suburban Arterials

The predictive models for urban and suburban roadway segments were as follows:

\[ N_{predicted\, rs} = C_r \times (N_{br} + N_{pedr} + N_{biker}) \]  
\[ N_{br} = N_{spf\, rs} \times (CMF_{1r} \times CMF_{2r} \times \ldots \times CMF_{nr}) \]  
\[ N_{spf\, rs} = N_{bmv} + N_{bsv} + N_{brdwy} \]  
\[ N_{bmv} = \exp (a + b \times \ln(AADT) + \ln(L)) \]

where,
\( N_{predicted\, rs} \) = predicted average crash frequency of an individual roadway segment,
\( C_r \) = calibration factor for roadway segments developed for use for a particular geographical area,
\( N_{br} \) = predicted average crash frequency of an individual roadway segment (excluding vehicle-pedestrian and vehicle-bicycle collisions),
\( N_{pedr} \) = predicted average crash frequency of vehicle-pedestrian collisions for an individual roadway segment,
\( N_{biker} \) = predicted average crash frequency of vehicle-bicycle collisions for an individual roadway segment,
\( N_{spf\, rs} \) = predicted total average crash frequency for roadway segment base conditions (excluding vehicle-pedestrian and vehicle-bicycle collisions),
\( CMF_{1r} \ldots CMF_{nr} \) = crash modification factors for roadway segments,
N_{brmv} = predicted average crash frequency of multiple-vehicle non-driveway collisions for base conditions,
N_{brsv} = predicted average crash frequency of single-vehicle crashes for base conditions,
N_{brdwy} = predicted average crash frequency of multiple-vehicle driveway collisions for base conditions,

AADT = average annual daily traffic volume (vehicles per day) on roadway segment,
L = length of roadway segment (miles), and
a,b = regression coefficients, which for total crashes are as follows

- For multiple vehicle non-driveway collisions
  - 2U: -15.22, 1.68
  - 3T: -12.40, 1.41
  - 4U: -11.63, 1.33
  - 4D: -12.34, 1.36
  - 5T: -9.70, 1.17

- For single vehicle collisions
  - 2U: -5.47, 0.56
  - 3T: -5.74, 0.54
  - 4U: -7.99, 0.81
  - 4D: -5.05, 0.47
  - 5T: -4.82, 0.54

Initially, and in the first three iterations, only two of the five urban/suburban SPFs were estimated, leaving out a majority of the computations required to predict average crash frequency along these segments. Multiple-vehicle non-driveway collisions and single-vehicle crashes were estimated, and the remaining three (multiple-vehicle driveway-related collisions, vehicle-pedestrian collisions, and vehicle-bicycle collisions) were excluded due to lack of data on number of driveways by land-use type and posted speed data. In Iterations 1-3, only the median width CMF was applied, and base conditions were assumed for the remaining CMFs because of insufficient data. These CMFs include on-street parking, roadside fixed objects, lighting, and automated speed enforcement. The missing data was then collected for the fourth iteration on the selected segments using Google Earth™. The driveways were counted individually by type; information was gathered on posted speed, on-street parking, roadside fixed objects, and lighting;
and it was assumed that there was no automated speed enforcement. Table 6 summarizes the data inclusion of each iteration by facility type.

3.7 Calculating Calibration Factors

For each facility type, the applicable SPFs and CMFs were multiplied to obtain the predicted crash frequency for each segment in the sample set. The number of actually observed crashes on the same segments over the three year analysis period was then divided by the predicted crash frequency to compute separate calibration factors for each segment. Finally, the resulting calibration factors for each of these individual segments were then averaged to determine an overall calibration factor for each of the eight facility types in each iteration. The final results from these processes can be found in the next section.
Table 6: Summary of Data Inclusion by Facility Type

<table>
<thead>
<tr>
<th>Data Element</th>
<th>Rural Two-Lane</th>
<th>Rural Multilane</th>
<th>Urban/Suburban Arterials</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Iterations 1-3</td>
<td>Iteration 4</td>
<td>Iterations 1-3</td>
</tr>
<tr>
<td>Segment Length</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average Annual Daily Traffic</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lane Width</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shoulder Type</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shoulder Width</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Presence of Lighting</td>
<td></td>
<td></td>
<td>★</td>
</tr>
<tr>
<td>Driveway Density</td>
<td></td>
<td></td>
<td>★</td>
</tr>
<tr>
<td>Presence of Center TWLTL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Presence of Centerline Rumble Strip</td>
<td></td>
<td></td>
<td>★</td>
</tr>
<tr>
<td>Roadside Hazard Rating</td>
<td></td>
<td></td>
<td>★</td>
</tr>
<tr>
<td>Use of Automated Speed Enforcement</td>
<td></td>
<td></td>
<td>★</td>
</tr>
<tr>
<td>Sideslope</td>
<td></td>
<td></td>
<td>★</td>
</tr>
<tr>
<td>Median Width</td>
<td></td>
<td></td>
<td>★</td>
</tr>
<tr>
<td>No. of Through Traffic Lanes</td>
<td></td>
<td></td>
<td>★</td>
</tr>
<tr>
<td>Presence of Median</td>
<td></td>
<td></td>
<td>★</td>
</tr>
<tr>
<td>No. of Driveways by Land-Use Type</td>
<td></td>
<td></td>
<td>★</td>
</tr>
<tr>
<td>Low Speed vs. Intermediate or High Speed</td>
<td></td>
<td></td>
<td>★</td>
</tr>
<tr>
<td>Presence of On-Street Parking</td>
<td></td>
<td></td>
<td>★</td>
</tr>
<tr>
<td>Type of On-Street Parking</td>
<td></td>
<td></td>
<td>★</td>
</tr>
<tr>
<td>Roadside Fixed Object Density</td>
<td></td>
<td></td>
<td>★</td>
</tr>
</tbody>
</table>

- : Not Applicable to this facility type
- : Data element not included in calibration
- : Data element available in Louisiana Roadway Database
- ★ : Data element gathered in additional data collection effort
CHAPTER 4. DISCUSSION OF RESULTS

The iterative development of calibration factors resulted in six sets of eight calibration factors to represent the effects of the various data inclusion and crash definitions. This chapter presents the results of the four primary objectives of the research. The first two are development of state-specific calibration factors and performing the calibration in an iterative process. Once the calibration factors had been developed, the next step was to estimate how well they predicted crashes on Louisiana roads. This was done using the HSM predictive method, including the calibration factors from Iteration 4 with crashes within 150 ft. of the intersection removed, to predict crashes for the year 2012 along the same segments used in Iteration 4. The 2012 observed crashes were then counted and compared to the predicted values. Finally, the fourth objective, development of new calibration factors based on 2012 crash data along the associated segments, is discussed.

4.1 Results of Calibration Factor Development

The results of the development of Louisiana calibration factors are shown in Table 7. The table presents the calibration factors by iteration number.

For Iteration 1, the most basic of all iterations, it can be seen that three of the four urban multilane roadways had significantly higher calibration factors compared to other facility types. Application of HSM guidance on these results suggests that urban roadways in Louisiana experienced crash rates that were four to six times greater than national averages. However, it is unlikely that these results were representative of actual conditions based on practical experience and observational evidence. While it has been recognized for some time that Louisiana crash rates are generally higher than national norms, rates in excess of four times national averages were not likely. Rather, it was speculated that the elevated values were more a reflection of data that was
not included, how crashes were reported by responding police officers, and how curved segments of road and their associated crashes were coded into the databases. For the urban facility types, the lack of readily-available data, made Iterations 1-3 seem inaccurate in their predictive capabilities.

Table 7: Calibration Factors by Facility Type

<table>
<thead>
<tr>
<th>Facility Type</th>
<th>1st Iteration</th>
<th>2nd Iteration</th>
<th>3rd Iteration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rural Two Lane</td>
<td>2.71</td>
<td>1.11</td>
<td>1.16</td>
</tr>
<tr>
<td>Rural Multilane Undivided</td>
<td>1.43</td>
<td>0.48</td>
<td>0.52</td>
</tr>
<tr>
<td>Rural Multilane Divided</td>
<td>2.88</td>
<td>1.68</td>
<td>1.48</td>
</tr>
<tr>
<td>Urban Two Lane</td>
<td>1.54</td>
<td>1.43</td>
<td>2.38</td>
</tr>
<tr>
<td>Urban Three Lane with TWLTL</td>
<td>4.53</td>
<td>0.14</td>
<td>0.14</td>
</tr>
<tr>
<td>Urban Four Lane Undivided</td>
<td>4.08</td>
<td>1.35</td>
<td>1.37</td>
</tr>
<tr>
<td>Urban Four Lane Divided</td>
<td>6.04</td>
<td>2.77</td>
<td>2.87</td>
</tr>
<tr>
<td>Urban Five Lane with TWLTL</td>
<td>0.38</td>
<td>0.02</td>
<td>0.02</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Facility Type</th>
<th>4th Iteration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rural Two Lane</td>
<td></td>
</tr>
<tr>
<td>Rural Multilane Undivided</td>
<td></td>
</tr>
<tr>
<td>Rural Multilane Divided</td>
<td></td>
</tr>
<tr>
<td>Urban Two Lane</td>
<td></td>
</tr>
<tr>
<td>Urban Three Lane with TWLTL</td>
<td></td>
</tr>
<tr>
<td>Urban Four Lane Undivided</td>
<td></td>
</tr>
<tr>
<td>Urban Four Lane Divided</td>
<td></td>
</tr>
<tr>
<td>Urban Five Lane with TWLTL</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Facility Type</th>
<th>Calibration Factor</th>
<th>Calibration Factor</th>
<th>Calibration Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rural Two Lane</td>
<td>1.19</td>
<td>1.05</td>
<td>0.98</td>
</tr>
<tr>
<td>Rural Multilane Undivided</td>
<td>1.04</td>
<td>0.68</td>
<td>0.49</td>
</tr>
<tr>
<td>Rural Multilane Divided</td>
<td>3.27</td>
<td>2.39</td>
<td>1.73</td>
</tr>
<tr>
<td>Urban Two Lane</td>
<td>3.23</td>
<td>2.00</td>
<td>1.48</td>
</tr>
<tr>
<td>Urban Three Lane with TWLTL</td>
<td>0.25</td>
<td>0.14</td>
<td>0.03</td>
</tr>
<tr>
<td>Urban Four Lane Undivided</td>
<td>3.72</td>
<td>1.70</td>
<td>1.03</td>
</tr>
<tr>
<td>Urban Four Lane Divided</td>
<td>6.20</td>
<td>3.73</td>
<td>2.54</td>
</tr>
<tr>
<td>Urban Five Lane with TWLTL</td>
<td>0.05</td>
<td>0.04</td>
<td>0.02</td>
</tr>
</tbody>
</table>

Notes:
(1) No additional data outside of DOTD database, no curve or “additional” intersection crashes removed
(2) No additional data outside of DOTD database, curves removed from rural two-lane segments, crashes from within 250 ft. of intersections removed
(3) No additional data outside of DOTD database, curves removed from rural two-lane segments, crashes from within 250 ft. of intersections removed, all segments included
(4) Additional data collected, curves removed, crashes occurring within 50 ft., 150 ft., and 250 ft. removed from database
To address the excessively high urban calibration factors from Iteration 1, the assumptions for Iteration 2 were modified, with curves removed for rural two-lane roads and crashes removed within 250 ft. of the center of an intersection instead of simply basing this on the coding in the crash report. Changes from Iteration 1 to Iteration 2 can be seen below in Table 8. The calibration factors on every one of the facility types declined and in some cases quite significantly, particularly on the urban area roadways. Although these declines were expected, the magnitude of the decreases highlight the relative effect of intersection crashes and the way in which their definition is established on the expected number of crashes along these segments.

Table 8: Percent Change between Iteration 1 and 2 Calibration Factors

<table>
<thead>
<tr>
<th>Facility Type</th>
<th>Iteration 1 Calibration Factor</th>
<th>Iteration 2 Calibration Factor</th>
<th>% Change in Calibration Factors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rural Two Lane</td>
<td>2.71</td>
<td>1.11</td>
<td>-59.0%</td>
</tr>
<tr>
<td>Rural Multilane Undivided</td>
<td>1.43</td>
<td>0.48</td>
<td>-66.4%</td>
</tr>
<tr>
<td>Rural Multilane Divided</td>
<td>2.88</td>
<td>1.68</td>
<td>-41.7%</td>
</tr>
<tr>
<td>Urban Two Lane</td>
<td>1.54</td>
<td>1.43</td>
<td>-7.1%</td>
</tr>
<tr>
<td>Urban Three Lane with TWLTL</td>
<td>4.53</td>
<td>0.14</td>
<td>-96.9%</td>
</tr>
<tr>
<td>Urban Four Lane Undivided</td>
<td>4.08</td>
<td>1.35</td>
<td>-66.9%</td>
</tr>
<tr>
<td>Urban Four Lane Divided</td>
<td>6.04</td>
<td>2.77</td>
<td>-54.1%</td>
</tr>
<tr>
<td>Urban Five Lane with TWLTL</td>
<td>0.38</td>
<td>0.02</td>
<td>-94.7%</td>
</tr>
</tbody>
</table>

To address the potential effects of outliers in the Iteration 2 selection group, the assumptions of Iteration 3 were modified to include all available segments in the DOTD roadway database. The results were consistent with expectations and can be seen below in Table 9.

In general, there were relatively modest differences between the inclusion of the minimum required number of segments and all available segments. Most of the calibration factors changed by only a small percentage. Interestingly, however, the urban two-lane roadway classification increased by well over 60 percent from 1.43 to 2.38. When reviews of the data and assessments of the results were completed, it was found that the random selection of Iteration 2 did not contain
any segments with extremely high calibration factors. Specifically, the selection process did not include any DOTD dataset segments among the highest three percent of calibration factors. When these segments with high calibration factors were included in the complete dataset of Iteration 3, the calibration factors increased significantly.

Table 9: Percent Change between Iteration 2 and 3 Calibration Factors

<table>
<thead>
<tr>
<th>Facility Type</th>
<th>Iteration 2 Calibration Factor</th>
<th>Iteration 3 Calibration Factor</th>
<th>% Change in Calibration Factors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rural Two Lane</td>
<td>1.11</td>
<td>1.16</td>
<td>4.5%</td>
</tr>
<tr>
<td>Rural Multilane Undivided</td>
<td>0.48</td>
<td>0.52</td>
<td>8.3%</td>
</tr>
<tr>
<td>Rural Multilane Divided</td>
<td>1.68</td>
<td>1.48</td>
<td>-11.9%</td>
</tr>
<tr>
<td>Urban Two Lane</td>
<td>1.43</td>
<td>2.38</td>
<td>66.4%</td>
</tr>
<tr>
<td>Urban Three Lane with TWLTL</td>
<td>0.14</td>
<td>0.14</td>
<td>0.0%</td>
</tr>
<tr>
<td>Urban Four Lane Undivided</td>
<td>1.35</td>
<td>1.37</td>
<td>1.5%</td>
</tr>
<tr>
<td>Urban Four Lane Divided</td>
<td>2.77</td>
<td>2.87</td>
<td>3.6%</td>
</tr>
<tr>
<td>Urban Five Lane with TWLTL</td>
<td>0.02</td>
<td>0.02</td>
<td>0.0%</td>
</tr>
</tbody>
</table>

In Iteration 4, a considerably increased level of effort went into the data collection, and the definition of “intersection-related” crashes was varied using three different distances, including 50 ft., 150 ft., and 250 ft. from the center of an intersection. The effect of the added data and modified crash definitions can be examined in several different ways.

First the effect of undertaking the “external” data collection effort was examined directly by comparing the results of Iteration 2 with the results of Iteration 4 (250 ft. Removed). Here the intersection-related crash definitions were the same (250 ft.) in both, but the fourth iteration included numerous additional data elements. Interestingly, however, the difference between the calibration factors was relatively small, which is shown in Table 10 below.

In most cases, the significantly increased data collection and evaluation effort only resulted in percent changes of several tenths, or 10 to 15 percent in relative terms. The percentage change in the Urban Three-Lane with TWLTL calibration factor was much larger, but only about a tenth.
in absolute terms. If it is assumed that the fourth iteration results were a more accurate reflection of the true calibration factor values, these results suggest that the three months of effort taken to identify and collect road feature details yielded only marginal benefits.

Table 10: Difference and Percent Change between Iteration 2 and 4 Calibration Factors

<table>
<thead>
<tr>
<th>Facility Type</th>
<th>Iteration 2 Calibration Factor</th>
<th>Iteration 4 (250') Calibration Factor</th>
<th>Difference between Calibration Factors</th>
<th>% Change in Calibration Factors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rural Two Lane</td>
<td>1.11</td>
<td>0.98</td>
<td>-0.13</td>
<td>-11.7%</td>
</tr>
<tr>
<td>Rural Multilane Undivided</td>
<td>0.48</td>
<td>0.49</td>
<td>0.01</td>
<td>2.1%</td>
</tr>
<tr>
<td>Rural Multilane Divided</td>
<td>1.68</td>
<td>1.73</td>
<td>0.05</td>
<td>3.0%</td>
</tr>
<tr>
<td>Urban Two Lane</td>
<td>1.43</td>
<td>1.48</td>
<td>0.05</td>
<td>3.5%</td>
</tr>
<tr>
<td>Urban Three Lane with TWLTL</td>
<td>0.14</td>
<td>0.03</td>
<td>-0.11</td>
<td>-78.6%</td>
</tr>
<tr>
<td>Urban Four Lane Undivided</td>
<td>1.35</td>
<td>1.03</td>
<td>-0.32</td>
<td>-23.7%</td>
</tr>
<tr>
<td>Urban Four Lane Divided</td>
<td>2.77</td>
<td>2.54</td>
<td>-0.23</td>
<td>-8.3%</td>
</tr>
<tr>
<td>Urban Five Lane with TWLTL</td>
<td>0.02</td>
<td>0.02</td>
<td>0.00</td>
<td>0.0%</td>
</tr>
</tbody>
</table>

In terms of the effects of changes of the definition of “intersection-related” crashes, it was not a surprise to see that the further the distance that crashes were regarded to be associated with the intersection, the lower the calibration factors became. The reductions were similar at 20 to 40 percent between the 50 to 150 ft. distance differences and the 150 to 250 ft. distance differences. These differences can be seen in tabular form in Table 11.

In the table, as anticipated, the greatest reductions were observed on urban roadways, especially on undivided roadways, both conditions under which higher number of crashes would be expected to be associated with intersections. Although the HSM does not explicitly define a distance criterion for an “intersection-related” crash, prior studies (Dixon et al., 2012) (Garber et al., 2010) (Srinivasan & Carter, 2011) have used 250 ft. as the cut-off distance. In Louisiana, DOTD officials suggested the use of 150 ft. as it is consistent with intersection influence definitions for other types of operational analyses conducted in the state. Because of this
recommendation by the DOTD, the 2012 validation dataset made use of the calibration factors from Iteration 4 with crashes within 150 ft. of an intersection removed.

<table>
<thead>
<tr>
<th>Facility Type</th>
<th>50' Removed Cal. Factor</th>
<th>150' Removed Cal. Factor</th>
<th>% Change b/t 50' and 150'</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rural Two Lane</td>
<td>1.19</td>
<td>1.05</td>
<td>-11.8%</td>
</tr>
<tr>
<td>Rural Multilane Undivided</td>
<td>1.04</td>
<td>0.68</td>
<td>-34.6%</td>
</tr>
<tr>
<td>Rural Multilane Divided</td>
<td>3.27</td>
<td>2.39</td>
<td>-26.9%</td>
</tr>
<tr>
<td>Urban Two Lane</td>
<td>3.23</td>
<td>2.00</td>
<td>-38.1%</td>
</tr>
<tr>
<td>Urban Three Lane with TWLTL</td>
<td>0.25</td>
<td>0.14</td>
<td>-44.0%</td>
</tr>
<tr>
<td>Urban Four Lane Undivided</td>
<td>3.72</td>
<td>1.70</td>
<td>-54.3%</td>
</tr>
<tr>
<td>Urban Four Lane Divided</td>
<td>6.20</td>
<td>3.73</td>
<td>-39.8%</td>
</tr>
<tr>
<td>Urban Five Lane with TWLTL</td>
<td>0.05</td>
<td>0.04</td>
<td>-20.0%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Facility Type</th>
<th>150' Removed Cal. Factor</th>
<th>250' Removed Cal. Factor</th>
<th>% Change b/t 150' and 250'</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rural Two Lane</td>
<td>1.05</td>
<td>0.98</td>
<td>-6.7%</td>
</tr>
<tr>
<td>Rural Multilane Undivided</td>
<td>0.68</td>
<td>0.49</td>
<td>-27.9%</td>
</tr>
<tr>
<td>Rural Multilane Divided</td>
<td>2.39</td>
<td>1.73</td>
<td>-27.6%</td>
</tr>
<tr>
<td>Urban Two Lane</td>
<td>2.00</td>
<td>1.48</td>
<td>-26.0%</td>
</tr>
<tr>
<td>Urban Three Lane with TWLTL</td>
<td>0.14</td>
<td>0.03</td>
<td>-78.6%</td>
</tr>
<tr>
<td>Urban Four Lane Undivided</td>
<td>1.70</td>
<td>1.03</td>
<td>-39.4%</td>
</tr>
<tr>
<td>Urban Four Lane Divided</td>
<td>3.73</td>
<td>2.54</td>
<td>-31.9%</td>
</tr>
<tr>
<td>Urban Five Lane with TWLTL</td>
<td>0.04</td>
<td>0.02</td>
<td>-50.0%</td>
</tr>
</tbody>
</table>

A final issue of note in the Iteration 4 results were in the calibration factors computed for urban roads with TWLTLs. Both of these facility types showed results that were a small fraction of what would be considered “average” nationally. The reasons for this are not completely understood, although two reasons that have been suggested for this are the relative infrequency of roads with center left turn lanes in Louisiana and the possibility that an overwhelming majority of crashes on these types of roads occur closer to intersections. It should be noted that in other states such as Missouri and North Carolina, when an insufficient number of sites or crashes was available for a particular facility type, those facility types were excluded from analysis (Sun et al., 2013) (Srinivasan & Carter, 2011).
4.2 Results of Calibration Factor Validation

While developing statewide calibration factors in an iterative process for implementation in Louisiana provided new and useful information, validating the predictive capability of these factors was a vital step before they could be used with confidence. The predicated average crash frequencies (SPF*CMFs*C) that were calculated using Iteration 4 (150 ft.) calibration factors were compared to the actual crashes that occurred on the same segments in 2012, and the results of the comparison can be found below in Table 12. An additional calculation was performed, comparing the predicted crashes (SPF*CMFs) without the multiplicative calibration factor applied and the actual 2012 crashes to determine whether or not the calibration factors were beneficial in reliably predicting crashes along Louisiana’s roads.

From these results, it is evident that the predicted crash frequency with the calibration factors applied performed more consistent with actual conditions than the predicted crashes without the calibration factor applied in all but two cases. The most notable of these were the urban three-lane with TWLTL and urban five-lane with TWLTL. In these two cases, crash prediction was still not within an acceptable range. This issue is similar to the conditions previously discussed where many crashes along these segments likely occurred close to intersections and because there were not as many segments of these type in Louisiana.

In three of the facility types, rural multilane divided, urban two-lane, and urban four-lane divided, crashes with the calibration factor applied were all predicted within 10 percent of the actual 2012 value. For rural two-lane and urban four-lane undivided segments, the predicted number of crashes was within 20 percent of their actual values. The rural multilane undivided calibration factor resulted in a crash prediction about 45 percent higher than the actual value. This
facility type, along with urban three-lane and urban five-lane segments, warranted additional evaluation due to the poor predictive capability of those calibration factors.

### Table 12: Percent Change between Predicted Crashes and Actual Crashes

<table>
<thead>
<tr>
<th>Facility Type</th>
<th>Predicted Crashes* (SPF<em>CMFs</em>C)</th>
<th>Actual 2012 Crashes</th>
<th>% Change with Calibration Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rural Two Lane</td>
<td>128</td>
<td>112</td>
<td>14.3%</td>
</tr>
<tr>
<td>Rural Multilane Undivided</td>
<td>86</td>
<td>59</td>
<td>45.8%</td>
</tr>
<tr>
<td>Rural Multilane Divided</td>
<td>218</td>
<td>208</td>
<td>4.8%</td>
</tr>
<tr>
<td>Urban Two Lane</td>
<td>151</td>
<td>138</td>
<td>9.4%</td>
</tr>
<tr>
<td>Urban Three Lane with TWLTL</td>
<td>16</td>
<td>143</td>
<td><strong>-88.8%</strong></td>
</tr>
<tr>
<td>Urban Four Lane Undivided</td>
<td>213</td>
<td>264</td>
<td>-19.3%</td>
</tr>
<tr>
<td>Urban Four Lane Divided</td>
<td>335</td>
<td>365</td>
<td>-8.2%</td>
</tr>
<tr>
<td>Urban Five Lane with TWLTL</td>
<td>57</td>
<td>1677</td>
<td><strong>-96.6%</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Facility Type</th>
<th>Predicted Crashes without Cal. Factor* (SPF*CMFs)</th>
<th>Actual 2012 Crashes</th>
<th>% Change without Calibration Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rural Two Lane</td>
<td>132</td>
<td>112</td>
<td>17.9%</td>
</tr>
<tr>
<td>Rural Multilane Undivided</td>
<td>138</td>
<td>59</td>
<td>133.9%</td>
</tr>
<tr>
<td>Rural Multilane Divided</td>
<td>114</td>
<td>208</td>
<td>-45.2%</td>
</tr>
<tr>
<td>Urban Two Lane</td>
<td>79</td>
<td>138</td>
<td>-42.8%</td>
</tr>
<tr>
<td>Urban Three Lane with TWLTL</td>
<td>60</td>
<td>143</td>
<td><strong>-58.0%</strong></td>
</tr>
<tr>
<td>Urban Four Lane Undivided</td>
<td>134</td>
<td>264</td>
<td>-49.2%</td>
</tr>
<tr>
<td>Urban Four Lane Divided</td>
<td>132</td>
<td>365</td>
<td>-63.8%</td>
</tr>
<tr>
<td>Urban Five Lane with TWLTL</td>
<td>995</td>
<td>1677</td>
<td><strong>-40.7%</strong></td>
</tr>
</tbody>
</table>

* Predicted crashes developed using data from 2009-2011

#### 4.3 Comparison of Calibration Factors

Next, the validation process of Objective 3 was taken a step further in the fourth primary objective of the research. New calibration factors were estimated using the 2012 data rather than the suggested use of multiple years of data and compared to the calibration factors that were developed using 2009-2011 data. This was completed to more closely examine significant calibration factor differences of the urban three-lane and urban five-lane facility types. Although it is preferred to use multiple years for developing calibration factors, the 2013 data was not
available yet. In the future, the calibration factors will need to be redeveloped with additional years of crash data. The actual 2012 crashes were divided by the predicted crashes without the calibration factor applied to estimate a new set of calibration factors. The results of these are shown in Table 13.

From this table, it is evident that the calibration factors changed most significantly for urban three-lane and urban five-lane roads. This is not surprising because these new calibration factors likely predicted crashes more reliably than the 2009-2011 calibration factors. Also, the updated rural multilane undivided calibration factor changed by 37 percent. This is likely an indication that the initial calibration factors for these three facility types were likely the least reliable.

<table>
<thead>
<tr>
<th>Facility Type</th>
<th>Calibration Factor (2009-2011 data)</th>
<th>New Calibration Factor (2012 data)</th>
<th>% Change in Calibration Factors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rural Two Lane</td>
<td>1.05</td>
<td>0.85</td>
<td>-19.2%</td>
</tr>
<tr>
<td>Rural Multilane Undivided</td>
<td>0.68</td>
<td>0.43</td>
<td>-37.1%</td>
</tr>
<tr>
<td>Rural Multilane Divided</td>
<td>2.39</td>
<td>1.82</td>
<td>-23.7%</td>
</tr>
<tr>
<td>Urban Two Lane</td>
<td>2.00</td>
<td>1.75</td>
<td>-12.7%</td>
</tr>
<tr>
<td>Urban Three Lane with TWLTL</td>
<td>0.14</td>
<td>2.38</td>
<td>1602.4%</td>
</tr>
<tr>
<td>Urban Four Lane Undivided</td>
<td>1.70</td>
<td>1.97</td>
<td>15.9%</td>
</tr>
<tr>
<td>Urban Four Lane Divided</td>
<td>3.73</td>
<td>2.77</td>
<td>-25.9%</td>
</tr>
<tr>
<td>Urban Five Lane with TWLTL</td>
<td>0.04</td>
<td>1.69</td>
<td>4113.6%</td>
</tr>
</tbody>
</table>

When combined, all of these results suggest the need to include as much data as possible when applying the HSM process for calibration factor development. This also shows the need for increased consistency in “intersection crash” definition, so that the results of the HSM safety analyses will be consistent within and between states and local jurisdictions. These and other key findings and conclusions will be discussed in additional detail in the following, final, chapter of this thesis.
CHAPTER 5. SUMMARY AND CONCLUSIONS

While the initial goal of this research was to develop HSM calibration factors for various roadway facility types in Louisiana, the results of these initial efforts suggested the need for more detailed analyses. After beginning the research it became apparent that calibration factors can vary fairly significantly based on the inclusion or exclusion of certain data elements as well as assumptions and interpretations of how crashes are reported by police and are classified within traffic safety databases. In Louisiana, similar to other states, this variability can become problematic because of the lack of formally established or widely accepted methods on which to base decisions of what can and should be included in HSM computations and how these decisions end up affecting outcomes of the predictive equations.

Based on the need for better information regarding HSM calibration factor development, this research demonstrated the effect of including or excluding various data elements and varying crash definition assumptions. To investigate these effects, HSM predictive model calibration factors were calculated using a series of iterations in which the amount of data and assumed crash conditions were varied from one iteration to the next. The results of these comparative assessments demonstrate the extent to which the variability and sensitivity of HSM calibration factors were impacted by the inclusion of data that may or may not be included in roadway databases and how crashes were included or excluded based on their distance from the intersection.

5.1 General Findings

In addition to the specific outcomes of the comparisons, several broader conclusions can be reached from the quantitative results. Among the general findings was that the more data that was included in the computational process, the lower the calibration factors became. Thus, if it is assumed that the most accurate calibration factors result from using the greatest amount of data,
the calibration factors from Iteration 4 represent the best estimate of actual conditions in Louisiana. A somewhat unexpected finding, however, was that reasonably accurate calibration factors may be attainable without enormous time expenditures. Specifically, it was found that calibration factors remained within 15 percent for five of the eight facility types and within several tenths on two of the other three facility types. The practical implications of these general findings are thought to be significant. For example, major efforts of data collection and process may not be necessary for sketch-plan levels of analysis where only rough estimates are required.

Another interesting general finding was the importance of defining “intersection-related” crashes. The comparative results for the three exclusion distances of Iteration 4 showed that calibration factors for nearly all of the facility types were reduced by 20 to 40 percent as crashes were excluded from 50 to 150 ft. A similar result was also observed when crashes were excluded from 150 to 250 ft.

In terms of the resulting practical implications of these broad findings, it is thought that they can have mixed results on safety analyses. Basically, the more crashes that are occurring, the greater the effect of the calibration factor. For example, if it is assumed that a sample of 50 segments had a total of 300 crashes over a three year period this would be an average of 2 crashes per year per segment. Thus, a calibration factor difference 1.37 to 1.70 would increase the average expected number of crashes from 2.74 to 3.40 per year on any particular segment, or a difference of 0.66 crashes per year. In a sample where many more crashes occurred along the same number of segments, for instance 50 segments with 500 crashes in three years would be approximately 3.33 crashes per year per segment, and the prediction estimate would be more pronounced. Under these conditions, the same calibration factor difference would increase the average expected number of crashes from 4.57 to 5.67 per year per segment, or a difference of 1.1 crashes per year.
On segments where no crashes occurred, the effect of using any calibration factors would always be zero. All of these calculations and findings from the iterative calibration process suggest that the substantial amounts of additional work may not be necessary or beneficial enough when trying to develop quick estimates of a segment’s predicted average crash frequency.

5.2 Calibration Factor Recommendations

Because the original calibration factors were developed using multiple years of data, it was hypothesized that reducing the effects of one year of particularly high or low crash frequency would provide more representative crash prediction results. To test this idea, calibration factors were calculated based solely on 2012 crash data to determine if actual conditions for the three facility types with the poorest predicting ability from the initial calibration factors, rural multilane undivided, urban three-lane with TWLTL, and urban five-lane with TWLTL segments, are more representative. Table 14 below shows the final set of calibration factors for the state of Louisiana. Consistent with empirical observation in Louisiana, for three of the facility types, shown in bold, the 2012 calibration factors likely represent a more realistic reflection of field conditions in Louisiana, although these should be updated when more years of data are available.

<table>
<thead>
<tr>
<th>Facility Type</th>
<th>Calibration Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rural Two Lane</td>
<td>1.05</td>
</tr>
<tr>
<td>Rural Multilane Undivided</td>
<td>0.43*</td>
</tr>
<tr>
<td>Rural Multilane Divided</td>
<td>2.39</td>
</tr>
<tr>
<td>Urban Two Lane</td>
<td>2.00</td>
</tr>
<tr>
<td>Urban Three Lane with TWLTL</td>
<td>2.38*</td>
</tr>
<tr>
<td>Urban Four Lane Undivided</td>
<td>1.70</td>
</tr>
<tr>
<td>Urban Four Lane Divided</td>
<td>3.73</td>
</tr>
<tr>
<td>Urban Five Lane with TWLTL</td>
<td>1.69*</td>
</tr>
</tbody>
</table>

* Based on 2012 Crash Data
Despite the extensive amount of work that was required to develop these calibration factors, aspects of the calibration process can likely be developed to tailor the calibration factors to Louisiana-specific conditions even further. Although these were not tested as part of this research, the following section includes suggestions for future improvement.

5.3 Future Work

Based on the significance of the differences between calibration factors developed in Iteration 4 and the lack of a specific HSM definition of “intersection-related” crashes, it is suggested that future studies may seek to review crash reports individually in detail or establish a defensible and consistent criteria to classify crashes as “intersection-related” or “non-intersection-related.” As the results of this thesis show, there is a trade-off between the level of effort for data inclusion and the perceived accuracy of the calibration factors. While crash reports would ideally be individually reviewed, this would be a highly time-consuming effort and still dependent upon the accuracy of the police reporting in the crash reports.

Another avenue of future research is the calibration of intersections. These locations require even more data elements than segments. Calculating the intersection calibration factors in a similar fashion, with the different methods of classifying crashes, would likely further explain the effects of classifying crashes as “intersection-related” or “non-intersection-related.”

Another area of need is the possibility of including more in-depth statistical analyses and incorporating standard deviation, over dispersion parameters, etc. In Louisiana, there was a suggestion to undertake further random selection of segments until the standard deviation reaches an “acceptable” range of error. This would require additional data collection on a new set of segments, similar to that of Iteration 4.
It should also be noted that the process of developing state-specific SPFs, a task already performed by a few other states, is currently underway in Louisiana. This effort is being undertaken by others, but once completed, the state-specific SPFs should be compared to the calibration factors developed in this research to determine if one method is preferred over the other.

A final area of potential future need is in assessing crash severity using HSM techniques. Due to the difference in crash reporting thresholds by state, it may be more beneficial to predict fatal and injury crashes separately without inclusion of Property Damage Only crashes. This would also allow for inclusion of Louisiana-specific crash type distributions, which would make the calibration even more inclusive of Louisiana data. Another option in the future would be to develop regional calibration factors, for example North Louisiana calibration factors, Baton Rouge calibration factors, etc. This would adjust the calibration factors for regional differences, including methods of crash reporting, driver behavior, and differences in weather.

While crashes occur randomly and cannot be predicted with 100 percent confidence, the methods documented in the Highway Safety Manual represent an improvement over previous methods of evaluating highway safety. As the HSM is still in its relative infancy, the results of this research will improve future applications of the HSM for location-specific safety analyses and support further refinement of future editions of the manual to address many of the issues discussed in this thesis.
REFERENCES


Garber, N., Haas, P., & Gosse, C. (2010). Development of Safety Performance Functions for Two-Lane Roads Maintained by the Virginia Department of Transportation (pp. 64).


Sun, C., Brown, H., Edara, P., Claros, B., Nam, K. (2013). Calibration of the Highway Safety Manual for Missouri. Missouri Department of Transportation; Missouri Department of Transportation; Mid-America Transportation Center; Research and Innovative Technology Administration (pp. 259).
APPENDIX A: MACROS USED FOR CRASH, CURVE, AND INTERSECTION ANALYSIS

Sub Intersections250ft()
    Dim i As Long
    Dim j As Long
    Dim ControlSection As String
    Dim LogMile As Double
    Dim IntersectionBook As Excel.Workbook
    Dim FileBook As Excel.Workbook

    Set FileBook = ActiveWorkbook
    Set IntersectionBook = Workbooks.Open("INTERSECTIONS.xlsx")

    i = 2
    j = 2

    Do Until 53698 = i - 1
        ControlSection = FileBook.Sheets("CRASHES").Cells(i, 4)
        LogMile = FileBook.Sheets("CRASHES").Cells(i, 5).Value
        Do Until 44002 = j - 1
            If IntersectionBook.Sheets("INT").Cells(j, 12) = ControlSection Then
                If LogMile >= IntersectionBook.Sheets("INT").Cells(j, 46) Then
                    If LogMile <= IntersectionBook.Sheets("INT").Cells(j, 47) Then
                        FileBook.Sheets("CRASHES").Cells(i, 23).Value = 1
                    End If
                End If
            End If
            j = j + 1
        Loop
        j = 2
        Do Until 4099 = j - 1
            If IntersectionBook.Sheets("INT2").Cells(j, 23) = ControlSection Then
                If LogMile >= IntersectionBook.Sheets("INT2").Cells(j, 46) Then
                    If LogMile <= IntersectionBook.Sheets("INT2").Cells(j, 47) Then
                        FileBook.Sheets("CRASHES").Cells(i, 23).Value = 1
                    End If
                End If
            End If
            j = j + 1
        Loop
        j = 2
        i = i + 1
    Loop
End Sub

Figure A-1: Excel® Macro Used to Remove the Crashes Occurring Within 250 ft. of an Intersection
Sub CrashOnCurve()
    Dim i As Long
    Dim j As Long
    Dim ControlSection As String
    Dim LogMile As Long
    Dim RangeFrom As Long
    Dim RangeTo As Long
    Dim CrashCurveTotal As Long
    Dim CrashBook As Excel.Workbook
    Dim CurveBook As Excel.Workbook

    Set CrashBook = ActiveWorkbook
    Set CurveBook = Workbooks.Open("C:\\Users\\jrodri57\\Desktop\\CRASH OUT\\All_Curves_Statewide_ARC.xlsx")

    i = 2
    j = 2

    Do Until 53698 = i - 1
        ControlSection = CrashBook.Sheets("CRASHES").Cells(i, 4)
        LogMile = CrashBook.Sheets("CRASHES").Cells(i, 5)

        Do Until 67111 = j - 1
            If CurveBook.Sheets("CURVES").Cells(j, 5) = ControlSection Then
                RangeFrom = CurveBook.Sheets("CURVES").Cells(j, 35).Value
                RangeTo = CurveBook.Sheets("CURVES").Cells(j, 36).Value
                If LogMile >= RangeFrom Then
                    If LogMile < RangeTo Then
                        CrashCurveTotal = CrashCurveTotal + 1
                        CrashBook.Sheets("CRASHES").Cells(i, 14) = "YES"
                    End If
                End If
            End If
            j = j + 1
        Loop
        j = 2

        CrashBook.Sheets("CRASHES").Cells(2, 17) = CrashCurveTotal
        If CrashBook.Sheets("CRASHES").Cells(i, 14) = "" Then
            CrashBook.Sheets("CRASHES").Cells(i, 14) = "NO"
        End If
    Loop
End Sub

Figure A-2: Excel® Macro Used to Remove the Crashes Occurring Within a Curve for Rural Two-Lane Roads
Sub CrashCount1500
Dim i As Long
Dim j As Long
Dim ControlSectionFile As String
Dim RangeFrom As Double
Dim RangeTo As Double
Dim CrashTotal As Long
Dim CrashBook As Excel.Workbook
Dim FileBook As Excel.Workbook

Set FileBook = ActiveWorkbook
Set CrashBook = Workbooks.Open("CRASH_DATA_FOR_RURAL2_LANE_150FT.xlsx")
i = 2
j = 2
Do Until 67111 = i - 1

ControlSectionFile = FileBook.Sheets("RURAL_TWO_LANE").Cells(i, 52)
RangeFrom = FileBook.Sheets("RURAL_TWO_LANE").Cells(i, 53).Value
RangeTo = FileBook.Sheets("RURAL_TWO_LANE").Cells(i, 54).Value

Do Until 67111 = j - 1
If CrashBook.Sheets("CRASHES").Cells(j, 4) = ControlSectionFile Then
  If CrashBook.Sheets("CRASHES").Cells(j, 5) >= RangeFrom Then
    CrashTotal = CrashTotal + 1
  End If
End If
j = j + 1
Loop

FileBook.Sheets("RURAL_TWO_LANE").Cells(i, 85) = CrashTotal
i = i + 1
CrashTotal = 0
Loop
End Sub

Figure A-3: Excel® Macro Used to Count Crashes for Rural Two-Lane Without Crashes on Curves or Within 150 ft. of Intersection
APPENDIX B: MACROS USED IN CALIBRATION

Rural Two-Lane Segments

The rural two-lane segment CMFs for lane width and shoulder width were calculated using the following Macros.

Lane Width Code

```
Select Case LaneWidth
    Case Is <= 9
        Select Case AADT
            Case Is < 400
                Cells(i, 68).Value = 1.05
            Case 400 To 2000
                Cells(i, 68).Value = 1.05 + 0.000281 * (AADT - 400)
            Case Is > 2000
                Cells(i, 68).Value = 1.5
        End Select
    Case 10
        Select Case AADT
            Case Is < 400
                Cells(i, 68).Value = 1.02
            Case 400 To 2000
                Cells(i, 68).Value = 1.02 + 0.000175 * (AADT - 400)
            Case Is > 2000
                Cells(i, 68).Value = 1.3
        End Select
    Case 11
        Select Case AADT
            Case Is < 400
                Cells(i, 68).Value = 1.01
            Case 400 To 2000
                Cells(i, 68).Value = 1.01 + 2.5e-05 * (AADT - 400)
            Case Is > 2000
                Cells(i, 68).Value = 1.05
        End Select
    Case Is >= 12
        Cells(i, 68).Value = 1
End Select
```

Shoulder Width Code

```
Select Case ShoulderWidth
    Case 0
```

56
Select Case AADT
  Case Is < 400
    Cells(i, 68).Value = 1.1
  Case 400 To 2000
    Cells(i, 68).Value = 1.1 + 0.00025 * (AADT - 400)
  Case Is > 2000
    Cells(i, 68).Value = 1.5
End Select
Case 1
Select Case AADT
  Case Is < 400
    Cells(i, 68).Value = 1.085
  Case 400 To 2000
    Cells(i, 68).Value = 1.085 + 0.0001965 * (AADT - 400)
  Case Is > 2000
    Cells(i, 68).Value = 1.4
End Select
Case 2
Select Case AADT
  Case Is < 400
    Cells(i, 68).Value = 1.07
  Case 400 To 2000
    Cells(i, 68).Value = 1.07 + 0.000143 * (AADT - 400)
  Case Is > 2000
    Cells(i, 68).Value = 1.3
End Select
Case 3
Select Case AADT
  Case Is < 400
    Cells(i, 68).Value = 1.045
  Case 400 To 2000
    Cells(i, 68).Value = 1.045 + 0.000112125 * (AADT - 400)
  Case Is > 2000
    Cells(i, 68).Value = 1.225
End Select
Case 4
Select Case AADT
  Case Is < 400
    Cells(i, 68).Value = 1.02
  Case 400 To 2000
    Cells(i, 68).Value = 1.02 + 0.00008125 * (AADT - 400)
  Case Is > 2000
    Cells(i, 68).Value = 1.15
End Select
Case 5
Select Case AADT
Case Is < 400
    Cells(i, 68).Value = 1.01
Case 400 To 2000
    Cells(i, 68).Value = 1.01 + 0.000040625 * (AADT - 400)
Case Is > 2000
    Cells(i, 68).Value = 1.075
End Select
Case 6
    Cells(i, 68).Value = 1
Case 7
    Select Case AADT
        Case Is < 400
            Cells(i, 68).Value = 0.99
        Case 400 To 2000
            Cells(i, 68).Value = 0.99 + 0.000034375 * (AADT - 400)
        Case Is > 2000
            Cells(i, 68).Value = 0.935
    End Select
Case Is >= 8
    Select Case AADT
        Case Is < 400
            Cells(i, 68).Value = 0.98
        Case 400 To 2000
            Cells(i, 68).Value = 0.98 + 0.00006875 * (AADT - 400)
        Case Is > 2000
            Cells(i, 68).Value = 0.87
    End Select
End Select

**Rural Multilane Highways**

The rural multilane undivided segment CMFs for lane width and shoulder width were calculated using the following Macros.

**Rural Multilane Undivided Lane Width Code**

Select Case LaneWidth
    Case Is <= 9.25
        Select Case AADT
            Case Is < 400
                Cells(i, 66).Value = (1.04 - 1) * 0.27 + 1
            Case 400 To 2000
                Val = 1.04 + 0.0000213 * (AADT - 400)
            Case Is > 2000
                Cells(i, 66).Value = 1.04 + 0.0000213 * (AADT - 400)
Val = (Val - 1) * 0.27
Cells(i, 66).Value = Val + 1
Case Is > 2000
Cells(i, 66).Value = (1.38 - 1) * 0.27 + 1
End Select
Case 9.26 To 9.75
Select Case AADT
Case Is < 400
Cells(i, 66).Value = (1.03 - 1) * 0.27 + 1
Case 400 To 2000
Val = 1.03 + 0.000172 * (AADT - 400)
Val = (Val - 1) * 0.27
Cells(i, 66).Value = Val + 1
Case Is > 2000
Cells(i, 66).Value = (1.305 - 1) * 0.27 + 1
End Select
Case 9.76 To 10.25
Select Case AADT
Case Is < 400
Cells(i, 66).Value = (1.02 - 1) * 0.27 + 1
Case 400 To 2000
Val = 1.02 + 0.00031 * (AADT - 400)
Val = (Val - 1) * 0.27
Cells(i, 66).Value = Val + 1
Case Is > 2000
Cells(i, 66).Value = (1.23 - 1) * 0.27 + 1
End Select
Case 10.26 To 10.75
Select Case AADT
Case Is < 400
Cells(i, 66).Value = (1.015 - 1) * 0.27 + 1
Case 400 To 2000
Val = 1.015 + 0.0000749 * (AADT - 400)
Val = (Val - 1) * 0.27
Cells(i, 66).Value = Val + 1
Case Is > 2000
Cells(i, 66).Value = (1.135 - 1) * 0.27 + 1
End Select
Case 10.76 To 11.25
Select Case AADT
Case Is < 400
Cells(i, 66).Value = (1.01 - 1) * 0.27 + 1
Case 400 To 2000
Val = 1.01 + 0.0000188 * (AADT - 400)
Val = (Val - 1) * 0.27
Cells(i, 66).Value = Val + 1
Case Is > 2000
   Cells(i, 66).Value = (1.04 - 1) * 0.27 + 1
End Select
Case Is >= 11.26
   Cells(i, 66).Value = 1
End Select

Rural Multilane Undivided Shoulder Width Code

Select Case ShoulderWidth
Case 0
   Select Case AADT
      Case Is < 400
         Cells(i, 67).Value = (1.1 - 1) * 0.27 + 1
      Case 400 To 2000
         Val = 1.01 + 0.00025 * (AADT - 400)
         Val = (Val - 1) * 0.27
         Cells(i, 67).Value = Val + 1
      Case Is > 2000
         Cells(i, 67).Value = (1.5 - 1) * 0.27 + 1
   End Select
End Select
Case 1
   Select Case AADT
      Case Is < 400
         Cells(i, 67).Value = (1.085 - 1) * 0.27 + 1
      Case 400 To 2000
         Val = 1.085 + 0.000197 * (AADT - 400)
         Val = (Val - 1) * 0.27
         Cells(i, 67).Value = Val + 1
      Case Is > 2000
         Cells(i, 67).Value = (1.4 - 1) * 0.27 + 1
   End Select
End Select
Case 2
   Select Case AADT
      Case Is < 400
         Cells(i, 67).Value = (1.07 - 1) * 0.27 + 1
      Case 400 To 2000
         Val = 1.07 + 0.000143 * (AADT - 400)
         Val = (Val - 1) * 0.27
         Cells(i, 67).Value = Val + 1
      Case Is > 2000
         Cells(i, 67).Value = (1.3 - 1) * 0.27 + 1
   End Select
End Select
Case 3
   Select Case AADT
      Case Is < 400

Cells(i, 67).Value = (1.045 - 1) * 0.27 + 1
Case 400 To 2000
Val = 1.045 + 0.000112 * (AADT - 400)
Val = (Val - 1) * 0.27
Cells(i, 67).Value = Val + 1
Case Is > 2000
Cells(i, 67).Value = (1.22 - 1) * 0.27 + 1
End Select
Case 4
Select Case AADT
Case Is < 400
Cells(i, 67).Value = (1.02 - 1) * 0.27 + 1
Case 400 To 2000
Val = 1.02 + 0.00008125 * (AADT - 400)
Val = (Val - 1) * 0.27
Cells(i, 67).Value = Val + 1
Case Is > 2000
Cells(i, 67).Value = (1.15 - 1) * 0.27 + 1
End Select
Case 5, 6, 7
Cells(i, 67).Value = 1
Case Is >= 8
Select Case AADT
Case Is < 400
Cells(i, 67).Value = (0.98 - 1) * 0.27 + 1
Case 400 To 2000
Val = 0.98 - 0.00006875 * (AADT - 400)
Val = (Val - 1) * 0.27
Cells(i, 67).Value = Val + 1
Case Is > 2000
Cells(i, 67).Value = (0.87 - 1) * 0.27 + 1
End Select
End Select

For the rural multilane divided segments, the lane width, right shoulder width, and median width CMFs were calculated using the following Macros.

**Rural Multilane Divided Lane Width Code**

Select Case LaneWidth
Case Is <= 9.5
Select Case AADT
Case Is < 400
Cells(i, 68).Value = (1.03 - 1) * 0.5 + 1
Case 400 To 2000
Val = 1.03 + 0.000138 * (AADT - 400)
Val = (Val - 1) * 0.5
Cells(i, 68).Value = Val + 1
Case Is > 2000
Cells(i, 68).Value = (0.87 - 1) * 0.5 + 1
End Select
End Select
Val = (Val - 1) * 0.5
Cells(i, 68).Value = Val + 1
Case Is > 2000
Cells(i, 68).Value = (1.25 - 1) * 0.5 + 1
End Select
Case 9.6 To 10.5
Select Case AADT
Case Is < 400
Cells(i, 68).Value = (1.01 - 1) * 0.5 + 1
Case 400 To 2000
Val = 1.01 + 8.75e-05 * (AADT - 400)
Val = (Val - 1) * 0.5
Cells(i, 68).Value = Val + 1
Case Is > 2000
Cells(i, 68).Value = (1.15 - 1) * 0.5 + 1
End Select
Case 10.6 To 11.5
Select Case AADT
Case Is < 400
Cells(i, 68).Value = (1.01 - 1) * 0.5 + 1
Case 400 To 2000
Val = 1.01 + 1.25e-05 * (AADT - 400)
Val = (Val - 1) * 0.5
Case Is > 2000
Cells(i, 68).Value = (1.03 - 1) * 0.5 + 1
End Select
Case Is >= 11.6
Cells(i, 68).Value = 1
End Select

**Rural Multilane Divided Shoulder Width Code**

Select Case ShoulderWidth
Case 0
Cells(i, 70).Value = 1.18
Case 1
Cells(i, 70).Value = 1.16
Case 2
Cells(i, 70).Value = 1.13
Case 3
Cells(i, 70).Value = 1.11
Case 4
Cells(i, 70).Value = 1.09
Case 5
Cells(i, 70).Value = 1.07
Case 6
Cells(i, 70).Value = 1.04
Case 7
  Cells(i, 70).Value = 1.02
Case Is >= 8
  Cells(i, 70).Value = 1
End Select

**Rural Multilane Divided Median Width Code**

Select Case MedianWidth
  Case 0 To 14
    Cells(i, 69).Value = 1.04
  Case 15 To 24
    Cells(i, 69).Value = 1.02
  Case 25 To 34
    Cells(i, 69).Value = 1
  Case 35 To 44
    Cells(i, 69).Value = 0.99
  Case 45 To 54
    Cells(i, 69).Value = 0.97
  Case 55 To 64
    Cells(i, 69).Value = 0.96
  Case 65 To 74
    Cells(i, 69).Value = 0.96
  Case 75 To 84
    Cells(i, 69).Value = 0.95
  Case 85 To 94
    Cells(i, 69).Value = 0.94
  Case Is >= 95
    Cells(i, 69).Value = 0.94
End Select

**Urban and Suburban Arterials**

The urban/suburban segment CMF for median width was calculated using the follow Macro.

Select Case MedianWidth
  Case 0 To 11
    Cells(i, 70).Value = 1.01
  Case 12 To 18
    Cells(i, 70).Value = 1
  Case 19 To 24
    Cells(i, 70).Value = 0.99
  Case 25 To 34
    Cells(i, 70).Value = 0.98
  Case 35 To 44
Cells(i, 70).Value = 0.97
Case 45 To 54
  Cells(i, 70).Value = 0.96
Case 55 To 64
  Cells(i, 70).Value = 0.95
Case 65 To 74
  Cells(i, 70).Value = 0.94
Case 75 To 84
  Cells(i, 70).Value = 0.93
Case 85 To 94
  Cells(i, 70).Value = 0.93
Case Is >= 95
  Cells(i, 70).Value = 0.92
End Select
VITA

Bridget Scheyd Robicheaux, a native of New Orleans, Louisiana, received her bachelor’s degree at Louisiana State University in 2007. After graduation, she worked as an Engineer Intern for ABMB Engineers, Inc. and the Louisiana Department of Transportation and Development on numerous traffic engineering projects. During an extended maternity leave, she decided to enroll in graduate school to further her knowledge of traffic engineering with an emphasis on highway safety. She is a candidate for the degree of Master of Science in the Department of Civil and Environmental Engineering in December 2014 and plans to begin work on her doctorate upon graduation.