VERMONT
AGENCY OF TRANSPORTATION

Calibration of the Highway Safety Manual Predictive Models for Rural Two-Lane Roads for Vermont

Dr. James L. Sullivan, Research Projects Director

University of Vermont Transportation Research Center

November 2019

Research Project

Reporting on Project VTRC 18-03

Final Report 2020-01

A Report from the University of Vermont Transportation Research Center

Calibration of the Highway Safety Manual Predictive Models for Rural Two-Lane Roads for Vermont

Final Report

November 2019

TECHNICAL DOCUMENTATION PAGE

<table>
<tr><td>1. Report No.
2020-01</td><td>2. Government Accession No.</td><td>3. Recipient's Catalog No.</td></tr>
<tr><td colspan="2" rowspan="2">4. Title and Subtitle
Calibration of the Highway Safety Manual Predictive Models for Rural Two-Lane Roads for Vermont: Final Report</td><td>5. Report Date
November 19, 2019</td></tr>
<tr><td>6. Performing Organization Code</td></tr>
<tr><td colspan="2">7. Author(s)
Sullivan, James L. (ORCID 0000-0002-4435-9002)</td><td>8. Performing Organization Report No. 19-005</td></tr>
<tr><td colspan="2" rowspan="2">9. Performing Organization Name and Address
University of Vermont Transportation Research Center
Mansfield House
25 Colchester Avenue, Burlington, VT 05405</td><td>10. Work Unit No.
VTRC 18-3</td></tr>
<tr><td>11. Contract or Grant No.
CA0500 VTRC 18-3</td></tr>
<tr><td colspan="2" rowspan="2">12. Sponsoring Agency Name and Address
Vermont Agency of Transportation (SPR)
Research Section
219 N. Main Street,
Barre, VT 05641</td><td>13. Type of Report and Period Covered
Final Report 2018-2019</td></tr>
<tr><td>14. Sponsoring Agency Code</td></tr>
<tr><td colspan="3">15. Supplementary Notes
Conducted in cooperation with the U.S. Department of Transportation, Federal Highway Administration.
https://vtrans.vermont.gov/sites/aot/files/planning/documents/research/publishedreports/2020-01%20Calibration%20of%20Highway%20Safety%20Manual.pdf</td></tr>
<tr><td colspan="3">16. Abstract
The 2010 Highway Safety Manual (HSM) developed by AASHTO provides predictive equations for quantifying the safety effects of planning and designing roadway alternatives. These equations have been developed based on data sets that are nearly 20 years old from a small number of states, so they must be calibrated to conditions in a specific state in order to ensure that the predicted number of crashes on a state's infrastructure are accurate. The purpose of this project was to develop calibration factors (CFs) and updated safety-performance functions (SPFs) for the undivided, two-lane, two-way rural road (TLTWRR) predictive models in the HSM for Vermont. Calibration-factor calculations were conducted for the entire state together, and also for two sets of geographic divisions, to investigate the effects of regional variations on crash prediction in Vermont. The calculated calibration factor for undivided TLTWRR segments (2U) in Vermont was found to be 0.298, while the regional CFs varied from 0.214 to 0.367. A trend toward decreasing crash rates over time may be reflected in the relatively low CFs calculated for Vermont's 2U segments. Calculated CFs for TLTWRR Intersections in Vermont were 0.448 for both three-leg (3ST) and four-leg (4ST) stop- or yield-controlled intersections, and 0.568 for the four-leg signalized (4SG) intersections. These CFs are similar to the average of the states reviewed in this study. Re-estimation of SPFs for the 2U, 3ST and 4ST site types resulted in new SPF equations that outperformed the use of CFs, as measured by the Freeman-Tukey (FT) R^2 measure. The re-estimated SPF for the 4SG site type was not statistically viable. Therefore, the default equation in the HSM should continue to be used with CMFs and the CF to estimate predicted number of crashes at 4SG sites.</td></tr>
</table>

<table>
<tr><td colspan="3">17. Key Words
Highway safety, highway safety manual, calibration factor, crash modification factor, safety performance function, undivided two-way, two-lane rural roads</td><td colspan="2">18. Distribution Statement
No restrictions. This document is available through the National Technical Information Service, Springfield, VA 22161.</td></tr>
<tr><td>19. Security Classif. (of this report)
Unclassified</td><td colspan="2">20. Security Classif. (of this page)
Unclassified</td><td>21. No. of Pages
67</td><td>22. Price</td></tr>
</table>

Acknowledgements

The author would like to acknowledge VTrans for providing funding for this work. The VTrans project sponsor, Mario Dupigny-Giroux, provided invaluable support and supervision, and the input of the project TAC members was greatly appreciated:

- Mario Dupigny-Giroux (Traffic Safety Engineer, Office of Highway Safety)
- Johnathan Croft (GIS Database Administrator, Mapping)
- Sai Sarepalli (CCRPC Transportation Planning Engineer)
- Ben Tietze (Traffic Safety Design)
- Zoe Neaderland (Coordinator, Policy and Planning Section)
- Emily Parkany (Manager, Research Section)

Disclaimer

The information contained in this report was compiled for the use of the Vermont Agency of Transportation. Conclusions and recommendations contained herein are based upon the research data obtained and the expertise of the researchers and are not necessarily to be construed as Agency policy. This report does not constitute a standard, specification, or regulation. The Vermont Agency of Transportation assumes no liability for its contents or the use thereof.

This material is based upon work supported by the Federal Highway Administration under SPR-B VTRC 18-3. Any opinions, findings and conclusions or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the views of the Federal Highway Administration.

Table of Contents

List of Tables

List of Figures

Executive Summary

The 2010 Highway Safety Manual (HSM) developed by AASHTO provides predictive equations for quantifying the safety effects of planning and designing roadway alternatives. These equations were developed from data sets that are nearly 20 years old. As a result, they must be calibrated for accurate use. Once VTrans has a calibrated set of equations, it will be able to accurately evaluate project alternatives for safety improvement. The purpose of this project was to calculate calibration factors (CFs) and update safety-performance functions (SPFs) for the undivided, two-lane, two-way rural road (TLTWRR) predictive models in the HSM. Rural two-lane, two-way road models were prioritized for calibration by VTrans as they represent the most common type of roads for which projects are being designed. The calibration of the models for the other conditions listed in the first edition of the HSM (i.e., rural multilane highways and urban and suburban arterials) may be conducted in the future.

As noted, the analysis calculated CFs to calibrate the default SPFs in the HSM to Vermont-specific conditions. The calculation of these CFs first required the determination of crash-modification factors (CMFs) for each site in the sample. CMFs are adjustment-factors that vary just above and below 1.0 to modify the predicted number of crashes at a site based on its specific physical characteristics and the characteristic of crashes in Vermont. For example, if a site has narrower lanes an/or shoulders, then a CMF higher than 1.0 is used to increase the predicted number of crashes that will occur at the site. However, if the same site has roadway lighting, then a CMF lower than 1.0 is used to decrease the predicted number crashes. For each site, a set of CMFs are determined, primarily from look-up tables and equations provided in the HSM.

These updates were conducted separately for the following site types on undivided rural two-way, two-lane roadway

- Roadway segments (2U)
- Signalized four-leg intersections on these segments (4SG)
- Unsignalized intersections on these segments:
    - Three-leg with minor-road yield- or stop-control (3ST)
    - Four-leg with minor-road yield- or stop-control (4ST)

These calculations were conducted for the entire state and also for two sets of geographic divisions to investigate the effects of regional variations on crash prediction in Vermont. The two ways of dividing the state are shown on Figure ES-1. The first set of divisions was northern, central, and southern municipalities based on climate, along with its related effects on travel patterns and tourism. The second set was based on geology and climate. It defined three physiologies:

A. The Vermont Lowlands, the Valley of Vermont, and the Taconic Mountains
B. The Green Mountains
C. The Vermont Piedmont and Northeast Highlands

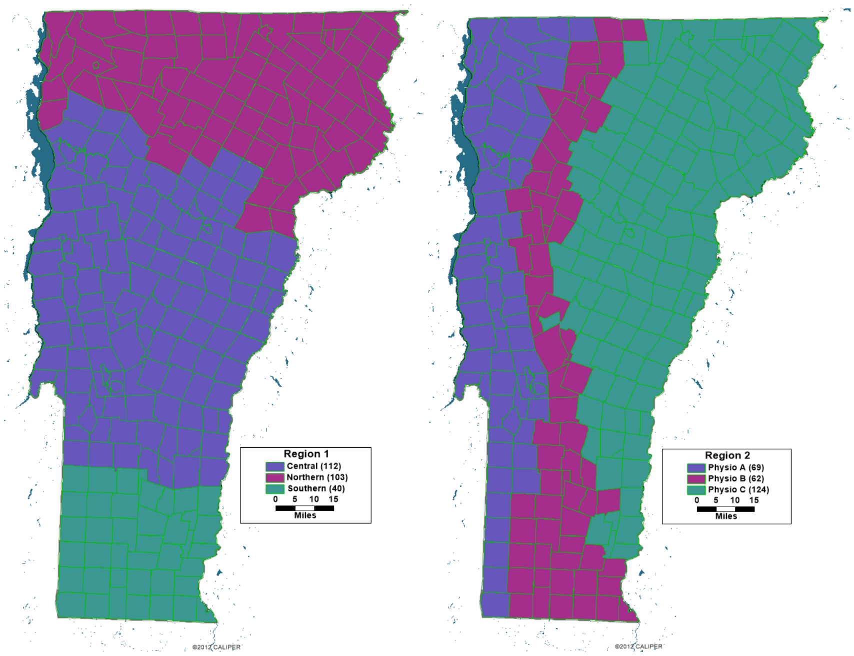


Figure ES-1 Categorizations of Towns Used in the Calculation of CFs

The regional breakdown indicates a slightly elevated crash rate in the southern region of the state, as opposed to the central and northern region, and in the Green Mountains (Physio B) and Vermont Piedmont (Physio C) as opposed to the western edge of the state (Physio A). A trend toward decreasing crash rates over time may be reflected in the relatively low CFs calculated for Vermont's 2U segments.

Statewide and regional CFs are provided. For the 4ST and 4SG site types, sample sizes for the regional breakdowns fell below the thresholds established in the HSM. The statewide CFs for all three of the intersection site types are similar to the average of the states reviewed in this study. The regional breakdown for the 3ST site type indicates a distinction between the slightly higher crash rate in the Green Mountains (Physio B) compared to the rest of Vermont (Physio A and C), which is not surprising considering the driving conditions that are frequently encountered in this mountainous region.

The data also supported re-estimation of the default SPFs for the 3ST and 4ST intersection site types and the 2U segments, and they are included in the report. The re-estimated SPF for the 4SG site type was not statistically viable. As a result, the default SPF in the HSM should continue to be used with CMFs and the new CF to estimate predicted number of crashes at 4SG sites. The Freeman-Tukey (FT) R^2 measure was used to compare the application of the calculated CFs and the re-estimated SPFs for goodness of fit. Based on the results of this comparison, the use of the re-estimated SPFs is recommended for the 3ST, 4ST and 2U site types for the calculation of the predicted number of crashes in Vermont.

The HSM recommends that these calibration factors be updated at least every three years and recommends combining all three years of data. It might be more effective in the future to use a Bayes approach with the individual years' data to arrive at a final SPF. This approach will take advantage of any possible trends in traffic safety that are influencing the data over time.

Crash data quality collection, management, and distribution can continue to improve. It is important to avoid empty data and to ensure consistency in location descriptions for data to be used for these estimations. Annual crash data on the open geodata portal (http://geodata.vermont.gov/) should contain all data from the original crash reports. In particular, the following fields are critical for the types of analyses dictated by the HSM:

- Location – latitude/longitude
- Date/Time
- Direction of collision
- Roadway characteristic
- Animal involved (wild; moose; deer; domestic; none/other)
- Impairment (alcohol; alcohol and drugs)
- Involving (pedestrian; motorcycle; heavy truck; bicycle; none/other)
- Crash type (fatal; injury; property damage only; unknown crash type)
- Crash injury (fatality, suspected serious injury, suspected minor injury, possible injury, no injury, unknown, and untimely death)
- Light conditions (dark – lighted roadway, dark – roadway not lighted, dark – unknown roadway lighting, dawn, daylight, dusk, not reported, other, unknown

Undefined entries in any data fields should be avoided, and the ID field should be uniform across reporting agencies and crash types.

1 Introduction

The 2010 Highway Safety Manual (HSM) developed by AASHTO provides predictive equations for quantifying the safety effects of planning and designing roadway alternatives. These equations have been developed based on data sets from a small number of states so they must be calibrated to local conditions in order to ensure that the results at the local levels are accurate. Once VTrans has a calibrated set of equations, VTrans will be able to predict crashes more accurately and be able to better evaluate project alternatives. The HSM equations are a great tool to quantify safety, but because they are not calibrated for Vermont, VTrans has not been able to fully benefit from their use.

The purpose of this project is to develop calibration factors and updated functions for the two-lane rural road predictive models in the HSM. Calibration factors will be developed for roadway segments and for intersections. Specifically, calibration factors will be computed for undivided-two lane, two-way rural roadway (TLTWRR) segments, unsignalized three-leg intersections (stop control on minor approach), unsignalized four-leg intersections (stop control on both minor road approaches) and signalized four-leg intersection:

- Undivided rural two-lane roadway segments (2U)
- Signalized Four-leg intersections (4SG)
- Unsignalized intersections
    - Three-leg with minor-road stop control (3ST)
    - Four-leg with minor-road stop control (4ST)

In addition, the default values in the HSM tables used to calculate crash modifications factors (CMFs) will be replaced with values specific to Vermont.

This project relates to a strategy listed in the 2017-2021 Vermont Strategic Highway Safety Plan under the Data Special Emphasis Area to improve crash data analysis. The specific title of this strategy is “Improve Crash Data Analysis to Support Data-Driven Decision Making”. The uniqueness of this project resides in the use of Vermont data to generate calibration factors and replacement for defaults that are unique and specific to Vermont. Several state DOTs have conducted similar research to calibrate the HSM predictive models to their state conditions. Some of these states have identified improvements to the calibration methodology proposed in the HSM. The calibration methodology presented in the Appendix A of Part C of the HCM includes the following steps:

- Step One – Identify facility types for which the applicable Part C predictive model is to be calibrated.

- Step Two – Select sites for calibration of the predictive model for each facility type
- Step Three – Obtain data for each facility type applicable to a specific calibration period
- Step Four – Apply the applicable Part C predictive model to predict total crash frequency for each site during the calibration period as a whole.
- Step Five – Compute calibration factors for use in Part C predictive model.

2 Literature Review

A comprehensive review of the literature documenting the experiences of other states and universities that have undertaken this type of work was conducted. Each source reviewed was either a technical report or a journal article documenting each state's development of calibration factors for the safety performance functions (SPFs) of undivided two-lane, two-way rural roadway (TLTWRR) predictive models from the HSM. Sources were reviewed for the following states:

- Alabama
- Idaho
- Illinois
- Kansas
- Louisiana
- Maine
- Maryland
- Missouri
- North Carolina (2)
- Utah
- Virginia
- Oregon
- South Carolina

2.1 Calibration Factor Calculations

North Carolina was the only state with evidence of having calculated these calibration factors twice, with an original calculation in 2011 and an update in 2017. With the exception of the Virginia study (Hass et. al., 2010), calibration factors were developed in each case for roadway segments (2U). Nine of the states included in this review also calculated calibration factors for TLTWRR intersections:

- Idaho
- Kansas
- Maine
- Maryland
- Missouri
- N. Carolina
- Oregon
- S. Carolina

Idaho, Kansas, and Missouri were unable to calculate calibration factors for signalized Four-leg intersections (4SG), but all of these states calculated calibration factors for unsignalized three-leg intersections with stop control on minor approach (3ST) and unsignalized Four-leg intersections with stop control on both minor approaches (4ST). Table 1 contains a summary of the calibration factors calculated for each state.

Table 1 Calibration Factors for States Reviewed in This Project

State	Year	2U	4SG	3ST	4ST
Alabama	2013	1.39[1]	--	--	--
Idaho	2015	0.87	--	0.56	0.62
Illinois	2010	1.40	--	--	--
Kansas	2013	1.48	--	0.21	
Louisiana	2015	0.97	--	--	--
Maine	2017	1.08	0.55	0.54	0.38
Maryland	2014	0.70	0.26	0.16	0.20
Missouri	2014	0.82	--	0.77	0.49
N. Carolina	2011	1.08	1.04	0.57	0.68
N. Carolina	2017	1.09[2]	0.77	0.58	0.63
Utah	2011	1.16	--	--	--
Virginia	2010	--	--	--	--
Oregon	2012	0.74	0.47	0.32	0.31
S. Carolina	2018	0.99	0.46	0.40	0.47

Notes:

1. Alabama (Mehta and Lou, 2013) also included a newly proposed approach that treated the estimation of the calibration factor as a special case of a negative binomial regression (1.52).
2. North Carolina (Smith et al, 2017) also included 2U calibration factors for three sub-regions of the state – Coast (1.78), Mountain (0.78), and Piedmont (1.21).

The calculation of calibration factors for signalized four-leg intersections (4SG) was problematic for all states, including the six where it was successfully calculated. This category routinely had the fewest possible sites for selection, so reaching a sample with enough crashes to conduct a statistically defensible calculation was difficult. The sensitivity of these calculations is perhaps best evidenced by the two calculations conducted for North Carolina in 2011 and 2017. The calculation for North Carolina conducted in 2011 was the first intersection-based calibration factor for TLTWRRs that was greater than 1.0, and then when it was recalculated in 2017, the calibration factor came down to 0.77. For the 2011 calculation, the sample consisted of only 19 sites, whereas the calculation for the stop controlled intersection types (3ST and 4ST) consisted of 133 and 59 sites, respectively. In 2017, the sample for the 4SG category consisted of 85 sites. This variation shows the effect that an adequate number of sites in the sample can have on the resulting calibration factor. The variation in the number of sites and the calculation factors for North Carolina between 2011 and 2017 is summarized in Table 2.

Table 2 Site Selection for North Carolina Calibration Factor Calculations

Year	Site Type	No. of Sites	CF
2011	**4SG**	**19**	**1.04**
	3ST	133	0.57
	4ST	59	0.68
2017	**4SG**	**85**	**0.77**
	3ST	173	0.58
	4ST	203	0.63

2.2 Re-Estimation of State-Specific SPFs

Finally, a different set of seven states in the review went the additional step of developing new state-specific SPFs by updating the default coefficients in the SPFs of the Part C Predictive Models:

- Alabama
- Idaho
- Illinois
- North Carolina
- Utah
- Virginia
- South Carolina

This process consisted of a negative binomial (NB) regression using a state-specific sample, which is the basis of the functional form of the SPFs in the HSM. Negative binomial regression is a type of generalized linear model in which the dependent variable Y is a count of the number of times an event occurs (Zwilling, 2013). The traditional NB regression model is given as:

$$\ln \mu = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p \quad (1)$$

Where μ is the mean of Y, the predictor variables are x, and the estimated coefficients are β. State-specific SPFs are expected to provide more accurate crash estimations than the HSM default equations. However, the sample sizes and the number of crashes necessary to yield defensible new SPFs is higher than what is necessary for simply calibrating the SPFs given in the HSM. For this reason, only Illinois (Tegge et al., 2010) and South Carolina (Ogle and Rajabi, 2018) were able to re-estimate a new SPF for the 4SG site type, and only one other state (Idaho) was able to re-estimate a new SPF for the 3ST and 4ST site types (Abdel-Rahim and

Sipple, 2015). The determination of the acceptability of the estimated SPFs were made through the evaluation of goodness-of-fit measures.

The methods and tools used to complete the NB regressions varied amongst the states reviewed. The Alabama study (Mehta and Lou, 2013) used Nlogit (http://www.limdep.com/products/nlogit/) and SPSS (https://www.ibm.com/analytics/spss-statistics-software) to alter the form of the NB model, estimating a total of three new functional forms to estimate crashes from AADT and segment length. They tested the fit of these three new functional forms against the HSM NB form using the median absolute deviation (MAD), mean square percent error (MSPE), mean prediction bias (MPB), log likelihood (LL), and Akaike information criterion (AIC). The Idaho study (2015) used R (https://www.r-project.org/) to estimate new functional forms for the 2U, 3ST, and 4ST types and tested the fit of each using the Pearson correlation coefficient (PCC), MSPE, and Freeman-Tukey R^2. The Illinois study (Tegge et al., 2010) used Excel visual basic for applications (VBA) and SAS (https://www.sas.com/en_us/home.html) to estimate state-specific NB equations for all site types separately for fatal crashes, injury crashes, and fatal + injury crashes. This study also incorporated three types of access control (uncontrolled, partial control, and full control) into the regression. North Carolina estimated a new SPF for the 2U site types in each of its studies (Srinivasan and Carter, 2011; Smith et al., 2017) using SAS. However, in the 2011 study, the new SPF was not recommended, due to poor fit. In the 2017 study (Smith et al., 2017), separate estimations were made for three geographic/climatologic regions in the state – Coast, Mountain, and Piedmont. The SPFs were found to be very different for the three regions, warranting a regional approach to estimation of crashes. In the Utah study (Saito et al., 2011), new SPFs were estimated for all four TLTWRR site types using NB regression with a stepwise approach. The variables maintained (in addition to AADT and Length) were driveway density, passing prohibition (yes/no), shoulder rumble strip (yes/no), % trucks, and speed limit. In the South Carolina study (Ogle and Rajabi, 2018), new SPFs were estimated for all four TLTWRR site types in a NB regression.

The Virginia study (Hass et al., 2010) consisted solely of the development of new SPFs for the 2U site type using only segment length and AADT, but the data was parsed for a variety of sub-classes, including secondary/primary roads, and geographic regions (North, East, and West). For this study, the AASHTOware tool SafetyAnalyst was used. SafetyAnalyst is a set of software tools used by state and local highway agencies for highway safety management. The new SPFs were then tested for fit along a variety of AADTs and compared to the fit of the HSM base model.

2.3 Crash Period and Segment Lengths

The primary issue with data used for these studies is a sparsity of sites, arising from the fact that all of the physical roadway characteristics needed are not available for all of the infrastructure in the state where crashes can occur. Therefore, the set of segments and intersections that can be used is limited to those with all attributes available. Data supplementation can include (1) imputing these attributes for road segments where data is not available or (2) adjusting the sizes of segments, by subdividing the original segment lengths. Imputing physical characteristics is only realistic for segments. Intersection data is too specific to be imputed, which explains why data sparsity for certain intersection types could not be addressed in most of the studies. Adjusting the segment lengths arbitrarily can create questions about autocorrelation and problems with later analyses. Additionally, the original segment lengths that are native to many GIS layers maintained by state DOTs are based on relevant delineations of roadway characteristics, so it is likely that altering these natural segmentations diminishes the quality of the data. This explains why most states do not take advantage of the HSM recommendation to combine segments to avoid those shorter than 0.10 mi. Segments shorter than 0.10-mile long can often embody meaningful roadway characteristics, like a bridge.

Physical changes to the infrastructure are also difficult to track in a systematic way, so that they can be readily connected to road segments, intersections and crashes spatially. This difficulty makes it less desirable to expand the time period of the analysis too far. Indeed, the HSM recommends not exceeding a three-year span for considering crashes, AADT, and physical characteristics. Illinois (Tegge et al., 2010) used a "sliding window" approach to analyzing the data for a variety of default segment lengths for urban (0.25 mi.) and rural (1.0 mi.) segments, and analysis period of five years. South Carolina (Ogle and Rajabi, 2018) used fixed lengths of 0.25 miles for urban segments and 1.0 miles for rural segments. Missouri and Utah (Sun, 2017; Saito et al., 2011) enforced minimum segment lengths of 0.5 miles and 0.2 miles, respectively, increasing the number of sample selected to reach the recommended minimum of 100 crashes.

Most of the studies reviewed (Schrock and Wang, 2013; Wolshon and Robicheaux, 2015; Belz, 2017; Shin et al., 2014; Dixon et al., 2012) used a three-year span with natural segment lengths from their state's official centerline road network GIS, identified any road segments where a significant improvement may have been made, and removed them from the population of data available for analysis. Many states, including Vermont, work continuously to improve and expand the attributes contained in the GIS of their statewide road network. These improvements include allowing segmentation at meaningful distinctions along a roadway. These distinctions will often include more than intersections, but also changes in speed, capacity, jurisdiction, or geography. In this case, it is not advisable to divide or combine these segments to make arbitrary default segment lengths.

Even using these variety of methods to increase the sample size, several states were unable to calculate calibration factors for one or more site types due to lack of data. Kansas (Schrock and Wang, 2013) could not develop a calibration factor for the 4SG site type due to lack of sites, and they also combined sites from the 3ST and 4ST types to get enough crashes for development of calibration factors. The Oregon data set consisted of only 25 sites of the 4SG type (Dixon et al., 2012).

A secondary issue with data is the sparsity of crashes. In order to develop models like a negative binomial, although it is based on outcomes with infrequent occurrence, a certain minimum number of occurrences are nonetheless needed. Therefore, the number of years used to gather crash data is often expanded to meet the recommended minimum of 100 observed crashes for each type. Table 3 provides a summary of the crash period used for each of the studies reviewed.

Table 3 Summary of Crash Periods for States Reviewed

State	Report Year	No. of Years in Crash Period
Alabama	2013	4
Idaho	2015	10
Kansas	2013	3
Louisiana	2015	3
Maine	2017	3
Maryland	2014	3
Missouri	2014	3
N. Carolina	2011	5
N. Carolina	2017	6
Utah	2011	3
Virginia	2010	5
Oregon	2012	3
S. Carolina	2018	3

Expanding the number of years in the crash period too far results in difficulty selecting the appropriate physical characteristics of the roadway. A larger temporal span undoubtedly includes changes to the physical infrastructure, making the crash observations before the change not comparable to the crash observations after the change. This type of change typically makes the site unusable, since the two periods are not totally independent, but the site has changed.

2.4 Lessons Learned

All of the studies reviewed reported problems assembling the data needed to conduct these analyses, noting that data collection and pre-processing for analysis was the most challenging task. The completeness, accuracy, format and interoperability of data sources were frequent issues. A good review of these issues is provided in the Maryland (Shin et al., 2014) study. The Maryland study also points out some of the ambiguities in the HSM guidance regarding data, particularly with the number of sites to use, the number of crashes needed, and the optimal length of roadway segments for analysis.

The Alabama study (Mehta and Lou, 2013) found that the default HSM method was underpredicting crashes for Alabama, and one particular new SPF model outperformed even the calibrated HSM method. The Idaho study (Abdel-Rahim and Sipple, 2015) also found that the HSM default method was over-predicting crashes for Idaho. One of the new SPFs for the 2U site type and the 3ST site type performed better than even the calibrated HSM method. However, for the 4ST site type in Idaho, the calibrated HSM method performed best.

The Maine study (Belz, 2017) noted that its unique crash rates may be due to Maine's heavily forested northern climate, hilly terrain, rural landscape, lifestyle, and older population. The Utah study (Saito et al., 2011) concluded that the data needed to allow the SPF model to perform best for 2U sites was not feasible, so the recommended model only includes AADT, length, % trucks, and speed limit. The Virginia study (Hass et al., 2010) refutes the use of CFs alone, since their fit can vary greatly across AADT. The Oregon study (Dixon et al., 2012) recommends stratifying the determination of CFs based on crash severity over regionality. The South Carolina study (Ogle and Rajabi, 2018) stresses the importance of calculating an error on the determination of CFs.

3 Data Acquisition, Processing, and Compilation

For each of the two-lane, two-way rural road (TLTWRR) facilities included in this study, crash data and site characteristics were gathered, checked, cleaned, and aligned spatially and temporally. Generally, the goal was to acquire site characteristics data for as many sites as possible, to increase the population size available for the calculation of CFs and re-estimation of SPFs. Physical and traffic characteristics were gathered in GIS and sites were grouped by type, according to the classification of TLTWRR facilities provided in the HSM:

- Undivided rural two-lane roadway segments (2U)
- Signalized Four-leg intersections (4SG)
- Unsignalized intersections:
    - Three-leg with minor-road stop control (3ST)
    - Four-leg with minor-road stop control (4ST)

The physical and traffic characteristics gathered for each site are intended to satisfy the list of required and optional characteristics identified in Table A-2 of the HSM (Table 4).

Table 4 Required and Optional Characteristics from Table A-2 of the HSM

Type	Required Characteristics	Optional Characteristics
2U	Length of segment AADT Length of horizontal curve Radius of horizontal curve Lane width Shoulder type Shoulder width Presence of center 2-way left-turn lane	Spiral transition for horizontal curve Superelevation variance of horizontal curve Percent grade Presence of lighting Driveway density Passing lane or short, four-lane section Presence of centerline rumble strip Roadside hazard rating Use of automated speed enforcement Shared-use path crossings Rail crossings Access-management plan
4SG, 3ST, 4ST	AADT for major road AADT for minor road No. of approaches with left-turn lanes No. of approaches with right-turn lanes Presence of lighting	Skew angle

A calibration period of three years was used for collecting crash data for these facilities. Each crash in the state over the calibration period was associated with a roadway segment (line segment) or intersection (node) in GIS. A calibration sample set of data was selected using a random sampling process, and a variety of sample sizes were tested. For each facility type, the desirable minimum sample size for the calibration sample is 30 sites, and that calibration set should have a minimum of 100 crashes over the calibration period, in accordance with the HSM. The use of sub-regions in the state was also explored for site types with large enough populations.

The following subsections describe the source of the data used to compile physical and traffic characteristics of the sites, and the source of the crash data for the calibration period. For the 2U sites, line segment data was collected in GIS. For the intersection site types (4SG, 4ST, and 3ST), point data was collected in GIS, along with an associated GIS of short, disconnected line segments representing intersection approaches. Then, the crash data, consisting of a collection of points with associated crash characteristics in GIS, is described for the three years selected for this study. Finally, a series of sample selections were conducted in order to determine the minimum sample size that would be needed to ensure that 100 crashes were available for each site type.

3.1 Line Segment Data

3.1.1 Roadway Centerlines

The Vermont roadway centerline layer (VRCL), formerly known as TransRoad_RDS, is a GIS of line segments representing all federal-aid highways, town highways, and many private roads in the state. The centerlines were originally developed by the Vermont Center for Geographic Information (VCGI) in 1992. VCGI was the steward of the VRCL between 1992 and 2004, with updates by the RPCs and VTrans. VTrans now stewards the data and has revised the layer to match its official highway mileage. This layer meets the requirements articulated in the Road Centerline Data Standard of VCGI (http://vcgi.vermont.gov/resources/standards). It is the most reliable source for VTrans road class (AOTCLASS) information for road centerlines. This layer does not include every private road in the state, and the road name information may not match perfectly with the Enhanced 9-1-1 (E911) roadway GIS. The E911 road centerline layer maintained by Vermont's E911 Board has the most up-to-date road name information. It was originally based on TransRoad_RDS, but it includes all private roads and most driveways with more reliable name and address-range data. There was a significant change in the schema in the June 2013 release as part of the effort between VTrans and E911 to merge their two roadway datasets. The data layer includes the field structure agreed to by both entities, but most of the fields that came from the E911 road centerlines have not been populated completely in this release. The fields that were critical for use in this study include:

- Paved (yes/no)
- Length
- Functional Class
- Urban (yes/no)
- Lanes_Each-Way
- OneWay (yes/no)
- Class (Public/Private)

This layer includes 71,639 individual segments.

3.1.2 Geometric Characteristics

Lane and shoulder width and configuration of Vermont's federal-aid highway system (interstates, federal highways, and state highways) are stored in a GIS layer for reporting to FHWA's Highway Performance Monitoring System (HPMS). This layer contains 13,571 line segments with fields detailing the number of lanes in each direction of travel, the widths of those lanes, the purpose of those lanes, and the widths and types of medians and shoulders. This GIS layer also indicates if the segment includes the presence of a center two-way left-turn lane.

A separate GIS layer contains the horizontal curvature geometry of Vermont's roadways. This layer is far more disaggregate than the HPMS layer, with 557,903 segments statewide. Segments are individuated at each change in any of the horizontal curvature characteristics. These characteristics include the length, radius, and degree of horizontal curvature, and the type of transition included in the curve (reverse, reverse spiral, spiral, or none).

Data on roadway grade and superelevation for the federal-aid system was taken from the data collected by Vermont's Automatic Road Analyzer (ARAN) to support the HPMS program. Bi-annually, the entire federal-aid highway network, containing approximately 3,900 miles of roadway, is driven with an ARAN vehicle to collect asset data, a videolog, and a variety of other parameters with sensors like GPS. Interstates are driven annually, state and federal highways in even years, and major-collector town highways and federal-aid urban routes in odd years. Superelevation and grade data are collected at regular intervals along the routes, resulting in a point layer of 126,332 points representing these roadways.

Data on passing, or climbing, lanes in Vermont was obtained from the VTrans Traffic Operations Section of the Operations and Safety Bureau for use in this project. The data includes the Town, Route ID, and from/to mile-markers for each passing lane in the state. The data, including 80 passing-lane segments, is only for state roads, but it was assumed that there are no climbing lanes on town roads, so this is assumed to be a complete data set for passing lanes statewide. It was also assumed that all of these passing lanes existed before 2014, so no exceptions were made.

Data on centerline rumblestrips was also obtained from the VTrans Traffic Operations Section of the Operations and Safety Bureau for this project. It includes detailed locations, with town, route ID, and start/end mile-markers for projects extending back to 2014 from VTrans Highway Design Section. Information for installations after 2014 are less detailed and came from data gathered by the VTrans sponsor directly from design plans or from other sources, with varying precision in the start/end mile-markers. This data also includes the estimated year of installation, which was used to select individual records for this analysis:

- For the 37 records with an installation year before 2014, it is assumed that these locations had a rumble strip for the entire calibration period (2014-2016) and they were used in the analyses for this project
- For 233 records with an installation year after 2016, it is assumed that these locations did not have a centerline rumble strip during the calibration period, so they were disregarded and removed from the data set.
- For 287 records with an installation year in 2014, 2015, or 2016, it is assumed that these installations constitute site improvements and they were added to the site improvements data.

Finally, the E911 GIS layer was used to calculate driveway densities for this project. The E911 driveway data is maintained for use in directing emergency response, so the information and geography are updated weekly. E911 defines a "driveway" as any private road which leads to less than three buildings. Generally, if an inhabitable building is not visible from the road, the driveway is digitized. Sites that cannot be seen from the road are driven with sub-meter GPS and differentially corrected. Driveways occurring on short segments (less than 0.5 miles long) were found to result in unduly high driveway densities (up to 254 driveways per mile for a segment 0.0039 miles long with one driveway). When this situation was examined further, it was discovered that the calculation of driveway densities used suffered from a variety of problems. The first was that the buffer area used to assign driveways to a line segment resulted in some driveways being counted more than once, where they were within the buffer distance of more than one segment (Figure 1A).

Figure 1 A - Driveway (Dashed Line) Assigned to Two Segments (in Yellow) and B - Driveway Already Represented in the Roadway Layer

The second problem was that some of the lines in the driveways layer were already represented in the roadways layer, resulting in additional over-counting of driveway features (Figure 1B). Due to the intractability of these errors, the use of the CMF for driveway density was excluded from this study.

3.1.3 Traffic

GIS line layers of annual average daily traffic (AADT) for 2014, 2015, and 2016 were collected for use in the analyses conducted in this project. These AADT values are calculated by the Traffic Research Section at VTrans using AASHTO-specified methods of aggregating and extrapolating continuous and short-term traffic counts throughout the state. This method computes an average day of week for each month, and then computes an annual average value from those monthly averages, before finally computing a single annual average daily value. These values are used in a variety of reporting programs to FHWA, including for the calculation of exposure for safety analyses and vehicle-miles of travel (VMT). The 2016 AADT GIS layer include 3,197 segments with AADT values ranging from 50 to 55,400 vehicles per day. Another critical attribute for use in this study is the "IsDivided" field, indicating whether the segment represents a divided highway.

Since the AASHTO –specified methods for calculating AADT and the FHWA-required reporting of AADTs only apply to federal-aid roadways, AADTs are not typically available for minor, local roads and streets. However, VTrans Traffic Research recently compiled an analysis of estimated AADTs on local roads and streets by town, functional class, and surface type (paved/unpaved). These values were used to supplement the AADTs on federal-aid roadways so AADTs became available on a larger proportion of the segments and intersections in the data used in this project.

3.2 Intersection Data

3.2.1 Intersections

The Vermont intersections layer (VIL) is a GIS point layer of every formal intersection of federal-aid highways, town highways, and of many private roads and driveways, including entrances to privately-owned commercial properties. The GIS data was developed from nodes in the VRCL and building out the necessary fields to support the Minimum Inventory of Roadway Elements (MIRE). Development of a MIRE is recommended by FHWA so that critical roadway data variables are available to make more effective and efficient safety-improvement decisions. Some of the data in the VIL is yet to be populated but all critical field definitions have been fully defined. Some of the critical fields that are included in this GIS and are relevant to this study include:

- IntersectionLegCount: The number of approaches from a data management perspective, generally the number of primary direction routes entering/leaving a virtual polygon encompassing all the nodes of an intersection. Exceptions include untraveled centerlines, and approaches not represented by the centerline data.
- IntersectionGeometry:
    - 1 - Tee intersection - Two or more roadways intersect at grade in a Tee intersection
    - 2 - Y intersection - Two or more roadways intersect at grade in a Y intersection
    - 3 - Four-leg intersection - Two or more roadways intersect at grade in a four-leg intersection
    - 4 - Traffic circle/roundabout - Two or more roadways intersect at grade in a traffic circle or roundabout

- 5 - Multileg intersection, five or more legs - two or more roadways intersect at grade in a multileg intersection of five or more legs
  - 0 - Other - Two or more roadways intersect at grade in another intersection type
  - 99 - Unknown - Two or more roadways intersect at grade in an unknown intersection type
- Rural/Urban
- Complex - identifies which nodes are part of a multi-node intersection
- TrafficControlType
  - 1 - No control
  - 2 - Stop signs on cross street only
  - 3 - Stop signs on mainline only
  - 4 - All-way stop signs
  - 5 - Two-way flasher (red on cross street)
  - 6 - Two-way flasher (red on mainline)
  - 7 - All-way flasher (red on all)
  - 8 - Yield signs on cross street only
  - 9 - Yield signs on mainline only
  - 10 - Other non-signalized
  - 11 - Signals pre timed (two phase)
  - 12 - Signals pre timed (multi-phase)
  - 13 - Signals semi-actuated (two phase)
  - 14 - Signals semi-actuated (multi-phase)
  - 15 - Signals fully actuated (two phase)
  - 16 - Signals fully actuated (multi-phase)
  - 17 - Other signalized
  - 18 - Roundabout
  - 99 - Unknown
- Major_AADT
- Minor_AADT
- IntersectionSkewAngle: as defined in Chapters 9 and 10 in the Highway Safety Manual, degrees departure from 90 degrees of the minor route's intersection with the major route. If two minor legs have different skew angles, their values are averaged.

This layer currently includes 64,016 points, but only 3,780 points have an IntersectionLegCount of three or higher. Many of the points came from existing nodes in the VRCL which delineate segments for distinctions like town boundaries that do not correspond to intersections.

3.2.2 Intersection Approaches

Some of the MIRE features related to intersection analyses are specific to intersection *approaches*, as opposed to the intersections themselves. In response to this, a separate GIS layer of uniform segments is maintained by VTrans representing the *approaches* to the intersections as independent sets of short segments. Critical attributes of the intersection approaches (or legs) used in this study include:

- Turn_Lanes_L: number of exclusive left turn lanes
- Turn_Lanes_R: number of exclusive right turn lanes

The GIS layer of intersection approaches includes 152,240 segments, each adjoining a node point from the VIL.

3.2.3 Lighting

Roadway lighting data was also collected in a point GIS from the VTrans Asset Management Section. These 159 data points indicate the locations of light poles, most of which are adjacent to intersections. These point features were used to indicate the presence of lighting at an intersection.

3.3 Crash Data

Crash data for 2014, 2015, and 2016 were downloaded as geographic shapefiles from the Vermont Open Geodata Portal (http://geodata.vermont.gov/). Key attributes provided in these files include:

- Location – latitude/longitude
- Date/Time
- Direction of collision:
    - Head On
    - Left Turn and Thru, Angle Broadside
    - Left Turn and Thru, Broadside
    - Left Turn and Thru, Head On
    - Left Turn and Thru, Same Direction Sideswipe/Angle Crash
    - Left Turns, Opposite Directions, Head On/Angle Crash
    - Left Turns, Same Direction, Rear End
    - Left and Right Turns, Simultaneous Turn Crash
    - No Turns, Thru moves only, Broadside
    - Opposite Direction Sideswipe
    - Other - Explain in Narrative
    - Rear End
    - Rear-to-Rear
    - Right Turn and Thru, Angle Broadside
    - Right Turn and Thru, Broadside
    - Right Turn and Thru, Head On
    - Right Turn and Thru, Same Direction Sideswipe/Angle Crash

- Right Turn, Same Direction, Rear End
- Same Direction Sideswipe
- Single Vehicle Crash

- Roadway characteristic
    - Crossover
    - Driveway
    - Five-point or more
    - Four-way intersection
    - Not at a junction
    - Off ramp
    - On ramp
    - Other - explain in narrative
    - Parking lot
    - Railway grade crossing
    - Shared-use path or trail
    - T – Intersection
    - Traffic circle / roundabout
    - Unknown
    - Y – Intersection
- Animal involved (wild; moose; deer; domestic; none/other)
- Impairment (alcohol; alcohol and drugs)
- Involving (pedestrian; motorcycle; heavy truck; bicycle; none/other)
- Crash type (fatal; injury; property damage only; unknown crash type)
- Surface condition (wet; water; snow; slush; sand/mud/dirt/oil/gravel; other; not reported; ice; dry)
- Road condition (worn, travel-polished surface; work zone; unknown; traffic control device inoperative, missing, or obscured; shoulders (none, low, soft, high); ruts, holes, bumps; road surface condition (wet, icy, snow, slush, etc.); other; obstruction in roadway; not reported; none)

This data set contained 11,926 records for 2014, 14,111 records for 2015, and 12,501 records for 2016. Due to the concerns of the TAC with thresholds for crash reporting, the consistency of the crash data during the three-year crash period was analyzed in further detail. Inconsistencies between years in the crash period could be an indication that thresholds or criteria for reporting crashes changed. Therefore, a series of plots were created to observe differences in crash trends between the three years in the crash period. Random selections were conducted for site types 2U and 3ST, and trends in the crash totals for these site types are illustrated in Figure 2 and 3, respectively.

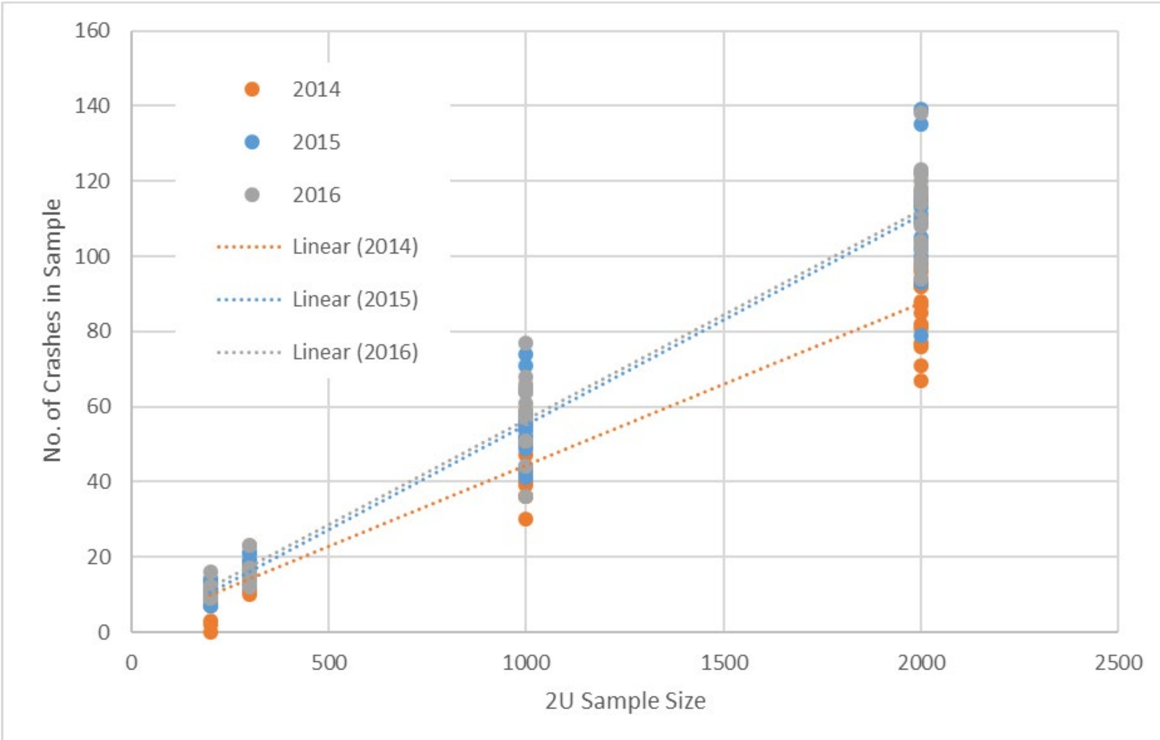


Figure 2 Trends in crash totals between years based on the sample size of the random selections conducted for 2U segments

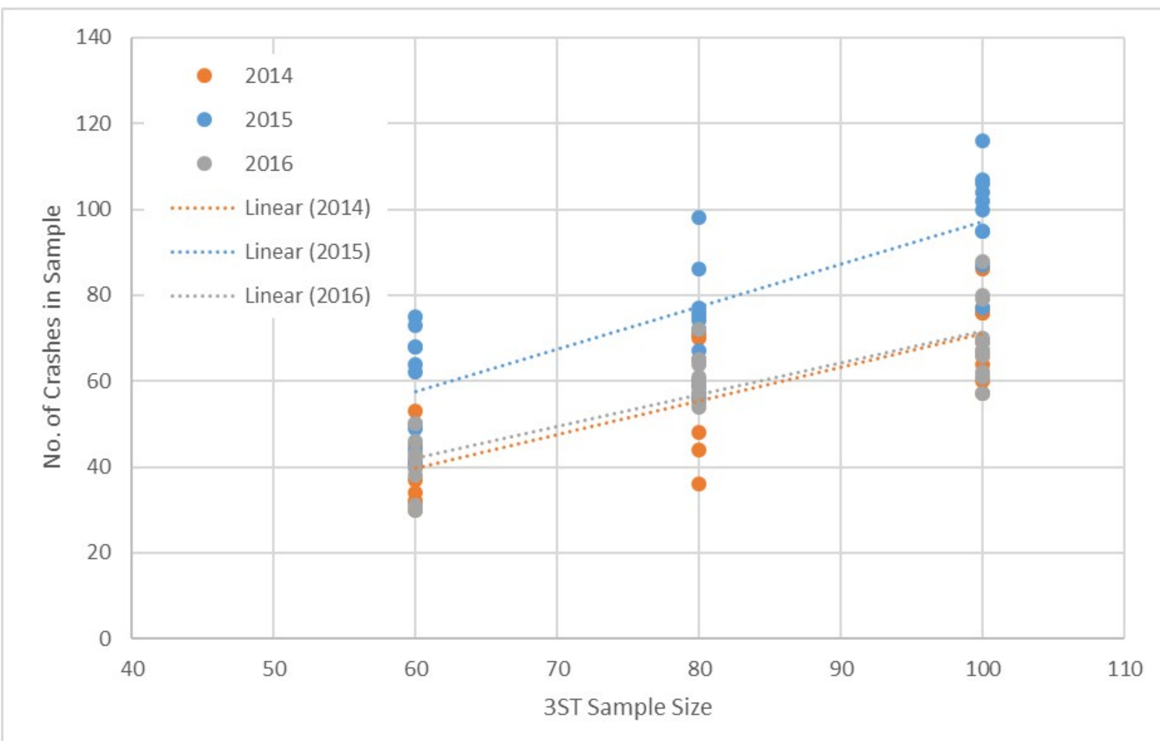
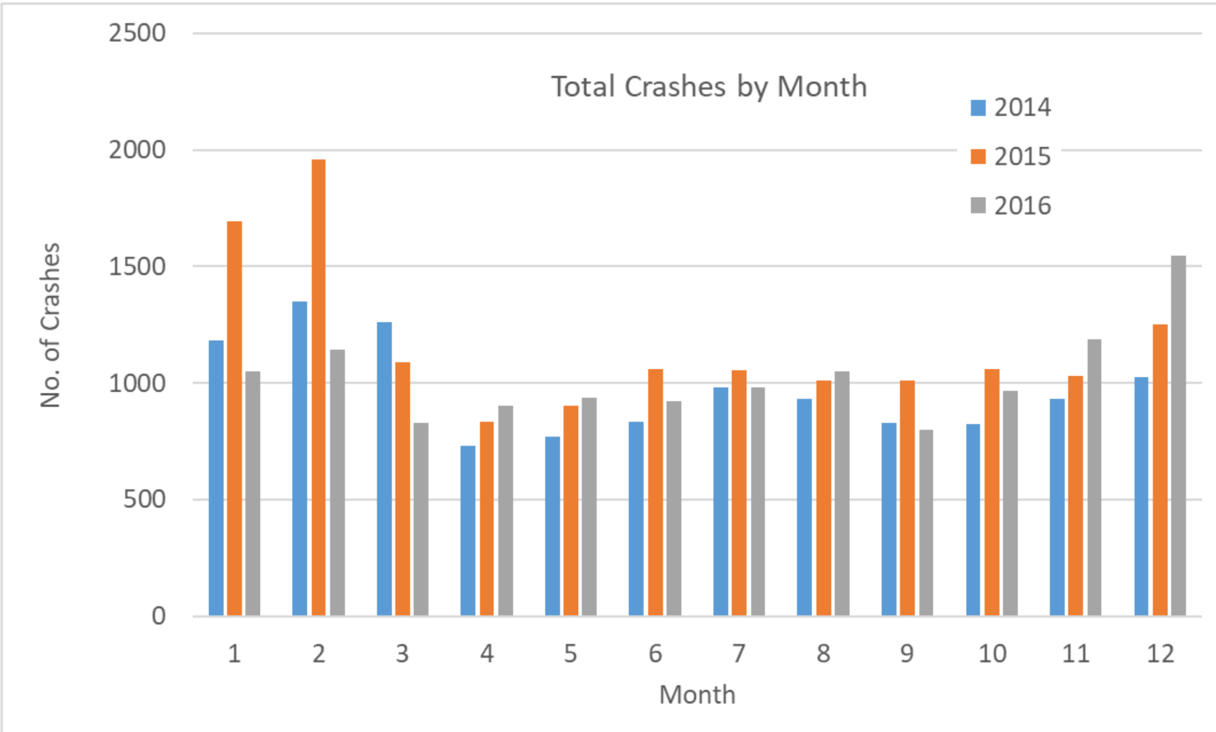
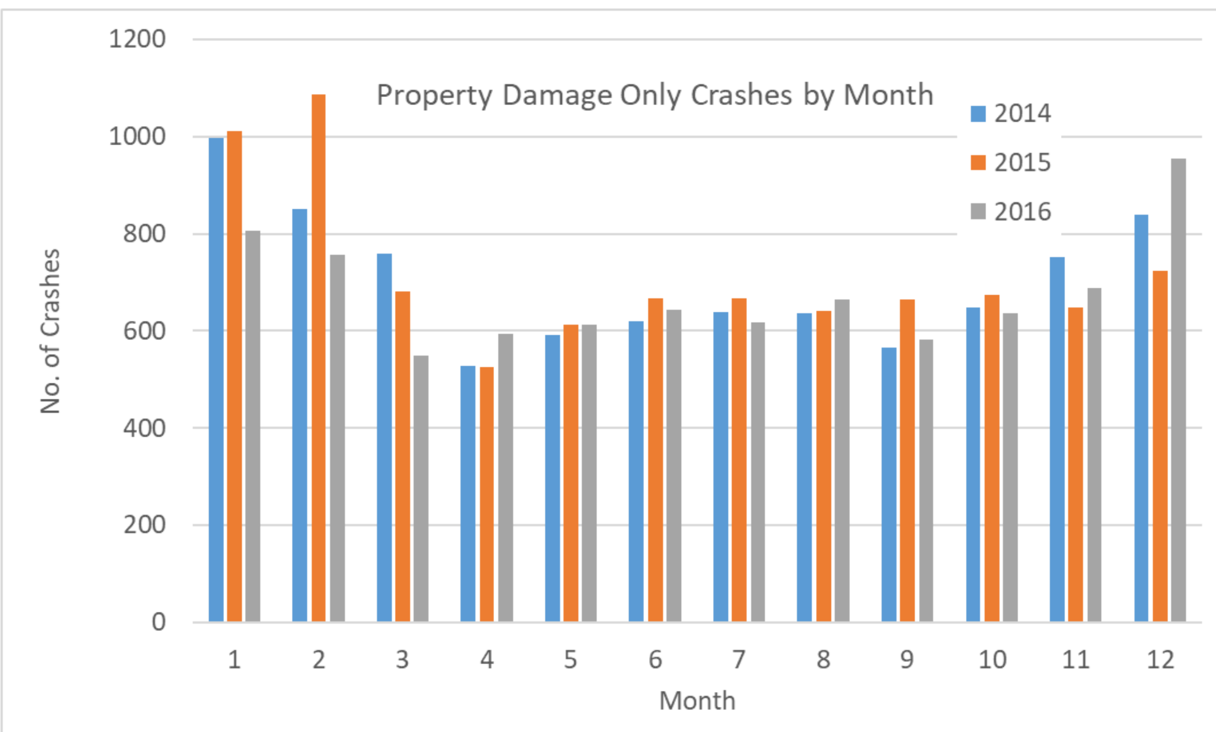


Figure 3 Trends in crash totals between years based on the sample size of the random selections conducted for 3ST intersections

Figure 2 indicates that crash reporting for 2U sites in 2014 may have been subjected to an *under*-reporting bias, but Figure 3 indicates a similarly strong bias for 3ST sites being *over*-reported in 2015. Potential monthly biases were explored further, as shown in the three charts below in Figure 4.





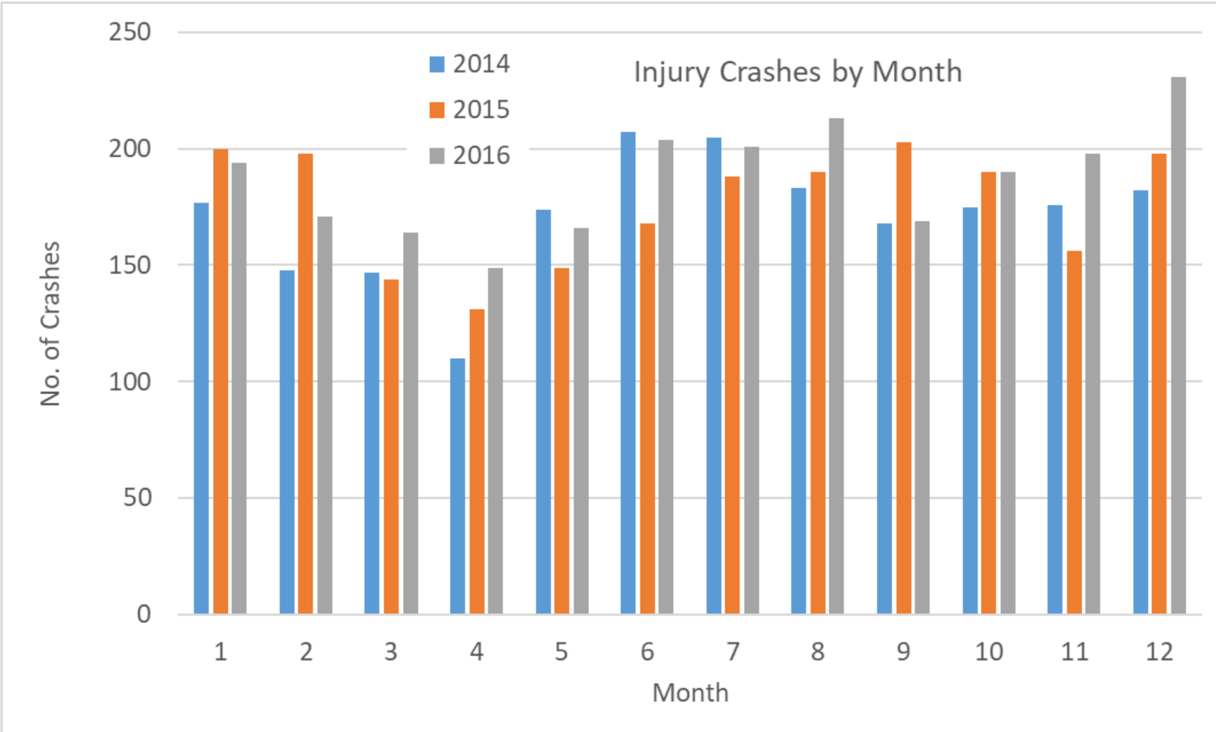


Figure 4 Monthly Crash Counts for Total Crashes, Property Damage Only Crashes, and Injury Crashes

The prevailing issue seems to be that an extensive *over*-reporting of minor crashes occurred in January and February of 2015. This issue was found to have stemmed from the inclusion of non-reportable crashes in the data, and the tendency for these non-reportable crashes to occur in winter-weather driving conditions. Consequently, these crashes are identifiable by the "unknown crash type" entry. Table 5 provides a summary of these non-reportable crashes entered into the crash database between 2012 and 2017.

Table 5 Monthly Summary of Non-Reportable Crashes in the Crash Database

Year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Total
2012			1					1		1			3
2013	279	254	204	114	108	124	2	1		79	311	608	**2,084**
2014		442	433	119			158	143	104				**1,399**
2015	**486**	**694**	269	180	141	221	201	185	146	195	226	330	**3,274**
2016	50	229	114	163	157	75	166	180	53	148	321	371	**2,027**
2017	221	325	320	97	90	155	141	108	118	118	165	563	**2,421**

These points were removed from the crash data for this project because Table 5 indicates that the reporting of these "non-reportable" crashes has been inconsistent between years in the calibration period. The highest monthly totals occurred in

January and February of 2015, but the values for the same months in 2014 and 2016 are much lower. After removal of these crashes from the data set, the number of crashes per year "flattens" somewhat, at 10,518 records for 2014, 10,767 records for 2015, and 10,411 records for 2016.

3.4 Site Improvements Data

Points and segments representing capital projects undertaken during the calibration period were accessed using the VTrans web map server at http://vtransmap01.aot.state.vt.us/arcgis/ rest/services/Master/AMP/FeatureServer. Points represent project information for projects that are located off the federal-aid system or on the federal-aid system but at a discrete location, and segments represent project information for projects that are located along the highway system and are best represented as a segment. These projects come from the VTrans Project Information and Navigation System (VPINS). A query was developed to access all projects for the 2014, 2015, and 2016, and the data was downloaded in a tabular format. The final data set consisted of 232 distinct projects from 2014-2016, each indicated by its starting and ending mile markers, route ID, and town.

Additionally, centerline rumble strips, spanning 95 miles, with an installation year in 2014, 2015, or 2016, are considered part of the site improvements made during the calibration period. Many of these data points overlap with known site improvements represented by capital projects, presumably because the rumble strips were installed as part of a larger repaving/resurfacing project. However, a few of them represent stand-alone rumble strip installations.

In order to transfer the site improvements features to the VRCL and the VIL, a point layer of mile markers along all federal-aid highways in Vermont was needed. This layer contains 52,201 point features for every 1/10th of a mile of centerline on the federal-aid system. Each point feature contains the mile-marker distance (beginning and ending at the town boundary), the town, and the route ID.

3.5 Site Population Selection Queries and Statistics

Physical and traffic characteristics were gathered in a common GIS environment, sites were selected from the VRCL and the VIL, and then other characteristics were transferred spatially to these sites. TransCAD spatial functions called "Tag" and "Aggregate" were used to transfer this data.

The "Tag" function fills data into a column of an attribute table for one layer with the name or ID of the nearest feature in another layer or with the distance to the nearest feature in the other layer. How a layer is tagged depends on the type of layer. Table 6 explains how different types of layers are tagged with this function.

Table 6 TransCAD "Tagging" Rules

Destination (layer type to be tagged)	Source (layer type to tag from)		
	Point	Line	Area
Point	Closest point	Closest line	Area that the point is in
Line	Point that is closest to the line	Line that is closest to a shape point on the line	Area that the midpoint of the lines is in
Area	Point that is closest to the area centroid	Line that is closest to the area centroid	Area that the centroid of the area is in

To use this method, the field type of the attributes in the source and the destination layers must be the same. The tagging provides a "match" between the identifying name or IDs of two layers. From this match, it is simple to transfer any data attributes between the layers.

However, in some cases, the relevant data to be transferred is best represented by a mathematical combination of multiple attributes in the source layer. In this case, the "Aggregate" function fills data into a column in the attribute table of a destination layer with aggregated data from another layer. With this function, a buffer area must be specific around the destination feature, around which source features will be sought. Attributes of features within the destination layer are then filled with aggregations of features in the source layer that are within the specified buffer area. Aggregation types available in this function include calculation of the sums, averages, minimum, or maximum of the attribute values of individual features in the source layer that are within a buffer area of the associated feature in the destination layer. Alternatively, the user can simply count the number of features in the source layer that are within the buffer area of the destination layer.

3.5.1 Road Segments

The population of road segments represented by the 2U site type for TLTWRRs was taken from a selection query on the VRCL. Segments in the TLTWRR class and the 2U site type were selected with the following query on the road centerline layer:

Lanes_Each-Way = 1 AND Urban = 0 AND OneWay = 0 AND CLASS <> "Private"

From the segments selected with this query, the following steps were taken to fill AADTs from the AADT layers:

1. The AADT layer was overlaid on the 2U segments and features were selected from it if they were touching the 2U selection and ISDIVIDED = “N”
2. The 2U selection was then filled by tagging it from the AADT selection
3. AADTs for segments whose functional class is 6 or 7 and had not already been filled were filled from the 2017 town-based AADT data received from VTrans Traffic Research using the Paved and Urban status of the segments as well

To fill horizontal curve attributes in the 2U segments selected with this query, the horizontal curve layer was overlaid and the “Aggregate” function was used to find the lowest value of the Radius and the lowest value of Length within 100 feet of each feature in the 2U selection. The “Tag” function was used to fill the Presence of Spiral Transition.

Since the data records for passing lanes and centerline rumble strips was only available in tabular format with mile-marker indicators, that data was first transferred to the point layer of mile-markers by flagging any mile-marker which sat between the start and end of a passing lane or rumble strip, with a matching town and route ID. Data from the mile-marker layer was then transferred to the “Roadway Width” layer, since it contains mile markers delineating the beginning and ending of each segment. Finally, to fill lane width, shoulder width, shoulder type, center turn lane presence, passing lane presence, and centerline rumble-strip presence in the 2U segments query, the “Roadway Width” layer was overlaid on the 2U segments and features were selected from it if they were touching the 2U selection and Divided_Se = “No”. The 2U selection was then filled by tagging from the “Roadway Width” selection. Following these steps, there were a total of 29,372 segments that fit the definition of 2U and also contain valid data for the required characteristics from Table 7.

To fill percent grade and superelevation in the 2U selection, the 2018 ARAN data was used. The GRADE field and the XFALL field were averaged for all points within 0.1 miles of each 2U segment using the Aggregation function. Of the 29,372 segments with valid data for all of the required characteristics, 9,664 also had valid values for grade and superelevation.

3.5.2 Intersections

The population of intersections for TLTWRRs was taken from the VIL and the VRCL. Intersections of the 4ST site type were selected if they were connected to at least one segment from the 2U site type and satisfied the following query on the intersections layer:

IntersectionLegCount = 4 AND IntersectionGeometry = 3 AND Rural Urban = “R” AND Complex = 0 AND (TrafficControlType = 2 OR 3 OR 5 OR 6 OR 8 OR 9)

Intersections of the 3ST site type were selected if they were connected to at least one segment from the 2U site type and satisfied the following query on the intersections layer:

IntersectionLegCount = 3 AND (IntersectionGeometry = 1 OR 2) AND Rural Urban = "R" AND Complex = 0 AND (TrafficControlType = 2 OR 3 OR 5 OR 6 OR 8 OR 9)

Intersections of the 4SG site type were selected if they were connected to at least one segment from the 2U site type and satisfied the following query on the intersections layer:

IntersectionLegCount = 4 AND IntersectionGeometry = 3 AND Rural Urban = "R" AND Complex = 0 AND (TrafficControlType = 11 OR 12 OR 13 OR 14 OR 15 OR 16 OR 17)

The functional classification and the paved status of minor roads were filled into the VIL from the VRCL by filling them with the maximum value of the functional classification and the minimum value of the paved status that are within 0.01 miles of the intersection. These fills were checked for QC, and minor corrections were made. It was assumed that any values lower than 6 that ended up in the functional classification field had come from the major road segment (because the functional class of the minor road segment was 0 or empty), so these were manually changed to a functional class of 7. These entries were used to populate the AADTs of minor roads for intersections in the TLTWRR population lacking those values.

3.5.3 Summary of Population Characteristics and Association with Crash Data

Finally, site improvements from capital projects and centerline rumble strips installed between 2014 and 2016 were transferred to the VRCL and the VIL by identifying points in the statewide mile-marker layer that fall within the start/end of each improvement's mile markers, while also matching the route ID and town. The set of mile markers that were identified as being within an improvement area were then selected for use in identifying TLTWRR segments and intersections where an improvement took place. To determine if a segment or an intersection was subject to an improvement, point features from the selection set of improvements in the mile-marker layer were counted if they were within 0.05 miles of a 2U segment or a 4SG, 4ST, or 3ST intersection, using the TransCAD "Aggregate" function. Any of these features with a count of improvement mile-markers greater than zero was identified as having been subjected to an improvement between 2014 and 2016. Note that only point features from the capital improvements data was available to count as an improvement for intersections, line-based capital improvements and centerline rumble strips were only counted as improvements for 2U segments.

Table 7 and Table 8 contain summaries of the required and optional characteristics for site type 2U (Table 7) and site types 4SG, 4ST, and 3ST (Table 8) compiled for use in this study.

Table 7 Required characteristics for the Predictive Method (CMFs) for TLTWRR segments (2U)

Characteristic	Units or Notes	Base for CMFs	No. of Sites
Length of segment	miles	NA	39,520
Segments w/out improvements		NA	37,769
2014 AADT	0-17,800 vpd	NA	30,140
2015 AADT	0-17,800 vpd	NA	30,140
2016 AADT	0-17,800 vpd	NA	30,475
Presence of horizontal curve	curve/tangent	tangent	36,411
Radius of horizontal curve	feet		
Length of horizontal curve	miles		
Presence of spiral trans. curve	yes or no		
Lane width	feet	12 ft	36,803
Shoulder type	paved/gravel/ composite/turf	paved	
Shoulder width	feet	6 ft	37,769
Presence of cntr two-way left-turn lane	yes or no	no	
Presence of auto. speed enforcement	yes or no	no	
Presence of passing lane	yes or no	no	
Presence of centerline rumble strip	yes or no	no	
		Set I - All of the Above	29,372
Superelevation on horizontal curve	ft/ft	< 0.01 difference	9,664
Max. superelev. allowed by jurisdiction	ft/ft		
Percent grade	Percent	0%	
		Set II - All of the Above	9,664
Presence of lighting	yes or no	no	--
Driveway density	#/mile	five/mile	--
Roadside hazard rating	1-7	3	--

Table 8 Required characteristics for the Predictive Method (CMFs) for TLTWRR intersections (4SG, 3ST, and 4ST)

Characteristic	Units or Notes	Base Condition for CMFs	No. of Sites 4SG	No. of Sites 4ST	No. of Sites 3ST
Total Sites in Selection Set		NA	14	136	1,298
Sites Remaining After Removing Improvements		NA	11	120	1,201
Intersection skew angle	degrees departure from 90 (0 for 4SG)	0	11	120	1,201
Number of approaches with left-turn lanes, not including stop-controlled approaches	0, 1, 2, 3, or 4	0	11	120	1,201
Number of approaches with right-turn lanes, not including stop-controlled approaches	0, 1, 2, 3, or 4	0	11	120	1,201
Presence of intersection lighting	yes or no	No	11	120	1,201
AADT for major road	0-19,500 for 3ST 0-14,700 for 4ST 0-25,200 for 4SG	NA	11	99	980
AADT for minor road	0-4,300 for 3ST 0-3,500 for 4ST 0-12,500 for 4SG	NA	9	99	977

The final sets of intersections is shown in Figure 5, along with the population of 2U segments in Set II. Five percent of the sites in each category were assessed visually using Google Streetview to confirm their type and physical characteristics.

Each crash in the state for 2014, 2015, and 2016 was associated with a 2U roadway segment or a 4SG, 3ST, or 4ST intersection for the analysis. A new field, Site_Type, was created within each of the crash layers and populated with one of the site types in this study if the following were satisfied:

- 4SG, 4ST, or 3ST: The crash is within 250 feet of one of the intersection types in this study
- 2U: The crash is NOT within 250 feet of ANY intersection (even those not in the scope of this study) AND is within 50 feet of a 2U roadway segment

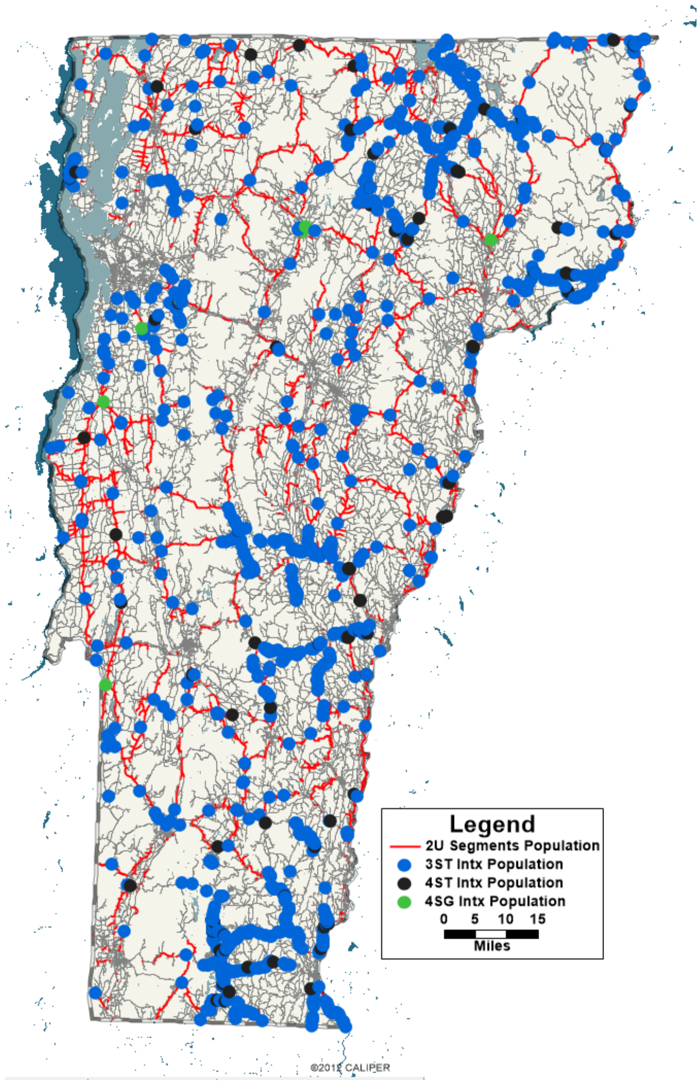


Figure 5 Final sets of intersections for this study

Using selection sets built from the new Site_Type field, crashes in the vicinity of each site were counted if they were within 250 feet of an intersection or 50 feet of a segment. After completing this step, a spot-check of all crashes within ½-mile of each segment in the more limited set of 2U segments was conducted. Crashes within ½-mile of each segment were selected, and those occurring on a highway that is not in the TLTWRR class were removed, along with those identified as occurring at an intersection. This process left about 300 crashes for each year. These crashes were inspected manually (in GIS) to determine if they should be included in the 2U data set. An additional 66 crashes across the three years in this study period were added to the 2U data set from this step. Five percent of the 2U segments (483) were inspected manually (in GIS) to identify unassociated crashes, and only one additional crash was found. Table 9 provides a summary of the crashes associated with each site type in the final data set.

Table 9 Summary of Crashes Associated with each Site Type

		2U Segments		Intersections		
		Set I	Set II	4SG	4ST	3ST
No. of Facilities with All Required Data		29,372	9,664	9	99	977
Crashes on Facilities w/ Required Data	**2014**	1,175	825	31	41	228
	2015	1,093	767	30	73	286
	2016	1,255	846	29	41	255
	Total	**3,523**	**2,438**	**90**	**155**	**769**

According to the HSM, for each facility type, the desirable minimum sample size for the calibration is 30 sites, although a number of studies suggest that 30 sites will not provide a defensible confidence interval on the resulting CFs (Shirazi and Geedipally, 2016; Banihashemi, 2012; Alluri and Gan, 2014; Trieu et al., 2014). Sites should be selected at random from a larger population of sites without regard to physical characteristics or the number of crashes at the sites. The selected sample should contain a minimum of 100 crashes over the calibration period. For our study, only the 3ST and 2U site types contain enough features to select a sample that will meet these criteria. For the 4SG and 4ST, the analysis will be performed on the entire population of sites, rather than a sample. Table 10 contains a summary of the composition of the 9,664 segments in the 2U Set II.

Table 10 Summary of the Segments in 2U Set II

	Segments			No. of Crashes		
Class	No.	Total Length (mi.)	Avg 2016 AADT	2014	2015	2016
Class 1 Undivided	145	16.3	5,196	10	11	9
Class 2 Undivided	720	245.1	2,088	36	41	44
Class3_Class4	4357	1227.6	1,868	53	48	68
Forest Hwy	35	13.5	1,617	0	0	0
Gov Hwy	3	3.7	3,233	0	0	0
State Hwy Undivided	3649	1315.6	2,871	583	536	591
US Hwy Undivided	755	259.0	4,479	144	131	133
Total	9,664	3,081		826	767	845

Using the set of 3,523 crashes on the 2U segments in Set I from 2014 to 2016, the default values in Tables 10-3, 10-4, and 10-12 of the HSM were updated to reflect Vermont-specific conditions for crash severity, collision type, and nighttime crash proportions. The following three tables contain these Vermont-specific values, alongside the HSM default values, which are based on data from Washington from 2002 to 2006.

Table 11 HSM Defaults and Vermont-Specific Crash Severity Percentages for Crashes on 2U TLTWRR Segments

Crash Severity	HSM Default	VT-2014	VT-2015	VT-2016	VT –Specific Values for allYears
Fatality	1.3%	1.0%	1.1%	1.5%	**1.2%**
Possible Injury	14.5%	7.8%	9.3%	9.5%	**8.9%**
Nonincapacitating Injury / Suspected Minor Injury	10.9%	19.7%	20.2%	21.7%	**20.6%**
Incapacitating Injury / Suspected Serious Injury	5.4%	4.0%	5.0%	4.8%	**4.6%**
Total Fatal Plus Injury	32.1%	31.7%	34.9%	36.7%	**34.5%**
No Injury	67.9%	61.9%	58.2%	55.5%	**58.4%**
Unknown	5.5%	5.7%	6.1%	6.7%	**6.2%**
Untimely Death	0.1%	0.1%	0.1%	0.1%	**0.1%**
Grand Total	**100.0%**	**100.0%**	**100.0%**	**100.0%**	**100.0%**

Table 12 HSM Defaults and Vermont-Specific Collision Type Percentages for Crashes on 2U TLTWRR Segments

Collision Type	HSM Default			VT-Specific Values, 2014-2016		
	Total Fatal Plus Injury	Property Damage Only	Total (All Severity Levels)	Total Fatal Plus Injury	Property Damage Only	Total (All Severity Levels)
Single-Vehicle Crashes						
Collision with animal	3.8%	18.4%	12.1%	1.9%	6.3%	4.6%
Collision with bicycle	0.4%	0.1%	0.2%	0.4%	0.0%	0.1%
Collision with pedestrian	0.7%	0.1%	0.3%	0.8%	0.0%	0.3%
Overturned	3.7%	1.5%	2.5%	13.3%	12.0%	12.5%
Ran off road	54.5%	50.5%	52.1%	49.5%	41.7%	44.7%
Other single-vehicle	0.7%	2.9%	2.1%	5.4%	4.5%	4.9%
Total single-vehicle	63.8%	73.5%	69.3%	71.3%	64.5%	67.1%
Multiple-Vehicle Crashes						
Angle collision	10.0%	7.2%	8.5%	2.6%	4.5%	3.8%
Head-on collision	3.4%	0.3%	1.6%	9.7%	5.1%	6.9%
Rear-end collision	16.4%	12.2%	14.2%	8.7%	12.6%	11.1%
Sideswipe collision	3.8%	3.8%	3.7%	5.5%	9.8%	8.2%
Other multiple-vehicle	2.6%	3.0%	2.7%	2.2%	3.4%	2.9%
Total multiple-vehicle crashes	36.2%	26.5%	30.7%	28.7%	35.5%	32.9%
Total Crashes	**100.0%**	**100.0%**	**100.0%**	**100.0%**	**100.0%**	**100.0%**

Table 13 HSM Defaults and Vermont-Specific Nighttime Crash Proportions for Unlighted TLTWRR 2U Roadway Segments

Crash Severity or Time of Day	HSM Default	VT-2014	VT-2015	VT-2016	VT – Specific Values for all Years
Fatality Plus Injury, p_{inr}	38.2%	39.1%	44.1%	41.2%	**41.3%**
Property-Damage Only, p_{pnr}	61.8%	60.9%	55.9%	58.8%	**58.7%**
Grand Total	**100.0%**	**100.0%**	**100.0%**	**100.0%**	**100.0%**
Proportion of Crashes that Occur at Night[1], p_{nr}	37.0%	31.4%	26.8%	31.5%	**30.0%**
Proportion of Crashes that Occur in Daytime	63.0%	68.6%	73.2%	68.5%	**70.0%**
Grand Total	**100.0%**	**100.0%**	**100.0%**	**100.0%**	**100.0%**

Notes:
1. Crashes that occurred at night were taken to be those with the following entries in the "Lighting" field: Dark-Lighted Roadway; Dark-Roadway Not Lighted; Dark - Unknown Roadway Lighting

Using the set of 1,014 crashes on the 3ST, 4ST, and 4SG intersections, the default values in Table 10-15 of the HSM were updated to reflect Vermont-specific conditions for nighttime crash proportions. Table 14 contains the Vermont-specific values, alongside the HSM default values, which are based on data from California from 2002 to 2006.

Table 14 HSM Defaults and Vermont-Specific Nighttime Crash Proportions for Unlighted TLTWRR Intersections

Intersection Site Type	HSM Defaults	VT-2014	VT-2015	VT-2016	VT – Specific Values for all Years
3ST	26.0%	26.5%	19.3%	20.3%	**21.8%**
4ST	24.4%	12.2%	26.5%	10.0%	**18.1%**
4SG	28.6%	25.8%	20.0%	13.8%	**20.0%**

3.6 Sub-Regions in Vermont

Depending on the size of the final data set available for calibration, the use of sub-regions in the state will be explored. Possible sub-regions include climatological zones and tourism destinations. Traffic corridors were disregarded as a potential sub-regional distinction, since traffic and roadway characteristics are incorporated into the HSM methods through the inclusion of variables in the SPFs and the CMFs.

3.6.1 Climatological/Tourism

Climatological conditions can vary dramatically from north to south in Vermont. This type of variation is illustrated in maps like the USDA's plant hardiness zones (Figure 6).

This variation has indirect influence on behavior of Vermonters and visitors to Vermont. Tourism travel trends are arguably more responsive to this north-south regionalization of Vermont than any other type of regionalization (https://www.vermontvacation.com/ towns-and-regions). One reason for this type of regionalization in tourist attractions in Vermont might be that Vermont attracts visitors primarily from the south (the New York and Boston mega-regions) and the north (the Montreal metropolitan area), so tourists' highway trips are constrained by the north-south distance of their destinations. Since tourism has been shown to have an influence on crash rates in Vermont, this type of regionalization can be significant for safety analyses. Based on this assessment, a three-region classification was determined:

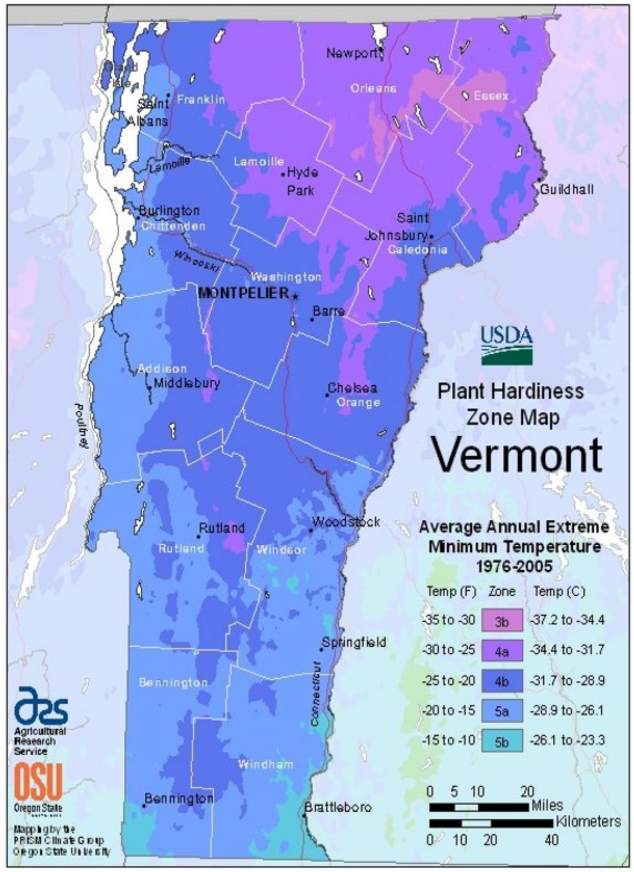


Figure 6 Plant Hardiness Zones in Vermont

- Northern (Grand Isle, Franklin, Lamoille, Orleans, Caledonia, and Essex Counties)
- Central (Addison, Chittenden, Washington, Orange, Rutland, and Windsor Counties)
- Southern (Bennington and Windham Counties)

Table 15 contains the site counts and crash counts for the 2U, 3ST, and 4ST site types for each of these three climatological/tourism regions in Vermont.

Table 15 Site Counts and Crash Counts for Climatological/Tourism Regions in Vermont

	Northern Vermont		Southern Vermont		Central Vermont	
Site Type	Total no. of sites	No. of crashes	Total no. of sites	No. of crashes	Total no. of sites	No. of crashes
2U	3,630	1,045	1,383	393	4,651	1,001
3ST	391	278	292	238	294	253
4ST	41	28	28	68	30	59

3.6.2 Physiographic

Climatologists, geographers, and natural-resource investigators might also divide our state according to physiographic regions. These regions are determined by a combination of the age and type of rock in, the natural landscape, (lowland, hills, mountains) and by the climate. Generally, Vermont is assumed to include six physiographic regions:

1. The Vermont Lowlands
2. The Green Mountains
3. The Taconic Mountains
4. The Valley of Vermont
5. The Vermont Piedmont
6. The Northeast Highlands

The approximate locations of these regions is shown in Figure 7.

These physiographic regions also affect travel behavior in a variety of ways. First, their climatological trends affect travelers on the roads, but these regions have also influenced settlement patterns in the state for many years, creating a proxy for major highway corridors and orientation of metropolitan

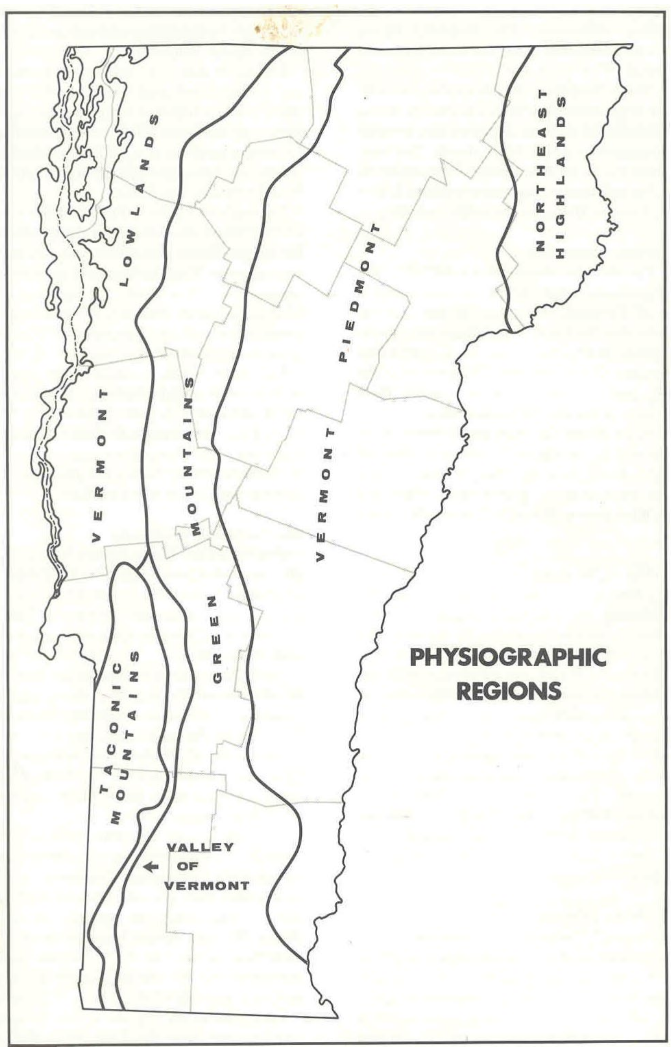


Figure 7 Physiographic Regions of Vermont

areas (http://academics.smcvt.edu/ vtgeographic/ textbook/physiographic /physiographic_regions_ of_vermont.htm).

Based on this assessment, a second classification was determined by grouping the six physiographic regions into three:

A. The Vermont Lowlands, the Valley of Vermont, and the Taconic Mountains
B. The Green Mountains
C. The Vermont Piedmont and Northeast Highlands

Table 16 contains the site counts and crash counts for the 2U, 3ST, and 4ST site types for each of these three physiographic regions in Vermont.

Table 16 Site Counts and Crash Counts for Physiographic Regions in Vermont

	Physio A		Physio B		Physio C	
Site Type	Total no. of sites	No. of crashes	Total no. of sites	No. of crashes	Total no. of sites	No. of crashes
2U	2,231	726	2,138	604	5,295	1,109
3ST	103	105	308	304	566	360
4ST	10	25	26	55	63	75

These classifications indicate that one or both of these sub-regional distinctions can be incorporated into the analysis of CFs and SPFs for the 2U and 3ST site types, but not for the 4ST and 4SG site types, due to lack of data.

3.7 Use of Injury Severity in the Analysis

Some of the studies reviewed from other states also stratified their calculations of CFs by crash severity. Crash severity, however, is related to many factors that are not the physical characteristics of the roadway, such as driver age, or are not included in the characteristics used in this study, such as the conditions surrounding the road, such as the presence of a guardrail or the Road Hazard Rating as mentioned in Table 4. Therefore, it makes less sense to stratify the analysis by crash severity. Without the data that specifically affects the severity of the crash in the analysis, it is also impossible to know how changes at the sites during the calibration period may have affected the use of those sites in the calculations.

4 Methods

4.1 Calibration Factors for 2U Segments

Once the calibration data sets were ready, calibration factors could be calculated for the 2U segments. The HSM predictive model was applied to predict total crash frequency for each site during the calibration period. To complete the predictive model, a series of crash modification factors (CMFs) were first determined for every site. Because CMFs could only be applied if the necessary data was available in the data set, Set II was used for the 2U segments. Valid CMFs applied and the HSM source for calculation of the CMFs are shown in Table 17.

Table 17 Summary of CMFs Applied for Calculation of CFs for 2U Segments

CMF No.	Characteristic	HSM Source
1	Lane width	Table 10-8
2	Shoulder width	Table 10-9
3	Shoulder type / width	Table 10-10
4	Percent grade	Table 10-11
5	Radius of horizontal curve	Equation 10-13
6	Length of horizontal curve	
7	Presence of spiral transition curve	
8	Superelevation variance	Equations 10-14 to 10-16
9	Driveway density	Equation 10-17
10	Presence of center two-way left-turn lane	Equation 10-18 and 10-19
11	Presence of automated speed enforcement	Yes is 0.93; No is 1.00
12	Presence of passing lane	Yes is 0.75; No is 1.00
13	Presence of centerline rumble strip	Yes is 0.94; No is 1.00

These CMFs are essentially a series of adjustment factors, most slightly higher or lower than 1.0, that are multiplied by the predicted number of crashes resulting from Equation 10-6 in the HSM:

$$N_{spf-2U,i} = AADT_i \times L_i \times .000365 \times e^{-0.312} \quad (2)$$

CMFs that are less than 1.0 indicate a physical characteristic that makes the segment generally safer, decreasing the predicted number of crashes at the site, whereas a CMF greater than 1.0 would indicates a characteristic that increases the predicted number of crashes at the site.

Computing the calibration factors involved the use of the predictive model, then the comparison of the resulting predicted number of crashes with the observed number of crashes. Generally, the calibration factor (CF) for site type r using sample i can be written as:

$$CF_{r,i} = \frac{\sum_i OC_r}{\sum_i PC_r} \tag{3}$$

Where OC_r are the observed no. of crashes on all sites of type r and PC_r are the predicted no. of crashes on all sites of type r:

$$PC_{r,i} = N_{spf-2U,i} \times (CMF_1 \times CMF_2 \times CMF_3 \times \dots) \tag{4}$$

4.2 Calibration Factors for Intersections

Computing calibration factors for the 3ST, 4ST, and 4SG intersections involved a similar process as it did for segments, except that the set of characteristics used to determine CMFs is more limited, as shown in Table 18.

Table 18 Summary of CMFs Applied for Calculation of CFs for Intersections

CMF No.	Characteristic	HSM Source
1	Intersection skew angle	Equation 10-22 and 10-23
2	Number of approaches with left-turn lanes, not including stop-controlled approaches	Table 10-13
3	Number of approaches with right-turn lanes, not including stop-controlled approaches	Table 10-14
4	Presence of intersection lighting	Table 10-15

Similar to the approach used for 2U segments, the CMFs are used to adjust the predicted number of crashes resulting from the type-specific SPFs:

$$N_{spf-3ST,i} = e^{[-9.86+0.79*\ln(AADT_{i,maj})+0.49*\ln(AADT_{i,min})]} \tag{5}$$

$$N_{spf-4ST,i} = e^{[-8.56+0.60*\ln(AADT_{i,maj})+0.61*\ln(AADT_{i,min})]} \tag{6}$$

$$N_{spf-4SG,i} = e^{[-5.13+0.60*\ln(AADT_{i,maj})+0.20*\ln(AADT_{i,min})]} \tag{7}$$

4.3 Re-Estimation of Safety Performance Functions

Re-estimating the safety performance functions involved the use of the sample sites to re-estimate the coefficients of the SPFs provided in the HSM, ignoring the application of the CFs. For the 2U site type, this analysis is conducted to estimate coefficients on AADT and L. NB regression is suited to data sets that consist primarily of 0s for the dependent variable. For the entire three-year period used for this study, the number of crashes is dominated by 0s, with 86% for the 2U site type and 63% for the 3ST site type. Although some states fixed the coefficients for AADT or L in their NB regression, all three were included in our estimation to improve model fit. The form of the function estimated for this study was:

$$N_{spf-2U,i} = e^{a*\ln(AADT_i)+b*\ln(L_i)+c} = AADT_i^{a} * L_i^{b} * e^{c} \quad (8)$$

The NB regression provides estimates for coefficients a, b, and c that minimize the log likelihood function, such that:

$$PC_{r,i} \approx OC_{r,i} \quad (9)$$

For the intersections, a similar approach is used, except that coefficients are estimated on the AADT of the major road, the AADT of the minor road, and a constant:

$$N_{spf,i} = e^{[a*\ln(AADT_{i,maj})+b*\ln(AADT_{i,min})+c]} \quad (10)$$

5 Results

5.1 Calibration Factors for 2U Segments

The calculated calibration factors for 2U segments in Vermont are:

- **Statewide: 0.298**
- Northern: 0.318
- Southern: 0.367
- Central: 0.285
- Physio A: 0.214
- Physio B: 0.316
- Physio C: 0.363

The regional breakdown indicates a slightly elevated crash rate in the southern region of the state, as opposed to the central and northern region, and in the Green Mountains (Physio B) and Vermont Piedmont (Physio C) as opposed to the western edge of the state (Physio A). Figure 8 provides a set of maps of Vermont illustrating the categorization of towns into each of the two region types.

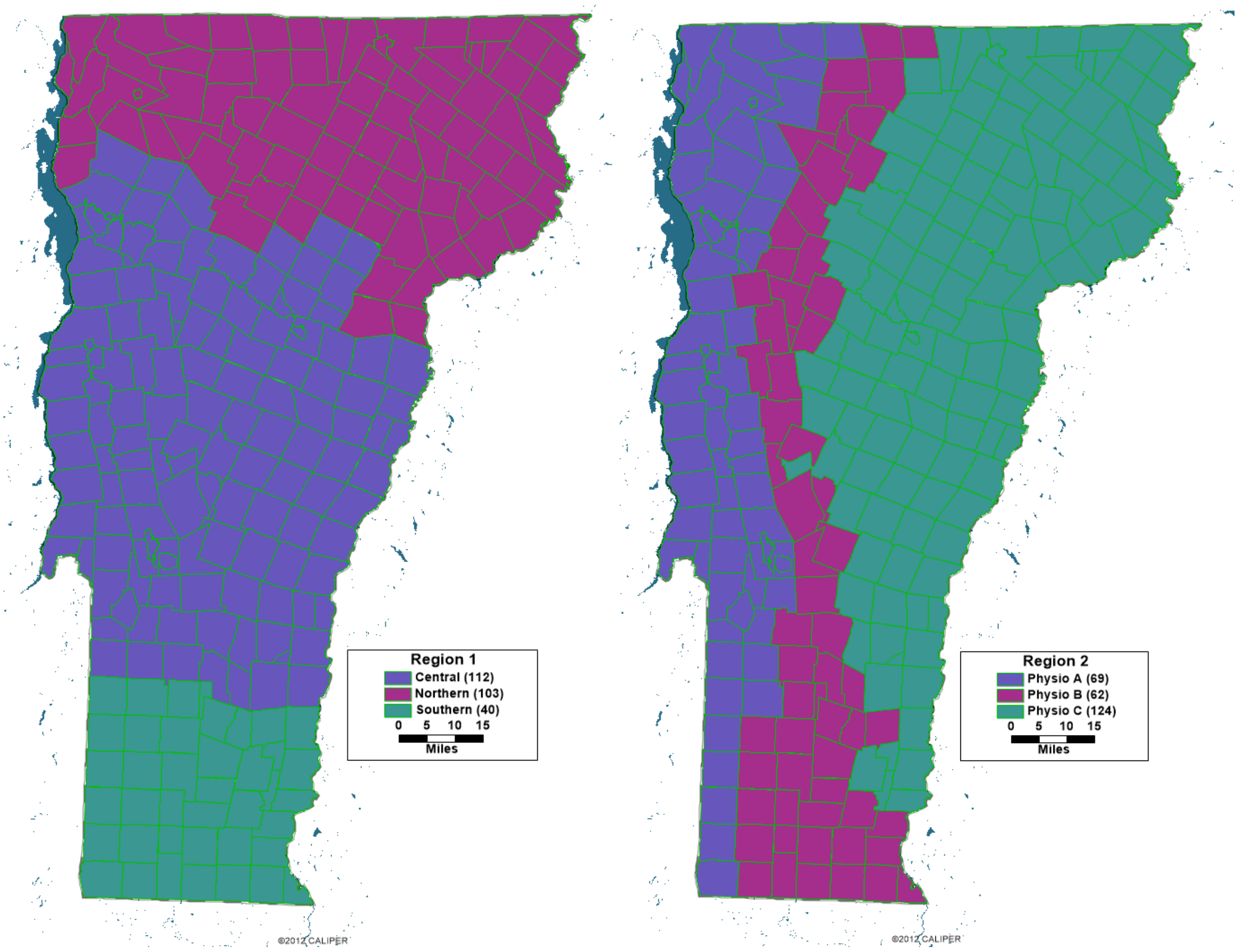


Figure 8 Categorizations of Towns Used in the Calculation of CFs

As shown in Table 1, the statewide CF for 2U segments in Vermont is *significantly* lower than the average CF of all states reviewed in this study (1.06). In fact, the CF for Vermont is lower than the lowest CF for any state reviewed (0.70 for Maryland). The explanation for these low CFs may be due to the approach used in this study to assign crashes to segments. The approach used was to assign crashes to a segment only if they had NOT previously been assigned to an intersection *and* were within 50 feet of the segment. This approach potentially leaves some crashes that were not located near the segment unassigned. To address this possibility, all crashes within ½-mile of the 2U segments were inspected individually and a few were added to the 2U data set. In addition, our removal of crashes denoted as "non-reportable" due to their inconsistent reporting, may have served to reduce this CF if those types of crashes were included in the calculation by other states.

Another possible explanation for the low CF that we found in this study is a temporal trend in decreasing crash rates in Vermont and nationwide between the years used to develop the HSM defaults and the years in our crash period. To examine this explanation further, Table 19 exhibits the 2U CFs for other states, sorted by the median of the years of crash data used to calculate the CF.

Table 19 2U Segment CFs for States Reviewed in this Study

State	2U CF	Years in Crash Period
Washington / Minnesota	1.00	1985-1995
Oregon	0.74	2004-2006
Kansas	1.48	2005-2007
Utah	1.16	2005-2007
N. Carolina	1.08	2004-2008
Alabama	1.39	2006-2009
Idaho	0.87	2003-2012
Maryland	0.70	2008-2010
Louisiana	0.97	2009-2011
Maine	1.08	2009-2011
Missouri	0.82	2009-2011
N. Carolina	1.09	2009-2015
S. Carolina	0.99	2013-2015

A linear trendline of these CFs across the years has only a slight downward slope (-0.002 per year). More revealing are data from the Bureau of Transportation

Statistics (BTS, 2017). Rates of all crashes per 100 million vehicle-miles traveled (VMT) in the U.S. has decreased generally since 1990, from 302 in 1990 to 203 in 2015. These rates show a stronger downward trend of -3.86 crashes per 100M VMT per year (Figure 9).

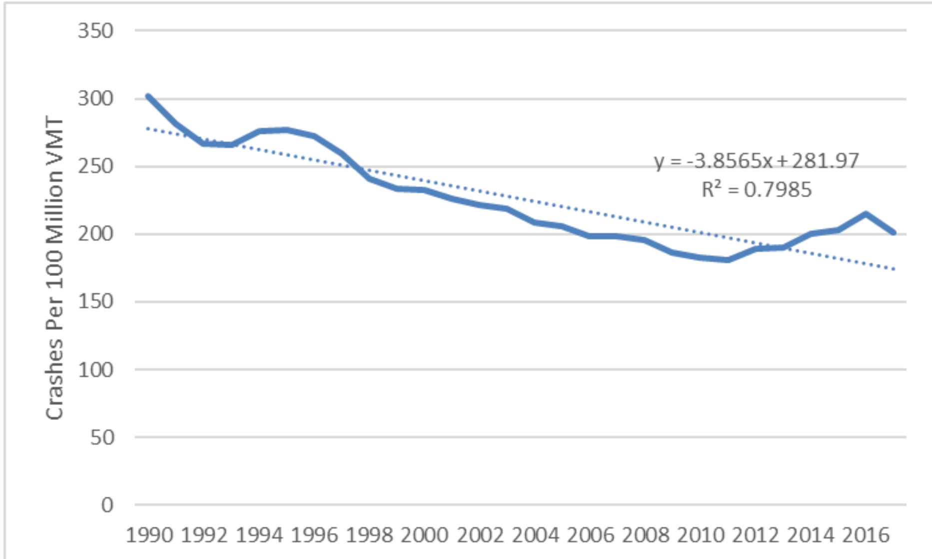


Figure 9 Downward Trend in Crash Rates in the U.S., 1990 - 2016

Crash rates in Vermont also attest to a general decrease over the study period. With "non-reportable" crashes removed, since 2014, Vermont's rate of crashes per 100 million VMT are shown in Table 20.

Table 20 Vermont Crash Rates, 2014 - 2017

Year	Total Crashes	Total VMT (100 millions)	Crashes per 100 million VMT
2014	10,518	7,059	1.49
2015	10,767	7,314	1.47
2016	10,411	7,382	1.41
2017	10,246	7,424	1.38

Finally, as of 2019, Vermonters own the newest vehicles in the U.S., according to the Alliance of Automobile Manufacturers (AAM, 2019). Vermont ranks last in the average age of its vehicles, at 9.7 years, with the nationwide average at 11.8 years. Having a newer passenger-vehicle fleet will create better safety outcomes for Vermont. Each of these explanations attests to a trend of decreasing crash rates over time that may be reflected in the low CF calculated for Vermont's 2U segments.

5.2 Re-Estimated SPFs for 2U Segments

A summary of the re-estimated SPFs for the 2U segments is provided in Table 21. The values reviewed by Idaho, Illinois, and Virginia, which all estimated the same functional form, are also provided for comparison.

Table 21 Re-Estimated SPFs for 2U Segments for Vermont, Idaho, Illinois, and Virginia

		a (AADT)	b (L)	c (e)
Vermont	All Years	**0.763**	**1.388**	**-6.634**
	2014	0.729	1.298	-6.665
	2015	0.815	1.355	-7.389
	2016	0.694	1.330	-6.370
Idaho		0.737	0.894	-5.800
Illinois		0.525	1	-4.435
Virginia		0.744	1	-5.710

The observed crashes in Vermont affected the resulting SPF in two ways. The first was to create a higher coefficient on the length of the segment (b), indicating that Vermont's predicted number of crashes on TLTWRR segments may be more affected by the length of the segment than in other states. However, Illinois and Virginia both held this coefficient at one, so it could not be determined how it would have changed if they had allowed it to be re-estimated. The biggest difference between Vermont and the other states was in the estimation of the coefficient on e (c), which is significantly lower than for any of the other states. This is not surprising, since this value tends to make the corresponding estimate of predicted number of crashes significantly lower. SPSS output for the NB regressions performed for this re-estimation are provided in Appendix A.

5.3 Calibration Factors for Intersections

A summary of the calculated statewide and regional calibration factors for TLTWRR intersections in Vermont is provided in Table 22. For the 4ST and 4SG site types, however, the regional CFs have been screened back because sample sizes for these regional breakdowns fell below the thresholds established in the HSM. The statewide CFs for all three site types are all similar to the average of the states reviewed in this study. The regional breakdown for the 3ST site type indicates a distinction between the slightly higher crash rate in the Green Mountains (Physio B) compared to the rest of Vermont (Physio A and C), which is not surprising considering the driving conditions that are frequently encountered in this mountainous region.

Table 22 Calculated CFs for TLTWRR Intersections in Vermont

Intersection Site Type	3ST	4ST	4SG
Statewide	0.448	0.448	0.568
Northern	0.432	0.322	0.456
Southern	0.463	0.597	0.771
Central	0.449	0.411	0.695
Physio A	0.375	0.616	0.277
Physio B	0.526	0.645	0.924
Physio C	0.419	0.343	0.306

5.4 Re-Estimated SPFs for Intersections

A summary of the coefficients of the re-estimated SPFs for the three intersection site types is provided in Table 23.

Table 23 Re-Estimated SPFs for TLTWRR Intersections in Vermont

Intx Site Type	a ($AADT_{maj}$)	b ($AADT_{min}$)	c (e)
3ST	0.936	0.357	-9.835
4ST	0.759	0.484	-8.665
4SG	1.452	0.667	-16.395

SPSS output for the NB regressions are provided in Appendix B (3ST), Appendix C (4ST), and Appendix D (4SG). The re-estimated SPF for the 4SG site type is not statistically viable, based on the confidence level for the individual parameters estimates provided in Appendix D.

5.5 Comparison of CFs and Re-Estimated SPFs

Based on the approach used by Idaho (Abdel-Rahim and Sipple, 2015) and Virginia (Hass et al., 2010), we used the Freeman-Tukey (FT) R^2 measure to compare the application of the calculated CFs and the re-estimated SPFs for goodness of fit. Equations 11 through13 show how the data are transformed to calculate the FT R^2 (Fridstrøm et al., 1994):

$$f_i = y_i^{0.5} + (y_i + 1)^{0.5} \tag{11}$$

$$\hat{e}_i = f_i - (4 * \hat{y}_i + 1)^{0.5} \quad (12)$$

$$R^2_{FT} = 1 - \frac{\sum \hat{e}_i^2}{\sum (f_i - f_m)^2} \quad (13)$$

where

f_i = Freeman-Tukey transformation statistic
y_i = observed data at site i
$\hat{e}_i$ = residual at site i
$\hat{y}_i$ = modeled (predicted) value at site i
f_m = mean of the FT transformation statistic across all sites

Table 24 provides the results of this calculation for the models developed in this study – first the re-estimated SPFs, and then the applied CFs.

Table 24 Freeman-Tukey R^2 for the Four Site Types in this Study

	Freeman-Tukey R^2	
Site Type	Re-Estimated SPF	Applied CF
2U	**0.240**	**0.247**
3ST	**0.265**	0.094
4ST	**0.383**	-0.387
4SG	0.271	-1.106

The re-estimated SPFs resulted in a better fitting model for all site types except the 2U segments. The HSM suggests that this will be the case when re-estimated SPFs are compared with the HSM default model applied with a CF. For the 2U segments, the newly-derived CF and CMFs results in a slightly better model. However, the re-estimated SPF requires only two parameters – AADT and length, whereas the CF (and CMFs) uses 13 additional characteristics. If these characteristics are unavailable or of questionable quality, then the new SPF can be used without significant loss of accuracy.

For the 3ST and 4ST site types, the re-estimated SPFs are superior to the applied CFs. These SPFs should be used in place of the HSM-prescribed method, with the traditional SPF and the application of a CF. For the 4SG site type, the re-estimated SPF should not be used in spite of its improved fit. The lack of a sufficient data set for this site type makes all of the results in this study for 4SG intersections questionable.

6 Conclusions and Recommendations

Based on the results of this research, we recommend using the re-estimated SPFs for 3ST, 4ST and 2U site types for the calculation of the predicted number of crashes in Vermont. For the 4SG site type, we recommend using the traditional HSM approach, with CMFs applied for the four characteristics shown in Table 18 and the HSM default SPF.

Since the SPF for the 2U segments only uses the AADT and length of the segment, a final re-estimated SPF was determined for the 2U segment site type using the larger collection of segments represented as Set I in Table 9. The SPSS output for this final NB regression is provided in Appendix A. Including this final re-estimation for the 2U site type, the final set of equations recommended for use in calculating the site-specific predicted number of crashes in Vermont for the TLTWRR class are:

$$N_{spf-2U,i} = AADT_i^{0.812} * L_i^{1.407} * e^{-7.036} \qquad (14)$$

$$N_{spf-3ST,i} = e^{[0.936*\ln(AADT_{i,maj})+0.357*\ln(AADT_{i,min})-9.835]} \qquad (15)$$

$$N_{spf-4ST,i} = e^{[0.759*\ln(AADT_{i,maj})+0.484*\ln(AADT_{i,min})-8.665]} \qquad (16)$$

$$N_{spf-4SG,i} = e^{[0.60*\ln(AADT_{i,maj})+0.20*\ln(AADT_{i,min})-5.13]} \times (CMF_{4SG-1} \times CMF_{4SG-2} \times CMF_{4SG-3} \times CMF_{4SG-4}) \qquad (17)$$

The HSM recommends that these calibration factors be updated at least every three years, and recommends combining all three years of data. It would be more effective in the future to use a Bayes approach with the individual years' data to arrive at a final SPF. This approach will take advantage of any possible trends in traffic safety that are influencing the data.

Crash data quality collection, management, and distribution can continue to improve. It is important to avoid empty data and to ensure consistency in location descriptions for data to be used for these estimations. Annual crash data on the open geodata portal (http://geodata.vermont.gov/) should contain all of the data from the original crash reports. In particular, the following fields are critical for the types of analyses dictated by the HSM:

- Location – latitude/longitude
- Date/Time
- Direction of collision
- Roadway characteristic
- Animal involved (wild; moose; deer; domestic; none/other)

- Impairment (alcohol; alcohol and drugs)
- Involving (pedestrian; motorcycle; heavy truck; bicycle; none/other)
- Crash type (fatal; injury; property damage only; unknown crash type)
- Crash injury (fatality, suspected serious injury, suspected minor injury, possible injury, no injury, unknown, and untimely death)
- Light conditions (dark – lighted roadway, dark – roadway not lighted, dark – unknown roadway lighting, dawn, daylight, dusk, not reported, other, unknown

Undefined entries in any data fields should be avoided. The ID field should be uniform across reporting agencies and crash types.

7 References

Alliance of Automotive Manufacturers, 2019. See What Your State Drives: Average Age of Vehicle Ranked by State. Accessed at https://autoalliance.org/economy/consumer-choice/average-age-of-vehicle/ on August 14, 2019.

Abdel-Rahim, Ahmed and Matthew C. Sipple, 2015. Calibrating the Highway Safety Manual Crash Prediction Models for Idaho's Highways. Prepared for Idaho Transportation Department Research Program, Contracting Services, Division of Engineering Services by the National Institute for Advanced Transportation Technology, University of Idaho, August 2015.

Alluri, P., D. Saha, and A. Gan. Minimum Sample Sizes for Estimating Reliable Highway Safety Manual (HSM) Calibration Factors. Journal of Transportation Safety & Security, 2014.

Banihashemi, M. Sensitivity Analysis of Data Set Sizes for Highway Safety Manual Calibration Factors. Transportation Research Record: Journal of the Transportation Research Board, No. 2279, Transportation Research Board of the National Academies, Washington, D.C., 2012, pp. 75-81.

Belz, Darryl, 2017. Local Calibration Factors for Implementing the Highway Safety Manual in Maine. Presented at the 2017 Northeast Transportation Safety Conference in Cromwell, Connecticut on October 24 -25, 2017.

Dixon, Karen, and Chris Monsere, Fei Xie and Kristie Gladhill, 2012. Calibrating the HSM Predictive Methods for Oregon Highways. Prepared by the Oregon Transportation Research and Education Center for the Oregon Department of Transportation, Report No. OTREC-RR-12-02, 2012.

Fridstrøm, L., Ifver, J., Ingebrigtsen, S., Kulmala, R., and Thomsen, L.K. Measuring the Contribution of Randomness, Exposure, Weather, and Daylight to the Variation in Road Accident Counts. *Accident Analysis and Prevention,* Vol. 27, 1994, pp. 1-20.

Hass, Phillip R., Conrad Gosse, and Nicholas J. Garber, 2010. Development of Safety Performance Functions for Two-Lane Roads Maintained by the Virginia Department of Transportation. Prepared for the Virginia Department of Transportation by the Virginia Transportation Research Council. VTRC 10-R25, June 2010.

Mehta, Gaurav, and Yingyan Lou, 2013. Safety Performance Function Calibration and Development for the State of Alabama: Two-Lane Two-Way Rural Roads and

Four-Lane Divided Highways. Transportation Research Record: Journal of the Transportation Research Board, No. 2398, Transportation Research Board of the National Academies, Washington, D.C., 2013, pp. 75–82.

Ogle, Jennifer H., and Mahdi Rajabi, 2018. Development of South Carolina Databases and Calibration Factors for the Highway Safety Manual: Final Report. Report FHWA-SC-18-05 prepared by the Glenn Department of Civil Engineering, Clemson University for the South Carolina Department of Transportation, December 2018.

Saito, Mitsuru, Bradford K. Brimley, and Grant G. Schultz, 2011. Report No. UT-10.12b, Transportation Safety Data & Analysis Volume 2: Calibration of the Highway Safety Manual and Development of New Safety Performance Functions. Prepared for the Utah Department of Transportation, Traffic and Safety Research Divisions by Brigham Young University, March 2011.

Schrock, Steven, Ming-Heng Wang, 2013. Evaluation of Interactive Highway Safety Design Model Crash Prediction Tools for Two-Lane Rural Roads on Kansas Department of Transportation Projects. Prepared by the University of Kansas, Report No. K-TRAN: KU-10-1, November 2013.

Shin, Hyeonshic, Young-Jae Lee, and Seyedehsan Dadvar, 2014. The Development of Local Calibration Factors for Implementing the Highway Safety Manual in Maryland. Prepared for the Maryland Department of Transportation, State Highway Administration by Morgan State University, Project No. SP209B4J, March 2014.

Shirazi, M, D. Lord, and S.R. Geedipally. Sample-Size Guidelines for Recalibrating Crash Prediction Models: Recommendations for the Highway Safety Manual. Presented at the 95th Annual Meeting of the Transportation Research Board, Washington, D.C., 2016.

Smith, Sarah, Daniel Carter, and Raghavan Srinivasan, 2017. Updated and Regional Calibration Factors for Highway Safety Manual Crash Prediction Models. Report no. 2016-09, prepared for the North Carolina Department Transportation by the University of North Carolina Highway Safety Research Center, February 2017.

Srinivasan, Raghavan, and Daniel Carter., 2011. Development of Safety Performance Functions for North Carolina. Prepared for the North Carolina Department Transportation by the University of North Carolina Highway Safety Research Center, December 2011.

Sun, Carlos (Principal Investigator), 2017. Missouri Highway Safety Manual Recalibration. Research In Progress. Publication expected by the start of this project.

Tegge, Robert A., Jang-Hyeon Jo, and Yanfeng Ouyang, 2010. Development and Application of Safety Performance Functions for Illinois. Prepared by Illinois Center for Transportation for the Illinois Department of Transportation, Bureau of Materials and Physical Research, Research Report ICT-10-066, March 2010.

Trieu, V., S. Park, and J. McFadden. Use of Monte Carlo Simulation for a Sensitivity Analysis of Highway Safety Manual Calibration Factors. Transportation Research Record: Journal of the Transportation Research Board, No. 2435, Transportation Research Board of the National Academies, Washington, D.C., 2014, pp. 1-10.

Wolshon, Brian, and Bridget Robicheaux, 2015. Calibration of the Louisiana Highway Safety Manual. Prepared by the Louisiana Transportation Research Center for the Louisiana Department of Transportation and Development, LTRC Project No. 12-3SA, State Project No. 300-00-603, and SWUTC Project No. 600451-00102, February 2015.

Zwilling, Michael, 2013. Negative Binomial Regression. The Mathematica Journal 15. © 2013 Wolfram Media, Inc.

Appendix A SPSS Output for 2U Segments

Dependent Variable: Three-Year Crashes

Model: (Intercept), ln(L), ln(Three-Year AADT)

Model Information

Dependent Variable	Three-Year Crashes
Probability Distribution	Negative binomial
Link Function	Log

Continuous Variable Information

		N	Minimum	Maximum	Mean	Std. Deviation
Dependent Variable	Three-Year Crashes	9664	0	19	.25	.818
Covariate	ln(L)	9664	-6.571283042360924	1.585227183719042	-1.772893622917250	1.207538608230885
	ln(AADT)	9664	5.480638923341991	10.815770263012745	8.293271015663308	1.236653671903057

Goodness of Fit[a]

	Value	df	Value/df
Deviance	4017.568	9661	.416
Scaled Deviance	4017.568	9661	
Pearson Chi-Square	8886.811	9661	.920
Scaled Pearson Chi-Square	8886.811	9661	
Log Likelihood[b]	-4286.362		
Akaike's Information Criterion (AIC)	8578.724		
Finite Sample Corrected AIC (AICC)	8578.727		
Bayesian Information Criterion (BIC)	8600.253		
Consistent AIC (CAIC)	8603.253		

Dependent Variable: Three-Year Crashes
Model: (Intercept), ln(L), ln(AADT)
a. Information criteria are in smaller-is-better form.
b. The full log likelihood function is displayed and used in computing information criteria.

Omnibus Test[a]

Likelihood Ratio Chi-Square	df	Sig.
3374.722	2	.000

Dependent Variable: Three-Year Crashes
Model: (Intercept), ln(L), ln(AADT)
a. Compares the fitted model against the intercept-only model.

Tests of Model Effects

	Type III		
Source	Wald Chi-Square	df	Sig.
(Intercept)	661.745	1	.000
ln(L)	1633.398	1	.000
ln(AADT)	667.650	1	.000

Dependent Variable: Three-Year Crashes
Model: (Intercept), ln(L), ln(AADT)

Parameter Estimates

			95% Wald Confidence Interval		Hypothesis Test		
Parameter	B	Std. Error	Lower	Upper	Wald Chi-Square	df	Sig.
(Intercept)	-6.634	.2579	-7.140	-6.129	661.745	1	.000
ln(L)	1.388	.0343	1.320	1.455	1633.398	1	.000
ln(AADT)	.763	.0295	.706	.821	667.650	1	.000
(Scale)	1[a]						
(Negative binomial)	1[a]						

Dependent Variable: Three-Year Crashes
Model: (Intercept), ln(L), ln(AADT)
a. Fixed at the displayed value.

Dependent Variable: 2014 Crashes

Model: (Intercept), ln(2014_AADT), ln(L)

Model Information

Dependent Variable	2014 Crashes
Probability Distribution	Negative binomial (1)
Link Function	Log

Continuous Variable Information

		N	Minimum	Maximum	Mean	Std. Deviation
Dependent Variable	2014 Crashes	9664	0	9	.08	.368
Covariate	ln(2014_AADT)	9664	4.382026634673881	9.711115659888671	7.148423429911131	1.266947222688559
	ln(L)	9664	-6.571283042360924	1.585227183719042	-1.772893622917250	1.207538608230885

Goodness of Fit[a]

	Value	df	Value/df
Deviance	2489.009	9661	.258
Scaled Deviance	2489.009	9661	
Pearson Chi-Square	7636.438	9661	.790
Scaled Pearson Chi-Square	7636.438	9661	
Log Likelihood[b]	-2174.491		
Akaike's Information Criterion (AIC)	4354.982		
Finite Sample Corrected AIC (AICC)	4354.985		
Bayesian Information Criterion (BIC)	4376.511		
Consistent AIC (CAIC)	4379.511		

Dependent Variable: 2014 Crashes
Model: (Intercept), ln(2014_AADT), ln(L)
a. Information criteria are in smaller-is-better form.
b. The full log likelihood function is displayed and used in computing information criteria.

Omnibus Test[a]

Likelihood Ratio Chi-Square	Df	Sig.
1296.826	2	.000

Dependent Variable: 2014 Crashes
Model: (Intercept), ln(2014_AADT), ln(L)
a. Compares the fitted model against the intercept-only model.

Tests of Model Effects

	Type III		
Source	Wald Chi-Square	df	Sig.
(Intercept)	385.337	1	.000
ln(2014_AADT)	274.721	1	.000
ln(L)	699.646	1	.000

Dependent Variable: 2014 Crashes
Model: (Intercept), ln(2014_AADT), ln(L)

Parameter Estimates

			95% Wald Confidence Interval		Hypothesis Test		
Parameter	B	Std. Error	Lower	Upper	Wald Chi-Square	df	Sig.
(Intercept)	-6.665	.3396	-7.331	-6.000	385.337	1	.000
ln(2014_AADT)	.729	.0440	.643	.815	274.721	1	.000
ln(L)	1.298	.0491	1.202	1.395	699.646	1	.000
(Scale)	1[a]						
(Negative binomial)	1[a]						

Dependent Variable: 2014 Crashes
Model: (Intercept), ln(2014_AADT), ln(L)
a. Fixed at the displayed value.

Dependent Variable: 2015 Crashes

Model: (Intercept), ln(2015_AADT), ln(L)

Model Information

Dependent Variable	2015 Crashes
Probability Distribution	Negative binomial (1)
Link Function	Log

Continuous Variable Information

		N	Minimum	Maximum	Mean	Std. Deviation
Dependent Variable	2015 Crashes	9664	0	7	.08	.363
Covariate	ln(L)	9664	-6.571283042360924	1.585227183719042	-1.772893622917250	1.207538608230885
	ln(2015_AADT)	9664	4.382026634673881	9.711115659888671	7.148423429911131	1.266947222688559

Goodness of Fit[a]

	Value	df	Value/df
Deviance	2323.781	9661	.241
Scaled Deviance	2323.781	9661	
Pearson Chi-Square	8192.290	9661	.848
Scaled Pearson Chi-Square	8192.290	9661	
Log Likelihood[b]	-2028.313		
Akaike's Information Criterion (AIC)	4062.626		
Finite Sample Corrected AIC (AICC)	4062.628		
Bayesian Information Criterion (BIC)	4084.154		
Consistent AIC (CAIC)	4087.154		

Dependent Variable: 2015 Crashes
Model: (Intercept), ln(L), ln(2015_AADT)
a. Information criteria are in smaller-is-better form.
b. The full log likelihood function is displayed and used in computing information criteria.

Omnibus Test[a]

Likelihood Ratio Chi-Square	df	Sig.
1329.002	2	.000

Dependent Variable: 2015 Crashes
Model: (Intercept), ln(L), ln(2015_AADT)
a. Compares the fitted model against the intercept-only model.

Tests of Model Effects

	Type III		
Source	Wald Chi-Square	df	Sig.
(Intercept)	409.379	1	.000
ln(L)	677.654	1	.000
ln(2015_AADT)	301.912	1	.000

Dependent Variable: 2015 Crashes
Model: (Intercept), ln(L), ln(2015_AADT)

Parameter Estimates

			95% Wald Confidence Interval		Hypothesis Test		
Parameter	B	Std. Error	Lower	Upper	Wald Chi-Square	df	Sig.
(Intercept)	-7.389	.3652	-8.104	-6.673	409.379	1	.000
ln(L)	1.355	.0521	1.253	1.457	677.654	1	.000
ln(2015_AADT)	.815	.0469	.723	.907	301.912	1	.000
(Scale)	1[a]						
(Negative binomial)	1[a]						

Dependent Variable: 2015 Crashes
Model: (Intercept), ln(L), ln(2015_AADT)
a. Fixed at the displayed value.

Dependent Variable: 2016 Crashes

Model: (Intercept), ln(2016_AADT), ln(L)

Model Information

Dependent Variable	2016 Crashes
Probability Distribution	Negative binomial (1)
Link Function	Log

Continuous Variable Information

		N	Minimum	Maximum	Mean	Std. Deviation
Dependent Variable	2016 Crashes	9664	0	5	.09	.353
Covariate	ln(L)	9664	-6.571283042360924	1.585227183719042	-1.772893622917250	1.207538608230885
	ln(2016_AADT)	9664	4.382026634673881	9.729134165391350	7.221092973724252	1.247796733761745

Goodness of Fit[a]

	Value	df	Value/df
Deviance	2477.703	9661	.256
Scaled Deviance	2477.703	9661	
Pearson Chi-Square	7238.167	9661	.749
Scaled Pearson Chi-Square	7238.167	9661	
Log Likelihood[b]	-2227.494		
Akaike's Information Criterion (AIC)	4460.988		
Finite Sample Corrected AIC (AICC)	4460.990		
Bayesian Information Criterion (BIC)	4482.516		
Consistent AIC (CAIC)	4485.516		

Dependent Variable: 2016 Crashes
Model: (Intercept), ln(L), ln(2016_AADT)
a. Information criteria are in smaller-is-better form.
b. The full log likelihood function is displayed and used in computing information criteria.

Omnibus Test[a]

Likelihood Ratio Chi-Square	df	Sig.
1318.718	2	.000

Dependent Variable: 2016 Crashes
Model: (Intercept), ln(L), ln(2016_AADT)
a. Compares the fitted model against the intercept-only model.

Tests of Model Effects

Source	Wald Chi-Square	Type III df	Sig.
(Intercept)	358.160	1	.000
ln(L)	719.477	1	.000
ln(2016_AADT)	254.510	1	.000

Dependent Variable: 2016 Crashes
Model: (Intercept), ln(L), ln(2016_AADT)

Parameter Estimates

			95% Wald Confidence Interval		Hypothesis Test		
Parameter	B	Std. Error	Lower	Upper	Wald Chi-Square	df	Sig.
(Intercept)	-6.370	.3366	-7.030	-5.710	358.160	1	.000
ln(L)	1.330	.0496	1.233	1.427	719.477	1	.000
ln(2016_AADT)	.694	.0435	.608	.779	254.510	1	.000
(Scale)	1[a]						
(Negative binomial)	1[a]						

Dependent Variable: 2016 Crashes
Model: (Intercept), ln(L), ln(2016_AADT)
a. Fixed at the displayed value.

Dependent Variable: Three-Year Crashes (for the larger set of 2U segments)

Model: (Intercept), ln(L), ln(Three-Year AADT)

Model Information

Dependent Variable	Three-Year Crashes
Probability Distribution	Negative binomial
Link Function	Log

Continuous Variable Information

		N	Minimum	Maximum	Mean	Std. Deviation
Dependent Variable	Three-Year Crashes	29371	0	19	.12	.544
Covariate	ln(L)	29371	-6.571283042360924	1.885259223321505	-1.636688530790841	1.157552322315696
	ln(AADT)	29371	5.010635294096256	10.815770263012745	7.111125381658716	1.189646359815400

Goodness of Fit[a]

	Value	df	Value/df
Deviance	8254.500	29368	.281
Scaled Deviance	8254.500	29368	
Pearson Chi-Square	29971.664	29368	1.021
Scaled Pearson Chi-Square	29971.664	29368	
Log Likelihood[b]	-7753.598		
Akaike's Information Criterion (AIC)	15513.195		
Finite Sample Corrected AIC (AICC)	15513.196		
Bayesian Information Criterion (BIC)	15538.058		
Consistent AIC (CAIC)	15541.058		

Dependent Variable: Three-Year Crashes
Model: (Intercept), ln(L), ln(AADT)
a. Information criteria are in smaller-is-better form.
b. The full log likelihood function is displayed and used in computing information criteria.

Omnibus Test[a]

Likelihood Ratio Chi-Square	df	Sig.
6887.849	2	.000

Dependent Variable: Three-Year Crashes
Model: (Intercept), ln(L), ln(AADT)
a. Compares the fitted model against the intercept-only model.

Tests of Model Effects

	Type III		
Source	Wald Chi-Square	df	Sig.
(Intercept)	3091.592	1	.000
ln(L)	2681.027	1	.000
ln(AADT)	2595.084	1	.000

Dependent Variable: Three-Year Crashes
Model: (Intercept), ln(L), ln(AADT)

Parameter Estimates

			95% Wald Confidence Interval		Hypothesis Test		
Parameter	B	Std. Error	Lower	Upper	Wald Chi-Square	df	Sig.
(Intercept)	-7.036	.1265	-7.284	-6.788	3091.592	1	.000
ln(L)	1.407	.0272	1.354	1.461	2681.027	1	.000
ln(AADT)	.812	.0159	.781	.844	2595.084	1	.000
(Scale)	1[a]						
(Negative binomial)	1[a]						

Dependent Variable: Three-Year Crashes
Model: (Intercept), ln(L), ln(AADT)
a. Fixed at the displayed value.

Appendix B SPSS Output for 3ST Intersections

Dependent Variable: Three-Year Crashes

Model: (Intercept), ln(AADTmaj), ln(AADTmin)

Model Information

Dependent Variable	Three-Year Crashes
Probability Distribution	Negative binomial (1)
Link Function	Log

Continuous Variable Information

		N	Minimum	Maximum	Mean	Std. Deviation
Dependent Variable	Three-Year Crashes	977	0	17	.79	1.520
Covariate	Ln AADTmaj	977	3.912023005428146	9.517825071724143	7.613399428178031	.841697009443622
	Ln AADTmin	977	4.382026634673881	8.699514748210191	5.760732558935749	.854861557840383

Goodness of Fit[a]

	Value	df	Value/df
Deviance	774.047	974	.795
Scaled Deviance	774.047	974	
Pearson Chi-Square	1141.192	974	1.172
Scaled Pearson Chi-Square	1141.192	974	
Log Likelihood[b]	-1041.183		
Akaike's Information Criterion (AIC)	2088.365		
Finite Sample Corrected AIC (AICC)	2088.390		
Bayesian Information Criterion (BIC)	2103.019		
Consistent AIC (CAIC)	2106.019		

Dependent Variable: Three-Year Crashes
Model: (Intercept), Ln AADTmaj, Ln AADTmin
a. Information criteria are in smaller-is-better form.
b. The full log likelihood function is displayed and used in computing information criteria.

Omnibus Test[a]

Likelihood Ratio Chi-Square	df	Sig.
313.267	2	.000

Dependent Variable: Three-Year Crashes
Model: (Intercept), Ln AADTmaj, Ln AADTmin
a. Compares the fitted model against the intercept-only model.

Tests of Model Effects

	Type III		
Source	Wald Chi-Square	df	Sig.
(Intercept)	254.692	1	.000
Ln AADTmaj	151.372	1	.000
Ln AADTmin	35.572	1	.000

Dependent Variable: Three-Year Crashes
Model: (Intercept), Ln AADTmaj, Ln AADTmin

Parameter Estimates

			95% Wald Confidence Interval		Hypothesis Test		
Parameter	B	Std. Error	Lower	Upper	Wald Chi-Square	df	Sig.
(Intercept)	-9.835	.6163	-11.043	-8.627	254.692	1	.000
Ln AADTmaj	.936	.0761	.787	1.085	151.372	1	.000
Ln AADTmin	.357	.0598	.240	.474	35.572	1	.000
(Scale)	1[a]						
(Negative binomial)	1[a]						

Dependent Variable: Three-Year Crashes
Model: (Intercept), Ln AADTmaj, Ln AADTmin
a. Fixed at the displayed value.

Appendix C SPSS Output for 4ST Intersections

Dependent Variable: Three-Year Crashes

Model: (Intercept), ln(AADTmaj), ln(AADTmin)

Model Information

Dependent Variable	Three-Year Crashes
Probability Distribution	Negative binomial (1)
Link Function	Log

Continuous Variable Information

		N	Minimum	Maximum	Mean	Std. Deviation
Dependent Variable	Three-Year Crashes	99	0	8	1.57	2.036
Covariate	Ln AADTmin	99	4.653960350157523	8.294049640102028	6.047309956718448	.882605881604689
	Ln AADTmaj	99	5.669880922980520	9.268609280100158	7.675055344601530	.826862722956947

Goodness of Fit[a]

	Value	df	Value/df
Deviance	73.090	96	.761
Scaled Deviance	73.090	96	
Pearson Chi-Square	66.446	96	.692
Scaled Pearson Chi-Square	66.446	96	
Log Likelihood[b]	-149.880		
Akaike's Information Criterion (AIC)	305.759		
Finite Sample Corrected AIC (AICC)	306.012		
Bayesian Information Criterion (BIC)	313.545		
Consistent AIC (CAIC)	316.545		

Dependent Variable: Three-Year Crashes
Model: (Intercept), Ln AADTmin, Ln AADTmaj
a. Information criteria are in smaller-is-better form.
b. The full log likelihood function is displayed and used in computing information criteria.

Omnibus Test[a]

Likelihood Ratio Chi-Square	df	Sig.
39.911	2	.000

Dependent Variable: Three-Year Crashes
Model: (Intercept), Ln AADTmin, Ln AADTmaj
a. Compares the fitted model against the intercept-only model.

Tests of Model Effects

	Type III		
Source	Wald Chi-Square	df	Sig.
(Intercept)	28.032	1	.000
Ln AADTmin	8.785	1	.003
Ln AADTmaj	13.026	1	.000

Dependent Variable: Three-Year Crashes
Model: (Intercept), Ln AADTmin, Ln AADTmaj

Parameter Estimates

			95% Wald Confidence Interval		Hypothesis Test		
Parameter	B	Std. Error	Lower	Upper	Wald Chi-Square	df	Sig.
(Intercept)	-8.665	1.6366	-11.873	-5.458	28.032	1	.000
Ln AADTmin	.484	.1634	.164	.805	8.785	1	.003
Ln AADTmaj	.759	.2102	.347	1.171	13.026	1	.000
(Scale)	1[a]						
(Negative binomial)	1[a]						

Dependent Variable: Three-Year Crashes
Model: (Intercept), Ln AADTmin, Ln AADTmaj
a. Fixed at the displayed value.

Appendix D SPSS Output for 4SG Intersections

Dependent Variable: Three-Year Crashes

Model: (Intercept), ln(AADTmaj), ln(AADTmin)

Model Information

Dependent Variable	Three-Year Crashes
Probability Distribution	Negative binomial (1)
Link Function	Log

Continuous Variable Information

		N	Minimum	Maximum	Mean	Std. Deviation
Dependent Variable	Three-Year Crashes	9	2	23	10.00	8.860
Covariate	Ln AADTmaj	9	8.216088098632316	9.400960731584833	9.066970241322060	.357871880414697
	Ln AADTmin	9	6.120297418950950	8.853665428037450	8.037825286864793	.830082943310368

Goodness of Fit[a]

	Value	df	Value/df
Deviance	4.344	6	.724
Scaled Deviance	4.344	6	
Pearson Chi-Square	5.289	6	.882
Scaled Pearson Chi-Square	5.289	6	
Log Likelihood[b]	-28.827		
Akaike's Information Criterion (AIC)	63.654		
Finite Sample Corrected AIC (AICC)	68.454		
Bayesian Information Criterion (BIC)	64.245		
Consistent AIC (CAIC)	67.245		

Dependent Variable: Three-Year Crashes
Model: (Intercept), Ln AADTmaj, Ln AADTmin
a. Information criteria are in smaller-is-better form.
b. The full log likelihood function is displayed and used in computing information criteria.

Omnibus Test[a]

Likelihood Ratio Chi-Square	df	Sig.
2.664	2	.264

Dependent Variable: Three-Year Crashes
Model: (Intercept), Ln AADTmaj, Ln AADTmin
a. Compares the fitted model against the intercept-only model.

Tests of Model Effects

	Type III		
Source	Wald Chi-Square	df	Sig.
(Intercept)	1.789	1	.181
Ln AADTmaj	1.315	1	.252
Ln AADTmin	1.447	1	.229

Dependent Variable: Three-Year Crashes
Model: (Intercept), Ln AADTmaj, Ln AADTmin

Parameter Estimates

			95% Wald Confidence Interval		Hypothesis Test		
Parameter	B	Std. Error	Lower	Upper	Wald Chi-Square	df	Sig.
(Intercept)	-16.395	12.2561	-40.416	7.627	1.789	1	.181
Ln AADTmaj	1.452	1.2666	-1.030	3.935	1.315	1	.252
Ln AADTmin	.667	.5545	-.420	1.754	1.447	1	.229
(Scale)	1[a]						
(Negative binomial)	1[a]						

Dependent Variable: Three-Year Crashes
Model: (Intercept), Ln AADTmaj, Ln AADTmin
a. Fixed at the displayed value.