MD-14-SP209B4J

Martin O'Malley, *Governor*
Anthony G. Brown, *Lt. Governor*

State Highway Administration
Maryland Department of Transportation

James T. Smith, Jr., *Secretary*
Melinda B. Peters, *Administrator*

STATE HIGHWAY ADMINISTRATION

RESEARCH REPORT

THE DEVELOPMENT OF LOCAL CALIBRATION FACTORS FOR IMPLEMENTING THE HIGHWAY SAFETY MANUAL IN MARYLAND

HYEONSHIC SHIN, PH.D.
YOUNG-JAE LEE, PH.D.
SEYEDEHSAN DADVAR

MORGAN STATE UNIVERSITY

PROJECT NUMBER SP209B4J
FINAL REPORT

March 2014

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
MD-14-SP209B4J		

4. Title and Subtitle	5. Report Date
The Development of Local Calibration Factors for Implementing the Highway Safety Manual in Maryland	March 2014
	6. Performing Organization Code
7. Author(s)	8. Performing Organization Report No.
Hyeonshic Shin, Young-Jae Lee, Seyedehsan Dadvar	
9. Performing Organization Name and Address	10. Work Unit No. (TRAIS)
Morgan State University 1700 E. Cold Spring Lane Baltimore, MD 21251	11. Contract or Grant No. SP209B4J
12. Sponsoring Agency Name and Address	13. Type of Report and Period Covered
Maryland State Highway Administration Office of Traffic and Safety (OOTS) 7491 Connelly Drive Hanover, MD 21076 Morgan State University 1700 E. Cold Spring Lane Baltimore, MD 21251	Final Report 14. Sponsoring Agency Code

15. Supplementary Notes

16. Abstract

The goal of the study was to determine local calibration factors (LCFs) to adjust predicted motor vehicle traffic crashes for the Maryland-specific application of the Highway Safety Manual (HSM). Since HSM predictive models were developed using data from states other than Maryland, the predictive models should be calibrated. LCFs for all 18 facility types were calculated using Maryland data. After data collection and compilation, samples were drawn based on 90% confidence level, which increased the overall sample size. After the comparison between HSM default crash proportion and Maryland specific data, the use of the Maryland data was suggested. In general, LCFs for all facilities were less than 1.0, implying Maryland had fewer crashes than predicted crashes estimated by HSM crash prediction methodology. Especially, LCFs for intersections were extremely low. Due to potential issues with unreported minor and property damage only crashes, the authors recommend using LCFs for fatal and injuries crashes where available. The report concludes with a discussion on data limitation, thoughts about HSM sampling requirement, and suggestions for future research. It should be noted that this study only considered roadways maintained by SHA and Baltimore City was not part of this study.

17. Key Words: Highway Safety Manual, Local Calibration Factor, Safety Performance Function	18. Distribution Statement No restrictions. This document is available from the Research Division upon request.

19. Security Classification (of this report)	20. Security Classification (of this page)	21. No. of Pages	22. Price
None	None	130	

Form DOT F 1700.7 (8-72) Reproduction of form and completed page is authorized.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF EQUATIONS

ACKNOWLEDGEMENTS

The authors would like to thank Morgan State University’s National Transportation Center for its funding support and continued encouragement of student participation in research projects. The authors also would like to thank the Maryland State Highway Administration (SHA) for funding this project and providing assistance in data collection. Special thanks go to Dr. Ruihua Tao of SHA for her guidance and constructive suggestions throughout this study. Finally, the authors are grateful to four undergraduate students who assisted with the data collection: Victoria Coles, Patrick Flanagan, Meahsha Makle, and April Thompson.

EXECUTIVE SUMMARY

Improving the safety of road users in Maryland has been one of the most important goals of the Maryland Department of Transportation (MDOT) and its modal agencies. It is clear that various traffic safety improvement strategies of MDOT have contributed to a dramatic decrease in traffic fatalities and injuries. Between 2007 and 2011, traffic fatalities in Maryland decreased by 20% (486 fatalities in 2011 from 615 in 2007) and injuries by 14% (44,489 injuries in 2011 from 51,729 in 2005) (Maryland Highway Safety Office 2012). More strategies should be implemented to keep the downward trend.

One of the best strategies to improve traffic safety and reduce motor vehicle traffic crashes is to provide well-planned engineering, education, and/or enforcement countermeasures that are tailored to given crash, traffic, and roadway characteristics. The implementation of Highway Safety Manual (HSM) methodologies is expected to provide a cost-effective approach to transportation safety planning and engineering.

Published in 2010, HSM provides "analytical tools and techniques for quantifying the potential effects on crashes as a result of decisions made in planning, design, operations, and maintenance (American Association of State Highway and Transportation Officials 2010)." It includes three facility types (rural two-lane road, rural multilane highways, and urban and suburban arterials) and each facility type is further divided into sub-types depending on other roadway characteristics, resulting in a total of 18 base predictive methods (eight roadway segment types and ten intersection types). Once a data set is prepared for HSM, safety planners should be able to conduct a system-wide safety analysis, identify potential problem locations, and evaluate effects of different countermeasures.

To apply HSM predictive method to the study area, one more step needs to be taken: the development of local calibration factors (LCFs). An LCF for a certain facility is a ratio of total predicted crashes to total observed crashes. It accounts for differences between Maryland local characteristics and HSM base model's data from select jurisdictions in the United States. In this sense, the primary goal of the study was to determine LCFs to adjust predicted motor vehicle crashes based on HSM for Maryland-specific application.

The study started with approximately 2.665 million data points for a three-year study period (2008-2010) collected from multiple resources of Maryland government agencies. Samples for additional data collection and crash prediction for 18 facility types were drawn. LCFs were developed for all 18 facility types. The summary tables are provided below. Table 1 and Table 2 show LCFs for total crashes including all types of crash severity (i.e., K-fatal; A-incapacitating injury; B-non-incapacitating injury; C-possible injury; and O-property damage only (PDO) crashes.) In general, LCFs for all facilities were less than 1.0, implying those facilities in Maryland had fewer crashes than predicted crashes estimated by HSM crash prediction methodology. Extremely lower LCFs for intersections probably imply that the intersections in Maryland are safer in general. However, the exclusion of the City of Baltimore, where, like other large cities, busy intersections are common, may be another reason for lower intersection LCFs.

Table 1. Maryland LCFs – Roadway Segments

Segments	R2U	R4U*	R4D	U2U	U3T	U4U	U4D	U5T
Population	9,519	19	1,410	7,215	537	741	5,328	276
Observed Crashes	8,938	43	1,818	7,859	973	2,491	12,105	2,098
Samples	251	19	160	252	138	145	244	115
Observed Crashes for Samples	458	43	315	360	330	592	654	1,257
Predicted Crashes for Samples	658	19	540	528	306	674	791	1,057
Local Calibration Factor	**0.6960**	**2.2632**	**0.5833**	**0.6818**	**1.0784**	**0.8783**	**0.8268**	**1.1892**

Note 1: There were only 19 R4U segments in the final data set. Thus, all of them were included in the study.
Note 2: The asterisk denotes that the facility did not meet HSM minimum sample size criteria of 30-50 sites or the minimum annual crash threshold of 100.

Table 2. Maryland LCFs – Intersections

Segments	R23ST*	R24ST*	R24ST*	RM3ST*	RM4ST*	RM4SG*	U3ST*	U4ST*	U3SG	U4SG
Population	579	219	69	33	7	39	492	160	488	960
Observed Crashes	307	290	267	50	29	238	306	297	2,455	7,271
Samples	162	115	67	26	10	35	152	90	167	244
Observed Crashes for Samples	103	142	262	36	30	231	103	173	789	1,763
Predicted Crashes for Samples	626	706	1,000	201	82	1,886	659	452	1981	3,842
Local Calibration Factor	**0.1645**	**0.2011**	**0.2621**	**0.1788**	**0.3667**	**0.1225**	**0.1562**	**0.3824**	**0.3982**	**0.4589**

Note: The asterisk denotes that the facility did not meet HSM minimum sample size criteria of 30-50 sites or the minimum annual crash threshold of 100.

Table 3 and Table 4 summarize LCFs by different combinations of crash severity. KABC (fatal and injury crashes) LCFs for two out of seven segment types and all intersection types were higher than LCFs for total crashes. If the purpose of the research is to predict fatal and injury crashes, KABC LCFs would be useful. The only disadvantage is that there is no LCF for KABC crashes for R2U.

Table 3. Maryland LCFs by Crash Severity – Roadway Segments

Segments	R2U	R4U*	R4D	U2U	U3T	U4U	U4D	U5T
Total Crashes	0.6956	2.3408	0.5838	0.6814	1.0785	0.8788	0.8269	1.1891
KABC Crashes	N.A.	1.9499	0.4193	0.6125	1.3053	0.7696	1.0665	1.1918
KAB Crashes	N.A.	1.9231	0.4565	N.A.	N.A.	N.A.	N.A.	N.A.
PDO Crashes	N.A.	N.A.	N.A.	0.7313	0.9362	0.9611	0.7310	1.1874

Note 1: There were only 19 R4U segments in the final data set. Thus, all of them were included in the study.
Note 2: The asterisk denotes that the facility did not meet HSM minimum sample size criteria of 30-50 sites or the minimum annual crash threshold of 100.
Note 3: N.A. means that no SPF is available in HSM.

Table 4. Maryland LCFs by Crash Severity – Intersections

Intersections	R23ST*	R24ST*	R24SG*	RM3ST*	RM4ST*	RM4SG*	U3ST*	U4ST*	U3SG	U4SG
Total Crashes	0.1645	0.2011	0.2634	0.1788	0.3667	0.1086	0.1562	0.3824	0.3982	0.4782
KABC Crashes	N.A.	N.A.	N.A.	0.2550	0.3923	0.1327	0.2273	0.4964	0.5967	0.6285
KAB Crashes	N.A.	N.A.	N.A.	0.2664	0.3953	0.1879	N.A.	N.A.	N.A.	N.A.
PDO Crashes	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	0.1138	0.3003	0.3427	0.3970

Throughout this study, the data collection was the biggest challenge. In this regard, all reviewed studies raised concerns about burdensome and laborious data collection tasks. To take full advantage of the existing HSM predictive models and/or to improve the current analysis process, more in-depth research should be conducted. First, a defendable sampling design method should be discussed. Sampling errors and confidence levels are often affected by the sample size. However, due to the lack of resources, a sample size cannot be increased beyond a certain point. In this respect, guidance on a sampling frame, which provides a certain level of confidence and minimum sample size, should be studied. Second, related to the sampling issue, there is no clear rationale for using a "long-enough" segment. The determination of minimum segment length depends on the roadway network of the study area. The procedure of deciding minimum length thresholds should be developed and provided in the manual. Third, sub-region specific LCFs will be helpful. Climate, rain history, population and other factors vary even within one state, so one single LCF for the entire state may not be reasonable and correct.

INTRODUCTION

Improving the safety of road users in Maryland has been one of the most important goals of the Maryland Department of Transportation (MDOT) and its modal agencies. Indeed, the 2006 Maryland Strategic Highway Safety Plan identified seven areas of emphasis for improving highway safety, and it set a performance target of 10% reduction in traffic fatalities and injuries, respectively, by 2010 from the 2005 level (Maryland Department of Transportation 2006, 6). It is clear that various traffic safety improvement strategies of MDOT have contributed to a dramatic decrease in traffic fatalities and injuries. Between 2007 and 2011, traffic fatalities in Maryland decreased by 20% (486 fatalities in 2011 from 615 in 2007) and injuries by 14% (44,489 injuries in 2011 from 51,729 in 2005) (Maryland Highway Safety Office 2012). More strategies should be implemented to keep the downward trend.

One of the best strategies to improve traffic safety and reduce motor vehicle fatalities and injuries is to provide well-planned engineering, education, and/or enforcement countermeasures that are tailored to given crash, traffic, and roadway characteristics. This process of providing countermeasures consists of at least five steps (Federal Highway Administration 2010): (1) identification of problem areas based on a system-wide crash analysis; (2) identification of appropriate countermeasures; (3) prioritization of improvement projects; (4) implementation of the selected countermeasures; and (5) evaluation of the improvement (before-and-after studies). This is a data-driven and resource-intensive process.

In many circumstances, however, conducting a complete system-wide analysis and defendable before-and-after studies on all implemented countermeasures is extremely difficult due to at least two reasons. First, the lack of resources often prevents transportation agencies from conducting a comprehensive approach similar to the five steps mentioned earlier. Second, it is difficult to assess the effects of countermeasures on crash reduction at a given location and over a reasonable time period (Zegeer, et al. 2004). This is because crashes are rare events and, thus, there is no guarantee that similar types of accidents will occur at the same location during the specified study period. More time-series data for a large number of locations with similar characteristics need to be collected, which again would be subject to resource constraints.

The implementation of Highway Safety Manual (HSM) predictive models may provide a more cost-effective approach to transportation safety planning and programming. Published in 2010 by the American Association of State Highway and Transportation Officials (AASHTO), HSM provides "analytical tools and techniques for quantifying the potential effects on crashes as a result of decisions made in planning, design, operations, and maintenance" (American Association of State Highway and Transportation Officials 2010, 1-1). The 18-step approach for computing predicted crash frequencies for the 18 facility types are provided in HSM. Once a data set is prepared for HSM, safety planners should be able to conduct a system-wide safety analysis, identify potential problem locations, and evaluate effects of different conditions and countermeasures (i.e., crash modification factors in HSM). According to Federal Highway Administration: the HSM will greatly advance state and local highway agencies' ability to incorporate explicit, quantitative consideration of safety into their planning and project development decision making (Federal Highway Administration n.d.).

To apply HSM predictive models to the study area of interest, one more step should be taken: the development of local calibration factors (LCFs). An LCF is a ratio of observed crashes to predicted crashes. LCFs are used to "account for differences between jurisdiction and time period for which the predictive models were developed and the jurisdiction and the time period to which they are applied by HSM users" (American Association of State Highway and Transportation Officials 2010, 3-16). This is because the crash prediction models of HSM were developed with data from select jurisdictions in the United States. Most data items used for roadway segments were from Washington and for intersections from California from 2002 to 2006 (Dixon, et al. 2012). Thus, the predicted crash frequencies from HSM models may not be directly transferred to the study area due to multiple factors that may vary across the country, such as climate, population, traffic, crash reporting system, and others. To be effective, LCFs for roadway segments and intersections with various roadway geometry configurations should be developed. The safety analysis using HSM shall involve collecting and compiling historical data on crashes, traffic volume, roadway characteristics, and land uses, as well as necessary procedures such as site selection, model estimation, and calibration.

This report describes procedures that the Morgan State University team took from data collection, compilation, frequency analysis, and computation of LCFs. Starting with approximately 2.665 million data points, derived from Maryland State Police (MSP) database, and additional data collection and estimation, the authors developed LCFs for the 18 facility types based on the methodologies in Part C of HSM. It should be noted that only SHA maintained roadways (excluding Interstates) were considered for this study. Also, note that the roadways within the Baltimore City boundaries were not part of this study.

Study Objectives

The primary goal is to determine LCFs to adjust HSM predicted vehicle crashes for Maryland-specific application. The specific objectives are

- to review available studies that apply and evaluate the suggested methodologies in HSM;
- to collect and compile all required data for the selected SHA maintained roadway segments and intersections;
- to estimate crash frequencies, severity, and types of crashes for roadway segments and intersections by different roadway facility types; and
- to develop LCFs for Maryland by comparing crash frequencies predicted by HSM methods to observed crashes.

Report Structure

The following chapter provides a review of literature that includes a brief introduction of HSM and its predictive methods and the commonly identified issues from previous local calibration factor development studies. Then, a detailed discussion on data collection, compilation, and limitations are discussed. After describing the local calibration development, research findings (LCFs based on Maryland conditions) are presented. The last chapter summarizes the discussion on the developed LCFs, findings of the study, barriers that the study team encountered, and future study suggestions.

LITERATURE REVIEW

This chapter includes a brief on HSM and its predictive methods and then a review of case studies.

Brief on the Highway Safety Manual and Its Predictive Methods

Highway Safety Manual (HSM)

The HSM is a culmination of decades-long efforts to provide a technical approach which is based on a system analysis frame. As stated in HSM, “The HSM assembles currently available information and methodologies on measuring, estimating and evaluating roadways in terms of crash frequency and crash severity (American Association of State Highway and Transportation Officials 2010).” The purpose of the development of HSM was to help provide transportation professionals with tools to facilitate decision making in roadway safety planning, design, operations, and maintenance decisions based on explicit consideration of their safety consequences (American Association of State Highway and Transportation Officials 2010). The application of HSM would fill the gap in the transportation safety study: the lack of a standardized and well-agreed upon manual. Unlike other transportation planning fields such as travel demand modeling and emissions modeling, the transportation safety field did not have a decision making tool that allowed professionals to evaluate substantive safety effects of safety planning (Dixson, et al. 2012). As a decision making tool, it is expected that ultimately HSM approach will help government agencies utilize limited resources more efficiently by prioritizing safety planning and engineering alternatives based on the quantification of the potential safety effects of government actions.

The HSM has four main sections (Parts A, B, C, and D). Part A discusses the purpose and scope of HSM. Part B consists of the roadway safety management process, including network screening, diagnosis, countermeasures selection, economic appraisal, prioritizing projects, and safety effectiveness evaluation. Part C provides the safety predictive methods for three roadway types: (1) rural two-lane road, (2) rural multilane highways, and (3) urban and suburban arterials. The three roadways are further divided into 8 segments and 10 intersection types (

Table 5). Please note that in the table, the "used acronyms" column is referring to the acronyms that are used in this report, since the authors thought that they are more intuitive and self-explanatory than those used in HSM. Part D includes crash modification factors (CMFs) for roadway segments, intersections, interchanges, special facilities and geometric situations, and road network. CMFs are average values of crash frequency changes as a result of geometric or operational modifications to a site that differs from given base conditions.

Table 5. The Facility Types Considered in Highway Safety Manual

(a) Rural Two-Lane, Two-Way Roads			
Type	**HSM**	**Used Acronyms**	**Definition**
Roadway Segments	2U	R2U	Undivided rural two-lane, two-way roadway segments
Intersections	3ST	R23ST	Un-signalized three-leg (stop control on minor-road approaches)
	4ST	R24ST	Un-signalized four-leg (stop control on minor-road approaches)
	4SG	R24SG	Signalized four-leg
(b) Rural Multilane Highways			
Type	**HSM**	**Used Acronyms**	**Definition**
Roadway Segments	4U	R4U	Rural four-lane undivided segments
	4D	R4D	Rural four-lane divided segments
Intersections	3ST	RM3ST	Un-signalized three-leg (stop control on minor-road approaches)
	4ST	RM4ST	Un-signalized four-leg (stop control on minor-road approaches)
	4SG	RM4SG	Signalized four-leg
(c) Urban and Suburban Arterials			
Type	**HSM**	**Used Acronyms**	**Definition**
Roadway Segments	2U	U2U	Two-lane undivided arterials
	3T	U3T	Three-lane arterials including a center two-way left-turn lane (TWLTL)
	4U	U4U	Four-lane undivided arterials
	4D	U4D	Four-lane divided arterials (i.e., including a raised or depressed median)
	5T	U5T	Five-lane arterials including a center TWLTL
Intersections	3ST	U3ST	Un-signalized three-leg intersection (stop control on minor-road approaches)
	4ST	U4ST	Un-signalized four-leg intersection (stop control on minor-road approaches)
	3SG	U3SG	Signalized three-leg intersection
	4SG	U4SG	Signalized four-leg intersection

Predictive Method

An HSM predictive model (or safety performance function; SPF) is used to estimate the predicted total crash frequency for a particular facility type, N_{SPF}, for a study year. An SPF function of traffic volume and a set of base site conditions. The HSM SPFs were developed based on site conditions and historic crash data from selected states (Dixon, et al. 2012).

Equation 1 shows the general model for HSM crash prediction.

Equation 1. Calibrated Predicted Crash Frequency

$$N_{Predicted\ (Adjusted)} = N_{SPF} \times (CMF_1 \times CMF_2 \times \ldots \times CMF_n) \times LCF$$

Where:

$N_{Predicted\ (Adjusted)}$ = Adjusted (i.e., LCF accounted) total predicted crash frequency,

N_{SPF} = Average crash frequency under base condition,

$CMF_1, \ldots, CMF_n$ = Crash Modification Factors, and

LCF = Local Calibration Factor.

Equation 1 consists of three parts. The first part, N_{SPF}, is the base SPF, which is used to estimate the average crash frequency for a certain facility type for a given base year with specified base geometric conditions. Each of the 18 facility types has its own base SPF. A base SPF for roadway segments is a function of segment length and AADT and a base SPF for intersections is a function of AADT values on the major and minor roadways at intersection. For example, $\mathrm{N_{SPF}}$ for rural two-way, two-lane roadway segments is computed using Equation 2. Similarly, $\mathrm{N_{SPF}}$ for rural two-way, two-lane four-leg signalized intersections, is computed using Equation 3.

Equation 2. N_{spf} for Rural Two-Way, Two-Lane Roadway Segments

$$N_{spf\ rs} = AADT \times L \times 365 \times 10^{(-6)} \times e^{(-0.312)}$$

Where:

$N_{spf\ rs}$ = Predicted average crash frequency for base condition using a statistical regression model,

AADT = Annual Average Daily Traffic, and

L = Length of roadway segments (miles).

Equation 3. N_{spf} for Rural Two-Way, Two-Lane, Four-Leg Signalized Intersections

$$N_{spf\ \mathrm{4SG}} = \exp[-5.13 + 0.60 \times \ln(AADT_{maj}) + 0.20 \times \ln(AADT_{min})]$$

Where:

$N_{spf\ 4SG}$ = Predicted average crash frequency for base condition using a statistical regression model,

$AADT_{maj}$ = Annual Average Daily Traffic on the major road, and

$AADT_{min}$ = Annual Average Daily Traffic on the major road.

The second part of

Equation 1 is a set of crash modification factors (CMFs) that are used to address local or regional site conditions that are different from the base conditions. A CMF is a multiplicative factor used for evaluating the impacts of countermeasures on crash frequency (Crash Modification Factors Clearninghouse n.d.). A CMF may have value either equal to, less than, or greater than 1.0. If a CMF is equal to 1.0, the associated countermeasure at the site has no impact on crash. A CMF less than 1.0 indicates that fewer crashes are expected than the base condition without the countermeasure. If a CMF is greater than 1.0, more crashes may occur at the site than the base condition. For example, the CMF of a left-turn lane at an intersection on a rural two-lane, two-way road is 0.31. Assuming an intersection with 100 annual crashes, the addition of a left-turn lane would result in 31 crashes (100 annual crashes multiplied by the CMF of 0.31). In other words, crashes would be reduced by 69%.

The last part of

Equation 1 is a local calibration factor (LCF) that is defined as:

"A factor to adjust frequency estimates produced from a safety prediction procedure to approximate local conditions. The factor is computed by comparing observed crash data at the state, regional, or local level to estimated crashes obtained from predictive models" (American Association of State Highway and Transportation Officials 2010)

The purpose of the LCF is to "account for differences between the jurisdiction and time period for which the predictive models were developed and the jurisdiction and time period to which they are applied by HSM users" (American Association of State Highway and Transportation Officials 2010).

Equation 4. Calculation of Calibration Factor

$$C = \frac{\sum_{All\ sites} N_{Observed}}{\sum_{All\ sites} N_{Predicted\ (Unadjusted)}}$$

Where:

$N_{Predicted\ (Unadjusted)}$ = Un-adjusted total predicted crash frequency, and

$N_{Objerved}$ = Total number of observed crashes during the study period.

An LCF of 1.0 means the predicted crashes are the same as the observed crash frequency. In other words, no calibration for local condition is required. An LCF less than 1.0 indicates that the observed crashes for a certain facility type of a region are fewer than the base model crash frequencies. An LCF of over 1.0 suggests that crash frequencies of a facility in a study location are greater than the base model. The computed LCFs are meant to be used to adjust HSM base model results to local conditions.

In the next section, case studies of LCF development in other states are reviewed, followed by a more detailed description of data requirements, computation process, and interpretation in later sections.

Review of Case Studies

Since HSM and LCF development is a relatively new area in transportation safety, not all states or local jurisdictions have developed LCFs and/or adopted HSM. Similarly, no well-agreed upon discussion on benefits and costs of HSM application has been published. According to discussions with other researchers at the 2013 TRB annual meeting and other venues, some states have already evaluated the applicability of HSM and have been moving toward agency-wide adoption. Other states, on the other hand, are moving toward developing their own SPFs. This study only considers the development of LCFs, not state-specific SPFs.

Numerous case studies conducted at HSM development stage and at the implementation stage were reviewed. A summary of selected case studies is provided in Appendix B. The LCFs of the case studies are presented in Appendix C. Since all studies followed HSM methodology, the literature review did not focus on study procedures. Instead, the main focus of the review was to identify types of challenges that were encountered in previous studies. The following sections provide issues that were commonly identified from the reviewed case studies.

Burdensome Data Requirement and Data Interoperability

Data collection posed the biggest challenge. As shown in HSM Data Needs table (Appendix D), there are 60 variables used in HSM predictive models. Forty-one variables are required and nineteen are desired. The use of HSM may gradually set data collection standards for state and local transportation agencies and lead to more efficient data management in the long run (Alluri 2010). The authors of the case studies, however, pointed out that states' data sets were not currently built for HSM. Many of the 60 variables used in HSM are not commonly collected items, such as AADT on minor roads, driveway density, liquor store density, and others; thus, collecting required/desired data was the most difficult task. There will be a long way to materialize HSM's benefits unless the data collection system is ready for HSM in the near future.

Pfefer et al. (1999, cited in (Alluri 2010, 19)) point to data limitations as a major impediment to the process of addressing the safety issues. The data limitations include (1) the lack of precision measurement, reporting, and data collection tools; (2) inadequate coverage of traffic data; (3) incomplete data; (4) lack of roadway inventory data; and (5) data integration and interoperability. Some of these limitations were also identified while reviewing the case studies and conducting the current study.

In terms of limitations (1), (2), and (3), AADT for minor roadways is a good example. It is a required variable for predicting intersection crash frequencies. However, in most states reviewed, AADT values on minor roadways were not available or, if available, incomplete. In the Oregon study, a regression analysis was conducted for estimating AADT for minor roadways. Because of the unavailability of AADT values for minor roadways, in some states the study teams limited their studies only to the roadways with available AADT data (e.g., in Florida, only intersections of two state roads were retained for analysis.) or several states did not consider intersection models. Another example is a signal phasing data set that is not completely stored for an entire state's intersections. In Oregon, the study team made an assumption that if the major street had protected or permissive phasing and the minor street had dedicated left-turn lanes, the same signal phasing existed on the minor approach.

In many states, curve data (lengths and radii) needed to be manually measured as curve data were not kept in databases or were not in a way to be readily used. This could be an example of limitation number (1), (3), and (4).

The last limitation, (5) data integration and interoperability, is also a big challenge. To perform HSM's data driven safety analysis, government agencies need to compile data from different agencies within the state DOT and from other agencies (e.g., land use, school locations, and liquor store locations). Thus, combining different data tables and maps to create an HSM-ready data set has become important. In this sense, data interoperability is becoming an important issue. Alluri (2010, 20) stated that "The process of linkage between the databases has not been given much attention in the past. Even within the same database, inconsistencies exist between the data items collected by local agencies, state officials and the federal requirements, mainly due to the flexibility within the agencies." The Florida case study is a good example (Figure 1). The roadway characteristics inventory (RCI) in Florida consists of multiple tables. Each table represents one attribute or variable. Unfortunately, the tables cannot be easily combined to create homogenous segments[1] for predictive crash estimation. A long and arduous GIS process with a Python script had to be developed to make sure that all the necessary attributes required for calibration were met.

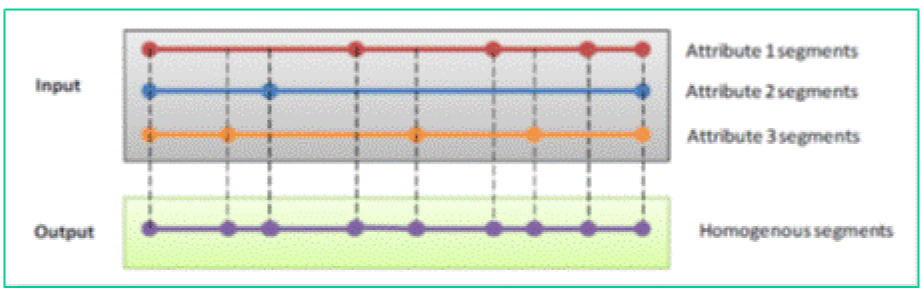


Figure 1. Creation of Homogenous Segments from the Florida RCI (Srinivasan, et al., 2011, p. 10)

To overcome data limitations, many studies used alternative data collection methods. Alternative data collection methods include the review of video logs, aerial photos, Google Earth images, the use of statistical estimation methods, and the development of an estimation tool. In a Louisiana study (Sun, et al. 2011), the team developed a methodology to measure curvature information from an aerial photo image. However, the study only considered one facility type, rural two-lane, two-way roads. Developing a new tool and applying it to statewide studies for all 18 facility types is not cost effective.

---

[1] Homogeneity of segments is required in safety analysis to improve the reliability of crash prediction. Within a same segment, some variables such as the number of lanes, lane width, median type and others should be same. Not all variables can be homogenous within the segment, but researchers should do their best in creating segments as homogenous as possible.

Property Damage Only Crash Reporting

In some studies, a crash reporting threshold was questioned. While the total number of crashes was used for computing SPFs, the different crash reporting threshold values by states may increase or decrease the reported number of property damage only crashes. For example, in Oregon, crashes with over $1,500 of property damage should be reported, while the threshold values are $700 and $750 in Washington and California, respectively (Dixon, et al. 2012, 4). As stated earlier, crash data from Washington and California were used in developing HSM predictive models. As a result of these differences, crash proportions by severity of a study area could be different from the proportions used in HSM. For this reason, the study team that developed LCFs for Oregon used crash severity proportion of Oregon in their computations. The HSM also suggest the use of local data when the data is available and the proportion is different from HSM model (American Association of State Highway and Transportation Officials 2010, A-10).

Sometimes, property damage only crash data is not readily available. In Florida, property damage only crashes are reported, but not stored in the state's electronic crash database. Thus, the Florida study only included LCFs for fatal and injury crashes (Srinivasan, et al. 2011, 31).

Ambiguous HSM Guideline

While HSM provides a very detailed guideline, some requirements are not clear. The HSM's guideline for selecting sites for calibration of SPFs is as follows;

"For each facility type, the <u>desirable minimum sample size</u> for the calibration data set is <u>30 to 50 sites</u>, with each site <u>long enough to adequately represent physical and safety conditions</u> for the facility. (…) Following site selection, the entire group of calibration sites should represent a total of <u>at least 100 crashes per year</u> (American Association of State Highway and Transportation Officials 2010, A-3)."

First, there is no clear rationale on minimum segment length criteria and minimum annual crashes. More evidence must have been provided. The HSM suggests that "when dividing roadway facilities into small homogenous roadway segments, limiting the segment length to no less than 0.10 miles will minimize calculation efforts and not affect results (American Association of State Highway and Transportation Officials 2010, C-8)." Following the guideline would be simple enough. However, depending on land use patterns and density of the built environment of a study area, the suggested 0.1 mile minimum length would be too long. In the case of Maryland, over 60% of rural roadway segments and over 80% of urban and suburban roadway segments considered for the current study are shorter than 0.1 mile. As shown in Table 6, each study area used different threshold values. The loss of information and an appropriate segment length threshold for a study area should be carefully studied as future research efforts.

Second, the sample size of 30-50 sites per facility type with minimum annual crashes of 100 is also confusing. While the minimum sample size criterion is somewhat agreed in statistics, the minimum annual crash threshold should be clearly discussed or at least references need be provided.

Table 6. Segment Length Used for Application of HSM

State	Segment length range	Reasons and method for segment length
GA	Segments with 0.1, 0.2, & 0.67 mile lengths	Predefined lengths by Safety Analyst (used in the study)
FL	0.1 mile for rural and 0.04 mile for urban segments	Because of homogenous segmentation
IA	Referring to minimum 0.1 mile in the several parts of this research.	Based on HSM suggestions
ID	Two averages: 0.47 and 0.57 mile	Combined shorter segments smaller than 0.1 mile
OR	0.1 mile for rural and 0.07 mile for urban segments (2-mile segment as maximum)	Based on HSM suggestions and creating homogenous segments.
UT	Minimum threshold: 0.2 mile	Using 0.2 mile threshold for minimum segment length that is longer than HSM recommendations

METHODOLOGY

This chapter discusses the methodology that the study team employed for the development of LCFs for the state of Maryland. After the discussion on the LCF development process, a brief introduction of the Interactive Highway Safety Design Model (IHSDM) is provided. The IHSDM was the main computer software for crash frequency computation for this study. Then, the detailed data collection and compilation methods along with different sampling strategies are presented.

Method of Study

This section includes the framework that the study team employed for the development of LCFs for the state of Maryland. Also, a brief introduction of the Interactive Highway Safety Design Model (IHSDM) is provided. The IHSDM was the main computer software for crash frequency computation for this study.

LCF Development Process

HSM's predictive method consists of 18 steps. Depending on the available data and the purpose of the study, all or part of the 18-step procedure are used.

Figure 2 illustrates a simplified procedure that the current study employed from HSM. First, the process starts with identifying study locations and initial data collection. As stated earlier, the study includes all 18 facility types included in HSM (see

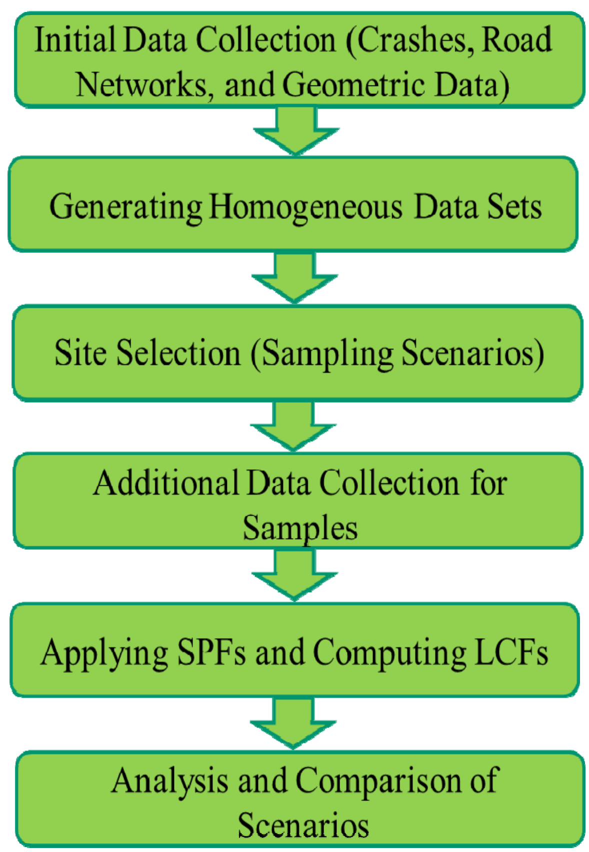


Figure 2. LCF Development Process

Table 5). Only SHA-maintained roadways were considered for the study. Both city and county roads were not included.

Second, after identification of study facilities and initial data sets, homogenous roadway segments or intersections were created. Homogeneity means geometric characteristics within a segment do not vary over the study period. As shown in Figure 1 earlier, a segment should be divided in order for a segment to have the same value for each segment. For example, there is a roadway segment that starts from mile point 0.1 mile and ends at mile point 0.2 mile. Within that 0.1 mile segment, all geometric characteristics, such as number of lanes, shoulder width, median type and other variables, should remain the same over the study period, or the researchers should redefine the segments to make them as homogeneous as possible.

Third, once the initial data set is ready, sites for analyses are sampled. The HSM suggests that for each facility type, at least 30-50 sites with at least 100 total annual crashes should be selected. The study team has increased the sample size by selecting sample sites based on the sample size suggested by 90% confidence interval criterion. Also, other sampling scenarios for roadway segments were evaluated.

Fourth, for the selected samples, additional data was collected, which involved extensive manual data coding work.

Finally, predicted crash frequencies were computed and compared with observed crashes to compute LCFs.

Interactive Highway Safety Design Model (IHSDM)

For the estimation of predicted crash frequencies and LCFs, the Interactive Highway Safety Design Model (IHSDM) was used. The IHSDM is analysis software for evaluating safety and operational effects of geometric design decisions on highways. Developed by the Federal Highway Administration's Safety Research and Development Program, IHSDM can be downloaded free of charge from [www.ihsdm.org.](www.ihsdm.org) The latest version (8.1.0) includes models for freeways and ramps that have been newly developed, which is expected to be included in the next HSM edition.

IHSDM currently includes six evaluation modules: (1) crash prediction module, (2) design consistency module, (3) intersection review module, (4) policy review module, (5) traffic analysis module, and (6) driver/vehicle module. The crash prediction module includes capabilities to evaluate rural two-lane highways, rural multilane highways, and urban/suburban arterials, while other evaluation modules are only applicable to rural two-lane highways.

IHSDM has featured a new calibration utility for crash prediction (from version 7.0.0 of IHSDM) since 2011. This feature was added to the IHSDM administration tool (Admin Tool) to assist agencies in implementing the calibration procedures. The IHSDM calibration utility provides all required steps to calculate local calibration factors (LCFs) for all facility types. The

IHSDM user tutorial[2] provides details of the IHSDM application for calculating local calibration factors (LCFs).

Data Collection and Compilation

Similar to previous studies, data collection and compilation were the most challenging tasks for the current study. Using HSM is not an easy matter due to the huge data requirement. Over 40 data items are required, while about 20 data items are desirable, but optional (See Appendix D).

For the desired data items, default assumptions given by HSM can be used when predicting crashes. For example, the number of schools within 1,000 feet of the intersection is a desirable variable. If no such data is available, HSM suggests assuming no school within 1,000 feet of that intersection. On the other hand, some of the desirable variables should be treated as required items. For example, roadway grade is a desirable variable for the rural two-lane, two-way road type. HSM suggests making an assumption based on terrain (level, rolling, and mountains). However, terrain data set for the current study was incomplete, which led to additional data collection. Another example is roadside fixed-object density for all urban and suburban arterials. The default assumption for roadside fixed-object density is “database default on fixed-object offset and density categories,” that must be collected from a state agency or manually measured.

About 60% of the required and desired data items were obtained from SHA. While some data sets were complete, some data sets needed to be augmented using additional data collection. About 70% of this study effort was put into this task.

Data Collection Steps

The data collection task consists of two steps. First, readily available data sets were collected. This included several must-have variables such as historical crash data, AADT, and roadway geometric information. Second, after selecting sample sites for computing local calibration factors, additional data was collected by counting features on aerial photos (i.e., Google Map) in most cases. Therefore, most of the data items were manually coded into an Excel table, which includes, for example, the number of right-turn lanes, the number of left-turn lanes, the existence of lighting, and driveway density by land uses. The additional data collection also involved manual estimation of curves and manual collection of signal phasing and regression analyses for estimating AADT. The initial study database consisted of 2.665 million data elements as presented in Table 7.

Table 7. Analyzed Data Items

Data	Year	Count
Network data used for roadway segments	2008	177,701
	2009	180,722
	2010	185,164
MAST data used for intersections	2008	543,964

[2] The tutorial is available at http://www.ihsdm.org/wiki/Download_Registration.

	2009	548,208
	2010	553,812
Crash data	2008-2010	282,310
ADA Driveway Crossings	2012	19,810
Data items from Asset Data Warehouse	The most recent	173,167
Manual (Side slope & Curve data)	The most recent	489
Total	**2008-2010**	**2,665,347**

Once initial data sets were collected, a list of available variables for HSM's SPFs was identified. In some cases, several variables should be combined together to create a new variable. For example, land use types and parking spot counts should be combined to find out the number of major/minor driveways based on different land use types. To obtain the median width, the left side and right side median widths have to be summed up. Then, data quality was checked for identifying missing, inconsistent, or counter-intuitive information. Vehicle crashes were assigned to segments and intersections. Most of this process was carried out in ArcGIS 10.1, a geographic information system (GIS) software for working with information and maps integrated.

Historical Crash Data

The study used historical 2008, 2009, and 2010 crash data of MSP. Crash data for the City of Baltimore was not included in this study because of the study scope.

The summary of the collected crash data of all roadways in Maryland is provided in Table 8. During the study period, 282,310 crashes occurred, resulting in 1,638 fatalities and 1,389,950 injuries. Of them, 78% of fatalities and 61% of injuries were from roadway crashes, and 22% of fatalities and 39% of injuries were from intersection crashes.

Table 8. Summary of Crashes in All Roadway Types: 2008 – 2010

Year	Total Crashes	Fatalities	Injuries	Roadways		Intersections	
				Fatalities	Injuries	Fatalities	Injuries
2008	95,354	592	48,148	439	28,013	153	20,135
2009	96,421	550	47,359	458	31,400	92	15,959
2010	90,535	496	44,443	379	25,487	117	18,956
Total	**282,310**	**1,638**	**139,950**	**1,276**	**849,00**	**362**	**55,050**

Table 9 summarizes crashes by route type. There are 11 route types in the database. As mentioned earlier, crashes that occurred on SHA-maintained roadways (route type MD and US) are of interest to this study. Approximately 34% (or 104,222 crashes) of the total vehicle crashes occurred on MD and US roads during the study period.

The crash data set was checked for consistency. Any data items with geocoding errors were removed from the data set, which were about 5.5% of crashes on MD and US roadways. After data cleaning, a total of 62,777 roadway segment crashes and 35,664 intersection crashes remained in the crash data set. Later these crashes were spatially joined to associated roadway segments and intersections in ArcGIS 10.1 in order to create a data set for sampling (site selection).

Table 9. Summary of Crashes by Road Type: 2008 – 2010

Route Types	Total Crashes	Fatalities	Injuries
CO	69,684	373	33,594
CY	52,927	116	20,439
GV	253	3	127
IS	22,765	160	11,783
MD	82,460	732	49,940
MU	12,856	18	5,089
OP	1,159	3	426
RP	222	1	79
SR	301	0	93
US	21,762	215	13,648
UU	17,921	17	4,732
Total	**282,310**	**1,638**	**139,950**

Roadway Characteristics Data

Roadway data was collected from four main sources: (1) SHA roadway network GIS maps, (2) mile point GIS maps, (3) additional data collection efforts, and (4) assumptions based on HSM suggestions for some desirable variables.

Roadway Network

GIS maps of the Maryland roadway network for the study period were provided by SHA. The roadway network maps included many variables required by HSM SPFs for roadway segments and intersections. While some variables, such as AADT and the total number of through lanes can be used without modification, some variables needed to be modified to obtain variables for HSM models. For example, to obtain effective median width, three columns (variables) should be summed up which are median width, middle shoulders and turning lanes if exist. Table 10 shows variables available from the roadway network maps. Also, the table provides notes on the ways of complementing incomplete data points that are discussed in detail later.

Table 10. Collected Variables from Roadway Network Maps

List of Variables	Required	Desirable	Notes
Area type (rural/suburban/urban)	●		
Annual average daily traffic volume	●		
Average annual daily traffic (AADT) for some minor roads	●		SHA database (regression estimation was used for those minors without actual data)
Segment length	●		
Number of through traffic lanes	●		
Lane width	●		
Shoulder width	●		
Shoulder type	●		
Presence of median (divided/undivided)	●		
Median width	●		
Presence of two-way left-turn lane	●		
Low-speed vs. intermediate or high speed	●		
Lengths of horizontal curves and tangents	●		SHA's eGIS does not have required data for all rural two-lane two-way roadways, so additional data were manually estimated by using Circle measurment tool in Google Earth Pro
Radii of horizontal curves	●		
Roadside slope (sideslope)	●		Manually gathered from SHA's eGIS.
Presence of lighting	●		Asset Data Warehouse (ADW) of SHA's eGIS has "Highway lighting". Final determination was made by manually double-checking Google Earth (StreetView in some cases).
Roadside fixed-object density		●	Asset Data Warehouse (ADW) of SHA's eGIS has "Signs" and "Traffic Barriers". This data was used as a base for density and then double-checked manually by Google Earth. Length measurement tool in Google Earth was used for estimating offset.
Roadside fixed-object offset		●	
Presence of centerline rumble strip		●	Asset Data Warehouse (ADW) of SHA's eGIS has "Rumble Strips". Final determination was made by manually double-checking Google Earth (StreetView in some cases).

Mile Point Data: Intersection and Traffic Signal Information

Obtaining intersection location was simple (using MAST database from SHA), but making the data useful for the study was somewhat challenging (see "Creating Intersection Database"). Table 11 shows variables available from the mile point data.

Table 11. Collected Variables from Mile Point Data

List of Variables	Required	Desirabl	Notes
Number of intersection legs	●		
Type of intersection traffic control	●		
Presence of left-turn phasing	●		Manually gathered by using the combination of "Signal Plan Locator" of SHA and Google Earth (StreetView in some cases)
Type of left-turn phasing	●		

Additional Data Collection

Some of the variables were not available from SHA or needed to be complemented to make complete data set. In these cases, additional data collection methods had to be employed. Table 12 and

Table 13 show additional variables collected for roadway segments and intersections, respectively. A more detailed discussion on additional data collection methods are discussed later.

Table 12. Additional Data Collected for Roadway Segments

List of Variables	Required	Desirabl	Notes
Driveway density	●		Manually counted using Google Earth
Number of major commercial driveways	●		Manually counted using Google Earth for numbers and commercial, Industrial/institutional, residential and other land uses for type from Maryland Department of Planning [for major/minor distinction, HSM guidelines of 50 parking space threshold used.)
Number of minor commercial driveways	●		
Number of major residential driveways	●		
Number of minor residential driveways	●		
Number of major industrial/institutional driveways	●		
Number of minor industrial/institutional driveways	●		
Number of other driveways	●		
Lengths of horizontal curves and tangents	●		SHA's eGIS does not have required data for all rural two-lane two-way roadways, so additional data manually was estimated by using Circle measurment tool in Google Earth Pro
Radii of horizontal curves	●		
Percent of on-street parking	●		Manually gathered by using length measurement tool in Google Earth
Type of on-street parking	●		Manually gathered using Google Earth for type and commercial, Industrial/institutional, residential and other land uses for type from Maryland Department of Planning
Presence of lighting	●		Asset Data Warehouse (ADW) of SHA's eGIS has "Highway lighting". Final determination was made by double-checking manually Google Earth (StreetView in some cases).
Percent grade		●	HSM default assumption: Base defaut on terrain, SHA database does not have all required data for all rural two-lane two-way roadways, so additional data was estimated by using elevation profile in Google Earth for majority of samples
Terrain (level, rolling and mountainous)		●	
Roadside fixed-object density		●	Asset Data Warehouse (ADW) of SHA's eGIS has "Signs" and "Traffic Barriers". This data was used as a base for density and then double-checked manually by Google Earth. Length measurement tool in Google Earth used for estimating offset.
Roadside fixed-object offset		●	
Presence of centerline rumble strip		●	Asset Data Warehouse (ADW) of SHA's eGIS has "Rumble Strips". Final determination was made by double-checking manually Google Earth (StreetView in some cases).

Table 13. Additional Data Collected for Intersections

List of Variables	Required	Desirabl	Notes
Average annual daily traffic (AADT) for minor road	●		SHA database (regression estimation was used for those minors without actual data)
Presence of left-turn phasing	●		Manually gathered by using the combination of "Signal Plan Locator" of SHA and Google Earth (StreetView in some cases)
Type of left-turn phasing	●		
Use of right-turn-on-red signal operation	●		Manually gathered by using Google Earth (StreetView in some cases).
Use of red-light cameras	●		
Presence of major-road left-turn lane(s)	●		Manually gathered by using Google Earth (StreetView in some cases). SHA's Visidata could provide more detailed information, but the research team's access to it was limited.
Presence of major-road right-turn lane(s)	●		
Presence of minor-road left-turn lane(s)	●		
Presence of minor-road right-turn lane(s)	●		
Presence of lighting	●		Asset Data Warehouse (ADW) of SHA's eGIS has "Highway lighting". Final determination was made by double-checking manually Google Earth (StreetView in some cases).
Intersection skew angle		●	Manually estimated using an uploaded compass on Google Earth 2D view
Pedestrian Volume		●	Research team used land uses from Maryland Department of Planning to translate pedestrian activities into volume and a method based on presence of left-turn and right-turn lanes, median type and width, and through lanes to find out the max. number of lanes crossed by pedestrians in each manuover.
Maximum number of lanes crossed by pedestrians on any approach		●	
Presence of Bus Stops		●	Spatially join of buffers (with 1000 ft radius) for each intersection and bus locations/schools from external source were used.
Presence of Schools within 1,000 ft		●	
Presence of alcohol sales establishments		●	Manually gathered by using buffers (with 1000 ft radius) added as KMZ file in Google Earth with search results for "Liquor store" and counting them.

Assumptions Based on HSM Suggestions

For some desirable variables, this study followed HSM's provided "default assumption" (Table 14). Given resource limitation, these data items could not be collected using alternative methods such as manual counting or manual measurement. It is suggested that having these data collected would help a future update of LCFs. Otherwise, sensitivity analysis would be of help to prioritize the variables that should be collected first when resources become available.

Table 14. Assumptions Made for Some Desirable Variables Based on HSM Suggestions

List of Variables	Required	Desirabl	Notes	Assumption
Superelecation variance for horizontal curves		●	HSM default assumption: No superelevation variance	Assumed no superelevation variance.
Presence of spiral transition for horizontal curves		●	HSM default assumption: Base default on agency design policy (SHA: No spiral transition)	Assumed no spiral transition.
Roadside hazard rating		●	HSM default assumption: Assume roadside hazard rating = 3	Assumed 3.
Presence of passing lane		●	HSM default assumption: Assume not present	Assumed not present.
Presence of short four-lane section		●	HSM default assumption: Assume not present	Assumed not present.
Use of automated speed enforcement		●	HSM default assumption: Base default on current practice	Assumed not present.

Data Generation

Two sets of master databases are necessary for computing predicted crashes and calibration factors using IHSDM: (1) homogenous roadway segments and (2) intersections. A set of selection criteria was used to create the two data sets. Most work was performed using ArcGIS 10.1 and some tasks were carried out using Microsoft Excel and Access.

Creating Homogenous Segment Database

For the application of HSM methodology, each roadway must be divided into homogenous segments. In general, a segment is a section of a roadway where traffic is not interrupted by an intersection and it consists of homogenous geometric and traffic control features. In Figure 3, a section between two intersections (A) is a segment. To be used for analysis, roadway geometric and traffic variables within the segment should be same, that is, homogenous. Not all variables can be homogenous within the segment, but researchers should do their best in creating segments as homogenous as possible. If a certain variable changes (e.g., adding a lane), the segment should be divided into two at the location where the number of lanes changes. Depending on how roadway geometry data is collected and maintained, the detailed steps to be taken would vary. As stated earlier, in the case of the Florida study, the information on each geometric attribute was kept with a different database table in Florida RCI[3] and the attributes did not change at the same locations. An algorithm to break down the link needed to be written in Python. Fortunately, the roadway geometry data provided by SHA was organized in a way that variables between two mile points did not change. According to email discussions with an SHA staff member, new mile points were added when new changes were made[4]. Thus, the study team did not need to go through a time-consuming segmentation process. However, data cleaning and correction were required to prepare final data sets. Figure 4 shows an example of the data table received from SHA for roadway data.

[3] Florida Roadway Characteristics Inventory (RCI)
[4] "*A mile point is added when there is a change in any physical or administrative attribute.*"

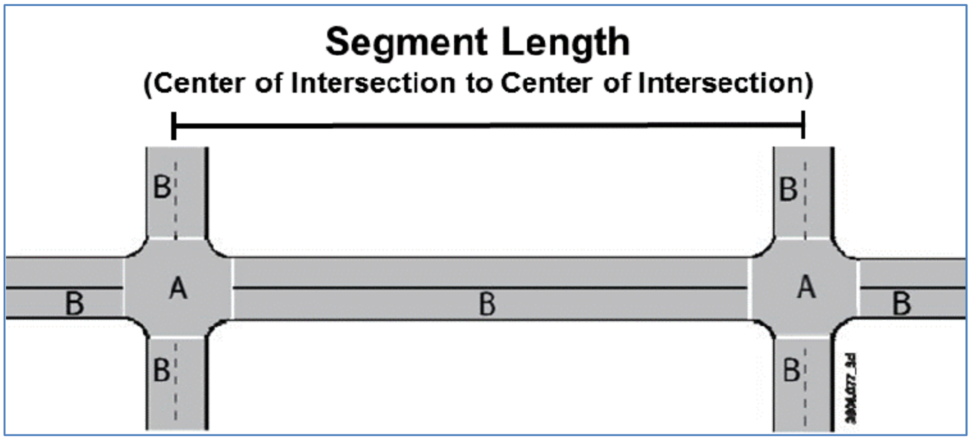
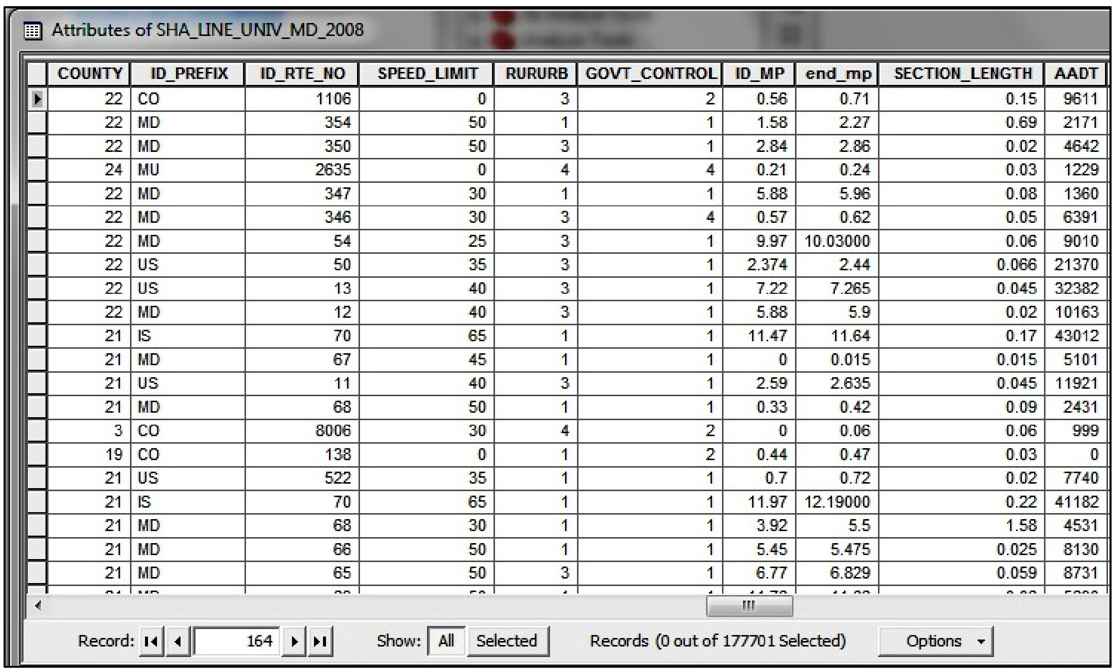


Figure 3. Segments and Intersections (Source: HSM)



Attributes of SHA_LINE_UNIV_MD_2008

COUNTY	ID_PREFIX	ID_RTE_NO	SPEED_LIMIT	RURURB	GOVT_CONTROL	ID_MP	end_mp	SECTION_LENGTH	AADT
22	CO	1106	0	3	2	0.56	0.71	0.15	9611
22	MD	354	50	1	1	1.58	2.27	0.69	2171
22	MD	350	50	3	1	2.84	2.86	0.02	4642
24	MU	2635	0	4	4	0.21	0.24	0.03	1229
22	MD	347	30	1	1	5.88	5.96	0.08	1360
22	MD	346	30	3	4	0.57	0.62	0.05	6391
22	MD	54	25	3	1	9.97	10.03000	0.06	9010
22	US	50	35	3	1	2.374	2.44	0.066	21370
22	US	13	40	3	1	7.22	7.265	0.045	32382
22	MD	12	40	3	1	5.88	5.9	0.02	10163
21	IS	70	65	1	1	11.47	11.64	0.17	43012
21	MD	67	45	1	1	0	0.015	0.015	5101
21	US	11	40	3	1	2.59	2.635	0.045	11921
21	MD	68	50	1	1	0.33	0.42	0.09	2431
3	CO	8006	30	4	2	0	0.06	0.06	999
19	CO	138	0	1	2	0.44	0.47	0.03	0
21	US	522	35	1	1	0.7	0.72	0.02	7740
21	IS	70	65	1	1	11.97	12.19000	0.22	41182
21	MD	68	30	1	1	3.92	5.5	1.58	4531
21	MD	66	50	1	1	5.45	5.475	0.025	8130
21	MD	65	50	3	1	6.77	6.829	0.059	8731

Record: 164 Show: All Selected Records (0 out of 177701 Selected) Options

Figure 4. An Example of DB Received from SHA (2008)

The first data generation and reduction step is to select roadway segments. SHA maintained roadways meeting the following criteria were selected. This process involved a series of steps in Microsoft Excel, Python, and ArcGIS.

1. SHA road type: MD and US
2. Government control type =1 (State highway)
3. One of the eight roadway segment types included in HSM
4. Segments with the same number of lanes for both directions (symmetric roadway segments)
5. Roadway variables consistent for the study period

Table 15 presents the data reduction procedure. There were nearly 200,000 segments per roadway network data set for each year. After three data reduction steps, approximately 45,000 to 47,000 segments maintained by SHA remained in the database. After 35 more data cleaning steps, 25,486 common roadway segments remained in the data set. More detailed procedures for the process of finding common roadway segments (step 6) are provided in Appendix E. It should also be noted that further cleaning was necessary during the process of crash data assignment to the network, which reduced the final number of the sampling pool for site selection and was discussed later.

Table 15. Summary of Data Reduction Procedure

Year		Roadway data			Total	Note
		2008	**2009**	**2010**		
Original data from MSP database		177701	180722	185164	543587	-
% of original data		32.69%	33.25%	34.06%	100.00%	
Step 1	"LOC_ERROR" = NO ERROR	177701	180146	184995	542842	In this step, 576 records of 2009 data and 169 records of 2010 data were deleted. There is no "LOC_ERROR" field for 2008 data. After contacting SHA: All considered as "NO ERROR".
	% reduction from original data	0.00%	0.32%	0.09%	0.14%	
Step 2	"GOVT_CONTROL" = 1	55254	56594	58711	170559	"GOVT_CONTROL" = 1 means State Highways. In this step, 122447 records of 2008 data, 124128 records of 2009 data and 126284 records of 2010 data were deleted.
	% reduction from original data	68.91%	68.68%	68.29%	68.62%	
Step 3	"ID_PREFIX" = MD or US	45252	46368	47536	139156	"ID_PREFIX" = MD or US means State routes based on the signed route number. In this step, 10002 records of 2008 data, 10226 records of 2009 data and 11175 records of 2010 data were deleted.
	% reduction from original data	74.53%	74.34%	74.33%	74.40%	
Step 4	"LT_THRU_LA" = "RT_THRU_LA"	41761	42767	43814	128342	Finding symmettric records. In this step, 3491 records of 2008 data, 3601 records of 2009 data and 3722 records of 2010 data were deleted.
	% reduction from original data	76.50%	76.34%	76.34%	76.39%	
Step 5	"Rd_Type" criteria	35621	36410	37320	109351	Filtering "Rd_Type" based on 8 roadway facility types of the HSM. In this step, 6140 records of 2008 data, 6357 records of 2009 data and 6494 records of 2010 data were deleted.
	% reduction from original data	79.95%	79.85%	79.84%	79.88%	
Step 6	**Common Segments (33 sub-steps)**	**25486**	**25486**	**25486**	76458	In this step, 10135 records of 2008 data, 10924 records of 2009 data and 11834 records of 2010 data were deleted.
	% reduction from original data	85.66%	85.90%	86.24%	85.93%	

Crash Data Assignment to Roadway Segments

After creating homogenous segments for the study years, crashes were assigned to the segment database. This task was more difficult than initially expected. The reason was that there is no unique identifier that connects two databases. While NLFID is provided in both databases, it is not a unique identifier, making it impossible to link crashes to each segment. For this reason, the study team used the ArcGIS spatial join tool for crash assignment.

While a crash (red triangle) should be assigned to only one segment (AB), there were some crashes assigned to two segments (Figure 5). This is the case when a crash occurred near the point B where segments AB and BC meet. Due to a default search range of a GIS spatial join tool, the crash is assigned to both segments. NLFIDs of segments and crashes, mile post information, and other variables were compared to remove incorrectly assigned crashes.

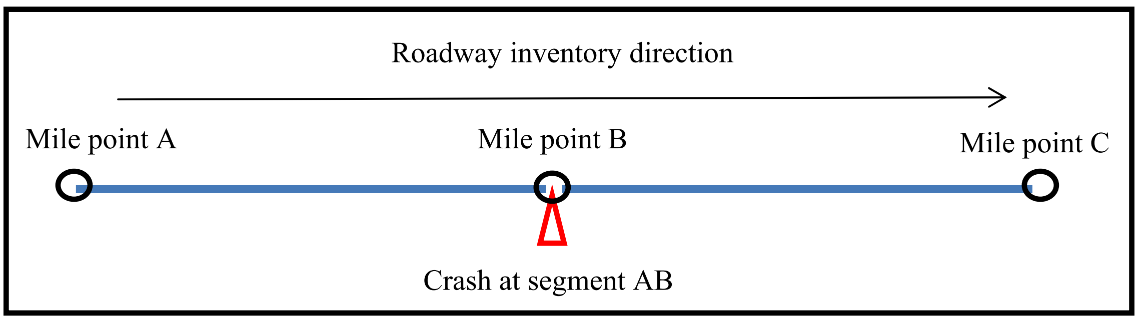


Figure 5. Potential Duplication of Assigned Crashes

Of 62,777 segment crashes on MD and US roadways, 36,325 crashes were successfully assigned to homogeneous MD and US segments. Due to the additional data cleaning work involved in this step, the number of segments that remained in the data set was reduced to 25,055 from 25,486. Approximately 52% of segments did not have crashes during the study period (Table 16). The remaining 26,452 crashes happened outside of the homogeneous segments.

Table 16. Summary of Crash Assignment to Roadway Segments

Segments	# Segments	% segments	# Crashes
With Crashes	12,044	48.1	36,325
Without Crashes	13,011	51.9	0
Total	**25,055**	**100.0**	**36,325**

A more detailed crash assignment result is summarized in Table 17. Among three rural roadway types, R2U (rural two lane undivided) is a dominant type for all variables. R2U accounts for 87% of segments and 93% of the total rural roadway length included in the study database. R2U has the longest average segment length of 0.2 mile per segment and the most crashes (approximately 83% of total rural crashes) occurred on R2U. However, in terms of crash frequency per mile, R4U (rural four lane undivided) has the highest average frequency of 25.29 crashes per mile, followed by R4D (rural four lane divided).

U2U (urban two lane undivided) and U4D (urban four lane divided) are the two dominant types for urban and suburban roadway types. The two roadway types together represent approximately 89% of segments and 89% of the total urban and suburban roadway length in the study database. Except for U5T (urban five-lane arterials including a center two-way left turn lane (TWLTL)), the average segment length is shorter than 0.08 mile, which is expected due to the more dense built environment in urban and suburban areas than in rural areas. The most crashes, almost half of all urban crashes, occurred on U4D, followed by U2U. In terms of the average crash frequency, U5T and U4U top the list with 62.59 and 61.19 crashes per mile, respectively.

Table 17. Crash Assignment to Roadway Segments by Facility Type

Facility type	# Segments	Segment Length (Miles)	Avg. Segment Length	# Crashes	Crashes per Mile
(a) Rural Roadway Segments					
R2U	9,519	1,874.34	0.20	8,938	4.77
R4U	19	1.7	0.09	43	25.29
R4D	1,410	130.18	0.09	1,818	13.97
Subtotal (Rural)	10,948	2,006.22	0.18	10,799	5.38
(b) Urban and Suburban Roadway Segments					
U2U	7,215	576.89	0.08	7,859	13.62
U3T	537	35.63	0.07	973	27.31
U4U	741	40.71	0.05	2,491	61.19
U4D	5,338	320.92	0.06	12,105	37.72
U5T	276	33.52	0.12	2,098	62.59
Subtotal (Urban & Suburban)	14,107	1,007.67	0.07	25,526	25.33
Total	**25,055**	**3,013.89**	**0.12**	**36,325**	**12.05**

Creating Intersection Database

The mile point GIS maps (MAST) included intersection information with traffic control types. The only and biggest problem was that the map database included lots of duplicated points. For example, a four-leg intersection may have four points that are located at the same intersection location: two for beginning or ending mile points of the intersecting roads, and two for traffic control.

The HSM provides models for ten intersection types. The differentiation is based on the number of the intersection legs (that is, 3 or 4 legs) and type of control (i.e., signalized or stop-controlled – stop signs on the minor roadway approaches) for three main facility types (i.e., rural two-lane, two-way roads; rural multilane highways; and urban and suburban arterials). Rural two-lane, two-way roads and rural multilane highways each similarly have three intersection types; 3-leg and 4-leg stop-controlled and only 4-leg signalized intersections; however, urban and suburban arterials have four intersection types which is 3-leg signalized intersections in addition to what the other two main groups have.

3,046 intersections meet HSM modeling requirements, including 1,490 stop-controlled and 1,556 signalized intersections.

Crash Data Assignment to Intersections

Once the MAST database was cleaned, intersection crashes were assigned to the intersection map. As was done in the segment crash assignment, ArcGIS spatial join tool was used for the database creation. There were 35,664 correctly geocoded crashes in MD and US intersections during the study period. However, after the data reduction process (i.e., selecting HSM facilities only and also assigning crashes to the intersections), 11,510 crashes were assigned to the study intersections (10,231 crashes for signalized and 1,279 crashes for stop-controlled intersections).

Roughly 61% of intersections had at least one crash during the study period, while 39% of them did not have crashes (See Table 18).

Table 18. Summary of Crash Assignment to Intersections

Intersections	# Intersections	% Intersection	# Crashes
With Crashes	1,858	61.0	11,510
Without Crashes	1,188	39.0	0
Total	**3,046**	**100.0**	**11,510**

Table 19 presents a summary of crash assignment by intersection type. U4SG (urban four-leg signalized intersection) was the most common intersection type. It accounted for roughly 32% (960 sites) of intersections included in the study database, followed by R3ST (rural two-lane, three-leg, stop-controlled intersection), U3ST (urban three-leg, stop-controlled intersection), and U3SG (urban three-leg signalized intersection). Three rural multilane intersections (RM3ST, RM4ST, and RM4SG) represented only 2.6% (or 79 intersections) of the all intersections. In terms of the average crash frequency per intersection, U4SG topped the list with 7.57 crashes per intersection, followed by RM4SG (6.10) and U3SG (5.03). Summarizing the data by traffic control type, the average crash frequency of signalized intersections was almost eight times as high as that of stop-controlled intersections.

Table 19. Crash Assignment to Intersections by Facility Type

Facility type	# Intersections	# Crashes	Average Crash per Intersection
R23ST	579	307	0.53
R24ST	219	290	1.32
R24SG	69	267	3.87
RM3ST	33	50	1.52
RM4ST	7	29	4.14
RM4SG	39	238	6.10
U3ST	492	306	0.62
U4ST	160	297	1.86
U3SG	488	2,455	5.03
U4SG	960	7,271	7.57
Subtotal (Stop-Controlled)	1,490	1,279	0.86
Subtotal (Signalized)	1,556	10,231	6.58
Total	**3,046**	**11,510**	**3.78**

Additional Data Collection for Samples

After the site selection (sampling) task was completed (See "Sampling (Site Selection)"), additional data items were collected. The additional data collection task was necessary because some data items were not readily available or complete. This task was put after the sampling to accomplish the main purpose of this study, development of LCFs, with the given resources. Since about 40% of the data was not readily available to the team, it was not possible to collect all data at one shot. So, the study team proposed to conduct additional data collection after the sampling task. For most data items, Google Earth was utilized for manually counting and measuring variables. Multiple regression models were developed for estimating AADT on minor roads. Five students (one graduate and four undergraduate students) spent four months collecting and measuring additional data items.

Data Items Collected by Counting

Several data items were collected by counting variables shown on Google Earth. XML files with KML format[5] were created for the sampled segments and superimposed on Google Earth. This made it easy to find the segments of interest.

Figure 6 shows an example of how data was overlaid to count driveways by land uses. A land use map downloaded from the Maryland Department of Planning website was transformed into an XML map and superimposed on the top of the XML segment map and Google Earth. Pink, yellow, and blue overlay layers represent commercial, residential, and industrial/institutional land uses, respectively. Red lines on Route 543 represent sampled segments. Driveways were counted by zooming into the location and counting driveways on both sides of the segment.

Using this method, the following variables were collected:

- Centerline rumble strip
- Driveway density (R2U)
- Driveway density by land uses for urban and suburban arterials
- Roadside fixed-object data including density and offset
- Presence of Lighting
- Left-turn phasing (complementary to “Signal Plan Locator” including PDF files of history of traffic control plans of intersections)
- Use of right-turn-on-red signal operation
- Use of red-light cameras

[5] Keyhole Markup Language (KML) is an XML notation for expressing geographic annotation and visualization within Internet-based maps (two-dimensional).

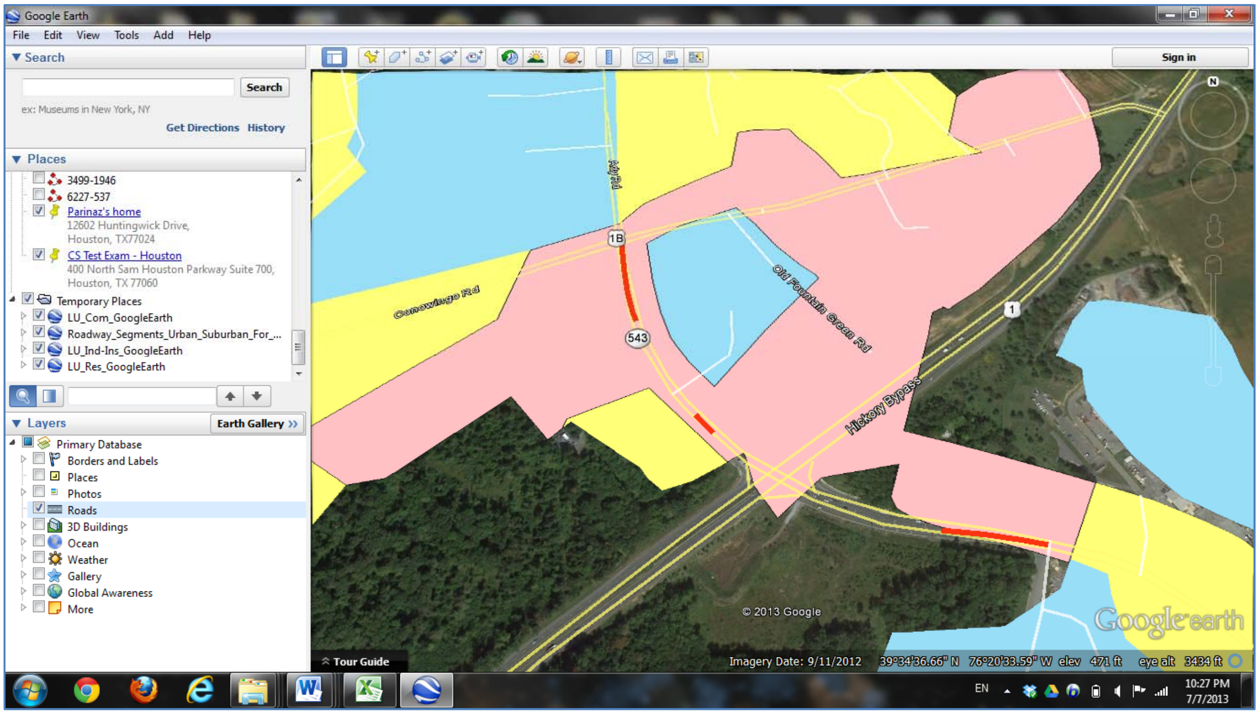


Figure 6. An Example of Driveway Data Collection

<u>Pedestrian Volume</u>

Pedestrian volume is a required variable for intersection models. Like other case studies reviewed, this variable was not available to the study team. While HSM provides HSM default assumption for crash frequency computation (Table 20), there was no direction for making a judgment on the level of pedestrian activity. The study team made an assumption, similar to the Oregon state case study, that pedestrian volume would be highly correlated with surrounding land uses.

Table 20. Estimations of Pedestrian Crossing Volumes for Signalized Intersections

Table 12-15. Estimates of Pedestrian Crossing Volumes Based on General Level of Pedestrian Activity

	Estimate of PedVol (pedestrians/day) for Use in Equation 12-29	
General Level of Pedestrian Activity	**3SG Intersections**	**4SG Intersections**
High	1,700	3,200
Medium-high	750	1,500
Medium	400	700
Medium-low	120	240
Low	20	50

Source: (American Association of State Highway and Transportation Officials 2010, 12-37)

Data Items Collected by Manual Measurement

Grade (elevation), curve data and skew angle(s) were collected by manual measurement. Google Earth Pro was used with XML files that were created by the study team.

Grade (Elevation)

The default assumption made by HSM for grade data in case of unavailability of actual data, is that grade values should be based on terrain type (level, rolling, and mountainous). However, terrain data was not available to the study team, so the Google Earth Elevation Profile was utilized as shown in Figure 7 to estimate terrain category (elevation profile includes average slope). Table 21 shows how to translate average slope values into terrain types and then grade values. Average slope ranges for the lower level and upper level of each category were selected on the basis of provided values in HSM (American Association of State Highway and Transportation Officials 2010, 10-28).

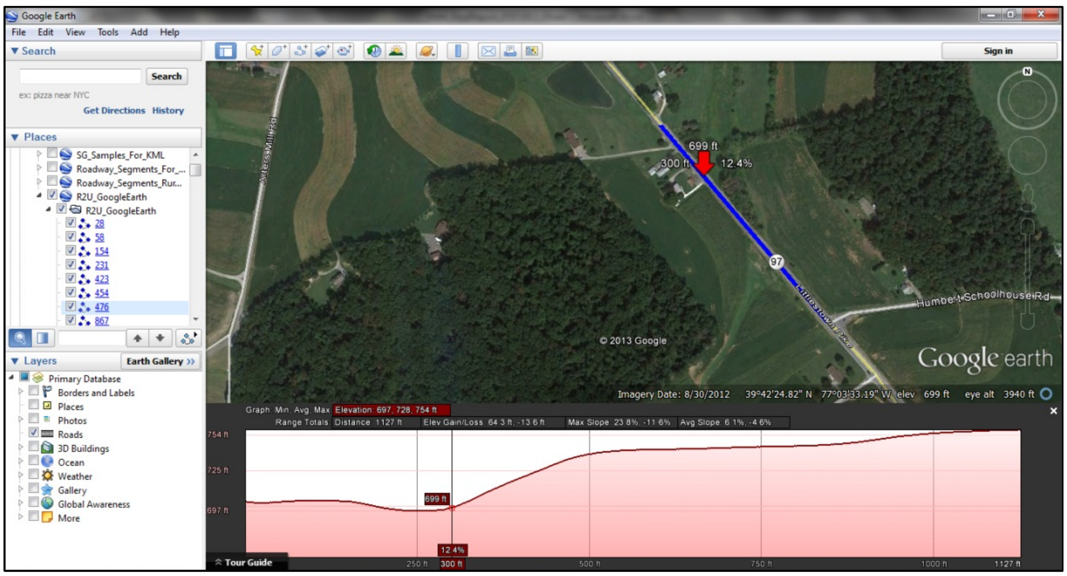


Figure 7. Using Google Earth Elevation Profile for Gathering Terrain Estimation

Table 21. Average Slope Translation into Terrain Categories

Average Slope	Terrain Category	Final Value for Grade
Average Slope ≤ 3%	Level	0%
3% < Average Slope ≤ 6%	Rolling	3%
Average Slope > 6%	Mountainous	6%

Source: (American Association of State Highway and Transportation Officials 2010, 10-28)

Curve Length and Radius Measurement

Curve radii and lengths for rural two-way, two-lane roadway (R2U) segments are “required data,” which need actual data. While a graduate research assistant of the study team reviewed the eGIS database of SHA, the curve information for roughly 40% (or 231 segments) of the sampled R2U segments were not available. To complement the SHA data, several alternative curve estimation methods were reviewed and tested. In conclusion, the study team decided to use Google Earth Pro to estimate curve lengths and radii.

The team reviewed the following five alternative methods to obtain curve data:

- Curve Calculator (COGO) by Environmental Systems Research Institute (ESRI), a vendor of ArcGIS
- Curvature Extension GIS application by the Florida Department of Transportation (Rasdorf, et al. 2012)
- Curve Finder by New Hampshire Department of Transportation (Rasdorf, et al. 2012)
- AutoCAD tool
- Google Earth Pro tool

Curve Calculator

Curve calculator is a command within the coordinate geometry (COGO) toolbar in ArcGIS developed by ESRI (

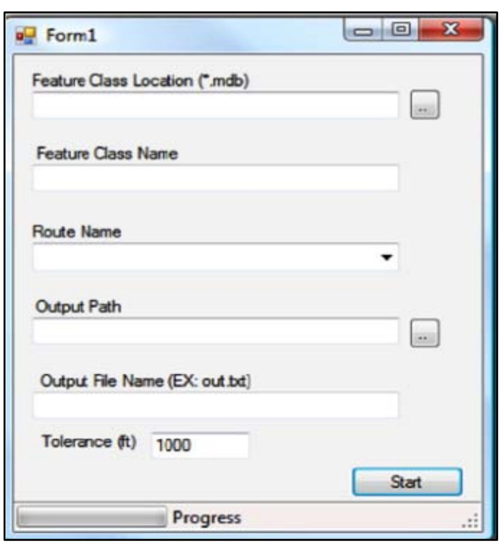


Figure 8). The curve calculator entails manual curve identification, requiring the user to define a PC (point of curvature, where the roadway begins to curve) and a PT (point of tangency, where the curve ends and the roadway returns to a tangent section). After identifying the PC and PT, the software can measure the chord and arc lengths. The user must input any two of four curve characteristics (chord length, angle, arc length, and radius) that are known to determine the remaining unknown characteristics and the chord height and tangent length. To determine the curve’s radius, the user should input the chord length (LC) and the arc length (distance along the curve

from the point of curvature to the point of tangency on the horizontal curve) (Rasdorf, et al. 2012). The study team was not able to use this tool because chord length and angle information were not readily available to the team.

Figure 8. Curve Calculator User Interface (Rasdorf, et al. 2012)

Curvature Extension

Curvature extension is a GIS plug-in tool that was developed by the Florida Department of Transportation (FDOT) in 2010 (Figure 9). The curvature extension also requires the user to manually identify a curve and to specify the limits of each curve. To execute the program, the user must select appropriate data layers for the program to reference, an output file for the results, the direction of the curve (clockwise or counter-clockwise), and the PC and PT points. The radius is determined by creating a circular arc utilizing the chord length, chord angle, and length of the curve along the route. The curve length is calculated on the basis of the end points. The calculated radius is displayed to the user on the existing GIS line work for visual confirmation of the suitability of the match (Rasdorf, et al. 2012). The team was not able to obtain this from the developer for testing.

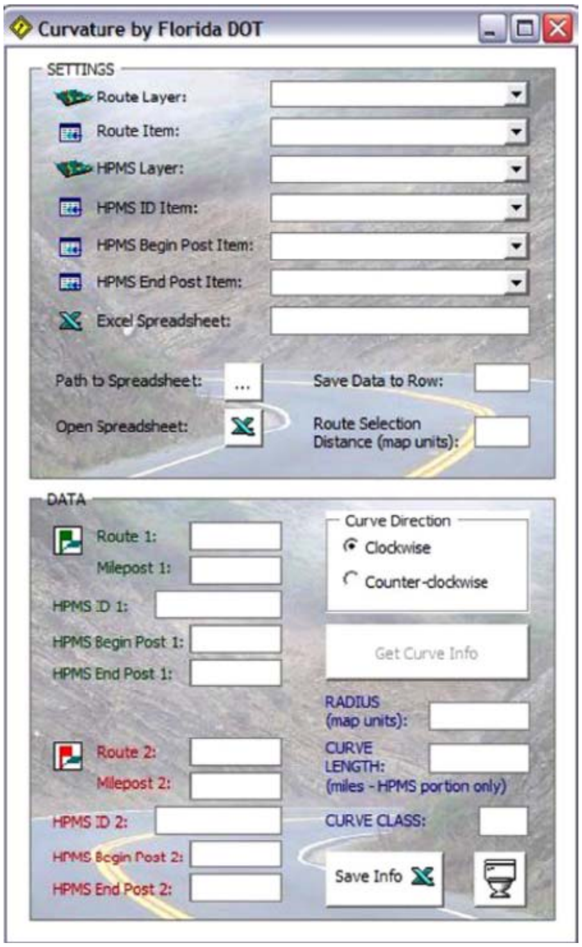


Figure 9. Curvature Extension User Interface (Rasdorf, et al. 2012)

Curve Finder

Curve finder is a program developed by the New Hampshire Department of Transportation (NHDOT) (Figure 10). The curve finder is an automated procedure that can be executed on a network of roadways. The curve finder uses GIS polylines to determine curve length and radius through coordinate data. Curves are identified as the curve finder scan the network, moving through every series of three points that together create a circle and determines if the points meet the curve tolerance. The advantage of the curve finder is that it works automatically. However, the margin of errors was huge, which is similar to COGO. In addition, curve finder is not compatible with ArcGIS 10 and Windows 7. It is an open-source program and can be developed for ArcGIS 10 and also Windows 7 or 8 in future.

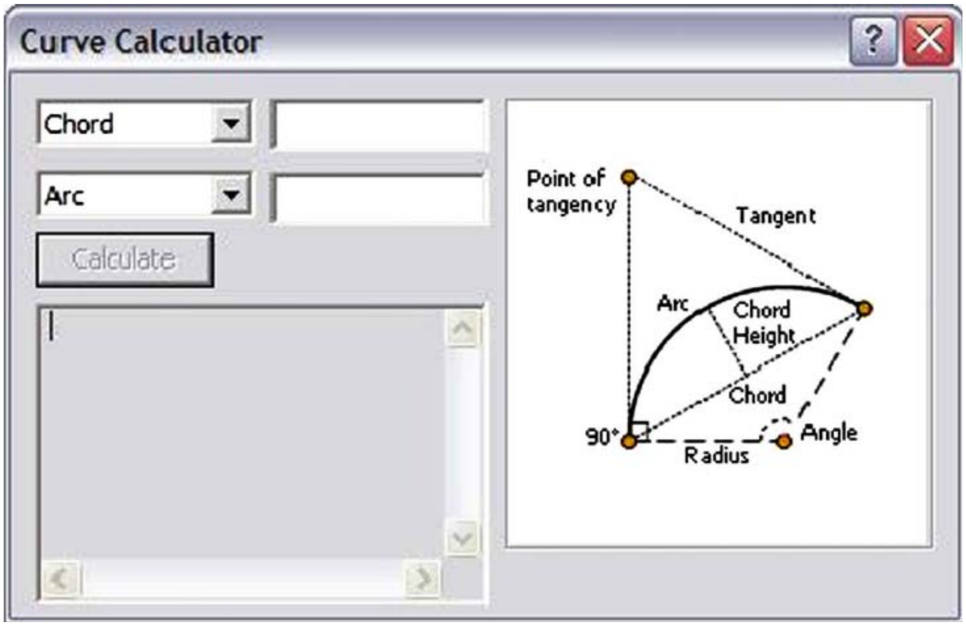


Figure 10. Curvature Finder User Interface (Rasdorf, et al. 2012)

<u>AutoCAD</u>

AutoCAD provides an application that can extract curve data from a network layer. Curves can be extracted when the nature of lines is curvature. If curvature lines are not available, the users have to re-draw all curves manually. Unfortunately, data from SHA was not in the curve format and re-drawing all curves in AutoCAD environment was a time-consuming process. Thus, this alternative was not selected.

<u>Google Earth Pro</u>

Google Earth Pro provides a tool that has the ability of drawing a circle by locating the center point and point and varying the radius (

Figure 11). A user can draw a circle that meets the roadway curvature to find out the curve radius. Then, using the path tool measurement on the curve portion of roadway segment, curve length is determined.

The team decided to use Google Earth Pro for two reasons. First, other software's required data formats were not available to the team. Second, the study team's measurement using Google Earth Pro was more accurate than a previous comparative study. Rasdorf, et al. (2012) compared the performance of COGO, curve extension, and curve finder compared. The margin of errors between the measurement and real data was as high as 230%. However, the study team's measurement error (comparison between manual measurement using Google Pro and SHA actual curve data) was around 5-10% in most cases. Some of the selective example locations are provided in Table 19.

Table 19. Comparison between eGIS Data and Google Earth Pro Estimation for Curve Data

SegID	eGIS Curve Radius (ft.)	Google Earth Pro Curve Radius (ft.)	Difference (%)	eGIS Curve Length (ft.)	Google Earth Pro Curve Length (ft.)	Difference (%)
3499	1,901	1,946	2.3	889	966	8.7
6227	573	537	6.3	588	627	6.6
6911	2,800	2,957	5.6	460	521	13.3
7826	2,154	2,089	3	1,361	1,546	13.6
8627	5,730	5,219	8.9	4,,898	4,390	10.4
15597	3,303	3,286	0.5	291	351	20.6

Note: SegID is a unique segment ID that was used for each facility type during the segmentation stage.

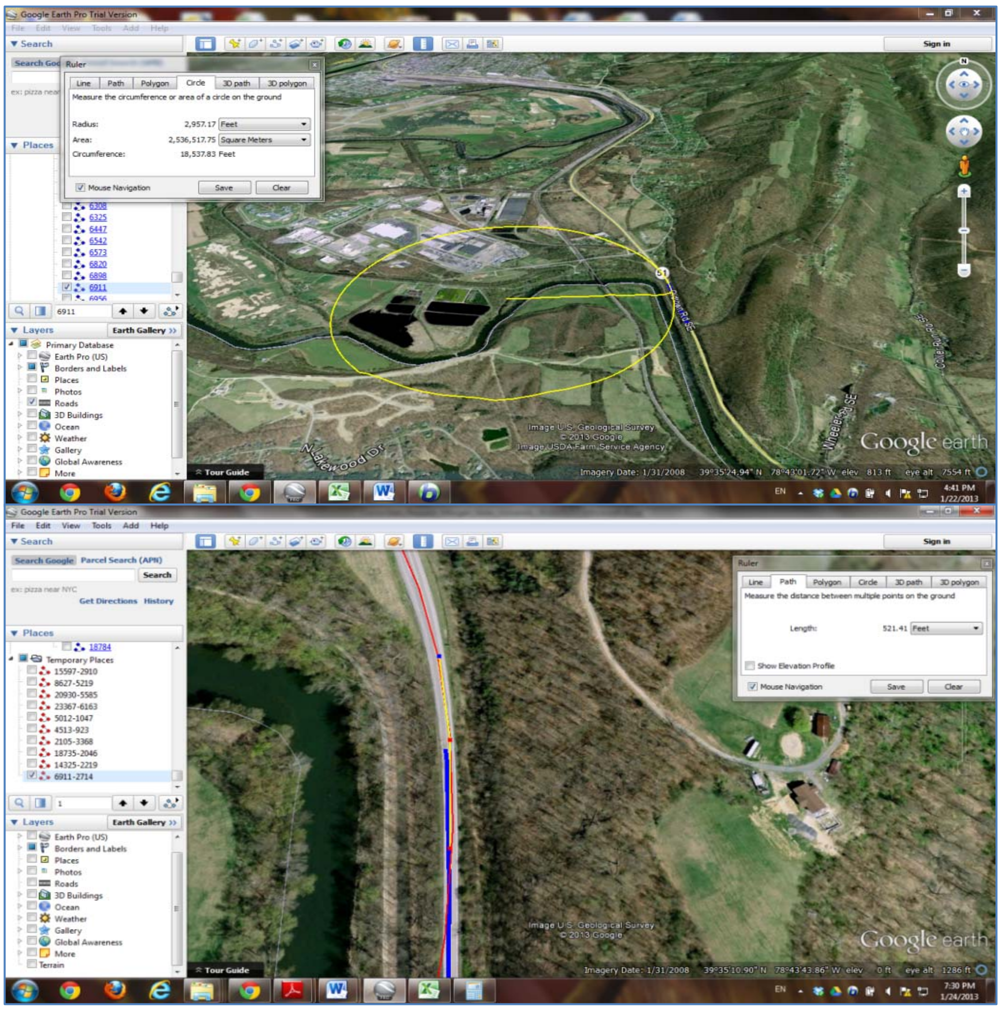


Figure 11. Curve Radius & Length Estimation Using Google Earth Pro

<u>Skew Angle</u>

The intersection skew angle(s), estimated manually by using an uploaded compass on Google Earth 2D view for R23ST, R24ST, RM3ST, and RM4ST, is shown in an example in Figure 12.

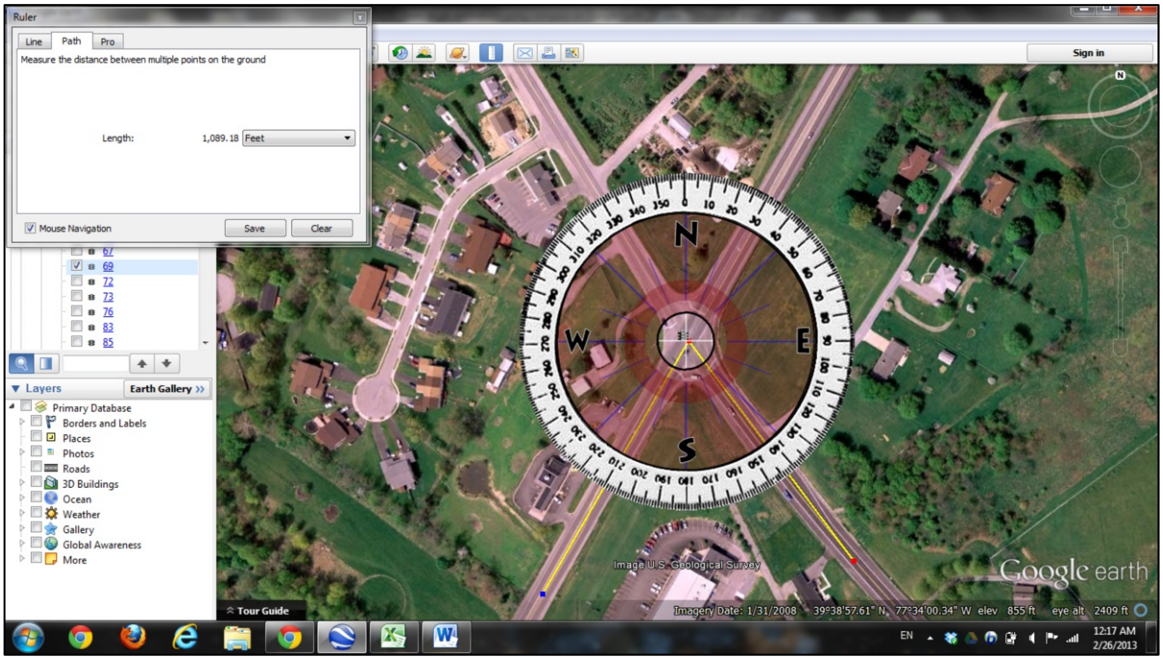


Figure 12. An Example of Intersection Skew Angle Estimation

Estimation of AADT on Minor Roads

The AADT on minor roads is a required variable for intersection SPFs. However, the AADT data table for minor roadways, intersecting roads with the major roads (i.e., MD and US), was not complete. Minor roads' AADT on approximately 45% of signalized intersections and 64% of stop-controlled intersections were missing. While actual field measurement would be the ideal method for collecting AADT, that was not a practical option for the study team.

Missing AADT (or any sorts of traffic count) has been an issue in most traffic studies. Thus, many studies have been carried out to estimate missing AADT values. Depending on computing software, data, and researchers' preferred skill sets, a variety of methods have been utilized, including multiple regression, data mining techniques, Neural Network approaches, and linear regression using geographic information systems (Gecchele, et al. 2011, Jin, Xu and Fricker 2008, Wang, Bai and Bao 2011, Lowry and Dixon 2012, Dixon, et al. 2012). The current study employed multiple regression analysis similar to the Oregon case study. (Dixon, et al. 2012, 5).

Multiple Regression

Multiple regression analysis is one of the most widely used and simple ways to estimate AADT due to its ease of application in many situations and straightforward interpretation of outputs (Washington, Karlaftis and Mannering 2003). Also, its statistical properties are well understood. It has a general form as described in Equation 5.

Equation 5. General Form of Multiple Regression

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_p X_{pi} + \varepsilon_i$$

Where,

y_i = Dependent variable for the i^{th} member of the observation,
x_{pi} = Independent variables,
ε_i = Error term, and
β_i = Coefficients.

For this study, AADT on minor roadways was the dependent variable. Land use categories and roadway geometry variables were used as independent variables.

Variable Selection

Each data set for signalized intersections and stop-controlled intersections contained over 70 independent variables that may have statistical influence on AADT. Since independent variables consist of different types, such as categorical variable (i.e., land uses), ordinal variables (i.e., speed limit – low, medium, high), and ratio variable (i.e., width, length), it was not easy to determine correlation between AADT and these potential independent variables. Unfortunately, there is no universally accepted standard that a researcher can use to determine how many and what variables should be included or removed from the model in order not to over-fit or under-fit the model (Gujarati 2003). As a matter of fact, developing a best fit regression model is often based on trial and error after finding out a base model.

To find out the best possible models, a set of variable selection criteria were used such as R-square, adjusted R-square, leaps and bounds, Akaike Information Criterion (AIC), and Bayesian Information Criterion (BIC). Since the discussion of each method is beyond the scope of this report, interested readers should refer to any college-level statistics books or, for a quick summary, Lindsey and Sheather (2010). Various combinations of independent variables were added, removed, or transformed to find the best fit models. Multicollinearity and model specification errors were also evaluated.

Selected Regression Models for Signalized Intersections

The final models for estimating AADT values on minor roadways without actual data for signalized intersections are presented below. Additional details for regression models are provided in Appendix G. Interestingly, the same independent variables were selected for all three years. While a variety of variables such as land uses, socio-economic characteristics, roadway geometry, and roadway types were considered, the selected modes showed that minor road AADT at signalized intersections is a function of roadway type. The R-squared values for the developed models are shown in Table 22. All models have R-squared values greater than 0.7 or 70%, meaning that over 70% of the data is explained by the selected regression models, which can be considered a good fit.

Equation 6. AADT Estimation Models for Minor Roadways of Signalized Intersections (2008 – 2010)

SG_AADT2008 = (-6050.305) + (4319.063*mnnumln) + (1444.975*mjnumln) + (6041.045*mnurothart) + (18318.07*mnurfree) + (-4821.744*mnurloc)

SG_AADT2009 = (-6753.022) + (4436.088*mnnumln) + (1608.455*mjnumln) + (5518.895*mnurothart) + (20185.76*mnurfree) + (-4866.872*mnurloc)

SG_AADT2010 = (-4901.359) + (3988.758*mnnumln) + (1409.611*mjnumln) + (5092.22*mnurothart) + (19502.15*mnurfree) + (-5107.076*mnurloc)

Where:

SG_AADT2008, 2009, and 2010 = Estimated value for AADT on minor roadway of signalized intersections for 2008, 2009, and 2010,

mnnumln = Number of through lanes for minor roadway,

mjnumln = Number of through lanes for major roadway,

mnurothart = 1 if functional class of minor road is “urban other arterials”, otherwise 0,

mnurfree = 1 if functional class of minor road is “urban freeway”, otherwise 0, and

mnurloc = 1 if functional class of minor road is “urban local street”, otherwise 0.

Table 22. R-Squared Values for AADT Estimation Models of Signalized Intersections

Year	R-Squared	Adjusted R-Squared
2008	0.7816	0.7713
2009	0.7781	0.7676
2010	0.7158	0.7024

<u>Selected Regression Models for Stop-Controlled Intersections</u>

Using the same procedure for signalized intersections, minor road AADT values for stop-controlled intersections were estimated. The developed models are presented below. Additional details for regression models are provided in Appendix G. The selected independent variables for the years 2008 and 2009 were the same. However, the 2010 model had somewhat different independent variables. The R-squared values for the developed models are shown in Table 23. Values for 2008 and 2009 were both greater than 0.5 or 50% and acceptable. However, the AADT model for the year 2010 had a weak fit. So, it was assumed that 2010 AADT trend was similar to the 2009 AADT and the 2009 regression model was used to estimate 2010 AADT. Like the regression models for signalized intersections, the roadway types were significant independent variables. In addition, population density at the census tract level and average household size at the census tract level were significant independent variables.

Equation 7. AADT Estimation Models for Minor Roadways of Stop-Controlled Intersections (2008-2010)

ST_AADT2008 = (-3014.581) + (-1009.298*mnruloc) + (818.4968*mnmedwd) + (4235.79*mnrumnart) + (1698.903*avghhsize) + (1300.952*mnrumjcol) + (0.7175917*popden) + (3796.183*mnruothart) + (3486.115*mnurothart)

ST_AADT2009 = (-3145.864) + (-1000.649*mnruloc) + (834.6083*mnmedwd) + (4283.924*mnrumnart) + (1748.293*avghhsize) + (1339.81*mnrumjcol) + (0.7199705*popden) + (3912.901*mnruothart) + (3494.391*mnurothart)

ST_AADT2010 = (-10010.94) + (2540.336*mnrumjcol) + (7432.884*avghhsize) + (-132.6905*pct_emp) + (705.4558*mnmedwd) + (982601.9*mu_l) + (3043.161*rm3st) + (-1437.878*mnruloc)

Where:

ST_AADT2008, 2009, and 2010 = Estimated value for AADT on minor roadway of stop-controlled intersections for 2008, 2009, and 2010,

mnruloc = 1 if functional class of minor road is “rural local street”, otherwise 0,

mnmedwd = Median width of minor road (ft.),

mnrumnart = 1if functional class of minor road is “rural minor arterials”, otherwise 0,

avghhsiz = Average household size,

mnrumjcol = 1 if functional class of minor road is “rural major collector”, otherwise 0,

popden = Population density,

mnruothart = 1 if functional class of minor road is “rural other arterials”, otherwise 0, and

mnurothart = 1 if functional class of minor road is “urban other arterials”, otherwise 0.

pct_emp = Percent of employment,

mu_l = Length of municipal road (mi), and

rm3st = 1 if type of intersection is “RM3ST”, otherwise 0.

Table 23. R-Squared Values for AADT Estimation Models of Stop-Controlled Intersections

Year	R-Squared	Adjusted R-Squared
2008	0.5511	0.5176
2009	0.5509	0.5173
2010	0.3664	0.3253

Sampling (Site Selection)

The sampling task followed the development of the homogeneous segments and intersection databases. The purpose of this task is to select candidate sites for calculating predicted crash frequencies and developing LCFs. In an ideal world, all segments and intersections developed in the previous task would be considered for a complete computation of crash frequencies and LCFs. In reality, however, about 40% of the required and desired data for HSM predictive models were not readily available to the team. Also, about 30% of the readily available data had to be complemented due to missing variables. Thus, the best strategy was to sample candidate sites to use given study resources efficiently. For this report, the two terms, “site selection” and “sampling”, will be used interchangeably.

The HSM provides four site selection criteria:

- The minimum samples size should be 30 to 50 sites per facility type.
- Samples should be drawn randomly.
- Each sample set should have at least 100 annual crashes.
- Short segments should be avoided to prevent biased crash prediction.

While HSM’s suggestion was simple enough to follow, the study team believed that some of the requirements were confusing since no clear discussion was provided for users. As stated in an earlier section, a sample size of 30-50 is a well-accepted criterion in statistics. However, the minimum annual crash requirement of 100 needs further guidance in HSM. Moreover, no clear guidance on the segment length was the most problematic to the study team. A minimum segment length of 0.1 mile stated in HSM could be too long or too short, depending on study areas. The minimum length of 0.1 mile for rural segments and 0.04 mile for urban segments were the thresholds used in the research efforts to develop HSM SPFs and were also used in FDOT study to develop the local calibration factors for Florida (Srinivasan, et al. 2011, 9). For this reason, this study followed the FDOT case study’s guide.

The study team came up with several scenarios for segments and intersections to see how they influenced the predicted crashes and LCFs. In the following sections, discussions about different scenarios and comparison of results by scenarios will be provided.

Roadway Segment Sampling

Four different sample strategies were used for the segment predictive models (Table 24). Developing different sampling scenarios is, in fact, beyond the study scope. This would be the first attempt among LCF case studies. The authors conducted this task in order to generate new ideas for a future LCF methodology refinement endeavor.

Table 24. Roadway Sampling Scenarios

Scenario	Description
1	HSM Base Scenario: Minimum sample of 30 segments and 100 total minimum annual crashes; minimum segment lengths of 0.1 mile for rural segments and 0.04 mile for urban and suburban segments
2	Modifying Scenario 1 by increasing sample size with 90% statistical confidence level and 5% margin of errors
3	Disproportionate Stratified Random Sampling: Stratification by segment length and minimum 100 annual crashes. No minimum length threshold
4	Combination of Scenarios 1-3

Roadway Segments Sampling - Scenario 1

The study team followed HSM guidelines for sampling and also ignoring short segments (0.1 mile and 0.04 mile thresholds for rural and urban and suburban segments, respectively). The study team started with 30 segments and then added more segments to reach a minimum 100 crashes per year for each facility (minimum 300 crashes for the three year study period).

Roadway Segments Sampling - Scenario 2

This scenario modified scenario 1. Three requirements—minimum sample size, segment length, and minimum annual crash criteria—are intact. The only change made was to increase the sample size by drawing a sample with 90% confidence level (C.L.). A sample size with a confidence level can be drawn by using Equation 8.

Equation 8. Minimum Sample Size

$$n = {(n_0 \times N)}\Big/{(n_0 + (N - 1))}$$

Where:

n = Minimum sample size,

N = Population, and

$$n_0 = {Z^2 \times P \times (1 - P)}\Big/{e^2}$$

Where:

Z = Area under normal curve corresponding to the desired confidence level

P = True proportion of factor in the population, or the expected frequency value, and

e = Margin of errors.

While higher confidence levels (95% or 99%) would be preferred, the 90% confidence level (i.e., Z=1.645) is also widely used. The decision was made because using higher confidence intervals significantly increased sample size. For additional data collection efforts, the team could not afford to deal with too large a sample size. To make a conservative assumption, the true proportion of factor in the population or the expected frequency value was set at 0.50 or 50%, which yields the maximum sample size at the 90% CL (i.e., the study team considered the value of P to be 0.50). The common value for e is 0.05 or 5%. Based on these assumptions and using Equation 6, the minimum sample sizes for roadway segments calculated are shown in Table 25. It should be noted that these values were only minimum sample sizes with 90%. Later for few facilities, the final sample size had to be increased to meet minimum annual crash counts of 100.

Table 25. Minimum Sample Size for Roadway Segments with 90% CL

Facility	Population	Minimum Sample Size
R2U	251	241
R4U*	19	19
R4D	297	142
U2U	3,569	252
U3T	278	138
U4U	311	145
U4D	2,470	244
U5T	197	115
Total	**10,455**	**1,306**

Roadway Segments Sampling - Scenario 3

Scenario 3 is based on disproportionate stratified random sampling. While a minimum annual crash criterion is used, no segment length requirement was considered. Disproportionate stratified random sampling is a subtype of stratified sampling strategies (See Figure 13). Briefly, (Daniel 2012) stated that "stratified sampling is a probability sampling procedure in which the target population is first separated into mutually exclusive, homogenous segments (strata), and then a simple random sample is selected from each segment (stratum) (Daniel 2012, 131)

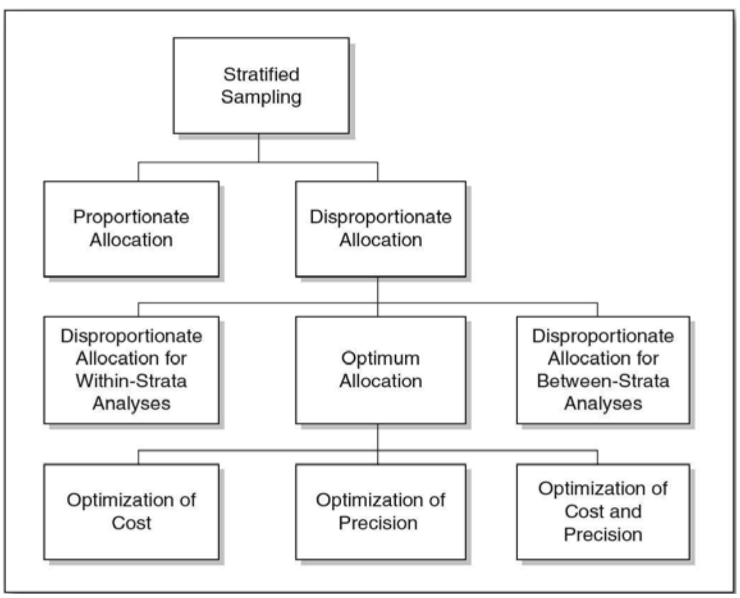


Figure 13. Subtypes of Stratified Sampling Based on Stratum Allocation (Daniel 2012, 133)

For this study, disproportionate stratified sampling with optimum allocation was used. Unlike proportionate stratified sampling, samples are drawn disproportionately. That is, a relatively higher proportion of samples are drawn from a stratum with fewer observations. This way over-representation of dominant strata can be avoided. A primary reason of using this scenario is to look at the impacts of segment length on the crash prediction and LCFs. Since fewer observations are selected from over-represented strata (in this case shorter segments), the sample size can be smaller than the one drawn by proportional stratified sampling, which required more segments to meet minimum annual crash requirement (Table 26). When disproportionate stratified sampling is designed carefully, this could be a cost effective way to sample. The minimum stratum sample size is usually set to at least two (in general five or six) in order to allow variances to be estimated (Daniel 2012, 131, Chambers and Clark 2012, 45). For the current study, ArcGIS Natural Breaks (Jenks) with five strata were used.

Table 26. Sample Sizes and Crashes for Proportionate and Disproportionate Stratified Sampling

Facility	Proportionate Stratified Sampling		Disproportionate Stratified Sampling	
	Samples	**Crashes**	**Samples**	**Crashes**
R2U	350	329	160	328
R4U	19	43	19	43
R4D	260	326	185	336
U2U	290	379	205	377
U3T	185	335	125	350
U4U	110	363	75	442
U4D	140	379	135	404
U5T	41	350	30	318
Total	**1395**	**2504**	**934**	**2598**

Roadway Segments Sampling - Scenario 4

The fourth scenario is a simple combination of scenarios 1 through 3. The intention is to examine the impact of the increased sample size on LCFs.

Roadway Sampling Summary

Table 27 summarizes roadway segments sampling by scenario. R4U was the only facility type that did not meet HSM criteria; there were only 19 R4U segments with 43 crashes for that three years of study. For the sampled segment, an additional data collection task was performed. In general, the sample size for scenario 1 was larger than scenarios 1 and 3. Obviously, the sample size for scenario 4 was the largest, since it includes all unique segments from scenarios 1, 2, and 3. While scenario 2 was used to increase the sample size from scenario 1, in the case of R4D, the same sample size was drawn. For U3T, the sample size for scenario 2 was fewer than scenario 1. This is due to the nature of random sampling. As described earlier, samples were drawn until the minimum annual crash threshold was met.

Table 27. Summary of Roadway Segment Sampling

Facility Type	Total Segments (Population)	Scenario 1		Scenario 2		Scenario 3		Scenario 4	
		Segments	Crashes	Segments	Crashes	Segments	Crashes	Segments	Crashes
R2U	9519	190	341	251	458	160	328	557	1058
R4U	19	19	43	19	43	19	43	19	43
R4D	1410	160	339	160	315	185	336	381	618
U2U	7215	235	361	252	360	205	377	653	1014
U3T	537	140	349	138	330	125	350	259	617
U4U	741	75	369	145	592	75	442	230	1039
U4D	5338	150	451	244	654	135	404	502	1402
U5T	276	30	357	115	1257	30	318	139	1370
Total	**25055**	**999**	**2610**	**1324**	**4009**	**934**	**2598**	**2730**	**7161**

Intersection Sampling

The intersection sampling task was less complicated than the segment sampling task. Only two scenarios were considered:

- Scenario 1 – HSM-Based Scenario: Minimum sample of 30 intersections and 100 total annual crashes per facility type.
- Scenario 2 – Modifying Scenario 1 by increasing sample size with 90% confidence level

Not all intersection types met annual crash criterion. Except for U3SG and U4SG, all other intersection types could not meet the minimum annual crash threshold of 100 (Table 28). For example, R23ST and U3ST had over 300 crashes during the three year period. However, at least for one year, these facilities had fewer than 100 crashes. Only U3SG and U4SG met HSM basic guidelines for sampling (scenario 1). In scenario 2, Equation 8 ($Z = 1.645$, $P = 0.5$ or 50%, and $e = 0.05$ or 5%) was used in a similar way to roadway segments sampling. However, since there were several intersections for some certain facility types, the study team decided to consider all of them for the study (R24SG, RM3ST, RM4ST, and RM4SG). The summary of the intersection samples is provided in Table 29.

Table 28. Total Number of Intersections and Total Crashes During the Study Period

Facility type	Number of Intersections	Total Crashes
R23ST	579	307
R24ST	219	290
R24SG	69	267
RM3ST	33	50
RM4ST	7	29
RM4SG	39	238
U3ST	492	306
U4ST	160	297
U3SG	488	2,455
U4SG	960	7,271
Total	**3,046**	**11,510**

Table 29. Summary of Intersection Sampling

Facility type	Number of Intersections	Scenario 1		Scenario 2	
		Intersections	Crashes	Intersections	Crashes
R23ST	579	N.A.	N.A.	162	103
R24ST	219	N.A.	N.A.	115	142
R24SG	69	N.A.	N.A.	67	262
RM3ST	33	N.A.	N.A.	26	36
RM4ST	7	N.A.	N.A.	10	30
RM4SG	39	N.A.	N.A.	35	231
U3ST	492	N.A.	N.A.	152	103
U4ST	160	N.A.	N.A.	90	173
U3SG	488	80	380	167	789
U4SG	960	40	354	244	1,763
Total	**3,046**	**120**	**734**	**1,068**	**3,632**

Computing Local Calibration Factors

All the collected data sets for all 18 facilities were saved as a Microsoft Excel database. Then the study team computed average predicted crashes per mile for each segment and intersection, and corresponding LCFs. The interactive highway safety design model (IHSDM) was used for the computation task. In this section, results of all scenarios using HSM default crash distribution as well as Maryland-specific crash distribution are presented. Sampling scenarios are compared and the best sampling scenario is discussed.

Local Calibration Factors

An LCF of a facility is a ratio of the total observed crashes at the study site to the total predicted crashes computed by an SPF (Equation 9). For example, there were a total of 9 crashes at a study facility and a predicted crash using an SPF was 10. Then, the LCF for the site is 0.9, meaning that for the same type of facility the predicted crashes using a corresponding SPF should be adjusted by multiplying 0.9.

Equation 9. Calculation of Local Calibration Factor

$$C = \frac{\sum_{All\ sites} N_{Observed}}{\sum_{All\ sites} N_{Predicted\ (Unadjusted)}}$$

Where:

$N_{Predicted\ (Unadjusted)}$ = Unadjusted total predicted crash frequency, and

$\mathrm{N}_{\mathrm{Objerved}}$ = Total number of observed crashes during the study period.

Computing LCFs

Local calibration factors can be calculated using two crash proportions: (1) HSM default proportion that was developed using the data sets from California and Washington states, and (2) study location specific (i.e., Maryland) crash proportion. It is recommended that researchers should compute SPFs and LCFs using two proportions and decide which one to use. If the crash proportion of the study area is significantly different from HSM-provided default crash proportion, the former should be used for calculating LCFs to produce more reliable results.

To compare computed predicted crashes and LCFs, the study team created crash severity and type proportion tables for Maryland. For example, Table 30 shows the comparison between HSM-default proportion and Maryland-specific crash proportion for different KABCO crash severity[6] for rural two-lane, two-way roads. Compared to HSM, the proportion of property damage only crashes in Maryland is much lower, while the proportion of non-incapacitating injury in Maryland is much higher than HSM-default value. Overall, the proportion of total fatal plus injury crashes are higher than HSM values. Table 31 compares crash type proportion of HSM to the Maryland data for rural two-lane, two-way roads. Again, one can observe that the crash type proportion of Maryland seems to be different from HSM-default value. For example, total single vehicle crashes in Maryland accounts for about 84% of total crashes, while the same type of crashes represents roughly 70% in HSM data. For this reason, SPFs and LCFs using data from both sources were calculated and compared to determine what data should be used for Maryland. More comparison tables are provided in Appendix F.

Table 30. Comparison of Crash Severity Proportion: Rural Two-Lane, Two-Way Roads

[6] KABCO scale is used to codify crash severity levels, which consists of fatal (K), incapacitating injury (A), non-Incapacitating injury (B), possible injury (C), and property damage only (O).

Crash severity level	Percentage of total roadway segment crashes	
	HSM-Provided Values	Maryland Values
Fatal	1.3	1.6
Incapacitating Injury	5.4	5.9
Nonincapacitating Injury	10.9	17.3
Possible Injury	14.5	16.8
Total Fatal Plus Injury	32.1	41.7
Property Damage Only	67.9	58.3
TOTAL	100.0	100.0

Note: HSM-provided crash severity data based on HSIS data for Washington (2002-2006)

Table 31. Comparison of Crash Type Proportion: Rural Two-Lane, Two-Way Roads

Collision type	Percentage of total roadway segment crashes by crash severity level					
	HSM-Provided Values			Maryland Values		
	Total fatal and injury	Property damage only	TOTAL (all severity levels combined)	Total fatal and injury	Property damage only	TOTAL (all severity levels combined)
SINGLE-VEHICLE CRASHES						
Collision with animal	3.8	18.4	12.1	27.7	35.2	32.1
Collision with bicycle	0.4	0.1	0.2	1.8	1.4	1.6
Collision with pedestrian	0.7	0.1	0.3	3.2	1.7	2.3
Overturned	3.7	1.5	2.5	13.7	10.8	12.0
Ran off road	54.5	50.5	52.1	30.9	30.1	30.4
Other single-vehicle crash	0.7	2.9	2.1	4.3	5.9	5.2
Total single-vehicle crashes	**63.8**	**73.5**	**69.3**	**81.6**	**85.1**	**83.7**
MULTIPLE-VEHICLE CRASHES						
Angle collision	10.0	7.2	8.5	1.3	1.1	1.2
Head-on collision	3.4	0.3	1.6	3.9	1.7	2.6
Rear-end collision	16.4	12.2	14.2	10.6	8.2	9.2
Sideswipe collision	3.8	3.8	3.7	1.0	1.3	1.2
Other multiple-vehicle collision	2.6	3.0	2.7	1.6	2.6	2.2
Total multiple-vehicle crashes	**36.2**	**26.5**	**30.7**	**18.4**	**14.9**	**16.3**
TOTAL CRASHES	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**

LCFs Based on HSM Default Crash Proportions

Using HSM-provided crash severity and type proportions, predicted crashes were computed and LCFs were calculated for each scenario. The results are summarized in Table 32. Please note that there were only 19 R4U roadway segments in the study database; thus, all of them were used for calculation. Due to a small sample size, the result may not be a correct reflection of the reality.

Except for R4U, all rural roadway segments (R2U and R4D) have LCFs less than 1.0. Depending on scenarios, LCFs of R2U range from 0.6997 to 0.8716 and LCFs of R4D range 0.5639 and

0.6139. For urban and suburban arterials, most LCFs are close to or over 1.0. It seems the LCFs of U5T seem to be highest in general, whose LCFs ranging from 0.9305 to 1.2482. On the other hand the, LCFs of U4D are generally lower (from 0.8783 to 0.9518).

Considering all roadway types together (except for R4U), LCFs for Maryland roads are less than 1.0 in 16 out of 28 cases. Also, there are no extremely high LCFs. Without conducting an in-depth, comparative study, it is not possible to draw any firm conclusion. However, several assertions can be made. First, drivers in Maryland drive more carefully than drivers in other states, and/or Maryland roadways are built safer than HSM-default data providers'. Second, crash severity and type proportions of HSM may be different from Maryland; thus, for a more reliable result, Maryland-provided crash proportion data may need to be used.

In the case of intersection LCFs (Table 33), all values are less than 1.0. RM4SG has the lowest LCF at 0.1086 and U4SG has the highest LCF at 0.4832. These values are a lot lower than those of segment LCFs (Table 32). Does this mean that intersections in Maryland are particularly safer than those in other states? This should be answered with a more in-depth analysis. One possible answer could be that the exclusion of the City of Baltimore may play a role. With a population around 620,000, the City of Baltimore is one of the 30 most populous cities in the Unites States (U.S. Census Bureau 2012). Like other populous cities, Baltimore's roadway network consists of lots of short blocks and intersections where many conflicts among vehicles and people occur. If crash data and roadway data from the City of Baltimore were included, LCF values would have been higher than the current values. Please note that low LCFs do not mean a low reliability of the computed LCFs.

Table 32. LCFs Based on HSM-Default Crash Distribution – Roadway Segments

Scenario	Facility	R2U	R4U*	R4D	U2U	U3T	U4U	U4D	U5T
1	Samples	190	19	160	235	140	75	150	30
	Observed Crashes	341	43	339	361	349	369	451	357
	Predicted Crashes	444	18	580	331	330	354	512	286
	Calibration Factor	0.7675	2.3408	0.5844	1.0905	1.0574	1.0416	0.8801	1.2482
2	Samples	251	N.A.	160	252	138	145	244	115
	Observed Crashes	458	N.A.	315	360	330	592	654	1257
	Predicted Crashes	655	N.A.	538	498	301	591	785	1127
	Calibration Factor	0.6997	N.A.	0.5853	0.7234	1.0960	1.0015	0.8336	1.1154
3	Samples	160	N.A.	185	205	125	75	135	30
	Observed Crashes	328	N.A.	336	377	350	442	404	318
	Predicted Crashes	376	N.A.	596	340	361	410	424	342
	Calibration Factor	0.8716	N.A.	0.5639	1.1088	0.9706	1.0778	0.9518	0.9305
4	Samples	577	N.A.	351	653	259	230	502	139
	Observed Crashes	1058	N.A.	618	1014	617	1039	1402	1370
	Predicted Crashes	1387	N.A.	1007	1079	602	943	1596	1291
	Calibration Factor	0.7626	N.A.	0.6139	0.9399	1.0247	1.1015	0.8783	1.0609

Note 1: There were only 19 R4U segments in the final data set. Thus, all of them were included in the study.
Note 2: The asterisk denotes that the facility did not meet HSM minimum sample size criteria of 30-50 sites or the minimum annual crash threshold of 100.

Table 33. LCFs Based on HSM-Default Crash Distribution – Intersections

Scenario	Facility	R23ST*	R24ST*	R24SG*	RM3ST*	RM4ST*	RM4SG*	U3ST*	U4ST*	U3SG	U4SG
1	Samples	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	80	40
	Observed Crashes	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	380	354
	Predicted Crashes	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	880	738
	Calibration Factor	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	0.4320	0.4798
2	Samples	162	115	67	26	10	35	152	90	167	244
	Observed Crashes	103	142	262	36	30	231	103	173	789	1763
	Predicted Crashes	610	672	995	201	82	2128	641	407	1921	3649
	Calibration Factor	0.1688	0.2113	0.2634	0.1788	0.3667	0.1086	0.1607	0.4249	0.4107	0.4832

Note: The asterisk denotes that the facility did not meet HSM minimum sample size criteria of 30-50 sites or the minimum annual crash threshold of 100.

LCFs Based on Maryland Crash Distribution

Using the same procedures used in the previous sections, LCFs were computed, this time, using Maryland crash severity and type proportions data. The HSM also suggest using the locally-driven distribution when it is available. Replacing HSM's default values with locally-derived values may improve the reliability of Part C predictive models (American Association of State Highway and Transportation Officials 2010, A-10). The results are shown in Table 34 and Table 35. The LCFs are a bit different from HSM-based ones, but the differences do not seem to be large.

Table 34. LCFs Based on the Maryland-Specific Crash Distribution – Roadway Segments

Scenario	Facility	R2U	R4U*	R4D	U2U	U3T	U4U	U4D	U5T
1	Samples	190	19	160	235	140	75	150	30
	Observed Crashes	341	43	339	361	349	369	451	357
	Predicted Crashes	447	18	581	347	337	395	515	262
	Calibration Factor	0.7625	2.3408	0.5833	1.0415	1.0358	0.9352	0.8753	1.3605
2	Samples	251	N.A.	160	252	138	145	244	115
	Observed Crashes	458	N.A.	315	360	330	592	654	1257
	Predicted Crashes	658	N.A.	540	528	306	674	791	1057
	Calibration Factor	0.6956	N.A.	0.5838	0.6814	1.0785	0.8788	0.8269	1.1891
3	Samples	160	N.A.	185	205	125	75	135	30
	Observed Crashes	328	N.A.	336	377	350	442	404	318
	Predicted Crashes	379	N.A.	598	352	366	465	427	324
	Calibration Factor	0.8665	N.A.	0.5623	1.0699	0.9567	0.9505	0.9453	0.9806
4	Samples	577	N.A.	351	653	259	230	502	139
	Observed Crashes	1058	N.A.	618	1014	617	1039	1402	1370
	Predicted Crashes	1396	N.A.	1010	1131	612	1080	1608	1206
	Calibration Factor	0.7579	N.A.	0.6119	0.8963	1.0090	0.9622	0.8722	1.1357

Note 1: There were only 19 R4U segments in the final data set. Thus, all of them were included in the study.
Note 2: The asterisk denotes that the facility did not meet HSM minimum sample size criteria of 30-50 sites or the minimum annual crash threshold of 100.

Table 35. LCFs Based on the Maryland-Specific Crash Distribution – Intersections

Scenario	Facility	R23ST*	R24ST*	R24SG*	RM3ST*	RM4ST*	RM4SG*	U3ST*	U4ST*	U3SG	U4SG
1	Samples	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	80	40
	Observed Crashes	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	380	354
	Predicted Crashes	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	908	745
	Calibration Factor	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	0.4184	0.4752
2	Samples	162	115	67	26	10	35	152	90	167	244
	Observed Crashes	103	142	262	36	30	231	103	173	789	1763
	Predicted Crashes	626	706	995	201	82	2128	659	452	1981	3687
	Calibration Factor	0.1645	0.2011	0.2634	0.1788	0.3667	0.1086	0.1562	0.3824	0.3982	0.4782

Note: The asterisk denotes that the facility did not meet HSM minimum sample size criteria of 30-50 sites or the minimum annual crash threshold of 100.

A visual representation of the data often helps researchers to understand the data. To identify differences between HSM default and Maryland data based models, predicted crashes were compared (Figure 14 and Figure 15). For all six graphs, the red color bars represent predicted crashes based on the Maryland data, and blue color bars represent the predicted crashes computed using HSM-default proportions. Almost all roadway facilities (except U5T) had higher predicted crashes when they were computed using Maryland-specific crash distribution than crashes predicted using HSM-default crash distribution. The same is true for intersections. All intersection facilities had higher predicted crashes when Maryland-specific crash distribution was used.

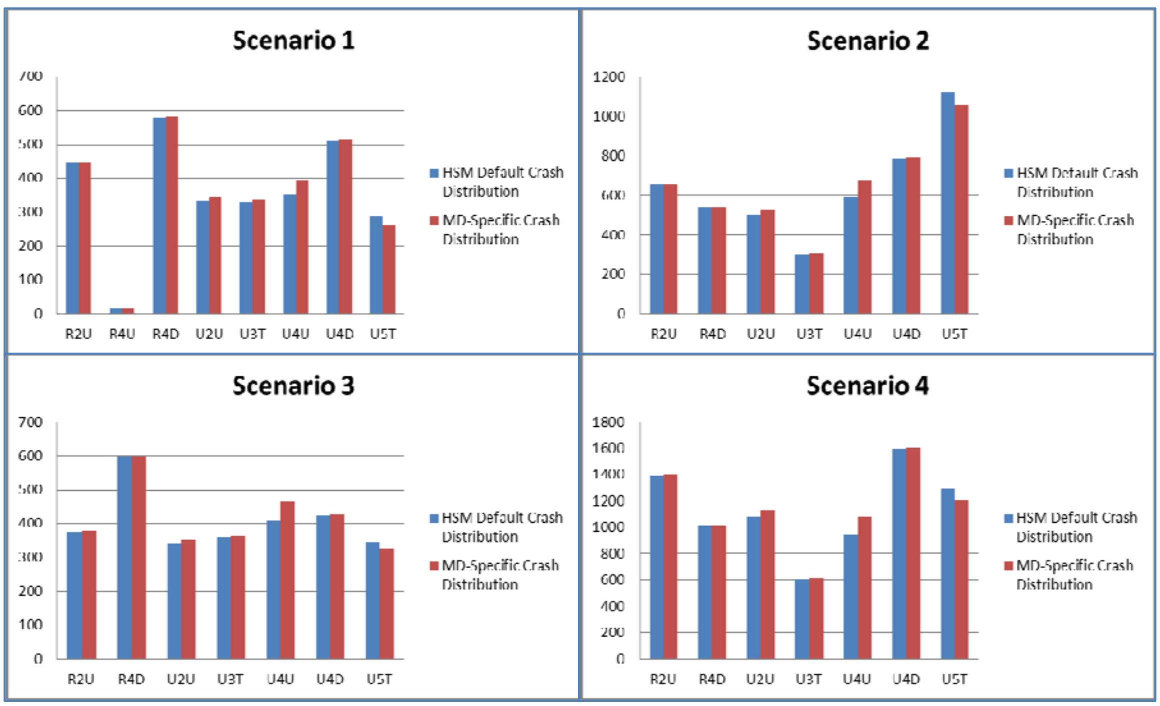


Figure 14. Comparison of Predicted Crashes Based on HSM-Default and Maryland-Specific Crash Distribution – Roadway Segments

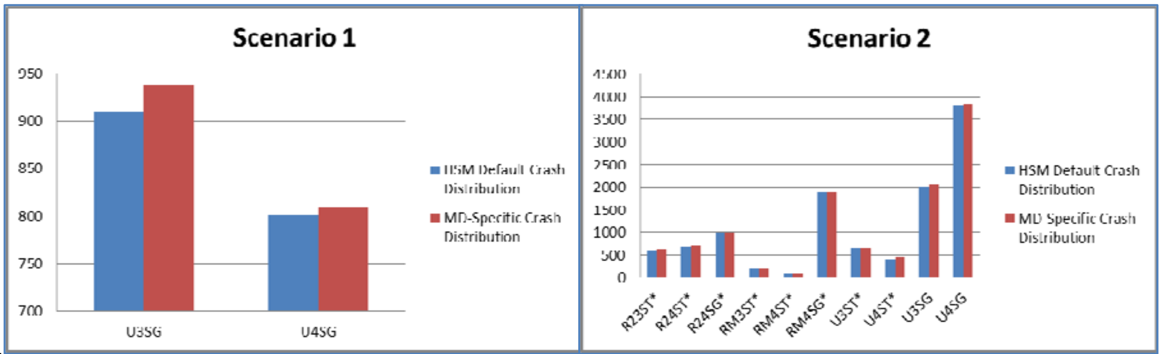


Figure 15. Comparison of Predicted Crashes Based on HSM-Default and Maryland-Specific Crash Distribution – Intersections

To determine whether differences were statistically significant, a two-sample t-test on predicted crashes was conducted (Table 36). The same procedure was employed in the Oregon case study (Dixon, et al. 2012). The Oregon study concluded that the differences between HSM and Oregon data were not statistically significant. The study suggested that "if agencies do not have available data to generate locally-derived values, they can confidently use default values" (Dixon, et al. 2012).

However, for the Maryland situation, P-value is statistically significant at 95% CL for the one-tail hypothesis test. It is statistically significant that predicted crashes using the Maryland distribution were larger than HSM data based prediction. For the two-tail test, the P-value was slightly larger than 0.05. In theory, it could be said that there is no statistical difference at 95% CL. However, the P-value was at the borderline. Considering the mixed t-test result and visual inspection of predicted crash comparison, the study team suggests that Maryland-driven crash data should be used for reliable results.

Table 36. T-Test Results for HSM-Default and Maryland-Specific Crash Distribution

Facility type	HSM	Maryland	Difference	Results
	Predicted Total Crashes	Predicted Total Crashes		
R2U	655	658	4	
R4D	538	540	1	
U2U	498	528	31	
U3T	301	306	5	
U4U	591	674	83	P-value (one-tail):
U4D	785	791	6	0.03026
U5T	1127	1057	70	P-value (two-tail):
R23ST	610	626	16	0.06052
R24ST	672	706	34	T Stat:
U3ST	641	659	19	-2.07152
U4ST	407	452	45	
U3SG	1921	1981	60	
U4SG	3649	3687	38	
		Mean Difference	32	

Comparing Sampling Scenarios

As discussed earlier, the study team developed several sampling scenarios to examine the impacts of segment length and sample size on the resulting LCFs. Now, the question is what scenario provides the best LCFs for Maryland? A clear conclusion may be possible by validating each sampling scenario by collecting additional data on the non-sampled locations and comparing the validation set to the study data set in order to determine which sampling scenario provides the best fit. However, this means an additional sampling task, which is beyond the current study scope. Thus, the study team tried a series of alternative evaluation methods to determine the best sampling scenario. The following section provides detailed discussion of each alternative evaluation method and provides the best scenario chosen for Maryland.

Alternative 1 – Range of Minimum and Maximum LCFs

The first alternative for comparing LCFs was a range of LCF, instead of using one fixed value of LCF for a facility. An example of this alternative is provided in Table 37. Providing a range could be the safest reporting method. However, the study team thought that the range was too large. For example, the range of LCFs for U2U is 0.3885. In other words, the largest LCF is 57% higher than the smallest LCF, which reduced the reliability of LCFs. Another reason is that scenario 3 used a disproportionate stratified sampling technique to include smaller segments that are not included in scenarios 1 and 2. Since populations where samples were drawn were different, using a range of LCFs from all scenarios could be problematic. This alternative was discarded.

Table 37. Summary of LCFs

Scenario		R2U	R4D	U2U	U3T	U4U	U4D	U5T
1	LCF	0.7625	0.5833	1.0415	1.0358	0.9352	0.8753	1.3605
2		0.6956	0.5838	0.6814	1.0785	0.8788	0.8269	1.1891
3		0.8665	0.5623	1.0699	0.9567	0.9505	0.9453	0.9806
4		0.7579	0.6119	0.8963	1.0090	0.9622	0.8722	1.1357
Summary	Min.	0.6956	0.5623	0.6814	0.9567	0.8788	0.8269	0.9806
	Average	0.7706	0.5853	0.9223	1.0200	0.9317	0.8799	1.1665
	Max.	0.8665	0.6119	1.0699	1.0785	0.9622	0.9453	1.3605
	Range	0.1710	0.0497	0.3885	0.1218	0.0834	0.1184	0.3799

Alternative 2 – Absolute Differences of LCFs from 1

The assumption of the second alternative was that the closer LCFs to value 1.0 (i.e., HSM default case), the better the LCFs are. Table 38 shows the results of this alternative for roadway segments. For roadway segments, the LCFs of scenario 3 had the lowest absolute difference from 1.0 and on the other hand, the LCFs of scenario 2 had the highest absolute difference with 1.0. However, this alternative has one big flaw. The value close to 1.0 does not mean the measurement of the LCFs is reliable. In other words, the value 1.0 means the computed LCFs are same as HSM's predicted numbers. The purpose of this study is not to find the LCFs close to HSM default values, so this alternative was discarded, too.

Table 38. Absolute Difference Between 1.0 and LCFs

Scenario		R2U	R4U*	R4D	U2U	U3T	U4U	U4D	U5T	Total
1	Abs. Value of (1-LCF)	0.2375	1.3408	0.4167	0.0415	0.0358	0.0648	0.1247	0.3605	2.6223
2		0.3044	1.3408	0.4162	0.3186	0.0785	0.1212	0.1731	0.1891	2.9420
3		0.1335	1.3408	0.4377	0.0699	0.0433	0.0495	0.0547	0.0194	2.1488
4		0.2421	1.3408	0.3881	0.1037	0.0090	0.0378	0.1278	0.1357	2.3850
5 (Average)		0.2555	1.3408	0.3816	0.0403	0.1084	0.0891	0.0828	0.0671	2.3657

Note: The asterisk denotes that the facility did not meet HSM minimum sample size criteria of 30-50 sites or the minimum annual crash threshold of 100.

Alternative 3 – Lower CV of Normalized Crashes

Another attractive method is the use of coefficient of variations (CV). CV is computed by dividing the standard deviation by the mean, showing a normalized measure of the dispersion of a sample distribution. The lower the CV, the narrower the sample distribution is. Observed crashes were normalized by segment length. It was assumed that the lower CV of normalized crashes mean the sampling distribution has a narrow range and, therefore, the quality of the sample is better than other samples with higher CVs. With this criterion, the gray shades are LCFs with lower CVs in Table 39.

While scenarios 1 and 2 perform well, this alternative was rejected. In theory, a sample with a smaller variation is considered reliable. However, this does not mean that a sample with a lower CV represents the true nature of the population.

Table 39. Lower CVs – Roadway Segments

Scenario	R2U	R4D	U2U	U3T	U4U	U4D	U5T
1	1.40	1.30	1.77	1.31	1.32	1.57	1.28
2	1.20	1.25	1.80	1.14	1.26	1.70	1.49
3	2.45	2.15	2.16	1.85	1.55	1.66	1.43
4	2.14	2.13	1.99	1.64	1.51	1.69	1.51

Alternative 4 – Similar CV of Normalized Crashes to Entire Population

After discarding alternative 3, the team did a further comparison of CVs. The CVs of the normalized crashes in samples were compared to the CVs of normalized crashes of the entire population (depending on scenario used for sampling). It was assumed that the sample CV close to the entire population CV was the best sampling scenario. The result is shown in Table 40. Like alternative 3, scenarios 1 and 2 look better than other scenarios. However, other scenarios also have at least one better LCF. That is, there is no scenario with results superior than the others. For this reason, this alternative was not considered further.

Table 40. Similar CVs – Roadway Segments

Segments		R2U	R4D	U2U	U3T	U4U	U4D	U5T
Population without Segment Length threshold		9519	1410	7215	537	741	5338	276
CV of Population without Segment Length threshold		2.59	2.57	2.48	1.84	1.52	2.18	1.76
Population with Segment Length threshold		3328	297	4123	320	394	2874	211
CV of Population with Segment Length threshold		1.47	1.29	1.85	1.36	1.29	1.99	1.34
Sc. 1	**# Samples**	190	160	235	140	75	150	30
	CV of Samples	1.40	1.30	1.77	1.31	1.317	1.57	1.28
Sc. 2	**# Samples**	251	160	252	138	145	244	115
	CV of Samples	1.20	1.25	1.80	1.14	1.261	1.70	1.49
Sc. 3	**# Samples**	160	185	205	125	75	135	30
	CV of Samples	2.45	2.15	2.16	1.85	1.55	1.66	1.43
Sc. 4	**# Samples**	577	351	653	259	230	502	139
	CV of Samples	2.14	2.13	1.99	1.64	1.51	1.69	1.51

Best Scenario - Conclusion

Even after evaluating four alternative methods, it is not clear what would be the best scenario. Without conducting a full-scale data collection on non-sampled sites and calculating LCFs for the whole population, it is difficult to choose one over another.

After a series of team discussions and consultation with the SHA technical contact, a conclusion was made that scenario 2 would provide the best LCFs. First, in general the larger the sample size, the more reliable the representation of the population is. As explained earlier, the sample size was increased in scenario 2 by applying 90% confidence level. Second, without a firm conclusion on the best case scenario, it would be safe to follow HSM's guidance. For scenario 2, the team followed HSM's minimum segment length guidance and minimum annual crash criteria. Third, a sample with 90% confidence level provides defendable statistical evidence. Except for scenario 4 (disproportionate stratified sample), other samples are not backed by statistical evidence.

More thoughts about Sampling

While scenario 2 is suggested as the best sampling strategy for this study, sampling method, sample size, minimum annual crash requirement, and segment length need to be closely examined if time and resources are available. Especially, researchers may need a more clear guidance of segment length. Even though the quality of LCFs based on disproportionate stratified sampling could not be evaluated, this sampling methodology could identify a more precise segment length threshold. Carefully creating each stratum will help researchers to find a minimum segment length threshold that can be excluded from the study data without affecting the reliability of final LCFs.

RESEARCH FINDINGS

For the simplicity of the discussion, LCFs based on total crashes was discussed. In fact, HSM provides more detailed methodologies (SPFs) to calculate LCFs by different crash severity categories. Using them, four types of LCFs can be computed: (1) LCFs for total crashes, (2) LCFs for KABC crashes, (3) LCFs for KAB crashes, and (4) LCFs for PDO crashes. The following sections provide a summary of each LCF type.

All LCFs tables provided in this chapter are based on the chosen best sampling scenario (Scenario 2) and Maryland-specific crash data. In addition, the results are reported with a range of LCFs using 90% CL. This means if the same size of sample is drawn repeatedly, 90% of these samples will include the true population value (in this case true LCFs), which will lie within the confidence interval. The confidence interval is computed using the Equation 10 (Diez, Barr and Cetinkaya-Rundel 2013, 165-172).

Equation 10. Calculation of Confidence Interval for LCF

$$C.I. = LCF \pm Z \times SEM$$

Where:

LCF = Local Calibration Factor computed in scenario 2,
Z = Standard normal variable (of 90% confidence level is 1.645), and
SEM = Standard error of the mean and is calculated based on Equation 11.

Equation 11. Calculation of SEM

$$SEM = {}^{S}/_{\sqrt{n}}$$

Where:

S = Sample standard deviation, and
n = Size of the sample.

LCFs for Total Crashes

Table 41 and Table 42 show the final LCFs for total crashes on segments and intersections, respectively. In general ranges of LCFs are small, except for a couple of facility types with smaller sample size due to insufficient population. For example, ranges for R4U, RM3ST, RM4ST, and RM4SG are wider than other facilities. They are the facilities that did not have large enough populations.

Interpreting the LCF and the confidence interval is rather straightforward. For example, the LCF for R2U is 0.6956 and the range of the confidence interval is 0.6888 to 0.7023. With the 90% CL, a repeated sampling of the same size for R2U, the true population LCF value would be between 0.6888 and 0.7023. Other values in the following tables can be interpreted same way.

Table 41. Final Maryland LCFs for Total Crashes – Roadway Segments

Facility		R2U	R4U*	R4D	U2U	U3T	U4U	U4D	U5T
Samples		251	19	160	252	138	145	244	115
Observed Crashes		458	43	315	360	330	592	654	1257
Predicted Crashes		658	18	540	528	306	674	791	1057
Calibration Factor		0.6956	2.3408	0.5838	0.6814	1.0785	0.8788	0.8269	1.1891
Calibration Factor Range	Minimum	0.6888	1.9375	0.5753	0.6674	1.0558	0.8606	0.8144	1.1673
	Maximum	0.7023	2.7441	0.5923	0.6954	1.1013	0.8969	0.8393	1.2109

Note 1: There were only 19 R4U segments in the final data set. Thus, all of them were included in the study.
Note 2: The asterisk denotes that the facility did not meet HSM minimum sample size criteria of 30-50 sites or the minimum annual crash threshold of 100.

Table 42. Final Maryland LCFs for Total Crashes – Intersections

Facility		R23ST*	R24ST*	R24SG*	RM3ST*	RM4ST*	RM4SG*	U3ST*	U4ST*	U3SG	U4SG
Samples		162	115	67	26	10	35	152	90	167	244
Observed Crashes		103	142	262	36	30	231	103	173	789	1763
Predicted Crashes		626	706	995	201	82	2128	659	452	1981	3687
Calibration Factor		0.1645	0.2011	0.2634	0.1788	0.3667	0.1086	0.1562	0.3824	0.3982	0.4782
Calibration Factor Range	Minimum	0.1611	0.1948	0.2581	0.1569	0.3124	0.1046	0.1502	0.3678	0.3915	0.4747
	Maximum	0.1678	0.2074	0.2688	0.2007	0.4211	0.1125	0.1622	0.3969	0.4049	0.4817

Note: The asterisk denotes that the facility did not meet HSM minimum sample size criteria of 30-50 sites or the minimum annual crash threshold of 100.

LCFs for KABC (Fatal + Injury) Crashes vs. Total Crashes

Another way to compute LCFs is based on the number of fatal (K) and injury crashes (i.e., A – incapacitating injury, B – non incapacitating injury, C – possible injury), excluding property dame only crashes. If the purpose of the research is the prediction of fatal and injury crashes, these LCFs will become useful. However, HSM does not provide methods for KABC crashes on R2U, R23ST, R24ST, and R24SG.

Table 43 and Table 44 show LCFs for KABC crashes as well as LCFs for total crashes. LCFs for three segment types (R4U, U3T and U4D) and all intersections types are higher than those for total crashes. This may imply that the proportion of fatal and injury crashes are higher in Maryland.

Table 43. Comparison of LCFs: Total vs. KABC Crashes – Roadway Segments

Roadway Segments		R4U*	R4D	U2U	U3T	U4U	U4D	U5T
LCF for Total Crash		2.3408	0.5838	0.6814	1.0785	0.8788	0.8269	1.1891
LCF for KABC		1.9499	0.4193	0.6125	1.3053	0.7696	1.0665	1.1918
Range for KABC	Minimum	1.5418	0.4100	0.5953	1.2778	0.7531	1.0479	1.1697
	Maximum	2.3579	0.4287	0.6298	1.3327	0.7861	1.0852	1.2140

Note: The asterisk denotes that the facility did not meet HSM minimum sample size criteria of 30-50 sites or the minimum annual crash threshold of 100.

Table 44. Comparison of LCFs: Total vs. KABC Crashes – Intersections

Intersections		RM3ST*	RM4ST*	RM4SG*	U3ST*	U4ST*	U3SG	U4SG
LCF for Total Crash		0.1788	0.3667	0.1086	0.1562	0.3824	0.3982	0.4782
LCF for KABC		0.2550	0.3923	0.1327	0.2273	0.4964	0.5967	0.6285
Range for KABC	Minimum	0.2164	0.3139	0.1279	0.2204	0.4806	0.5899	0.6247
	Maximum	0.2935	0.4707	0.1374	0.2343	0.5121	0.6035	0.6324

Note: The asterisk denotes that the facility did not meet HSM minimum sample size criteria of 30-50 sites or the minimum annual crash threshold of 100.

LCFs for KAB Crashes vs. other LCFs

LCFs are also computed using proportions of only three types of crash severities (KAB). Like LCFs for KABC crashes, not all facilities have LCFs for KAB only crashes. Table 45 and Table 46 show LCFs for KAB crashes and their comparison to the previous two LCFs. While KAB LCFs for both R4U and R4D are significantly lower than LCFs for total crash, differences between KABC and KAB LCFs are marginal for both facilities. In the case of intersections, LCFs for KAB crashes are significantly higher in all three intersection types, implying that the proportion of severe crashes may be higher in those intersection types. A more in-depth analysis on rural multilane intersections may be necessary.

Table 45. Comparison of LCFs: Total vs. KABC vs. KAB Crashes – Roadway Segments

Roiadway Segments		R4U*	R4D
LCF for Total Crash		2.3408	0.5838
LCF for KABC		1.9499	0.4193
LCF for KAB		1.9231	0.4565
Range for KAB LCF	Minimum	1.4299	0.4444
	Maximum	2.4163	0.4685

Note: The asterisk denotes that the facility did not meet HSM minimum sample size criteria of 30-50 sites or the minimum annual crash threshold of 100.

Table 46. Comparison of LCFs: Total vs. KABC vs. KAB Crashes – Intersections

Intersections		RM3ST*	RM4ST*	RM4SG*
LCF for Total Crash		0.1788	0.3667	0.1086
LCF for KABC		0.2550	0.3923	0.1327
LCF for KAB		0.2664	0.3953	0.1879
Range for KAB LCF	Minimum	0.2272	0.3294	0.1784
	Maximum	0.3056	0.4611	0.1973

Note: The asterisk denotes that the facility did not meet HSM minimum sample size criteria of 30-50 sites or the minimum annual crash threshold of 100.

LCFs for PDO Crashes vs. other LCFs

The last type of LCF is for PDO (property damage only) crashes. The comparison tables are provided below. In the case of intersections, LCFs for PDO crashes are lower than total crash-based LCFs in all types. However, the result is mixed in the case of segments.

Table 47. Comparison of LCFs: Total vs. KABC vs. KAB vs. PDO Crashes – Roadway Segments

Roadway Segments		R4U*	R4D	U2U	U3T	U4U	U4D	U5T
LCF for Total Crash		2.3408	0.5838	0.6814	1.0785	0.8788	0.8269	1.1891
LCF for KABC		1.949861	0.4193	0.6125	1.3053	0.7696	1.0665	1.1918
LCF for KAB		1.9231	0.4565	N.A.	N.A.	N.A.	N.A.	N.A.
LCF for PDO		N.A.	N.A.	0.7313	0.9362	0.9611	0.7310	1.1874
Range for PDO LCF	Minimum	N.A.	N.A.	0.7200	0.9199	0.9403	0.7214	1.1708
	Maximum	N.A.	N.A.	0.7427	0.9526	0.9819	0.7406	1.2039

Note: The asterisk denotes that the facility did not meet HSM minimum sample size criteria of 30-50 sites or the minimum annual crash threshold of 100.

Table 48. Comparison of LCFs: Total vs. KABC vs. KAB vs. PDO Crashes – Intersections

Intersections		RM3ST*	RM4ST*	RM4SG*	U3ST*	U4ST*	U3SG	U4SG
LCF for Total Crash		0.1788	0.3667	0.1086	0.1562	0.3824	0.3982	0.4782
LCF for KABC		0.2550	0.3923	0.1327	0.2273	0.4964	0.5967	0.6285
LCF for KAB		0.2664	0.3953	0.1879	N.A.	N.A.	N.A.	N.A.
LCF for PDO		N.A.	N.A.	N.A.	0.1138	0.3003	0.3427	0.3970
Range for PDO	Minimum	N.A.	N.A.	N.A.	0.1092	0.2898	0.3378	0.3946
	Maximum	N.A.	N.A.	N.A.	0.1183	0.3108	0.3475	0.3994

Note: The asterisk denotes that the facility did not meet HSM minimum sample size criteria of 30-50 sites or the minimum annual crash threshold of 100.

CONCLUSIONS

This study computed Maryland-specific local calibration factors to apply HSM's predictive models. This chapter provides final thoughts about interpreting LCFs, challenges that the team faced, and plans for future research to improve safety prediction methods as well as HSM.

Maryland Local Calibration Factors

LCFs for all 18 facility types have been calculated. Roadway network and crash data sets of state-owned and maintained roadways for the study period of 2008-2010 were used. After collection and compilation, samples were drawn and additional required/desirable data gathered for sampled sites. To address some unclear parts of HSM's sampling procedure, different sampling scenarios were developed. At last, the IHSDM was used to compute After the comparison of HSM default crash proportion and Maryland-specific data, the use Maryland data was suggested. Table 49 and

Table 50 summarize the final LCFs for roadway segments and intersections. Each table provides four LCF types and ranges.

Table 49. Maryland LCFs – Roadway Segments

Segments		R2U	R4U*	R4D	U2U	U3T	U4U	U4D	U5T
Total Crashes		0.6956	2.3408	0.5838	0.6814	1.0785	0.8788	0.8269	1.1891
Range	Minimum	0.6888	1.9375	0.5753	0.6674	1.0558	0.8606	0.8144	1.1673
	Maximum	0.7023	2.7441	0.5923	0.6954	1.1013	0.8969	0.8393	1.2109
KABC Crashes		N.A.	1.9499	0.4193	0.6125	1.3053	0.7696	1.0665	1.1918
Range	Minimum	N.A.	1.5418	0.4100	0.5953	1.2778	0.7531	1.0479	1.1697
	Maximum	N.A.	2.3579	0.4287	0.6298	1.3327	0.7861	1.0852	1.2140
KAB Crashes		N.A.	1.9231	0.4565	N.A.	N.A.	N.A.	N.A.	N.A.
Range	Minimum	N.A.	1.4299	0.4444	N.A.	N.A.	N.A.	N.A.	N.A.
	Maximum	N.A.	2.4163	0.4685	N.A.	N.A.	N.A.	N.A.	N.A.
PDO Crashes		N.A.	N.A.	N.A.	0.7313	0.9362	0.9611	0.7310	1.1874
Range	Minimum	N.A.	N.A.	N.A.	0.7200	0.9199	0.9403	0.7214	1.1708
	Maximum	N.A.	N.A.	N.A.	0.7427	0.9526	0.9819	0.7406	1.2039

Note 1: There were only 19 R4U segments in the final data set. Thus, all of them were included in the study.
Note 2: The asterisk denotes that the facility did not meet HSM minimum sample size criteria of 30-50 sites or the minimum annual crash threshold of 100.

Table 50. Maryland LCFs – Intersections

Intersections		R23ST*	R24ST*	R24SG*	RM3ST*	RM4ST*	RM4SG*	U3ST*	U4ST*	U3SG	U4SG
Total Crashes		0.1645	0.2011	0.2634	0.1788	0.3667	0.1086	0.1562	0.3824	0.3982	0.4782
Range	Minimum	0.1611	0.1948	0.2581	0.1569	0.3124	0.1046	0.1502	0.3678	0.3912	0.4747
	Maximum	0.1678	0.2074	0.2688	0.2007	0.4211	0.1125	0.1622	0.3969	0.4052	0.4817
KABC Crashes		N.A.	N.A.	N.A.	0.2550	0.3923	0.1327	0.2273	0.4964	0.5967	0.6285
Range	Minimum	N.A.	N.A.	N.A.	0.2164	0.3139	0.1279	0.2204	0.4806	0.5899	0.6247
	Maximum	N.A.	N.A.	N.A.	0.2935	0.4707	0.1374	0.2343	0.5121	0.6035	0.6324
KAB Crashes		N.A.	N.A.	N.A.	0.2664	0.3953	0.1879	N.A.	N.A.	N.A.	N.A.
Range	Minimum	N.A.	N.A.	N.A.	0.2272	0.3294	0.1784	N.A.	N.A.	N.A.	N.A.
	Maximum	N.A.	N.A.	N.A.	0.3056	0.4611	0.1973	N.A.	N.A.	N.A.	N.A.
PDO Crashes		N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	0.1138	0.3003	0.3427	0.3970
Range	Minimum	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	0.1092	0.2898	0.3378	0.3946
	Maximum	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	0.1183	0.3108	0.3475	0.3994

Note 1: There were only 19 R4U segments in the final data set. Thus, all of them were included in the study.
Note 2: The asterisk denotes that the facility did not meet HSM minimum sample size criteria of 30-50 sites or the minimum annual crash threshold of 100.

Interpretation of LCFs

While the computation and the LCF concept are simple, practitioners should be careful in interpreting LCFs. LCFs do not indicate good or bad about the level of safety of a certain facility. The primary purpose of computing LCFs is to adjust predicted average crashes from HSM base SPFs to the local conditions. This is because local traffic, network, population and other factors are different from the states on which HSM SPFs are based. Thus, the developed LCFs only indicate whether crashes on a certain facility are lower or higher than the base model. So, the factors should be multiplied by HSM-generated predicted crash frequency.

In addition, LCFs are the average value of all sampled sites and they may or may not accurately predict site-specific crashes. Rather, the predicted crashes and LCFs represent long-term trends. Crashes are rare events, so the same type of crashes at the same location may take a long time to occur again. However, if one observes historical data for a long time, overall crash trends will be regressed to the mean.

In general, LCFs for all intersections were less than 1.0 and some were even less than 0.5, which means none of the intersection facility types reached 50% of predicted crashes. Also, only 3 out of 18 facilities had LCFs values higher than 1.0 total crashes. While lower LCFs means fewer crashes occurred in Maryland, please be advised that the following limitations exist with the data:

Lack of enough data for sampling **-** In the case of R4U segments and many intersection types, there were not enough data for sampling (as mentioned, some facilities did not meet HSM criteria for sampling). This has probably influenced the final LCFs.

Availability of Baltimore City data **-** While many intersection crashes occur at busy urban centers, Baltimore City was not part of this study. Including the Baltimore City may increase LCFs for intersections.

Self-reporting system for property damage only (PDO) crashes **-** According to Maryland state sources, property damage during a car accident must be reported only when there is an injury or when an involved vehicle needs to be towed. This means lots of minor crashes were not reported. In this sense, where available, the use of LCFs for KABC crashes would be a better reflection of reality.

Different urban population **–** Differences in urban population and distribution between Maryland and the states whose data were used for developing HSM may not be fully reflected. There are many cities with population over 100,000 in Washington (e.g., Seattle [608,660], Spokane [208,916], Tacoma [198,397], and Vancouver [161,791]) and California (e.g., Los Angeles [3,792,621], San Diego [1,301,617], San Jose [945,942], San Francisco [805,235], Fresno [494,665], and Sacramento [466,488]). By contrast, excluding Baltimore City from this study, the most populous city in Maryland is Frederick [65,239], followed by Rockville [61,209] (U.S. Census Bureau 2012). Perhaps the large population difference between urban and suburban areas in Maryland and aforementioned states might cause the lower LCFs values for Maryland urban and suburban intersections.

Challenges: Data collection burden

Throughout the study, there were many challenges that the study team had to overcome. The following section discusses several challenges.

Data Collection Issue

Data collection and compilation consumed most of the study team's time. The HSM procedure is not too difficult, but there could be some room for improvement. On the other hand, the data requirement is too demanding. This study shares data issues similar to other states that have already developed LCFs. That is, the transportation database was not built for HSM adoption. Some data items were not collected, incomplete, or not readily available. Such data items are listed below.

- For roadway segments
    - Curve data (length and radius)
    - Grade or terrain
    - On-street parking data (length and type)
    - Driveway data (density and type)
    - Centerline rumble strip
    - Roadside fixed-object data (density and average distance)
    - Presence of lighting
- For intersections

- AADT for minor roads
- Pedestrian data (volume and maximum number of crossing lanes on any approach)
- Some intersection control and warning data (left-turn phasing, right-turn-on-red operation, red light camera)
- Intersection skew angle(s)
- Left-turn and right-turn lanes
- Presence of alcohol sale establishments within 1000 ft.
- Presence of lighting

For a full adoption of HSM, several strategies may need to be considered. There strategies are interrelated to each other.

First, a new architecture of a centralized data warehouse would become necessary. One of the difficulties of the data collection was to identify divisions responsible for data. Often the authors were referred to different divisions/offices to find out the availability of certain data items. If the data is stored in one central location with appropriate indices, data collection time would be saved in the future.

Second, in relation to the first strategy, it should be made sure that the data sets need to be ready for HSM. As discussed earlier, AADT on minor roadways is one example of an incomplete information that is a “must-have” variable for estimating intersection crashes.

Third, the automation of data generation for HSM could be a good direction for regular updates of LCFs. With the current data, about 50% of the procedures could be automated using ArcGIS ModelBuilder. That is, the generation of the homogeneous segments and crash assignment were generated using ArcGIS ModelBuider tool, while some of the verification process should be done manually. Once SHA variables for HSM are collected and kept electronically, an automation of data generation would become easier. While there is not a clear guide on the frequency of LCF updates, the fully developed data automation would allow safety practitioners/researchers to update LCFs every year (e.g., updates of LCFs using the most recent 3 years of data).

Fourth, to be able to accomplish the first three strategies, there is one thing that needs to be figured out. One of the barriers in data generation was merging crash data sets to segments and intersections. This was because there were not common unique identifiers in both the crash data table and the roadway network table, which can be used to table-to-table merge. So, the task was done by ArcGIS spatial join with manual inspection for error checking. However, due to ArcGIS’s default merge criteria in the spatial join tool, crashes were often assigned to multiple segments that later were manually removed.

Unclear Rationale of HSM about Sampling

There is currently unclear rationale of HSM about sampling. One concern with sampling segments and intersections for the application of HSM is to ensure “what the target variable is.” In other words, researchers should remember for what samples are drawn. According to HSM, it seems that the number of observed crashes is the primary important threshold value in addition to the 30-50 minimum sample size. While such a notion is not totally incorrect, to the practitioners who develop LCFs for their local jurisdictions, what is important is how to draw samples in a way to minimize errors of the developed LCF. A theoretical discussion is well beyond the current scope of the study. However, it is worth pursuing later to refine the current HSM methodology.

Another concern about the sample is segment length. While HSM recommends not using a shorter segment length, doing that may result in the loss of information and the reliability of predicted crashes. It should be clearly addressed in HSM how to appropriately determine the segment length threshold in order not to lose too much information while producing reliable LCFs.

Future Research

The experience from this study generated several interesting study topics that would be helpful for day-to-day operations of transportation agencies as well as for methodological improvement of safety analysis.

First, from the state transportation agencies' perspective, there is one missing puzzle in HSM: SPFs for interstate highways. It is expected that the newer edition of HSM will include models for interstate highways. A further study including interstate highways will provide a complete LCF table for Maryland.

Second, a defendable sampling design method should be discussed. The study team has tried to come up with different scenarios. Unfortunately, a thorough analysis was not possible due to resource limitation. While HSM provides a minimum sample size requirement, this requirement does not provide any rationale for computing a reliable range of LCFs for the segments. Using a state-of-the-art statistical technique, a more in-depth analysis of crash distribution with additional data collection needs to be conducted to develop a more reliable and transferable sampling method. Some modification of the sampling technique by employing finite population correction factor would provide a starting point to address sampling frame issues.

Third, related to the sampling issue, there is not a clear rationale for using "long-enough" segment as discussed in the previous section. A more specific methodological guidance should be developed. The study team's attempt to use a disproportionate stratified sampling is a good starting point to identify the influence of small segments to LCFs.

Fourth, sub-region-specific LCFs will be helpful. Climate, rain history, population, and other factors vary even within one state, so political boundaries may not be the only factor for developing LCFs. Considering different homogeneous zones in Maryland based on at least climate and population will enhance the accuracy of LCFs.

Fifth, the current HSM procedures can be applied to estimate costs of different types of crashes by facility. The inclusion or exclusion of CMFs used in SPFs will show changes in crash frequencies, which can be translated into the costs and benefits of different types of engineering safety countermeasures.

Appendix A List of Abbreviations

Table 51. List of Abbreviations

Abbreviation	Description
AADT	Annual Average Daily Traffic
AASHTO	American Association of State Highway and Transportation Officials
CI	Confidence Interval
CL	Confidence Level
CMF	Crash Modification Factor
FHWA	Federal Highway Administration
GIS	Geographic Information System
HSM	Highway Safety Manual
IHSDM	Interactive Highway Safety Design Model
KML	Keyhole Markup Language
LCF	Local Calibration Factor
MDOT	Maryland Department of Transportation
MP	Mile point
OOTS	Office of Traffic and Safety (of SHA)
PDO	Property Damage Only
R23ST	Rural Two-Lane, Two-Way Road with Un-signalized Three-leg Intersection (Stop Control on Minor-road Approaches)
R24SG	Rural Two-Lane, Two-Way Road with Signalized Four-leg Intersection
R24ST	Rural Two-Lane, Two-Way Road with Un-signalized Four-leg Intersection (Stop Control on Minor-road Approaches)
R2U	Undivided Rural Two-lane, Two-way Roadway Segments
R4D	Rural Four-lane Divided Segments
R4U	Rural Four-Lane Undivided Segments
RM3ST	Rural Multilane Highway with Un-signalized Three-leg Intersection (Stop Control on Minor-road Approaches)
RM4SG	Rural Multilane Highway with Signalized Four-leg Intersection
RM4ST	Rural Multilane Highway with Un-signalized Four-leg Intersection (Stop Control on Minor-road Approaches)
SHA	Maryland State Highway Administration
SPF	Safety Performance Function
TWLTL	Two-Way Left-Turn Lane
U2U	Two-lane Undivided Urban and Suburban Arterial Segments
U3SG	Urban and Suburban Arterial with Signalized Three-leg Intersection
U3ST	Urban and Suburban Arterial with Un-signalized Three-leg Intersection (Stop Control on Minor-road Approaches)
U3T	Three-lane Urban and Suburban Arterials including a Center TWLTL
U4D	Four-lane Divided Urban and Suburban Arterials (i.e., Including a Raised or Depressed Median)
U4SG	Urban and Suburban Arterial with Signalized four-leg intersection
U4ST	Un-signalized four-leg intersection (stop control on minor-road approaches)
U4U	Four-lane undivided arterials
U5T	Five-lane arterials including a center TWLTL
XML	Extensible Markup Language

Appendix B Summary of Selective Case Studies

Table 52. Summary of Selected Case Studies

State	Year of Publication	Study Period	Calibrated Facilities	Challenges	Solutions
FL	2011	2005-2008	All but R4U, RM3ST, RM4ST, U3ST and U4ST.	Creating homogenous segments	In Florida, collected variables are kept by different segmentation criteria. Thus, using GIS and a Python script language, homogenous roadway segments were created.
				Driveway density and roadside hazard rating for rural segments and driveway density and roadside fixed-object for urban segments were not available.	The study team made assumptions for the missing variables based on HSM default assumptions.
				Unavailability of crash data and AADT for non-state roads [required for intersections SPFs]	Only intersections of two state roads were retained for analysis because AADT for other types of roads was not available.
				Finding locally derived crash type and severity distributions	Due to the similarity of many Florida segments to the base conditions of developing SPFs, LCFs with and without state-specific collision type distributions were similar.
IA	2011	2005-2009	None[7]	Data collection issues - Using some common and some unique solutions, such as As-Built Plans, Aerial Imagery and LiDAR from the Photogrammetry Section within the Iowa DOT's Office of Design, MicroStation, and Geopak, were used in conjunction with the as-built plans, aerial imagery, and LiDAR to generate an approximate alignment and profile for the corridor. Microsoft Excel was used to calculate crash frequencies, crash rates, AADT, etc. Spreadsheet from NCHRP 17-38.	
ID	2012	2008	None[8]	Limitation of standard ITD (Idaho Transportation Department) databases, which contained only data on lane widths and shoulder width and type	Additional data was collected by using ITD video logs, aerial photos, and field visits for other required data.
				Information on horizontal curves, such as curve length and radius, and also for vertical curvature on grades	The study team developed a text report that could be generated using the video log software for gathering the required data.
				Information for the intersection CMFs	Required data were obtained by using a combination of aerial photos and the video logs.
				Intersection skew angle	The study team estimated the skew angles based on aerial photos.

[7] In the Iowa case study, only predicted crashes were estimated and no LCFs were developed.
[8] In the Idaho case study, only predicted crashes were estimated and no LCFs were developed.

State	Year of Publication	Study Period	Calibrated Facilities	Challenges	Solutions
				The presence/absence of turn lanes on the major road approaches of the intersections and the existence of intersection lighting	Available ITD video logs were used to address this issue.
LA	2010	2003-2007	Only rural multilane highways	The current highway database does not have information on lighting and automated speed enforcement.	The lack of data led to the assumption of no lighting and the assumption that automated speed enforcement had not been implemented in Louisiana in recent years.
MN	2010	2004-2008	None[9]	Fixed-object density and offset distance, driveway information, and information related to the presence of schools and alcohol sales establishments in the vicinity of signalized intersections	The project team collected data primarily by field measurements, scaled aerials, Google StreetView, and concept plans.
NC	2012	2007-2009	All but R2U, R4U, RM3ST and RM4ST.	Calibration factors had already been developed in a recent NCDOT project for R2U, RM3ST, and RM4ST, so the study team did not calibrate them in this study. They also did not calibrate R4U because of insufficient mileage in the state to prioritize this facility type.	
				Some inconsistencies between HSIS (Highway Safety Information System) indication of facility type and "real" facility type for segments	This issue was solved by double checking and making confirmation.
				Some segments in HSIS encompassed two or more non-homogenous sections.	The problem was addressed by redefinition of the beginning or ending milepost of a segment.
OR	2012	2004-2006	All 18 facilities	Unreliable database for driveway density, the presence of centerline rumble strips, the presence of TWLTL, the roadside hazard rating, the side slope and lighting status	The solution was using ODOT Digital Video Log, aerial photos, and Google Earth.
				Signal phasing information on the minor approaches for Oregon highways	The study team made an assumption that if the major street had protected or permissive phasing and the minor street had dedicated left-turn lanes, the same signal phasing existed on the minor approach.
				Information about right-turn-on-red restriction	Data were found via ODOT DVL or Google StreetView.
				AADT on minor roads	The study team developed an AADT estimation model to estimate $AADT_{minor}$ for rural and urban

[9] In the Montana case study, only predicted crashes were estimated and no LCFs were developed.

State	Year of Publication	Study Period	Calibrated Facilities	Challenges	Solutions
					intersections.
				Pedestrian volumes counts at urban intersections	The study team explored the sensitivity of the various HSM defaults in the pedestrian predictive and assumed a medium level of pedestrian activity for all intersections.
				Creating homogeneous segments - For rural segments, the study team determined that one alternative method for compiling homogeneous segments was to take the wide range of segments already available in the ODOT databases and then combine them into homogeneous segments at least 0.1 mile long. In the urban areas, segmentation is required at each intersection. In many of the urban environments, this resulted in sections that were less than 0.1 mile in length so they came up with 0.07 mile threshold for urban and suburban roadway segments.	
				Finding locally-derived crash type and severity distributions	The study team calculated these values based on Oregon conditions.
GA	2010	2004-2006	Only R2U	Data collection was a common issue for almost all studies and researchers of GA, KS, UT, MI, VA, and Italy addressed them with some common and local solutions.	
KS	2012	2005-2007			
UT	2011				
MI	2012	2005-2010 (yearly)	All roadway segments but U3T and U5T Intersections: U3ST, U4ST, U3SG, and U4SG + some non-HSM facilities		
VA	2010	2003-2007	All roadway segments but U3T & U5T		
Italy	2012	2004-2008	Only R4D[10]		

[10] The researchers developed calibration factor based on Italian conditions for R4D.

Appendix C LCFs of Case Studies and Maryland

Table 53. LCFs of the Case Studies and Maryland

Facility Types		State (Year)									
		MD (2008-2010)	FL (2007-2008)*	LA (2003-2007)**	NC (2007-2009)	OR (2004-2006)	GA (2004-2006)	KS (2005-2007)	UT (2005-2007)	MI (2010)***	Italy (2004-2008)****
Segments	**R2U**	0.6956	1.005	N.A.	N.A.	0.74	0.74	1.48	1.16	1.278	N.A.
	R4U	2.3408	N.A.	1.25	0.97	0.37	N.A.	N.A.	N.A.	1.0503	N.A.
	R4D	0.5838	0.683	N.A.	N.A.	0.77	N.A.	N.A.	N.A.	1.5128	1.26
	U2U	0.6814	1.025	N.A.	1.54	0.62	N.A.	N.A.	N.A.	2.5665	N.A.
	U3T	1.0785	1.038	N.A.	3.62	0.81	N.A.	N.A.	N.A.	N.A.	N.A.
	U4U	0.8788	0.729	N.A.	4.04	0.63	N.A.	N.A.	N.A.	0.7326	N.A.
	U4D	0.8269	1.628	N.A.	3.87	1.411	N.A.	N.A.	N.A.	N.A.	N.A.
	U5T	1.1891	0.669	N.A.	1.72	0.64	N.A.	N.A.	N.A.	1.3017	N.A.
Intersections	**R23ST**	0.1645	0.8	N.A.	0.57	0.31	N.A.	N.A.	N.A.	N.A.	N.A.
	R24ST	0.2011	0.8	N.A.	0.68	0.31	N.A.	N.A.	N.A.	N.A.	N.A.
	R24SG	0.2634	1.21	N.A.	1.04	0.45	N.A.	N.A.	N.A.	N.A.	N.A.
	RM3ST	0.1788	N.A.	N.A.	N.A.	0.15	N.A.	N.A.	N.A.	N.A.	N.A.
	RM4ST	0.3667	N.A.	N.A.	N.A.	0.39	N.A.	N.A.	N.A.	N.A.	N.A.
	RM4SG	0.1086	0.37	N.A.	0.49	0.15	N.A.	N.A.	N.A.	N.A.	N.A.
	U3ST	0.1562	N.A.	N.A.	1.72	0.35	N.A.	N.A.	N.A.	1.3021	N.A.
	U4ST	0.3824	N.A.	N.A.	1.32	0.45	N.A.	N.A.	N.A.	1.4249	N.A.
	U3SG	0.3982	1.41	N.A.	2.47	0.73	N.A.	N.A.	N.A.	1.3459	N.A.
	U4SG	0.4782	1.84	N.A.	2.79	1.05	N.A.	N.A.	N.A.	1.6981	N.A.

Notes:

* FL does not include PDO crashes.

** LA only includes KAB crashes.

*** MI developed LCFs for each year from 2005 to 2010. Here only 2010 LCFs are presented.

**** This is an application in Italy. The study used only KAB crashes.

Appendix D The HSM Data Needs

Table 54. The HSM Data Needs

Facility Type	List of Variables	Required	Desirable	Source	Notes
Roadway Segments	Observed crashes	●		MSP database	2008-2010 crash data
	Area type (rural/suburban/urban)	●		SHA	
	Annual average daily traffic volume	●		SHA	
	Segment length	●		SHA	
	Number of through traffic lanes	●		SHA	
	Lane width	●		SHA	
	Shoulder width	●		SHA	
	Shoulder type	●		SHA	
	Presence of median (divided/undivided)	●		SHA	
	Median width	●		SHA	
	Presence of two-way left-turn lane	●		SHA	
	Low-speed vs. intermediate or high speed	●		SHA	
	Driveway density	●		Manually counted using Google Earth	
	Number of major commercial driveways	●		These data items were manually counted using Google Earth for numbers and commercial, Industrial/institutional, residential and other land uses for type from Maryland Department of Planning [for major/minor distinction, HSM guidelines of 50 parking space threshold was used.)	
	Number of minor commercial driveways	●			
	Number of major residential driveways	●			
	Number of minor residential driveways	●			
	Number of major industrial/institutional driveways	●			
	Number of minor industrial/institutional driveways	●			
	Number of other driveways	●			
	Lengths of horizontal curves and tangents	●		SHA's eGIS does not have required data for all rural two-lane, two-way roadways, so additional data was manually estimated by using the circle measurement tool in Google Earth Pro.	
	Radii of horizontal curves	●			
	Roadside slope (side slope)	●		Manually gathered from eGIS of SHA	

Facility	List of Variables	Required	Desirable	Source	Notes
	% of on-street parking	●		Manually gathered by using length measurement tool in Google Earth	
	Type of on-street parking	●		This data item was manually gathered using Google Earth for type and commercial, Industrial/institutional, residential and other land uses for type from Maryland Department of Planning	
	Presence of lighting	●		Asset Data Warehouse (ADW) of SHA's eGIS has "highway lighting." Final determination was made by manually double-checking Google Earth (StreetView in some cases).	
	% grade		●	HSM default assumption: Base default on terrain, SHA database does not have all required data for all rural two-lane, two-way roadways, so additional data was manually estimated by using elevation profile in Google Earth for majority of samples	
	Terrain (level, rolling and mountainous)		●		
	Roadside fixed-object density		●	Asset Data Warehouse (ADW) of SHA's eGIS has "signs" and "traffic barriers." This data was used as a base for density and then manually double-checked by Google Earth. Length measurement tool of Google Earth used for estimating offset.	
	Roadside fixed-object offset		●		
	Presence of centerline rumble strip		●	Asset Data Warehouse (ADW) of SHA's eGIS has "rumble strips." Final determination was made by manually double-checking Google Earth (StreetView in some cases).	
	Superelevation variance for horizontal curves		●	HSM default assumption: No superelevation variance	Assumed no superelevation variance.
	Presence of spiral transition for horizontal curves		●	HSM default assumption: Base default on agency design policy (SHA: No spiral transition)	Assumed no spiral transition.
	Roadside hazard rating		●	HSM default assumption: Assume roadside hazard rating = 3	Assumed 3.
	Presence of passing lane		●	HSM default assumption: "Assume not present."	Assumed not present.
	Presence of short four-lane section		●	HSM default assumption: "Assume not present."	Assumed not present.

Facility	List of Variables	Required	Desirable	Source	Notes
	Use of automated speed enforcement		●	HSM default assumption: "Base default on current practice."	Assumed not present.
Intersections	Observed crashes	●		MSP database	2008-2010 crash data used.
	Area type (rural/suburban/urban)	●		SHA	
	Average annual daily traffic (AADT) for major road	●		SHA	
	Number of intersection legs	●		SHA	
	Type of traffic control	●		SHA	
	Average annual daily traffic (AADT) for minor road	●		SHA	
	Presence of left-turn phasing	●		Manually gathered by using the combination of "signal plan locator" of SHA and Google Earth (StreetView in some cases)	
	Type of left-turn phasing	●			
	Use of right-turn-on-red signal operation	●		Manually gathered by using Google Earth (StreetView in some cases)	
	Use of red-light cameras	●			There is an online database (http://www.photoenforced.com/) that has the location of different types of traffic control cameras, but as the process of using that source was manual, too, the research team decided to use Google Earth.
	Presence of major-road left-turn lane(s)	●		Manually gathered by using Google Earth (StreetView in some cases)	Visidata of SHA also could be used but as that data is only available via Intranet in SHA offices and the chance of availability of more than one workstation was small, so the project team decided to use Google Earth.
	Presence of major-road right-turn lane(s)	●			
	Presence of minor-road left-turn lane(s)	●			
	Presence of minor-road right-turn lane(s)	●			
	Presence of lighting	●		Asset Data Warehouse (ADW) of SHA's eGIS has "highway lighting." Final determination was made by manually double-checking Google Earth (StreetView in some cases).	
	Presence of median on		●	SHA database	

Facility	List of Variables	Required	Desirable	Source	Notes
	major road				
	Intersection skew angle		●	Manually estimated by using an uploaded compass on Google Earth 2D view	
	Presence of median on major road		●	SHA database	
	Pedestrian volume		●	Project team used land uses from the Maryland Department of Planning to translate pedestrian activities into volume and a method based on the presence of left-turn and right-turn lanes, median type and width, and through lanes to find out the maximum number of lanes crossed by pedestrians in each maneuver	
	Maximum number of lanes crossed by pedestrians on any approach		●		
	Presence of bus stops		●	These data items were collected by spatially joining of buffers with 1000 ft. radius for each intersection and bus locations/schools from external source.	
	Presence of schools within 1,000 ft.		●		
	Presence of alcohol sales establishments		●	This data item was collected by manually counting the Google Earth search results for "liquor store" points within buffers with 1000 ft. radius added as KMZ file in Google Earth.	
	Sum	41	19		
	Total	60			

Appendix E Finding Final Homogeneous Roadway Segments

Table 55. Required Steps for Finding Homogeneous Roadway Segments

Year		Roadway data 2008	Roadway data 2009	Roadway data 2010	Note
HSM facility types		35621	36410	37320	Based on 8 HSM facility types
Sub-Step 1	Intersecting 2008 & 2009	40623			Using "Analysis Tools" => Overlay => Intersect
Sub-Step 2	Intersecting (2008 & 2009) & 2010	42818			Using "Analysis Tools" => Overlay => Intersect
Sub-Step 3	Section length	31131			Adding new fields to 2008, 2009 and 2010 databases with "FLOAT" type and with 3 digits of numbers of decimal places. 11,687 records deleted that did not maintain the same section length for three years.
	% reduction from previous step	27.29%			
Sub-Step 4	RURURB	31130			1 record was deleted.
	% reduction from former step	0.00%			
Sub-Step 5	THROUGH_LANES	31124			6 records were deleted.
	% reduction from former step	0.02%			
Sub-Step 6	LT_ROADWAY_WD	31078			46 records were deleted.
	% reduction from former step	0.15%			
Sub-Step 7	RT_ROADWAY_WD	30843			235 records were deleted.
	% reduction from former step	0.76%			
Sub-Step 8	LT_OUT_SHLD_WD	30160			683 records were deleted.
	% reduction from former step	2.26%			
Sub-Step 9	RT_OUT_SHLD_WD	29600			560 records were deleted.
	% reduction from former step	1.86%			
Sub-Step 10	LT_IN_SHLD_TY	29527			73 records were deleted.
	% reduction from former step	0.25%			
Sub-Step 11	LT_OUT_SHLD_TY	29383			144 records were deleted.
	% reduction from former step	0.49%			
Sub-Step 12	RT_IN_SHLD_TY	29355			28 records were deleted.
	% reduction from former step	0.10%			
Sub-Step 13	RT_OUT_SHLD_TY	29274			81 records were deleted.
	% reduction from former step	0.28%			
Sub-Step 14	MEDIAN_TY	29107			167 records were deleted.
	% reduction from former step	0.57%			
Sub-Step 15	MEDIAN_WD	28858			249 records were deleted.
	% reduction from former step	0.86%			
Sub-Step 16	SPEED_LIMIT	28306			552 records were deleted.
	% reduction from former step	1.91%			
Sub-Step 17	LT_IN_AUX_NUMIA	28261			45 records were deleted.
	% reduction from former step	0.16%			
Sub-Step 18	LT_IN_AUX_TY	28220			30 records were deleted.
	% reduction from former step	0.15%			
Sub-Step 19	LT_OUT_AUX_NUMIA	28173			47 records were deleted.
	% reduction from former step	0.17%			
Sub-Step 20	LT_OUT_AUX_TY	28118			55 records were deleted.
	% reduction from former step	0.20%			
Sub-Step 21	RT_IN_AUX_NUMIA	28105			13 records were deleted.
	% reduction from former step	0.05%			
Sub-Step 22	RT_IN_AUX_TY	28097			8 records were deleted.
	% reduction from former step	0.03%			
Sub-Step 23	RT_OUT_AUX_NUMIA	28073			24 records were deleted.

Year		Roadway data 2008	2009	2010	Note
	% reduction from former step	0.09%			
Sub-Step 24	RT_OUT_AUX_TY	28029			44 records were deleted.
	% reduction from former step	0.16%			
Sub-Step 25	ROUTEID	27955			74 records were deleted.
	% reduction from former step	0.26%			
Sub-Step 26	FUNC_CL	27944			11 records were deleted.
	% reduction from former step	0.04%			
Sub-Step 27	NLFID	27944			No records were deleted.
	% reduction from former step	0.00%			
Sub-Step 28	LRS_ID	27944			No records were deleted.
	% reduction from former step	0.00%			
Sub-Step 29	ID_RTE_NO	27944			No records were deleted.
	% reduction from former step	0.00%			
Sub-Step 30	URBAN_AREA	27944			No records were deleted.
	% reduction from former step	0.00%			
Sub-Step 31	ID_MP	27638			Adding new fields to 2008, 2009 and 2010 databases with "FLOAT" type and with 3 digits of numbers of decimal places. 306 records were deleted.
	% reduction from former step	1.13%			
Sub-Step 32	STATE_MP	25486			Adding new fields to 2008, 2009 and 2010 databases with "FLOAT" type and with 3 digits of numbers of decimal places. 2152 records were deleted.
	% reduction from former step	7.79%			
Sub-Step 33	END_MP	25486			Adding new fields for 2008, 2009 and 2010 databases with "FLOAT" type and with 3 digits of numbers of decimal places. No records were deleted.
	% reduction from former step	0.00%			

The following tables include descriptive statistics of common roadway segments based on 8 types of HSM roadway facilities:

Table 56. Descriptive Statistics of Common Roadway Segments for R2U

R2U Segments		R2U Crashes		R2U Crashes (Normalized)	
Mean	0.196904822	Mean	0.938964177	Mean	11.53913104
Standard Error	0.004246704	Standard Error	0.019522115	Standard Error	0.305946815
Median	0.06	Median	0	Median	0
Mode	0.02	Mode	0	Mode	0
Standard Deviation	0.414331192	Standard Deviation	1.904682251	Standard Deviation	29.84981198
Sample Variance	0.171670337	Sample Variance	3.627814476	Sample Variance	891.0112755
Kurtosis	39.06668693	Kurtosis	31.38572863	Kurtosis	82.20245832
Skewness	5.164598664	Skewness	4.429447153	Skewness	6.624976884
Range	7.25	Range	28	Range	700
Minimum	0.01	Minimum	0	Minimum	0
Maximum	7.26	Maximum	28	Maximum	700
Sum	1874.337	Sum	8938	Sum	109840.9884
Count	9519	Count	9519	Count	9519

Table 57. Descriptive Statistics of Common Roadway Segments for R4U

R4U Segments		R4U Crashes		R4U Crashes (Normalized)	
Mean	0.089157895	Mean	2.263157895	Mean	78.02544713
Standard Error	0.026233127	Standard Error	0.620517164	Standard Error	26.22373753
Median	0.05	Median	1	Median	27.02702703
Mode	0.02	Mode	0	Mode	0
Standard Deviation	0.114347552	Standard Deviation	2.704771612	Standard Deviation	114.3066218
Sample Variance	0.013075363	Sample Variance	7.315789474	Sample Variance	13066.00379
Kurtosis	9.456544481	Kurtosis	1.357959591	Kurtosis	2.708924599
Skewness	2.856591806	Skewness	1.36402567	Skewness	1.788773506
Range	0.489	Range	9	Range	400
Minimum	0.01	Minimum	0	Minimum	0
Maximum	0.499	Maximum	9	Maximum	400
Sum	1.694	Sum	43	Sum	1482.483496
Count	19	Count	19	Count	19

Table 58. Descriptive Statistics of Common Roadway Segments for R4D

R4D Segments		R4D Crashes		R4D Crashes (Normalized)	
Mean	0.092326241	Mean	1.289361702	Mean	28.21394178
Standard Error	0.004795155	Standard Error	0.054078021	Standard Error	1.929911765
Median	0.05	Median	1	Median	3.849983542
Mode	0.01	Mode	0	Mode	0
Standard Deviation	0.1800579	Standard Deviation	2.030627873	Standard Deviation	72.46812252
Sample Variance	0.032420847	Sample Variance	4.123449557	Sample Variance	5251.628781
Kurtosis	97.45782724	Kurtosis	11.46585994	Kurtosis	81.40506137
Skewness	8.24616529	Skewness	2.838133683	Skewness	7.588836966
Range	3.225	Range	18	Range	1100
Minimum	0.01	Minimum	0	Minimum	0
Maximum	3.235	Maximum	18	Maximum	1100
Sum	130.18	Sum	1818	Sum	39781.65792
Count	1410	Count	1410	Count	1410

Table 59. Descriptive Statistics of Common Rural Roadway Segments

Rural Segments		Rural Segments Crashes		Rural Segments Crashes (Normalized)	
Mean	0.183249087	Mean	0.986390208	Mean	13.80207616
Standard Error	0.003759018	Standard Error	0.018417522	Standard Error	0.371442652
Median	0.057	Median	0	Median	0
Mode	0.02	Mode	0	Mode	0
Standard Deviation	0.393316122	Standard Deviation	1.927074888	Standard Deviation	38.8650442
Sample Variance	0.154697572	Sample Variance	3.713617625	Sample Variance	1510.491661
Kurtosis	43.25503975	Kurtosis	27.79004132	Kurtosis	164.020372
Skewness	5.431132496	Skewness	4.162993866	Skewness	9.493978564
Range	7.25	Range	28	Range	1100
Minimum	0.01	Minimum	0	Minimum	0
Maximum	7.26	Maximum	28	Maximum	1100
Sum	2006.211	Sum	10799	Sum	151105.1298
Count	10948	Count	10948	Count	10948

Table 60. Descriptive Statistics of Common Roadway Segments for U2U

U2U Segments		U2U Crashes		U2U Crashes (Normalized)	
Mean	0.079957588	Mean	1.089258489	Mean	22.72806987
Standard Error	0.001377717	Standard Error	0.024480426	Standard Error	0.664077163
Median	0.04	Median	0	Median	0
Mode	0.02	Mode	0	Mode	0
Standard Deviation	0.117024842	Standard Deviation	2.079395716	Standard Deviation	56.40748197
Sample Variance	0.013694814	Sample Variance	4.323886545	Sample Variance	3181.804022
Kurtosis	37.33559248	Kurtosis	54.74857947	Kurtosis	537.8786438
Skewness	4.897038579	Skewness	5.158693203	Skewness	14.95314872
Range	1.9	Range	42	Range	2500
Minimum	0.01	Minimum	0	Minimum	0
Maximum	1.91	Maximum	42	Maximum	2500
Sum	576.894	Sum	7859	Sum	163983.0241
Count	7215	Count	7215	Count	7215

Table 61. Descriptive Statistics of Common Roadway Segments for U3T

U3T Segments		U3T Crashes		U3T Crashes (Normalized)	
Mean	0.066346369	Mean	1.811918063	Mean	35.22746117
Standard Error	0.003674835	Standard Error	0.117560941	Standard Error	2.797340189
Median	0.044	Median	1	Median	15.625
Mode	0.02	Mode	0	Mode	0
Standard Deviation	0.085157904	Standard Deviation	2.724270313	Standard Deviation	64.82349278
Sample Variance	0.007251869	Sample Variance	7.42164874	Sample Variance	4202.085216
Kurtosis	103.2940106	Kurtosis	11.50669955	Kurtosis	38.3177919
Skewness	7.919641841	Skewness	2.875774067	Skewness	5.006513018
Range	1.35	Range	22	Range	750
Minimum	0.01	Minimum	0	Minimum	0
Maximum	1.36	Maximum	22	Maximum	750
Sum	35.628	Sum	973	Sum	18917.14665
Count	537	Count	537	Count	537

Table 62. Descriptive Statistics of Common Roadway Segments for U4U

U4U Segments		U4U Crashes		U4U Crashes (Normalized)	
Mean	0.054935223	Mean	3.361673414	Mean	80.72122237
Standard Error	0.002549567	Standard Error	0.19386716	Standard Error	4.49582594
Median	0.04	Median	2	Median	35.29411765
Mode	0.03	Mode	0	Mode	0
Standard Deviation	0.069402575	Standard Deviation	5.277319078	Standard Deviation	122.3822949
Sample Variance	0.004816717	Sample Variance	27.85009666	Sample Variance	14977.4261
Kurtosis	44.12429085	Kurtosis	32.48422496	Kurtosis	22.42645761
Skewness	5.618513865	Skewness	4.3138399	Skewness	3.512099079
Range	0.81	Range	62	Range	1400
Minimum	0.01	Minimum	0	Minimum	0
Maximum	0.82	Maximum	62	Maximum	1400
Sum	40.707	Sum	2491	Sum	59814.42578
Count	741	Count	741	Count	741

Table 63. Descriptive Statistics of Common Roadway Segments for U4D

U4D Segments		U4D Crashes		U4D Crashes (Normalized)	
Mean	0.06012027	Mean	2.26770326	Mean	62.38258607
Standard Error	0.001074756	Standard Error	0.056826074	Standard Error	1.86545099
Median	0.04	Median	1	Median	18.63425926
Mode	0.03	Mode	0	Mode	0
Standard Deviation	0.078523429	Standard Deviation	4.151804907	Standard Deviation	136.2928673
Sample Variance	0.006165929	Sample Variance	17.23748399	Sample Variance	18575.74568
Kurtosis	49.27632764	Kurtosis	54.06430216	Kurtosis	57.9701214
Skewness	5.82356272	Skewness	5.352774673	Skewness	5.922887286
Range	1.211	Range	78	Range	2650
Minimum	0.01	Minimum	0	Minimum	0
Maximum	1.221	Maximum	78	Maximum	2650
Sum	320.922	Sum	12105	Sum	332998.2445
Count	5338	Count	5338	Count	5338

Table 64. Descriptive Statistics of Common Roadway Segments for U5T

U5T Segments		U5T Crashes		U5TCrashes (Normalized)	
Mean	0.121434783	Mean	7.601449275	Mean	62.73892941
Standard Error	0.009193214	Standard Error	0.805859302	Standard Error	4.949171748
Median	0.07	Median	3	Median	33.33333333
Mode	0.06	Mode	0	Mode	0
Standard Deviation	0.152729134	Standard Deviation	13.38794022	Standard Deviation	82.22181629
Sample Variance	0.023326188	Sample Variance	179.2369433	Sample Variance	6760.427074
Kurtosis	20.55648371	Kurtosis	14.96297976	Kurtosis	9.939007687
Skewness	3.808928992	Skewness	3.331051476	Skewness	2.580026252
Range	1.321	Range	108	Range	600
Minimum	0.01	Minimum	0	Minimum	0
Maximum	1.331	Maximum	108	Maximum	600
Sum	33.516	Sum	2098	Sum	17315.94452
Count	276	Count	276	Count	276

Table 65. Descriptive Statistics of Common Urban and Suburban Roadway Segments

Urban Segments		Urban Segments Crashes		Urban Segments Crashes (Normalized)	
Mean	0.071430283	Mean	1.809456298	Mean	42.03790923
Standard Error	0.000861476	Standard Error	0.032714767	Standard Error	0.84901219
Median	0.04	Median	1	Median	7.547169811
Mode	0.02	Mode	0	Mode	0
Standard Deviation	0.102319997	Standard Deviation	3.885627544	Standard Deviation	100.8396336
Sample Variance	0.010469382	Sample Variance	15.09810141	Sample Variance	10168.6317
Kurtosis	45.11656872	Kurtosis	103.2051385	Kurtosis	108.4786561
Skewness	5.395023913	Skewness	7.402737328	Skewness	7.698690586
Range	1.9	Range	108	Range	2650
Minimum	0.01	Minimum	0	Minimum	0
Maximum	1.91	Maximum	108	Maximum	2650
Sum	1007.667	Sum	25526	Sum	593028.7855
Count	14107	Count	14107	Count	14107

Table 66. Descriptive Statistics of All Common Roadway Segments

All Common Segments		Crashes (All Common Segments)		Normalized Crashes (All Common Segments)	
Mean	1.449810417	Mean	1.449810417	Mean	29.70001657
Standard Error	0.020265529	Standard Error	0.020265529	Standard Error	0.512518435
Median	0	Median	0	Median	0
Mode	0	Mode	0	Mode	0
Standard Deviation	3.207784177	Standard Deviation	3.207784177	Standard Deviation	81.12537077
Sample Variance	10.28987932	Sample Variance	10.28987932	Sample Variance	6581.325783
Kurtosis	130.6401164	Kurtosis	130.6401164	Kurtosis	155.503782
Skewness	8.010648289	Skewness	8.010648289	Skewness	9.116891372
Range	108	Range	108	Range	2650
Minimum	0	Minimum	0	Minimum	0
Maximum	108	Maximum	108	Maximum	2650
Sum	36325	Sum	36325	Sum	744133.9152
Count	25055	Count	25055	Count	25055

Appendix F Maryland Crash Distribution

The presence of **X** in some cells denotes that the corresponding values of those cells were not developed for Maryland due to insufficient data; in those cases, the default values available in HSM were used.

<u>Rural Two-Lane, Two-Way Roadways – *Segments*</u>

Table 67. Distribution for Crash Severity Level for Rural Two-Lane, Two-Way Roadways

Distribution for Crash Severity Level on Rural Two-Lane, Two-Way Roadway Segments plus Locally Derived Values (HSM: Table 10-3)		
Crash severity level	**Percentage of total roadway segment crashes**	
	HSM-Provided Values	**Locally Derived Values (Maryland)**
Fatal	1.3	1.6
Incapacitating Injury (K)	5.4	5.9
Non-Incapacitating Injury (A)	10.9	17.3
Possible Injury (B)	14.5	16.8
Total Fatal Plus Injury (C)	32.1	41.7
Property Damage Only (O)	67.9	58.3
TOTAL	100.0	100.0

Note: HSM-provided crash severity data based on HSIS data for Washington (2002-2006)

Table 68. Distribution by Collision Type for Rural Two-Lane, Two-Way Roadways

Default Distribution by Collision Type for Specific Crash Severity Levels on Rural Two-Lane, Two-Way Roadway Segments plus Locally Derived Values (HSM: Table 10-4)						
Collision type	**Percentage of total roadway segment crashes by crash severity level**					
	HSM-Provided Values			**Locally Derived Values (Maryland)**		
	Total fatal and injury	**Property damage only**	**TOTAL (all severity levels combined)**	**Total fatal and injury**	**Property damage only**	**TOTAL (all severity levels combined)**
SINGLE-VEHICLE CRASHES						
Collision with animal	3.8	18.4	12.1	27.7	35.2	32.1
Collision with bicycle	0.4	0.1	0.2	1.8	1.4	1.6
Collision with pedestrian	0.7	0.1	0.3	3.2	1.7	2.3
Overturned	3.7	1.5	2.5	13.7	10.8	12.0
Ran off road	54.5	50.5	52.1	30.9	30.1	30.4
Other single-vehicle crash	0.7	2.9	2.1	4.3	5.9	5.2
Total single-vehicle crashes	63.8	73.5	69.3	81.6	85.1	83.7
MULTIPLE-VEHICLE CRASHES						
Angle collision	10.0	7.2	8.5	1.3	1.1	1.2
Head-on collision	3.4	0.3	1.6	3.9	1.7	2.6
Rear-end collision	16.4	12.2	14.2	10.6	8.2	9.2
Sideswipe collision	3.8	3.8	3.7	1.0	1.3	1.2
Other multiple-vehicle collision	2.6	3.0	2.7	1.6	2.6	2.2
Total multiple-vehicle crashes	36.2	26.5	30.7	18.4	14.9	16.3
TOTAL CRASHES	100.0	100.0	100.0	100.0	100.0	100.0

Note: HSM-provided values based on crash data for Washington (2002-2006); includes approximately 70 percent opposite-direction sideswipe and 30 percent same-direction sideswipe collisions

Table 69. Nighttime Crash Proportions for Unlighted Rural Two-Lane, Two-Way Roadways

Nighttime Crash Proportions for Unlighted Roadway Segments plus Locally Derived Values (HSM: Table 10-12)			
HSM Default Values			
Roadway type	**Proportion of total nighttime crashes by severity level**		**Proportion of crashes that occur at night**
	Fatal and Injury	**PDO**	
R2U	0.382	0.618	0.370
Locally Derived Values (Maryland)			
Roadway type	**Proportion of total nighttime crashes by severity level**		**Proportion of crashes that occur at night**
	Fatal and Injury	**PDO**	
R2U	0.408	0.592	0.221

Note: HSM-provided values based on HSIS data for Washington (2002-2006)

<u>Rural Two-Lane, Two-Way Roadways – *Intersections*</u>

Table 70. Crash Severity Level at Rural Two-Lane, Two-Way Intersections

Default Distribution for Crash Severity Level at Rural Two-Lane, Two-Way Intersections plus Locally Derived Values (HSM: Table 10-5)			
Collision type	**Percentage of total crashes**		
	HSM-Provided Values		
	3ST	**4ST**	**4SG**
Fatal	1.7	1.8	0.9
Incapacitating Injury (K)	4.0	4.3	2.1
Non-Incapacitating Injury (A)	16.6	16.2	10.5
Possible Injury (B)	19.2	20.8	20.5
Total Fatal Plus Injury (C)	41.5	43.1	34.0
Property Damage Only (O)	58.5	56.9	66.0
TOTAL	100.0	100.0	100.0
Collision type	**Locally Derived Values (Maryland)**		
	3ST	**4ST**	**4SG**
Fatal	0.7	X	1.8
Incapacitating Injury (K)	5.2	X	5.3
Non-Incapacitating Injury (A)	25.2	X	22.8
Possible Injury (B)	17.0	X	31.6
Total Fatal Plus Injury (C)	48.1	X	61.4
Property Damage Only (O)	51.9	X	38.6
TOTAL	100.0	X	100.0

Note: HSM-Provided values based on HSIS data for California (2002-2006)

Table 71. Distribution by Collision Type at Rural Two-Lane, Two-Way Intersections

Default Distribution for Collision Type and Manner of Collision at Rural Two-Way Intersections plus Locally Derived Values (HSM: Table 10-6)

Collision type	Percentage of total crashes by collision type (HSM Default Values)								
	Three-leg stop-controlled intersections			Four-leg stop-controlled intersections			Four-leg signalized intersections		
	Fatal and Injury	Property damage only	Total	Fatal and injury	Property damage only	Total	Fatal and injury	Property damage only	Total
SINGLE-VEHICLE CRASHES									
Collision with animal	0.8	2.6	1.9	0.6	1.4	1.0	0.0	0.3	0.2
Collision with bicycle	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
Collision with pedestrian	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
Overturned	2.2	0.7	1.3	0.6	0.4	0.5	0.3	0.3	0.3
Ran off road	24.0	24.7	24.4	9.4	14.4	12.2	3.2	8.1	6.4
Other single-vehicle crash	1.1	2.0	1.6	0.4	1.0	0.8	0.3	1.8	0.5
Total single-vehicle crashes	28.3	30.2	29.4	11.2	17.4	14.7	4.0	10.7	7.6
MULTIPLE-VEHICLE CRASHES									
Angle collision	27.5	21.0	23.7	53.2	35.4	43.1	33.6	24.2	27.4
Head-on collision	8.1	3.2	5.2	6.0	2.5	4.0	8.0	4.0	5.4
Rear-end collision	26.0	29.2	27.8	21.0	26.6	24.2	40.3	43.8	42.6
Sideswipe collision	5.1	13.1	9.7	4.4	14.4	10.1	5.1	15.3	11.8
Other multiple-vehicle collision	5.0	3.3	4.2	4.2	3.7	3.9	9.0	2.0	5.2
Total multiple-vehicle crashes	71.7	69.8	70.6	88.8	82.6	85.3	96.0	89.3	92.4
TOTAL CRASHES	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Collision type	Percentage of total crashes by collision type (Locally Derived Values - Maryland)								
	Three-leg stop-controlled intersections			Four-leg stop-controlled intersections			Four-leg signalized intersections		
	Fatal and Injury	Property damage only	Total	Fatal and injury	Property damage only	Total	Fatal and injury	Property damage only	Total
SINGLE-VEHICLE CRASHES									
Collision with animal	X	X	X	X	X	X	X	X	X
Collision with bicycle	X	X	X	X	X	X	X	X	X
Collision with pedestrian	X	X	X	X	X	X	X	X	X
Overturned	X	X	X	X	X	X	X	X	X
Ran off road	X	X	X	X	X	X	X	X	X
Other single-vehicle crash	X	X	X	X	X	X	X	X	X
Total single-vehicle crashes	X	X	X	X	X	X	X	X	X
MULTIPLE-VEHICLE CRASHES									
Angle collision	53.7	39.2	46.7	59.3	60.0	59.6	44.1	33.3	40.0

Head-on collision	13.0	7.8	10.5	13.6	17.1	14.9	38.2	38.1	38.2
Rear-end collision	22.2	35.3	28.6	20.3	14.3	18.1	11.8	14.3	12.7
Sideswipe collision	1.9	0.0	1.0	0.0	0.0	0.0	2.9	0.0	1.8
Other multiple-vehicle collision	9.3	17.6	13.3	6.8	8.6	7.4	2.9	14.3	7.3
Total multiple-vehicle crashes	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
TOTAL CRASHES	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Note: HSM-Provided values based on HSIS data for California (2002-2006)

Table 72. Nighttime Crash Proportions at Unlighted Rural Two-Lane, Two-Way Intersections

Nighttime Crash Proportions for Unlighted Intersections (Table 10-15)		
Intersection Type	**Proportion of crashes that occur at night**	
	HSM-Provided Values	**Locally Derived Values (Maryland)**
3ST	0.26	0.084
4ST	0.244	0.015
4SG	0.286	X

Note: HSM-Provided values based on HSIS data for California (2002-2006)

Rural Multilane Highways – *Segments*

Table 73. Distribution by Collision Type and Crash Severity Level for Undivided Rural Multilane Highways

Distribution of Crashes by Collision Type and Crash Severity Level for Undivided Roadway Segments (HSM: Table 11-4)								
Collision type	**Proportion of crashes by collision type and crash severity level**							
	HSM-Provided Values				**Locally Derived Values (Maryland)**			
	Total	**Fatal and injury**	**Fatal and injury***	**PDO**	**Total**	**Fatal and injury**	**Fatal and injury***	**PDO**
Head-on	0.009	0.029	0.043	0.001	X	X	X	X
Sideswipe	0.098	0.048	0.044	0.12	X	X	X	X
Rear-end	0.246	0.305	0.217	0.22	X	X	X	X
Angle	0.356	0.352	0.348	0.358	X	X	X	X
Single	0.238	0.238	0.304	0.237	X	X	X	X
Other	0.053	0.028	0.044	0.064	X	X	X	X
SV run-off-rd, Head-on, Sideswipe	0.27				X			

NOTE: * Using the KABCO scale, these include only KAB crashes. Crashes with severity level C (possible injury) are not included.

Table 74. Distribution by Collision Type and Crash Severity Level for Divided Rural Multilane Highways

Distribution of Crashes by Collision Type and Crash Severity Level for Divided Roadway Segments (HSM: Table 11-6)								
Collision type	**Proportion of crashes by collision type and crash severity level**							
	HSM-Provided Values				**Locally Derived Values (Maryland)**			
	Total	**Fatal and injury**	**Fatal and injury***	**PDO**	**Total**	**Fatal and injury**	**Fatal and injury***	**PDO**
Head-on	0.006	0.013	0.018	0.002	0.010	0.018	0.018	0.007
Sideswipe	0.043	0.027	0.022	0.053	0.015	0.021	0.020	0.013
Rear-end	0.116	0.163	0.114	0.088	0.083	0.105	0.099	0.071
Angle	0.043	0.048	0.045	0.041	0.012	0.016	0.020	0.010
Single	0.768	0.727	0.778	0.792	0.869	0.831	0.830	0.889
Other	0.024	0.022	0.023	0.024	0.010	0.010	0.013	0.011
SV run-off-rd, Head-on, Sideswipe	0.500				0.605			

NOTE: * Using the KABCO scale, these include only KAB crashes. Crashes with severity level C (possible injury) are not included.

Table 75. Nighttime Crash Proportions for Unlighted Undivided Rural Multilane Highways

Night-time Crash Proportions for Unlighted Roadway Segments (HSM: Table 11-15)						
Roadway type	**HSM-Provided Values**			**Locally Derived Values (Maryland)**		
	Proportion of total night-time crashes by severity level		**Proportion of crashes that occur at night**	**Proportion of total night-time crashes by severity level**		**Proportion of crashes that occur at night**
	Fatal and injury	**PDO**		**Fatal and injury**	**PDO**	
R4U	0.361	0.639	0.255	X	X	X

Table 76. Nighttime Crash Proportions for Unlighted Divided Rural Multilane Highways

Night-time Crash Proportions for Unlighted Roadway Segments (HSM: Table 11-19)						
Roadway type	**HSM-Provided Values**			**Locally Derived Values (Maryland)**		
	Proportion of total night-time crashes by severity level		**Proportion of crashes that occur at night**	**Proportion of total night-time crashes by severity level**		**Proportion of crashes that occur at night**
	Fatal and injury	**PDO**		**Fatal and injury**	**PDO**	
R4D	0.323	0.677	0.426	0.326	0.674	0.186

<u>Rural Multilane Highways – *Intersections*</u>

Table 77. Distribution by Collision Type and Crash Severity Level at Rural Multilane Intersections

Distribution of Intersection Crashes by Collision Type and Crash Severity (HSM: Table 11-9)								
Collision type	**Proportion of crashes by collision type and crash severity level**							
	HSM-Provided Values				**Locally Derived Values (Maryland)**			
	Total	**Fatal and injury**	**Fatal and injury***	**PDO**	**Total**	**Fatal and injury**	**Fatal and injury***	**PDO**
Three-leg intersections with minor road stop control								
Head-on	0.029	0.043	0.052	0.020	X	X	X	X
Sideswipe	0.133	0.058	0.057	0.179	X	X	X	X
Rear-end	0.289	0.247	0.142	0.315	X	X	X	X
Angle	0.263	0.369	0.381	0.198	X	X	X	X
Single	0.234	0.219	0.284	0.244	X	X	X	X
Other	0.052	0.064	0.084	0.044	X	X	X	X
SV run-off-rd, Head-on, Sideswipe	0.500				X			
Four-leg intersections with minor road stop control								
Head-on	0.016	0.018	0.023	0.015	X	X	X	X
Sideswipe	0.107	0.042	0.040	0.156	X	X	X	X
Rear-end	0.228	0.213	0.108	0.240	X	X	X	X
Angle	0.395	0.534	0.571	0.292	X	X	X	X
Single	0.202	0.148	0.199	0.243	X	X	X	X
Other	0.052	0.045	0.059	0.054	X	X	X	X
SV run-off-rd, Head-on, Sideswipe	0.500				X			
Four-leg signalized intersections								
Head-on	0.054	0.083	0.093	0.034	0.314	0.250	0.429	0.400
Sideswipe	0.106	0.047	0.039	0.147	0.029	0.000	0.000	0.067
Rear-end	0.492	0.472	0.314	0.505	0.200	0.200	0.143	0.200
Angle	0.256	0.315	0.407	0.215	0.314	0.400	0.429	0.200
Single	0.062	0.041	0.078	0.077	0.057	0.100	0.000	0.000
Other	0.030	0.042	0.069	0.022	0.086	0.050	0.000	0.133
SV run-off-rd, Head-on, Sideswipe	0.500				0.330			

NOTE: * Using the KABCO scale, these include only KAB crashes. Crashes with severity level C (possible injury) are not included

Table 78. Nighttime Crash Proportions at Rural Multilane Intersections

Night-time Crash Proportions for Unlighted Intersections (HSM: Table 11-24)		
Intersection Type	**HSM-Provided Values**	**Locally Derived Values (Maryland)**
	Proportion of crashes that occur at night	
3ST	0.276	X
4ST	0.273	X

Urban & Suburban Arterials – *Segments*

Table 79. Distribution of Multiple-Vehicle Non-Driveway Collision Type and Crash Severity Level for Urban and Suburban Arterials

Distribution of Multiple-Vehicle Non-driveway Collisions for Roadway Segments by Manner of Collision Type (HSM: Table 12-4)										
Collision type	**Proportion of crashes by severity level for specific road types**									
	HSM-Provided Values									
	U2U		**U3T**		**U4U**		**U4D**		**U5T**	
	FI	**PDO**	**FI**	**PDO**	**FI**	**PDO**	**FI**	**PDO**	**FI**	**PDO**
Rear-end collision	0.730	0.778	0.845	0.842	0.511	0.506	0.832	0.662	0.846	0.651
Head-on collision	0.068	0.004	0.034	0.020	0.077	0.004	0.020	0.007	0.021	0.004
Angle collision	0.085	0.079	0.069	0.020	0.181	0.130	0.040	0.036	0.050	0.059
Sideswipe, same direction	0.015	0.031	0.001	0.078	0.093	0.249	0.050	0.223	0.061	0.248
Sideswipe, opposite direction	0.073	0.055	0.017	0.020	0.082	0.031	0.010	0.001	0.004	0.009
Other multiple-vehicle collision	0.029	0.053	0.034	0.020	0.056	0.080	0.048	0.071	0.018	0.029
Collision type	**Proportion of crashes by severity level for specific road types**									
	Locally Derived Values (Maryland)									
	U2U		**U3T**		**U4U**		**U4D**		**U5T**	
	FI	**PDO**	**FI**	**PDO**	**FI**	**PDO**	**FI**	**PDO**	**FI**	**PDO**
Rear-end collision	0.713	0.748	0.686	0.684	0.814	0.831	0.782	0.794	0.816	0.775
Head-on collision	0.113	0.057	0.117	0.083	0.036	0.028	0.036	0.025	0.054	0.069
Angle collision	0.059	0.049	0.036	0.062	0.033	0.035	0.031	0.023	0.020	0.034
Sideswipe, same direction	0.020	0.033	0.066	0.036	0.023	0.034	0.054	0.059	0.041	0.054
Sideswipe, opposite direction	0.038	0.030	0.000	0.021	0.000	0.004	0.002	0.004	0.007	0.020
Other multiple-vehicle collision	0.056	0.083	0.095	0.114	0.094	0.069	0.094	0.096	0.061	0.049

Note: HSM-Provided values based on HSIS data for Washington (2002-2006)

Table 80. Distribution of Single-Vehicle Collision Type and Crash Severity Level for Urban and Suburban Arterials

Distribution of Single-Vehicle Collisions for Roadway Segments by Collision Type (HSM: Table 12-6)										
Collision type	**Proportion of crashes by severity level for specific road types**									
	HSM-Provided Values									
	U2U		**U3T**		**U4U**		**U4D**		**U5T**	
	FI	**PDO**	**FI**	**PDO**	**FI**	**PDO**	**FI**	**PDO**	**FI**	**PDO**
Collision with animal	0.026	0.066	0.001	0.001	0.001	0.001	0.001	0.063	0.016	0.049
Collision with fixed-object	0.723	0.759	0.688	0.963	0.612	0.809	0.5	0.813	0.398	0.768
Collision with other object	0.01	0.013	0.001	0.001	0.02	0.029	0.028	0.016	0.005	0.061
Other single-vehicle collision	0.241	0.162	0.31	0.035	0.367	0.161	0.471	0.108	0.581	0.122
Collision type	**Proportion of crashes by severity level for specific road types**									
	Locally Derived Values (Maryland)									
	U2U		**U3T**		**U4U**		**U4D**		**U5T**	
	FI	**PDO**	**FI**	**PDO**	**FI**	**PDO**	**FI**	**PDO**	**FI**	**PDO**
Collision with animal	0.049	0.064	0.026	0.022	0.007	0.006	0.033	0.047	0.024	0.025
Collision with fixed-object	0.652	0.639	0.707	0.676	0.672	0.718	0.606	0.617	0.64	0.632
Collision with other object	0.015	0.019	0.004	0.005	0.047	0.039	0.015	0.013	0.007	0.008
Other single-vehicle collision	0.284	0.278	0.264	0.297	0.275	0.237	0.347	0.323	0.328	0.334

Note: HSM-Provided values based on HSIS data for Washington (2002-2006)

Table 81. Proportion of Fixed-Object Collisions for Urban and Suburban Arterials

Proportion of Fixed-Object Collisions (HSM: Table 12-21)		
Roadway Type	**HSM-Provided Values**	**Locally Derived Values (Maryland)**
	Proportion of Fixed-Object Collisions	
U2U	0.059	0.072
U3T	0.034	0.033
U4U	0.037	0.060
U4D	0.036	0.036
U5T	0.016	0.003

Table 82. Nighttime Crash Proportions for Unlighted Urban and Suburban Arterials

Nighttime Crash Proportions for Unlighted Roadway Segments (HSM: Table 12-23)						
Roadway Type	**HSM-Provided Values**			**Locally Derived Values (Maryland)**		
	Proportion of Total Nighttime Crashes by Severity Level		**Proportion of Crashes that Occur at Night**	**Proportion of Total Nighttime Crashes by Severity Level**		**Proportion of Crashes that Occur at Night**
	Fatal and Injury	**PDO**		**Fatal and Injury**	**PDO**	
U2U	0.424	0.576	0.316	0.414	0.586	0.273
U3T	0.429	0.571	0.304	0.438	0.562	0.166
U4U	0.517	0.483	0.365	0.363	0.637	0.170
U4D	0.364	0.636	0.41	0.360	0.640	0.321
U5T	0.432	0.568	0.274	0.391	0.609	0.405

Urban & Suburban Arterials – *Intersections*

Table 83. Distribution of Multiple-Vehicle Collision Type and Crash Severity Level at Urban and Suburban Intersections

Distribution of Multiple-Vehicle Collisions for Intersections by Collision Type (HSM: Table 12-11)								
Collision type	**Proportion of crashes by severity level for specific intersection types**							
	HSM-Provided Values							
	U3ST		**U3SG**		**U4ST**		**U4SG**	
	FI	**PDO**	**FI**	**PDO**	**FI**	**PDO**	**FI**	**PDO**
Rear-end collision	0.421	0.44	0.549	0.546	0.338	0.374	0.45	0.483
Head-on collision	0.045	0.023	0.038	0.02	0.041	0.03	0.049	0.03
Angle collision	0.343	0.262	0.28	0.204	0.44	0.335	0.347	0.244
Sideswipe	0.126	0.04	0.076	0.032	0.121	0.044	0.099	0.032
Other multiple-vehicle collision	0.065	0.235	0.057	0.198	0.06	0.217	0.055	0.211
Collision type	**Proportion of crashes by severity level for specific intersection types**							
	Locally Derived Values (Maryland)							
	U3ST		**U3SG**		**U4ST**		**U4SG**	
	FI	**PDO**	**FI**	**PDO**	**FI**	**PDO**	**FI**	**PDO**
Rear-end collision	0.145	0.298	0.302	0.324	0.143	0.257	0.238	0.284
Head-on collision	0.159	0.149	0.317	0.202	0.143	0.143	0.300	0.249
Angle collision	0.580	0.340	0.309	0.271	0.667	0.543	0.359	0.284
Sideswipe	0.000	0.043	0.022	0.059	0.000	0.000	0.018	0.063
Other multiple-vehicle collision	0.116	0.170	0.050	0.144	0.048	0.057	0.085	0.120

Note: HSM-Provided values based on HSIS data for California (2002-2006)

Table 84. Distribution of Single-Vehicle Collision Type and Crash Severity Level at Urban and Suburban Intersections

Distribution of Single-Vehicle Crashes for Intersections by Collision Type (HSM: Table 12-13)								
Collision type	**Proportion of crashes by severity level for specific intersection types**							
	HSM-Provided Values							
	U3ST		**U3SG**		**U4ST**		**U4SG**	
	FI	**PDO**	**FI**	**PDO**	**FI**	**PDO**	**FI**	**PDO**
Collision with parked vehicle	0.001	0.003	0.001	0.001	0.001	0.001	0.001	0.001
Collision with animal	0.003	0.018	0.001	0.003	0.001	0.026	0.002	0.002
Collision with fixed object	0.762	0.8341	0.653	0.895	0.679	0.847	0.744	0.87
Collision with other object	0.09	0.092	0.091	0.069	0.089	0.07	0.072	0.07
Other single-vehicle collision	0.039	0.023	0.045	0.018	0.051	0.007	0.04	0.023
Non-collision	0.105	0.03	0.209	0.014	0.179	0.049	0.141	0.034
Collision type	**Proportion of crashes by severity level for specific intersection types**							
	Locally Derived Values (Maryland)							
	U3ST		**U3SG**		**U4ST**		**U4SG**	
	FI	**PDO**	**FI**	**PDO**	**FI**	**PDO**	**FI**	**PDO**
Collision with parked vehicle	X	X	0.000	0.000	X	X	0.000	0.000
Collision with animal	X	X	0.000	0.048	X	X	0.000	0.033
Collision with fixed object	X	X	0.421	0.857	X	X	0.179	0.667
Collision with other object	X	X	0.000	0.000	X	X	0.000	0.033
Other single-vehicle collision	X	X	0.579	0.095	X	X	0.821	0.267
Non-collision	X	X	0.000	0.000	X	X	0.000	0.000

Source: HSM-Provided values base on HSIS data for California (2002-2006)

Table 85. Nighttime Crash Proportions at Unlighted Urban and Suburban Intersections

Nighttime Crash Proportions for Unlighted Intersections (HSM: Table 12-27)		
Intersection Type	**Proportion of crashes that occur at night**	
	HSM-Provided Values	**Locally Derived Values (Maryland)**
U3ST	0.238	0.118
U4ST	0.229	0.106
U3SG	0.235	0.180
U4SG	0.235	0.172

***

Table 86. Conversion of Maryland Collision Codes to HSM Collision Types

Conversion of MD Collision Codes to HSM Collision Types		
Collision Type [MD Database]	**HARM_EVENT1&2 [MD Database]**	**HSM Collision Type**
17 [Single vehicle]	8 [Animal]	Collision with Animal
	4 [Bicycle]	Collision with Bicycle
	3 [Pedestrian]	Collision with Pedestrian
	11 [Overturn]	Overturned
	16 [Off Road]	Run Off Road
	All other values except 0 [Not Applicable], 01 [Other Vehicle], 02 [Parked Vehicle] and 99 [Unknown]	Other Single-vehicle Collision
11, 12, 13, 14	-	Angle Collision
1, 2	-	Head-on Collision
3	-	Rear-end Collision
6, 7	-	Sideswipe
All other values except 0 [Not Applicable] and 99 [Unknown]	-	Other Multiple-vehicle Collision

Appendix G Details for Regression Models for AADT Estimation

The following tables include details of the regression models developed for estimation of AADT on minor roadways for signalized and stop-controlled intersections.

Table 87. Details for AADT Estimation for Minor Roadways of Signalized Intersections – 2008

Source	SS	df	MS	Number of obs	=	112
				F(5, 106)	=	75.85
Model	7.7964e+09	5	1.5593e+09	Prob > F	=	0.0000
Residual	2.1790e+09	106	20556593	R-squared	=	0.7816
				Adj R-squared	=	0.7713
Total	9.9754e+09	111	89868516.3	Root MSE	=	4533.9

Table 88. Regression for AADT Estimation for Minor Roadways of Signalized Intersections – 2008

SG_AADT2008	Coef.	Std. Err.	t	P>t	[95% Conf.	Interval]
mnnumln	4319.063	509.4321	8.48	0.000	3309.065	5329.062
mjnumln	1444.975	334.5442	4.32	0.000	781.7088	2108.242
mnurothart	6041.045	1411.781	4.28	0.000	3242.053	8840.038
mnurfree	18318.07	4872.303	3.76	0.000	8658.261	27977.89
mnurloc	-4821.744	1786	-2.70	0.008	-8362.661	-1280.826
_cons	-6050.305	1294.599	-4.67	0.000	-8616.973	-3483.636

Table 89. Details for AADT Estimation for Minor Roadways of Signalized Intersections – 2009

Source	SS	df	MS	Number of obs	=	112
				F(5, 106)	=	74.34
Model	8.2657e+09	5	1.6531e+09	Prob > F	=	0.0000
Residual	2.3571e+09	106	22236638.2	R-squared	=	0.7781
				Adj R-squared	=	0.7676
Total	1.0623e+10	111	95700459.5	Root MSE	=	4715.6

Table 90. Regression for AADT Estimation for Minor Roadways of Signalized Intersections – 2009

SG_AADT2009	Coef.	Std. Err.	t	P>t	[95% Conf.	Interval]
mnnumln	4436.088	529.8406	8.37	0.000	3385.628	5486.549
mjnumln	1608.455	347.9465	4.62	0.000	918.6176	2298.293
mnurothart	5518.895	1468.339	3.76	0.000	2607.771	8430.019
mnurfree	20185.76	5067.494	3.98	0.000	10138.96	30232.56
mnurloc	-4866.872	1857.549	-2.62	0.010	-8549.644	-1184.1
_cons	-6753.022	1346.463	-5.02	0.000	-9422.515	-4083.528

Table 91. Details for AADT Estimation for Minor Roadways of Signalized Intersections – 2010

Source	SS	df	MS	Number of obs	=	112
				F(5, 106)	=	53.40
Model	6.8455e+09	5	1.3691e+09	Prob > F	=	0.0000
Residual	2.7175e+09	106	25636875.7	R-squared	=	0.7158
				Adj R-squared	=	0.7024
Total	9.5630e+09	111	86153281.2	Root MSE	=	5063.3

Table 92. Regression for AADT Estimation for Minor Roadways of Signalized Intersections – 2010

SG_AADT2010	Coef.	Std. Err.	t	P>t	[95% Conf.	Interval]
mnnumln	3988.758	568.9096	7.01	0.000	2860.839	5116.676
mjnumln	1409.611	373.6031	3.77	0.000	668.9061	2150.315
mnurothart	5092.22	1576.61	3.23	0.002	1966.437	8218.002
mnurfree	19502.15	5441.157	3.58	0.001	8714.526	30289.77
mnurloc	-5107.076	1994.52	-2.56	0.012	-9061.405	-1152.746
_cons	-4901.359	1445.747	-3.39	0.001	-7767.693	-2035.025

Table 93. Details for AADT Estimation for Minor Roadways of Stop-Controlled Intersections – 2008

Source	SS	df	MS	Number of obs	=	116
				F(8, 107)	=	16.42
Model	296256605	8	37032075.6	Prob > F	=	0.0000
Residual	241289765	107	2255044.53	R-squared	=	0.5511
				Adj R-squared	=	0.5176
Total	537546370	115	4674316.26	Root MSE	=	1501.7

Table 94. Regression for AADT Estimation for Minor Roadways of Stop-Controlled Intersections – 2008

ST_AADT2008	Coef.	Std. Err.	t	P>t	[90% Conf.	Interval]
mnruloc	-1009.298	350.8155	-2.88	0.005	-1591.378	-427.2184
mnmedwd	818.4968	132.7649	6.17	0.000	598.2107	1038.783
mnrumnart	4235.79	782.5554	5.41	0.000	2937.359	5534.221
avghhsize	1698.903	632.826	2.68	0.008	648.9063	2748.901
mnrumjcol	1300.952	387.2622	3.36	0.001	658.3991	1943.505
popden	.7175917	.2671666	2.69	0.008	.2743037	1.16088
mnruothart	3796.183	1522.822	2.49	0.014	1269.487	6322.878
mnurothart	3486.115	1521.306	2.29	0.024	961.9349	6010.296
_cons	-3014.581	1699.695	-1.77	0.079	-5834.748	-194.4137

Table 95. Details for AADT Estimation for Minor Roadways of Stop-Controlled Intersections – 2009

Source	SS	df	MS	Number of obs	=	116
				F(8, 107)	=	16.41
Model	304742323	8	38092790.4	Prob > F	=	0.0000
Residual	248421639	107	2321697.56	R-squared	=	0.5509
				Adj R-squared	=	0.5173
Total	553163961	115	4810121.4	Root MSE	=	1523.7

Table 96. Regression for AADT Estimation for Minor Roadways of Stop-Controlled Intersections – 2009

ST_AADT2009	Coef.	Std. Err.	t	P>t	[90% Conf.	Interval]
mnruloc	-1000.649	355.9624	-2.81	0.006	-1591.269	-410.0295
mnmedwd	834.6083	134.7127	6.20	0.000	611.0903	1058.126
mnrumnart	4283.924	794.0363	5.40	0.000	2966.444	5601.405
avghhsize	1748.293	642.1102	2.72	0.008	682.8909	2813.694
mnrumjcol	1339.81	392.9438	3.41	0.001	687.8301	1991.79
popden	.7199705	.2710862	2.66	0.009	.2701791	1.169762
mnruothart	3912.901	1545.163	2.53	0.013	1349.136	6476.665
mnurothart	3494.391	1543.625	2.26	0.026	933.1785	6055.604
_cons	-3145.864	1724.632	-1.82	0.071	-6007.406	-284.322

Table 97. Details for AADT Estimation for Minor Roadways of Stop-Controlled Intersections – 2010

Source	SS	df	MS	Number of obs	=	116
				F(8, 108)	=	8.92
Model	603370293	7	86195756.1	Prob > F	=	0.0000
Residual	1.0436e+09	108	9662857.69	R-squared	=	0.3664
				Adj R-squared	=	0.3253
Total	1.6470e+09	115	14321381.9	Root MSE	=	3108.5

Table 98. Regression for AADT Estimation for Minor Roadways of Stop-Controlled Intersections – 2010

ST_AADT2010	**Coef.**	**Std. Err.**	**t**	**P>t**	**[90% Conf.**	**Interval]**
mnrumjcol	2540.336	768.0097	3.31	0.001	1266.142	3814.529
avghhsize	7432.884	1803.085	4.12	0.000	4441.412	10424.36
pct_emp	-132.6905	54.93674	-2.42	0.017	-233.8353	-41.54581
mnmedwd	705.4558	277.0559	2.55	0.012	245.7964	1165.115
mu_1	982601.9	362158.2	2.71	0.008	381750.7	1583453
rm3st	3043.161	1019.376	2.99	0.004	1351.929	4734.393
mnruloc	-1437.878	706.4304	-2.04	0.044	-2609.906	-265.85
_cons	-10010.94	4019.836	-2.49	0.014	-16680.19	-3341.692

Appendix H Sample Sites

Roadway Segments

In the scenario 2 (the best sampling scenario), 1,324 roadway segments were selected. The following figures depicted them.

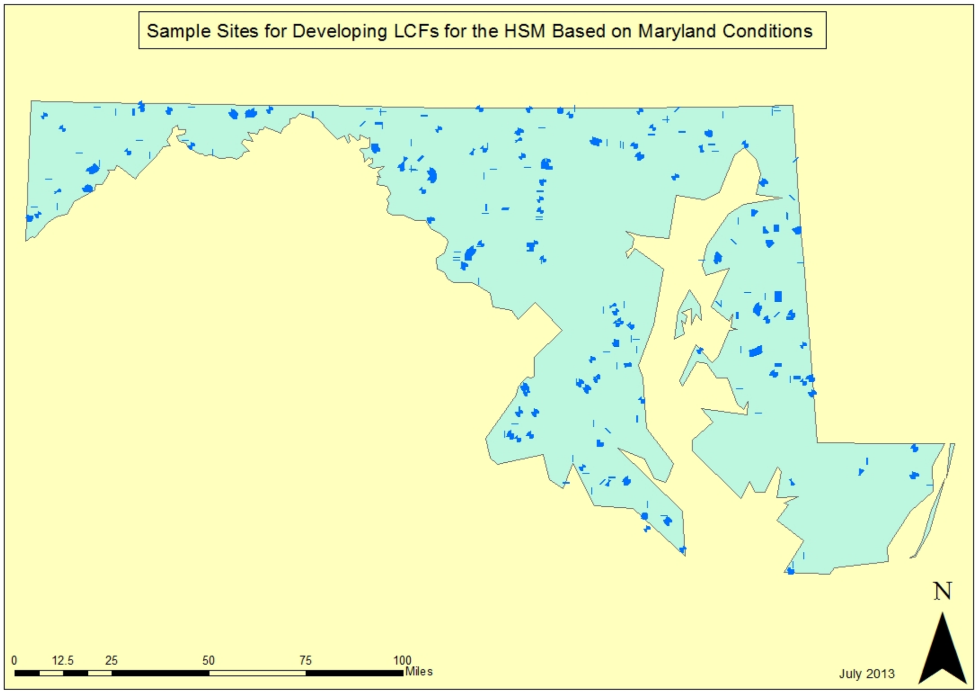


Figure 16. R2U Samples (251 Roadway Segments)

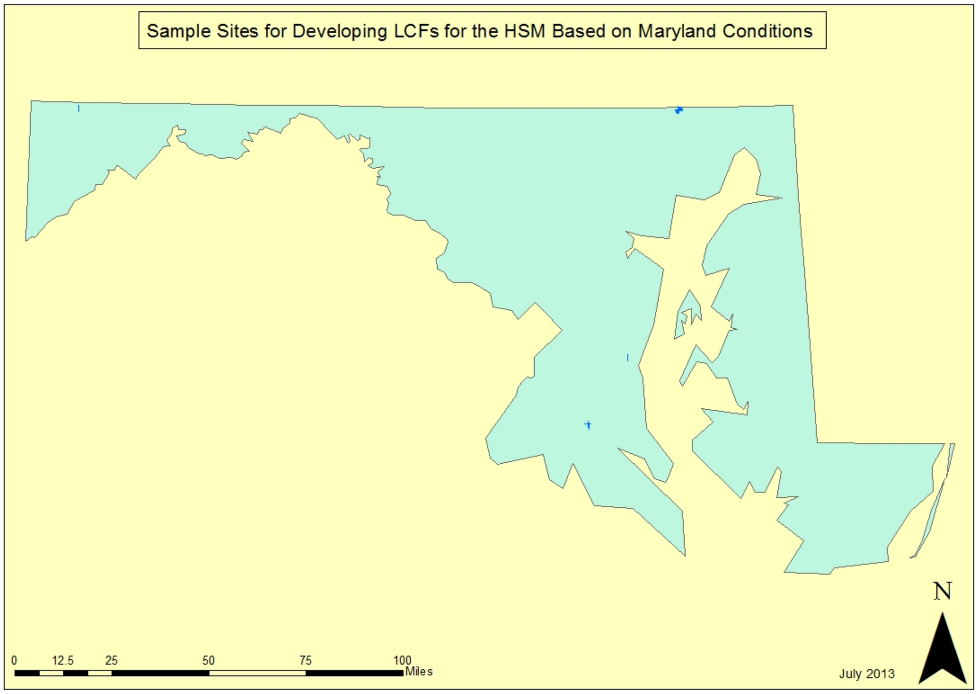


Figure 17. R4U Samples (19 Roadway Segments)

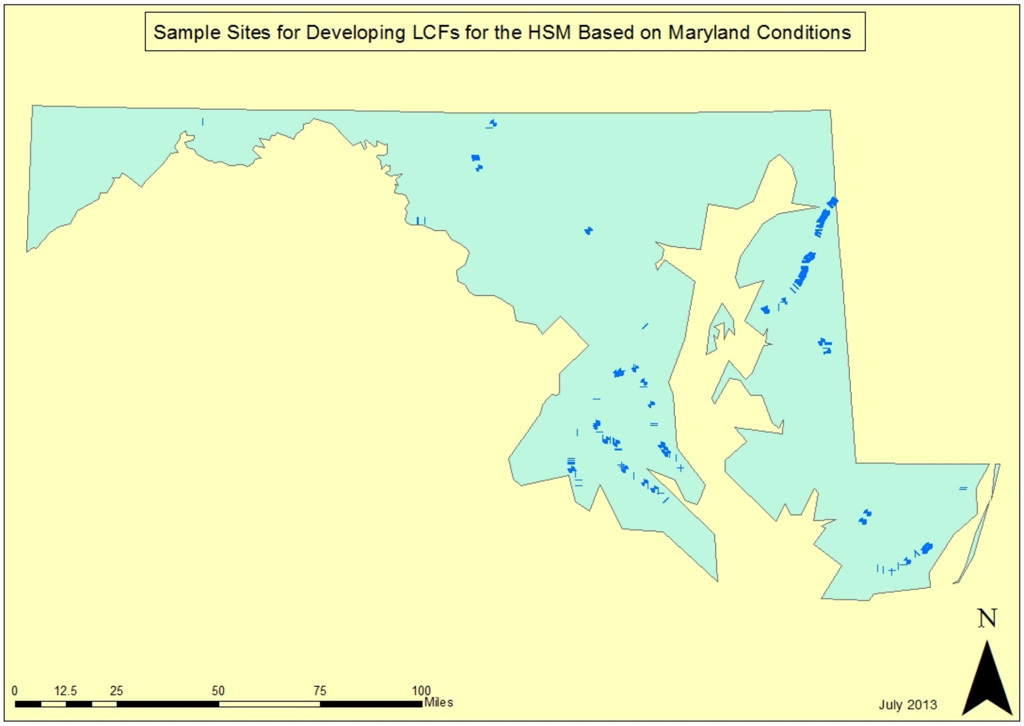


Figure 18. R4D Samples (160 Roadway Segments)

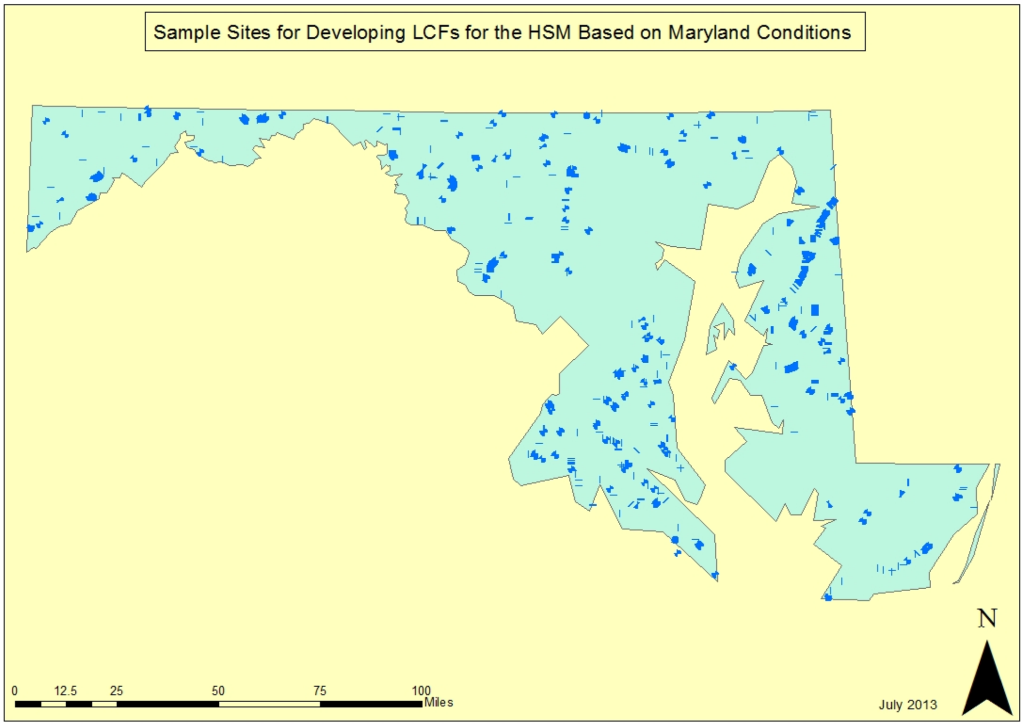


Figure 19. All Rural Samples (430 Roadway Segments)

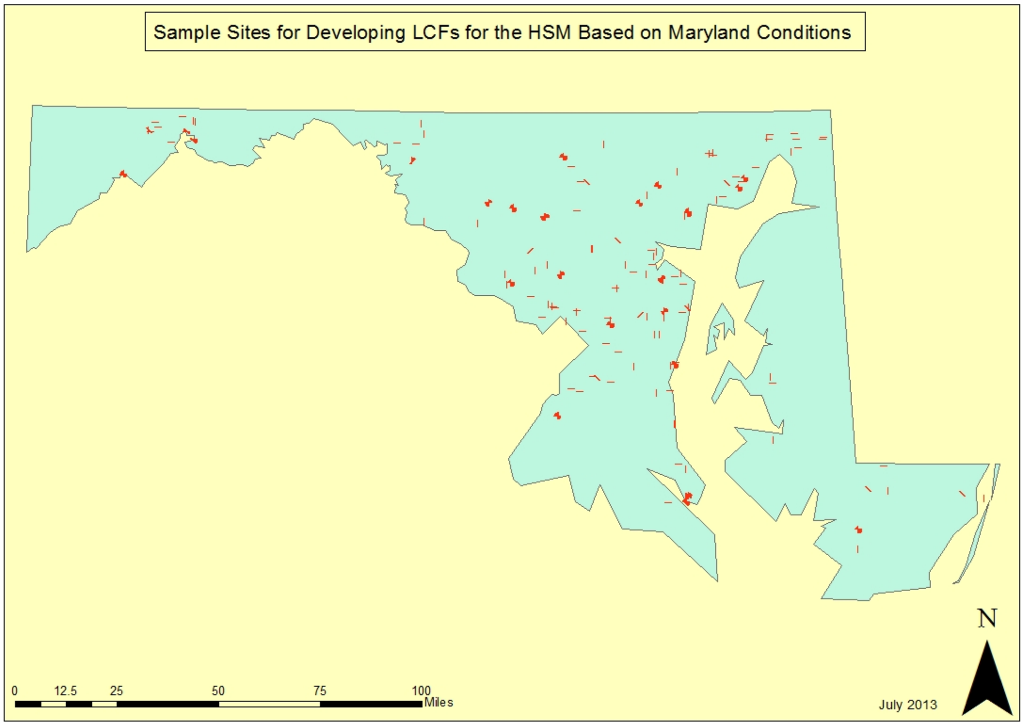


Figure 20. U2U Samples (252 Roadway Segments)

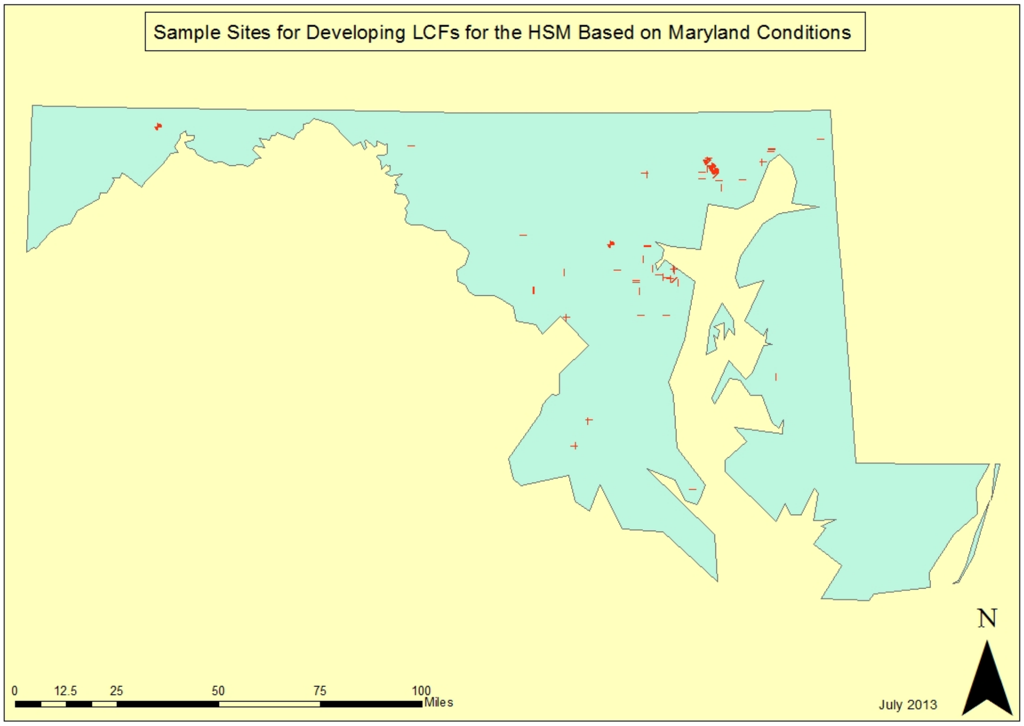


Figure 21. U3T Samples (138 Roadway Segments)

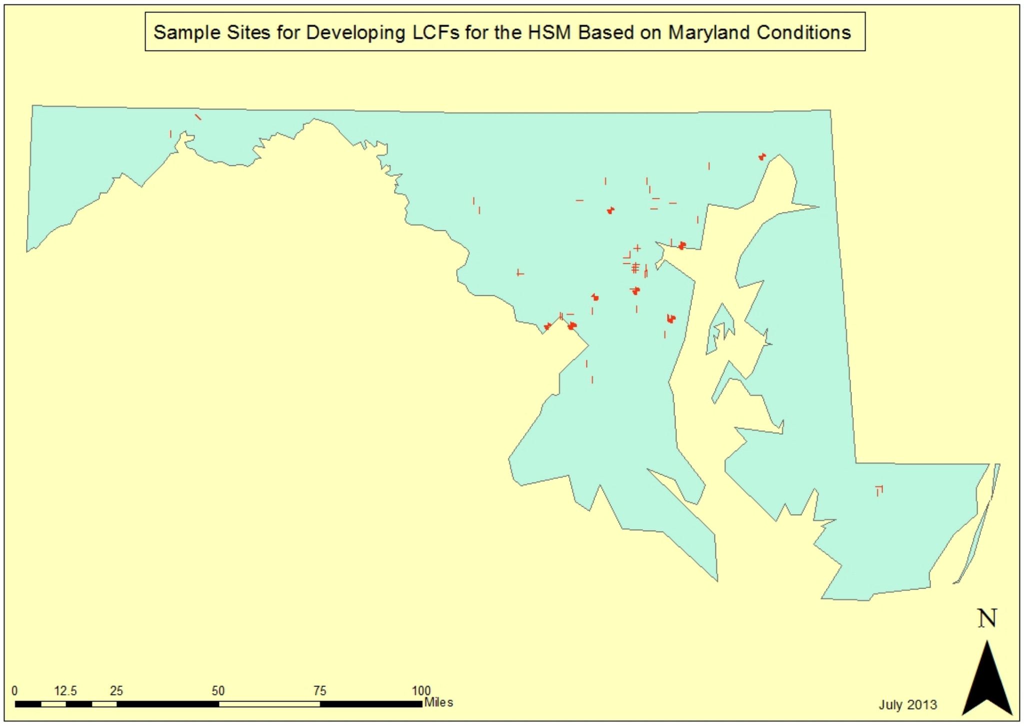


Figure 22. U4U Samples (145 Roadway Segments)

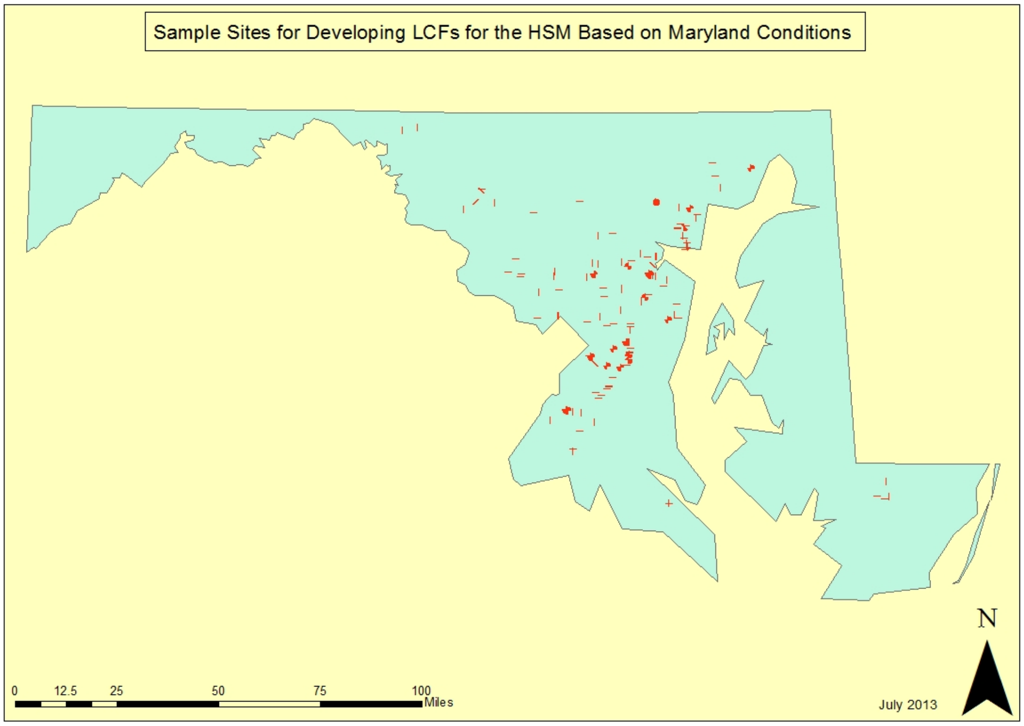


Figure 23. U4D Samples (244 Roadway Segments)

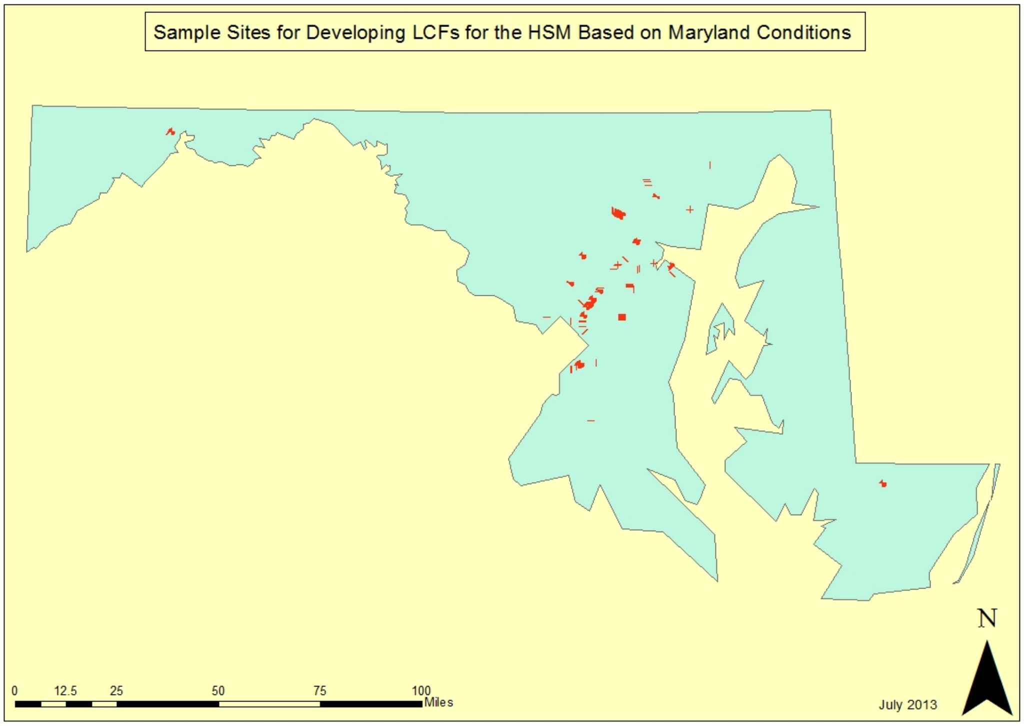


Figure 24. U5T Samples (115 Roadway Segments)

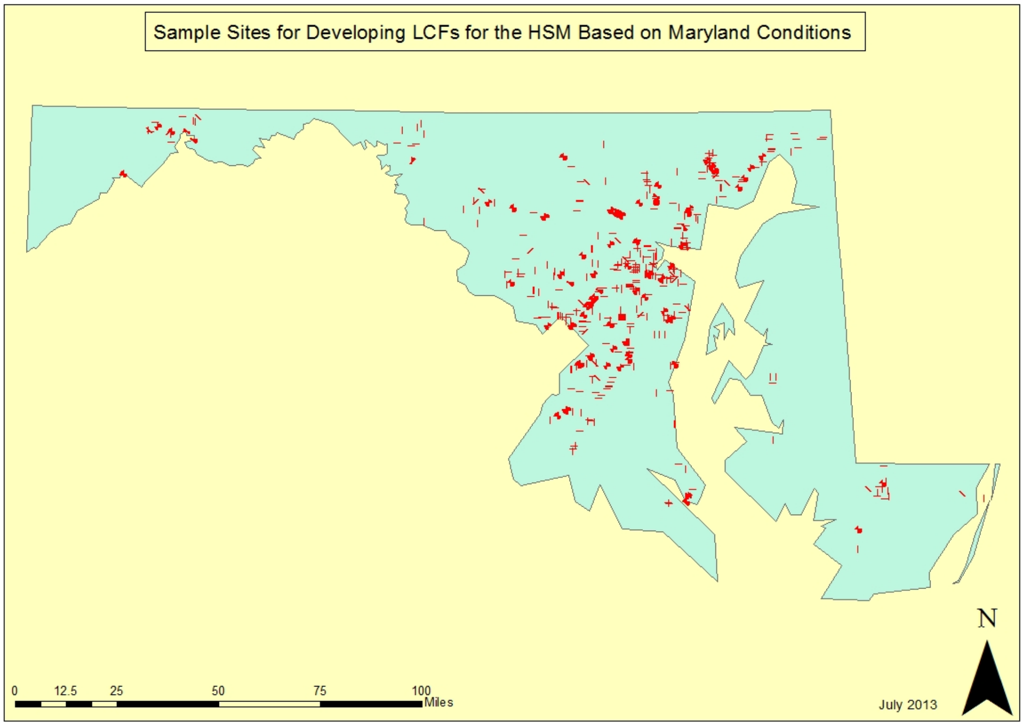


Figure 25. All Urban and Suburban Samples (894 Roadway Segments)

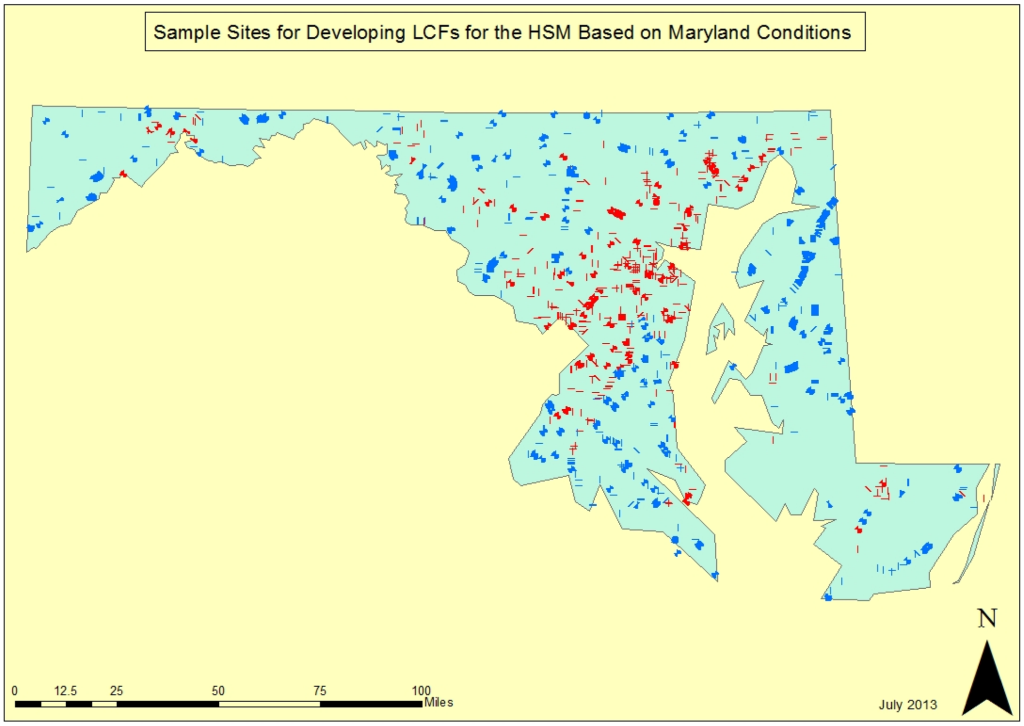


Figure 26. All Roadway Samples (1324 Roadway Segments)

Intersections

In the scenario 2 (the best sampling scenario), 1068 intersections were selected. The following figures depicted them.

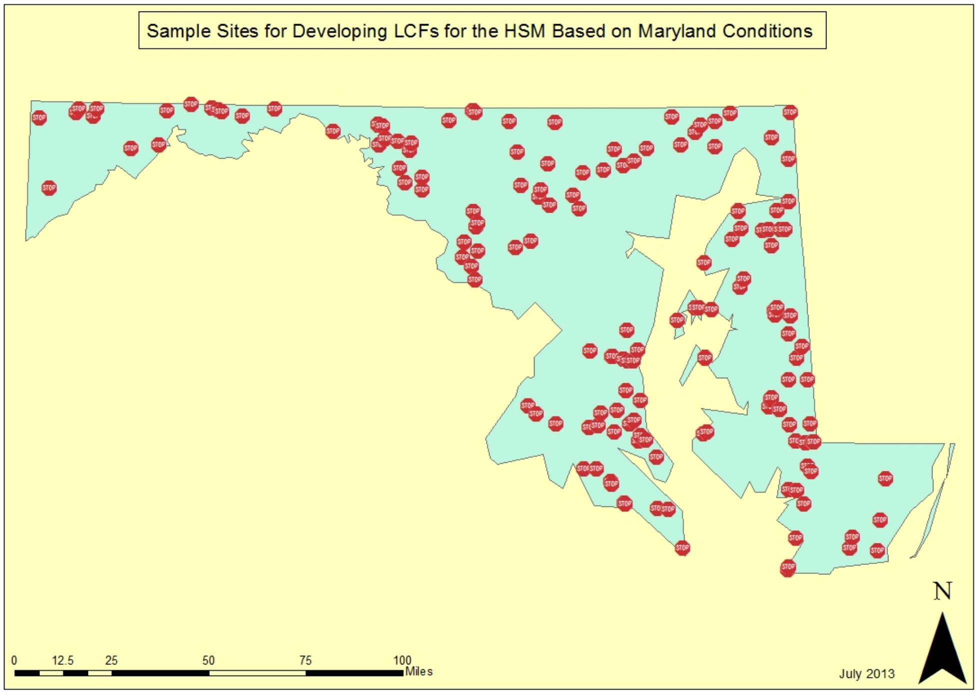


Figure 27. R23ST Samples (162 Intersections)

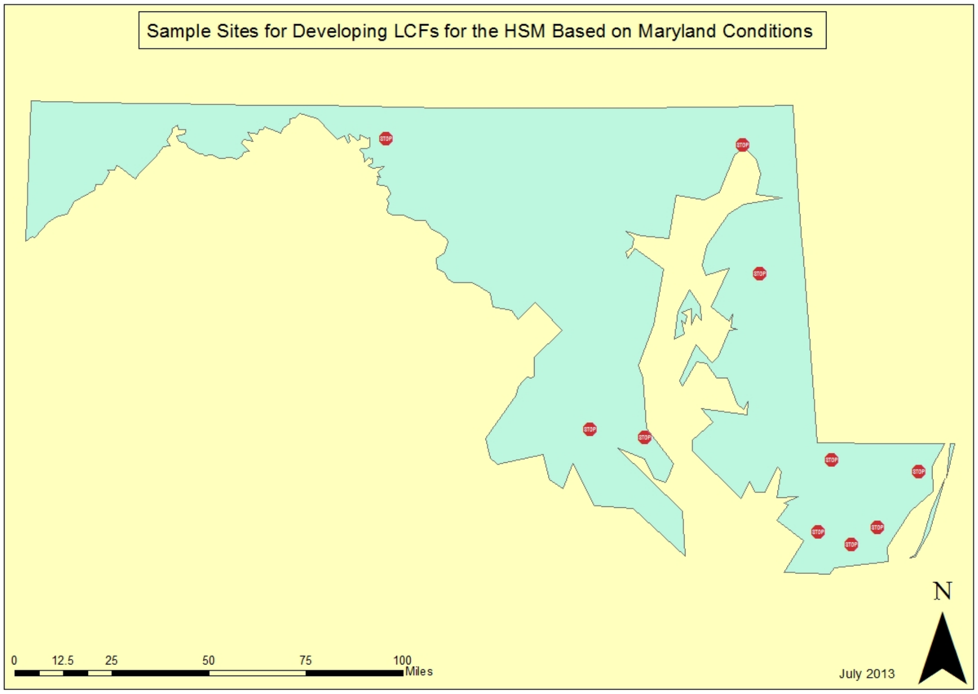


Figure 28. R24ST Samples (115 Intersections)

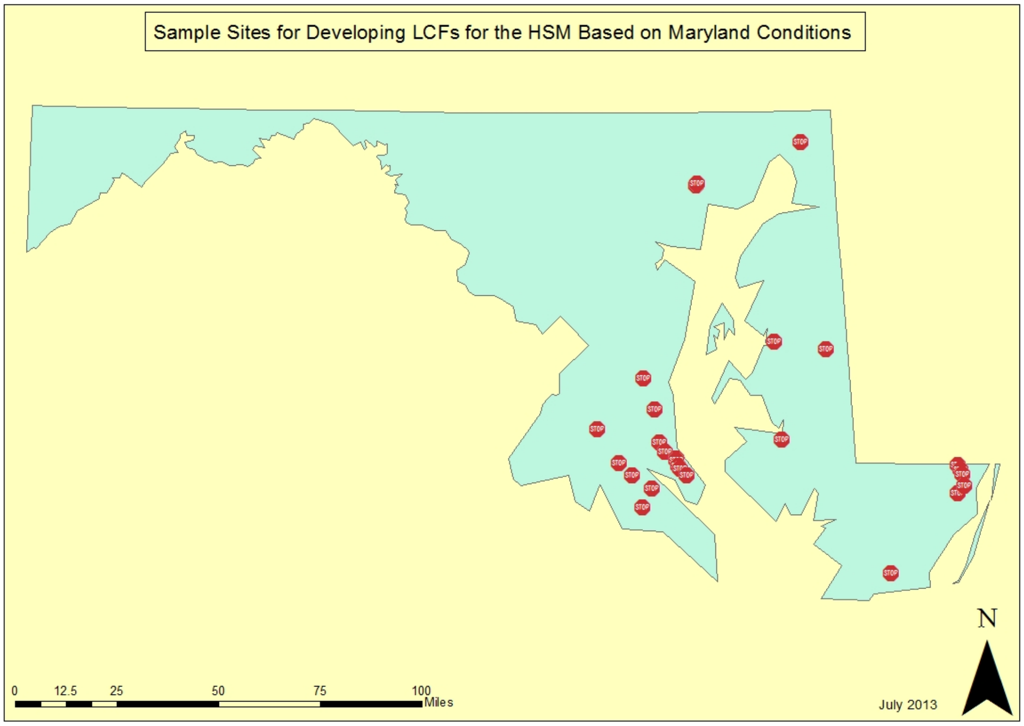


Figure 29. RM3ST Samples (26 Intersections)

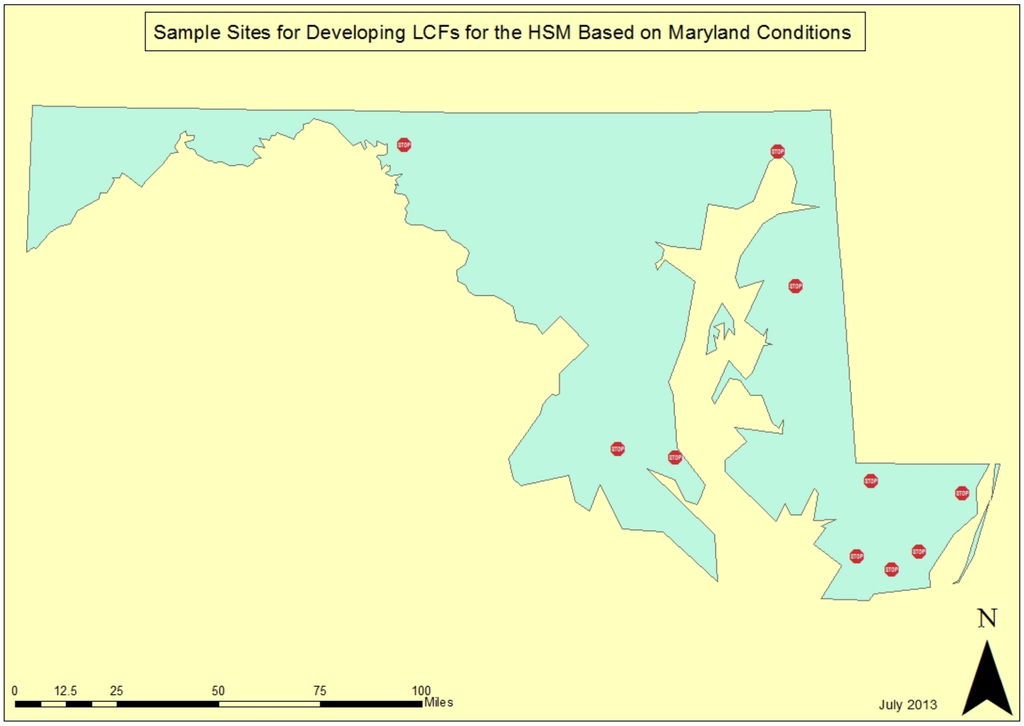


Figure 30. RM4ST Samples (10 Intersections)

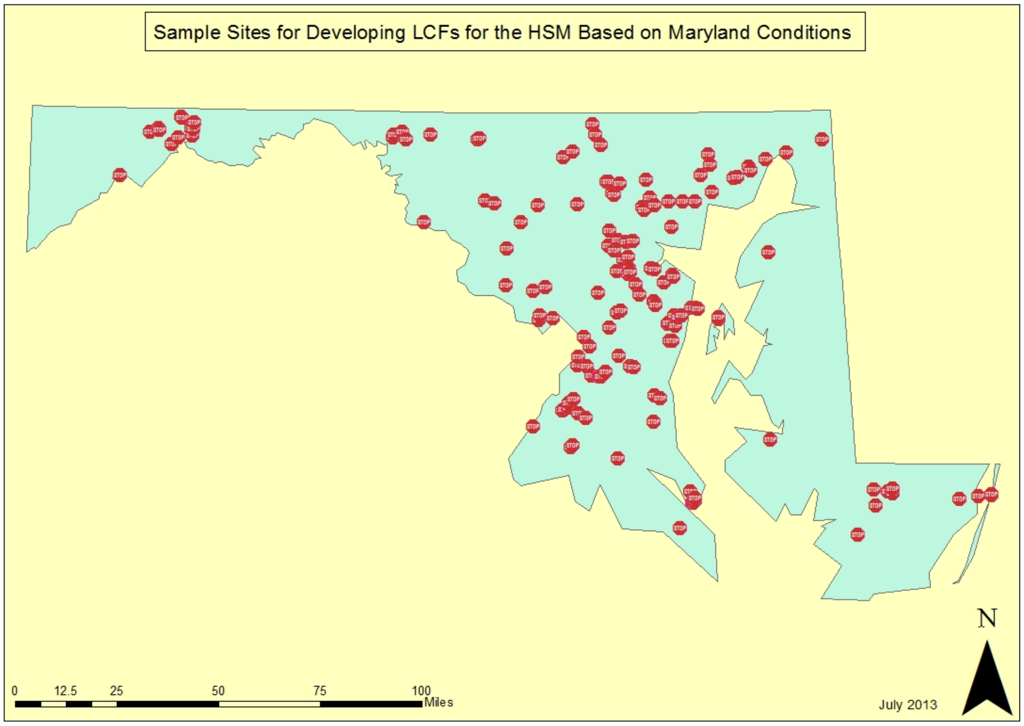


Figure 31. U3ST Samples (152 Intersections)

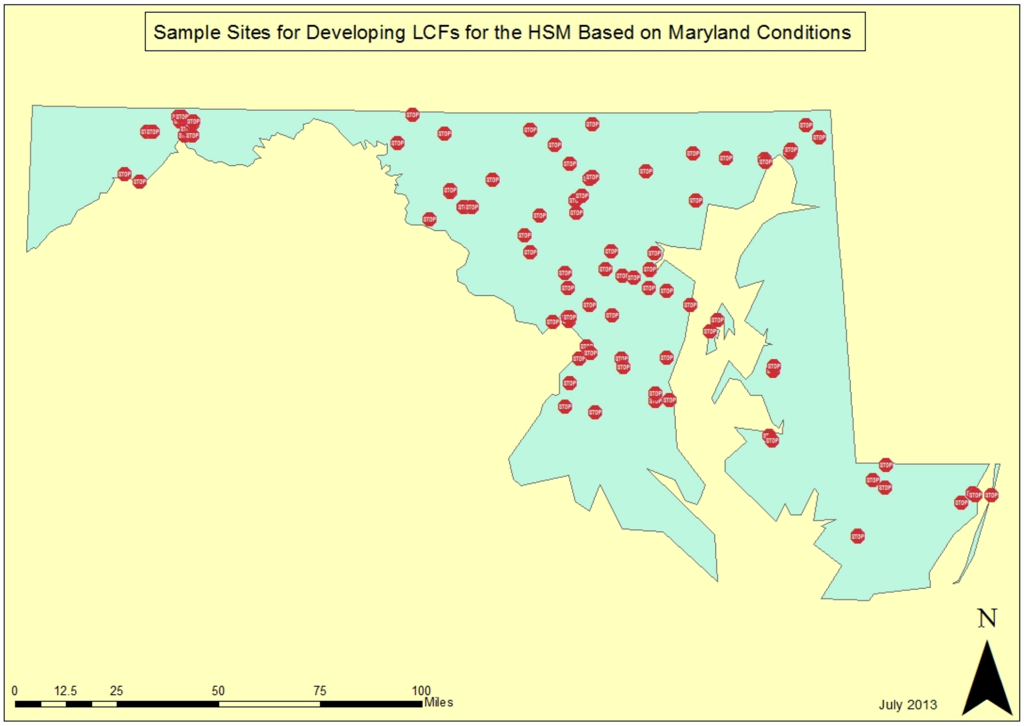


Figure 32. U4ST Samples (90 Intersections)

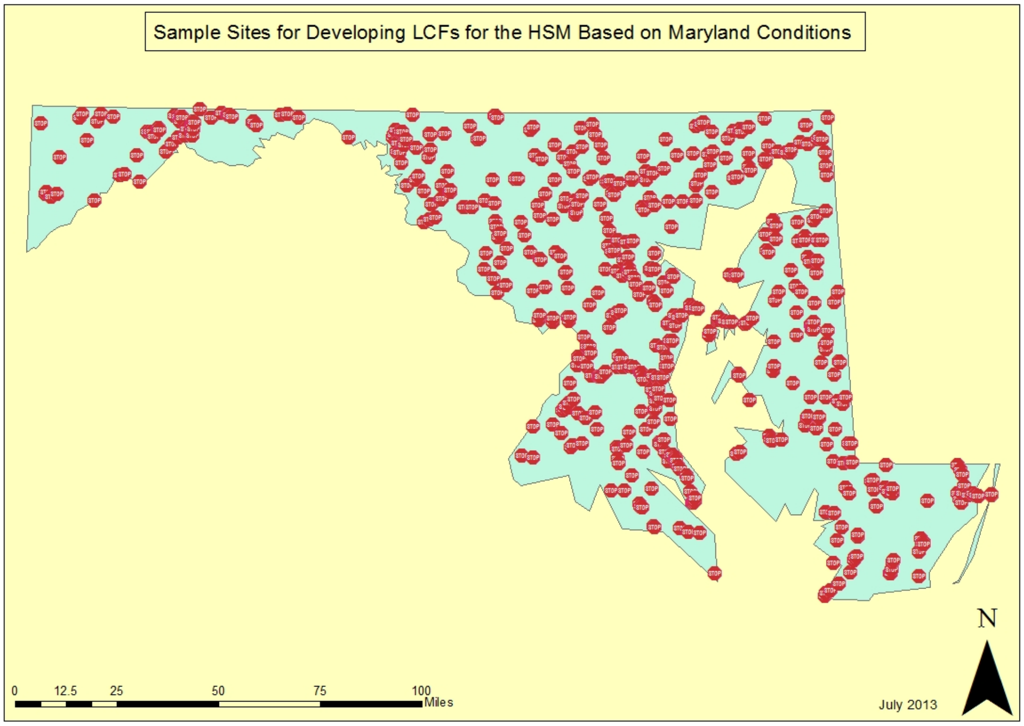


Figure 33. All Stop-Controlled Samples (555 Intersections)

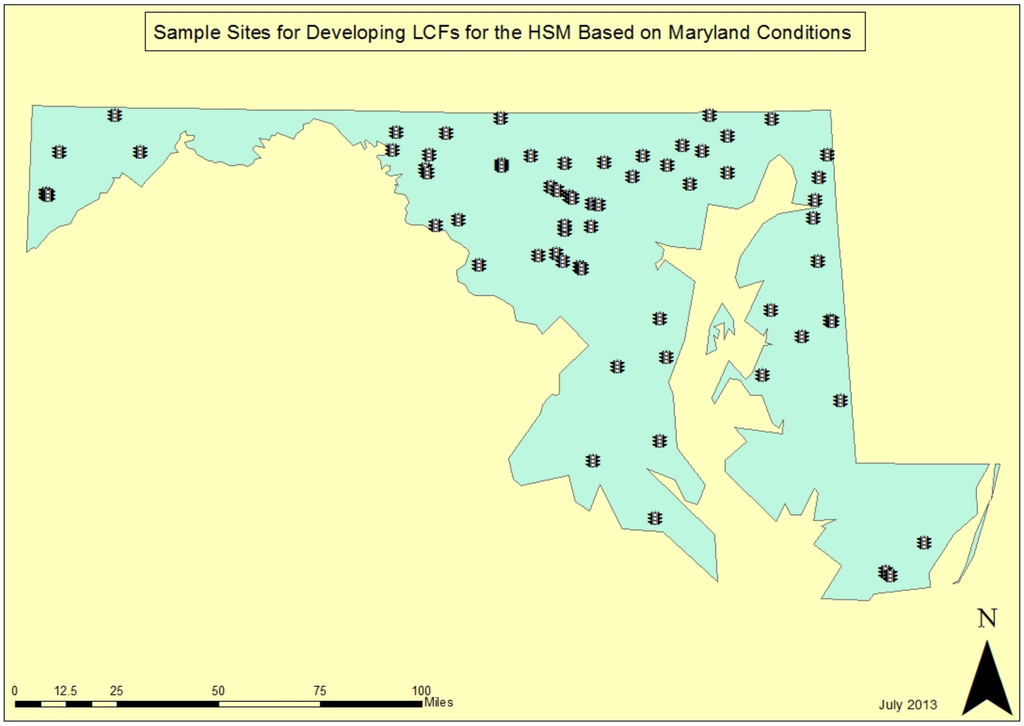


Figure 34. R24SG Samples (67 Intersections)

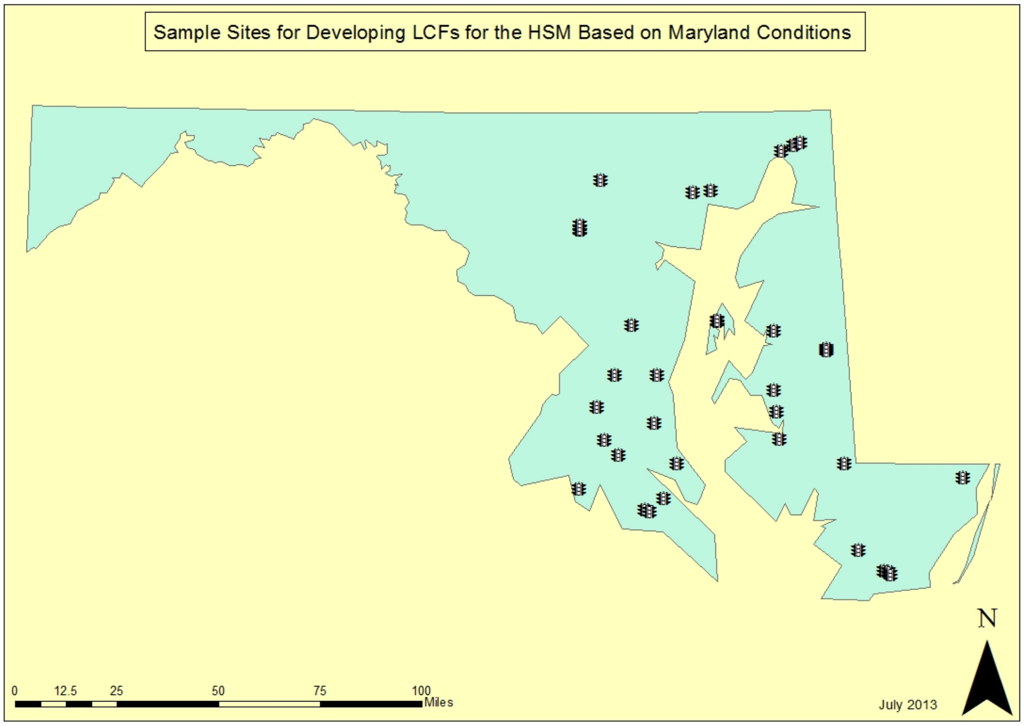


Figure 35. RM4SG Samples (35 Intersections)

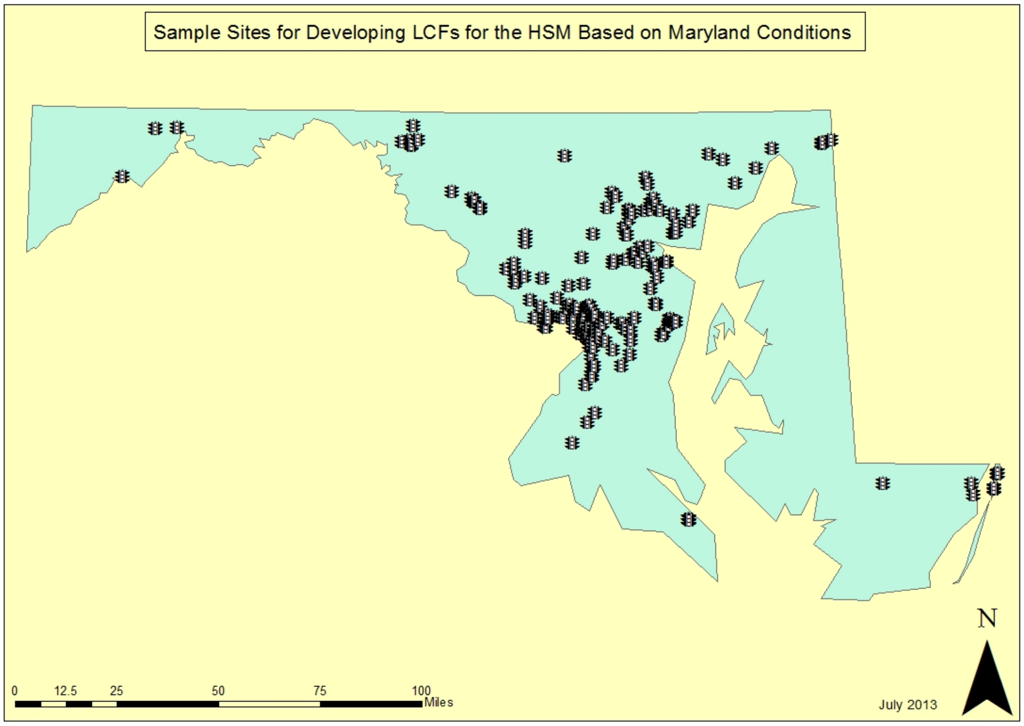


Figure 36. U3SG Samples (167 Intersections)

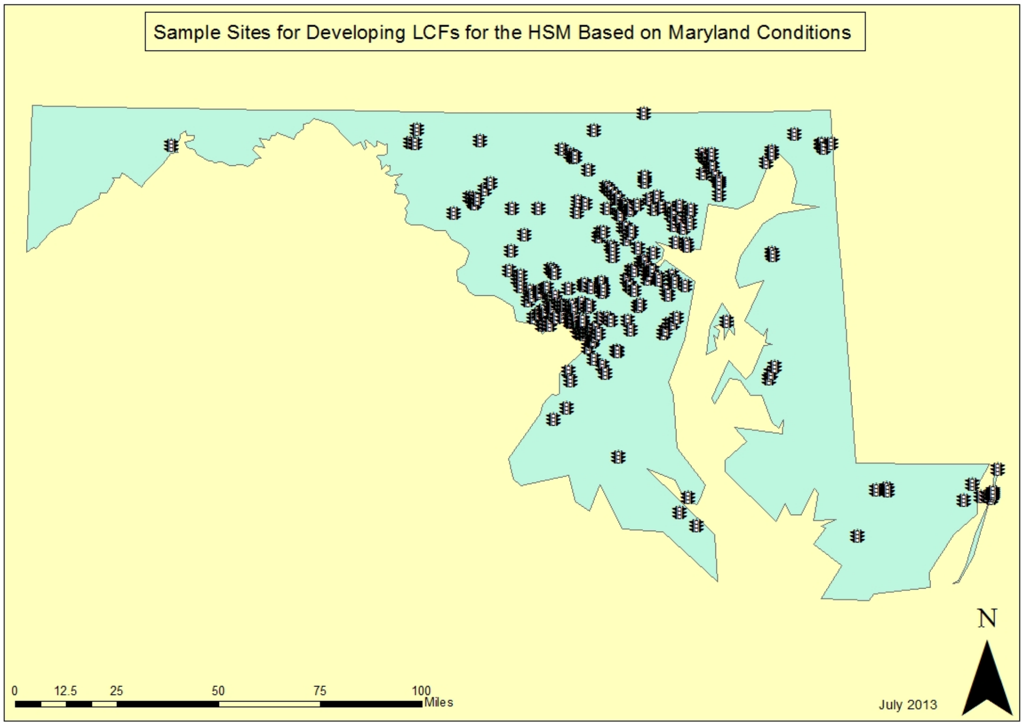


Figure 37. U4SG Samples (244 Intersections)

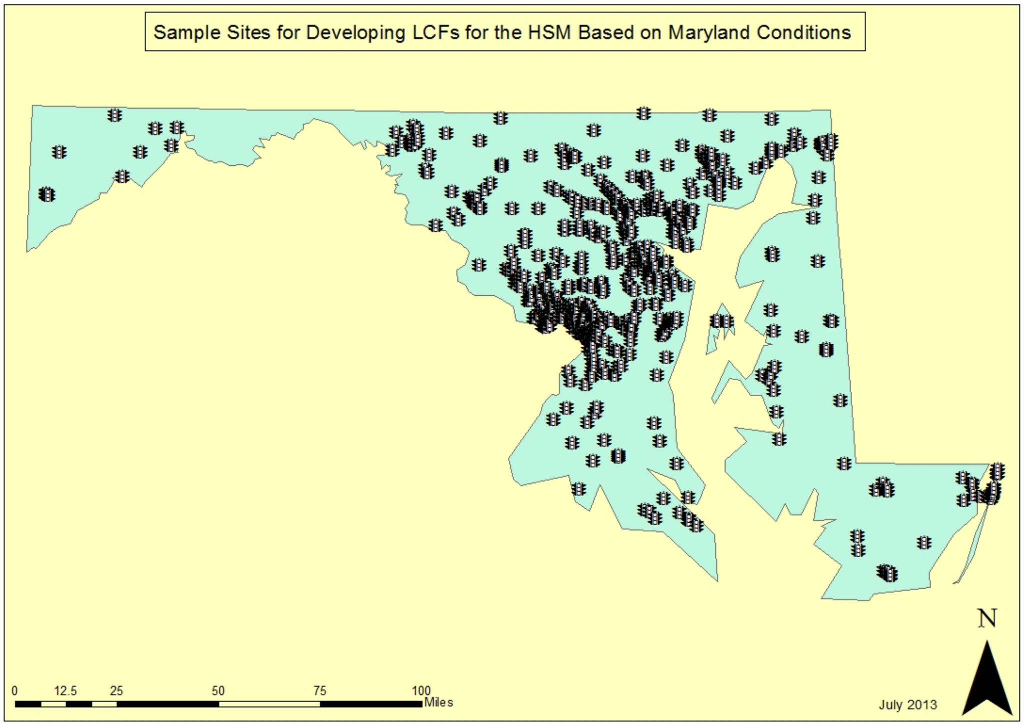


Figure 38. All Signalized Samples (555 Intersections)

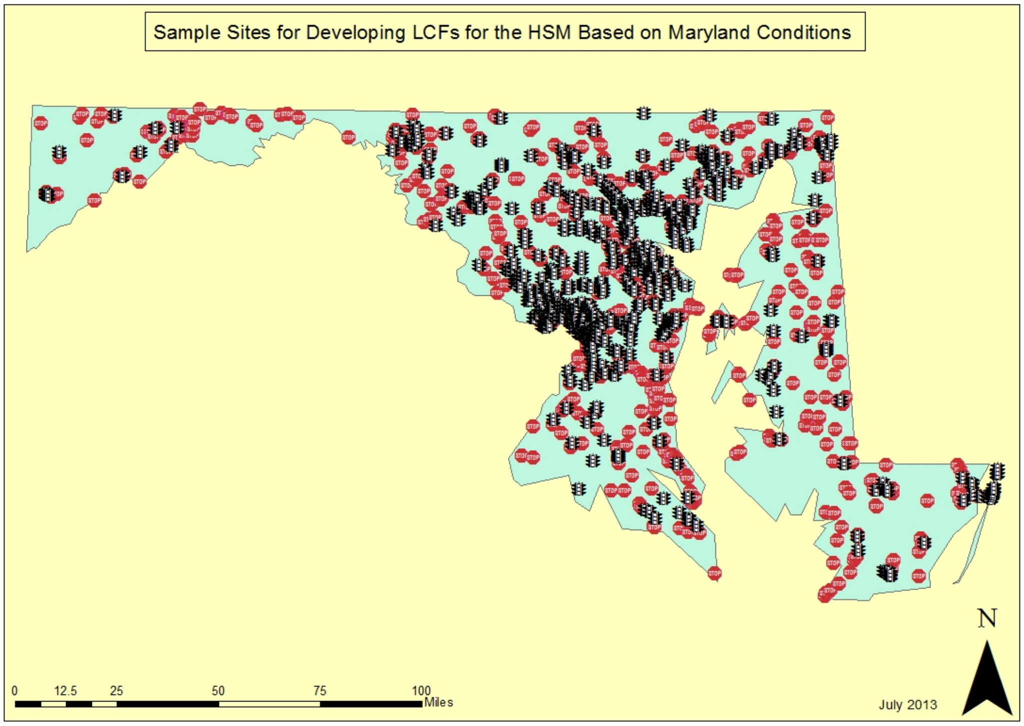


Figure 39. All Intersection Samples (1068 Intersections)

REFERENCES

Alluri, Priyanka. *Development of Guidance for States Transitioning to New Safety Analysis Tools.* Doctoral Dissertation, Clemson University, 2010.

American Association of State Highway and Transportation Officials. *Highway Safety Manual.* 1st. Washington, D.C.: AASHTO, 2010.

Banihashemi, Mohammad. "Highway Safety Manual Calibration Dataset Sensitivity Analysis." *TRB: 91st Annual Meeting.* Washington, D.C., 2012.

Chambers, Ray, and Robert Clark. "Chapter 4: Stratified Population." In *An Introduction to Model-Based Survey Sampling with Applications*. Oxford University Press, 2012.

Crash Modification Factors Clearninghouse. *Crash Modification Factors Clearninghouse.* n.d. www.cmfclearinghouse.org (accessed July 2, 2013).

Daniel, Johnnie. "Chapter 5: Choosing the type of Probability Sampling." In *Sampling Essential, Practical Guidelines for Making Sampling Choices*. Howard University, SAGE Publications, Inc., 2012.

Diez, D. M., C. D. Barr, and M. Cetinkaya-Rundel. *OpenIntro Statistics.* Second. OpenIntro, 2013.

Dixon, Karen, Chris Monsere, Fei Xie, and Kristie Gladhill. *Calibrating the Future Highway Safety Manual Predictive Methods for Oregon State Highways.* Oregon Department of Transportation, 2012.

Dixson, Karen K., et al. *Highway Safety Manaual Training Materials.* NCHRP Report 715, Washington, D.C.: TRB, 2012.

Federal Highway Administration. *Highway Safety Improvement Program Manual.* Washington, D.C., January 2010.

—. *Highway Safety Manual.* n.d. http://safety.fhwa.dot.gov/hsm/ (accessed June 12, 2013).

Federal Highway Administration. "Highway Safety Manual Case Study 1: Using Predictive Methods for a Corridor Study in Idaho." Monograph, 2011.

Gecchele, Gregorio, Riccardo Rossi, Massimiliano Gastaldi, and Armando Caprini. "Data Mining Methods for Traffic Monitoring Data Analysis: A Case Study." *Procedia Social and Behavioral Sciences* 20 (2011): 455-464.

Gujarati, Damodar N. *Basic Econometrics.* 4th. Boston, MA: McGraw-Hill, 2003.

Hamidi, A., M. Fontaine, and M. Demetksy. *A Planning-Level Methodology for Identifying High-Crash Sections of Virginia's Primary System.* Virginia Department of Transportation, 2010.

Jin, Li, Chengying Xu, and Jon D. Fricker. "Comparison of Annual Average Daily Traffic Estimates: Traditional Factor, Statistica, Artifical Network, and Fuzzy Basis Neural Network Approach." *TRB 2008 Annual Meeting.* Washington, DC, 2008.

Lindsey, Charles, and Simon Sheather. "Variable Selection in Linear Regression." *The Stata Journal* 10, no. 4 (2010): 650-669.

Lowry, Michael, and Michael Dixon. *GIS Tools to Estimate Average Annual Daily Traffic.* Moscow, ID: Natioanl Institute for Advanced Transportation Technology, University of Idaho, 2012.

Lubliner, H., and S., D. Schrok. "Calibration of the Highway Safety Manual Prediction Method for Rural Kansas Highways." *TRB: 91st Annual Meeting.* Washington, D.C., 2012.

Maryland Department of Transportation. *The Maryland Strategic Highway Safety Plan 2006-2010.* Hanover, MD, September 2006.

Maryland Highway Safety Office. "Satewide Crash Summary." 2012.

Mohamad, D., K.C. Sinha, T. Kuczek, and C.F. Scholer. "Annual Average Daily Traffic Prediction Model for County Roads." *Transportation Research Record, No. 1617*, 1998: 69-77.

Mulholland, D. *Evaluation of an existing corridor using chapter 10 of the Highway Safety Manual; A feasibility study for Iowa based on two case studies of rural portions of the US 69 corridor between Ankeny and Ames.* Dissertation for Master of Science degree, Iowa State University, 2011.

Rasdorf, W., D. Findley, C. Zegeer, C. Sundstorm, and J. Hummer. "Evaluation of GIS Applications for Horizontal Curve Data Collection." *Journal of Computing in Civil Engineering*, 2012: 191–203.

Saito, Mitsuru, Bradford, K. Brimley, and Grant, G. Schultz. *Transportation Safety Data and Analysis Volume 2: Calibration of The Highway Safety Manual and Development of New Safety Performance Functions.* Utah Department of Transportation, 2011.

"Sample Stratification." *ReStore.* n.d. http://www.restore.ac.uk/PEAS/sratheory.php (accessed October 15, 2012).

Schulte, B., and B. Loudon. "Application of New Highway Safety Manual Procedures for Evaluation and Prioritization of Safety Improvement Investment in Rural Idaho." *ITE.* 2012.

Srinivasan, R., and D. Carter. *Development of Safety Performance Functions for North Carolina.* North Carolina Department of Transportation, FHWA/NC/2010-09, 2011.

Srinivasan, Raghavan, Craig Lyon, Bhagwant Persaud, Carol Martell, and Jongdae Baek. *Methods for Identifying High Collision Concentration Locations (HCCL) for Potential Safety Improvements – Phase II: Evaluation of Alternative Methods for Identifying HCCL.* California Department of Transportation, 2011.

Srinivasan, Sivaramakrishnan, Phillip Haas, Nagendra, Singh Dhakar, Ryan Hormel, Darren Torbic, and Douglas Harwood. *Development and Calibration of Highway Safety Manual Equations for Florida Conditions.* Florida Department of Transportation, 2011.

Sun, Xiaoduan, Dan Magri, Hadi H. Shirazi, Samrat Gillella, and Long Li. "Application of the Highway Safety Manual: Louisiana Experience with Rural Multilane Highways." *The 90th TRB Annual Meeting.* Washington, D.C., 2011.

Thompson, Steven. "Chapter 11: "Stratified Random Sampling"." In *Sampling*. Wiley, 2012.

U.S. Census Bureau. "2010 Decennial Census." 2012.

Wang, Sheng-Guo, Libin Bai, and Yuanlu Bao. "Modeling and Algorithms of VMT and AADT Estimation for Community Area Traffic Netowrks." *The 18th IFAC World Congress.* Milano, Italy, 2011.

Washington, Simon P., Matthew G Karlaftis, and Fred L. Mannering. *Statistical and Econometric Methods for Transportation Data Analysis.* 1st. Boca Raton, FL: Chapman & Hall/CRC, 2003.

Wemple, E., N. Foster, and C. Bergh. "Application of the Highway Safety Manual to Predict Crash Frequency." *Australasian Transport Research Forum.* Canberra, Australia, 2010.

Zegeer, Charles V., et al. *Guidance for Implementation of the AASHTO Strategic Highway Safety Program - Volume 10: A Guide for Reducing Collisions Involving Pdestrians.* NCHRP Report 500, Volume 10, Washington, D.C.: Transportation Research Board, 2004.