Final Report

Contract No. BDK80-977-37

Improved Processes for Meeting the Data Requirements for Implementing the Highway Safety Manual (HSM) and *SafetyAnalyst* in Florida

Prepared for:

Research Center
Florida Department of Transportation
605 Suwannee Street, M.S. 30
Tallahassee, FL 32399-0450

Prepared by:

Priyanka Alluri, Ph.D., P.E., Research Associate
Dibakar Saha, M.S., Graduate Research Assistant
Kaiyu Liu, Ph.D., Senior Research Associate
Albert Gan, Ph.D., Professor

Lehman Center for Transportation Research
Florida International University
10555 West Flagler Street, EC 3680
Miami, FL 33174
Phone: (305) 348-1896
Fax: (305) 348-2802
E-mail: palluri@fiu.edu

March 2014

DISCLAIMER

The opinions, findings, and conclusions expressed in this publication are those of the authors and not necessarily those of the State of Florida Department of Transportation.

METRIC CONVERSION CHART

SYMBOL	WHEN YOU KNOW	MULTIPLY BY	TO FIND	SYMBOL
		LENGTH		
in	inches	25.400	millimeters	mm
ft	feet	0.305	meters	m
yd	yards	0.914	meters	m
mi	miles	1.610	kilometers	km
mm	millimeters	0.039	inches	in
m	meters	3.280	feet	ft
m	meters	1.090	yards	yd
km	kilometers	0.621	miles	mi
SYMBOL	**WHEN YOU KNOW**	**MULTIPLY BY**	**TO FIND**	**SYMBOL**
		AREA		
in^2	square inches	645.200	square millimeters	mm^2
ft^2	square feet	0.093	square meters	m^2
yd^2	square yard	0.836	square meters	m^2
ac	acres	0.405	hectares	ha
mi^2	square miles	2.590	square kilometers	km^2
mm^2	square millimeters	0.0016	square inches	in^2
m^2	square meters	10.764	square feet	ft^2
m^2	square meters	1.195	square yards	yd^2
ha	hectares	2.470	acres	ac
km^2	square kilometers	0.386	square miles	mi^2
SYMBOL	**WHEN YOU KNOW**	**MULTIPLY BY**	**TO FIND**	**SYMBOL**
		VOLUME		
fl oz	fluid ounces	29.570	milliliters	mL
gal	gallons	3.785	liters	L
ft^3	cubic feet	0.028	cubic meters	m^3
yd^3	cubic yards	0.765	cubic meters	m^3
mL	milliliters	0.034	fluid ounces	fl oz
L	liters	0.264	gallons	gal
m^3	cubic meters	35.314	cubic feet	ft^3
m^3	cubic meters	1.307	cubic yards	yd^3
NOTE: volumes greater than 1,000 L shall be shown in m^3.				

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Improved Processes for Meeting the Data Requirements for Implementing the Highway Safety Manual (HSM) and *SafetyAnalyst* in Florida		5. Report Date March 2014 6. Performing Organization Code
7. Author(s) Priyanka Alluri, Dibakar Saha, Kaiyu Liu, and Albert Gan		8. Performing Organization Report No.
9. Performing Organization Name and Address Lehman Center for Transportation Research Florida International University 10555 West Flagler Street, EC 3680, Miami, FL 33174		10. Work Unit No. (TRAIS) 11. Contract or Grant No. BDK80-977-37
12. Sponsoring Agency Name and Address Research Center State of Florida Department of Transportation 605 Suwannee Street, M.S. 30, Tallahassee, Florida 32399-0450		13. Type of Report and Period Covered Final Report September 2012 – March 2014 14. Sponsoring Agency Code 99700-3596-119

15. Supplementary Notes

Mr. Joseph Santos, P.E., of the State Safety Office at the Florida Department of Transportation served as the Project Manager for this project.

16. Abstract

Recent research in highway safety has focused on the more advanced and statistically proven techniques of highway safety analysis. This project focuses on the two most recent safety analysis tools, the Highway Safety Manual (HSM) and *SafetyAnalyst*. Meeting the data requirements is considered the most challenging task in implementing these tools.

In the case of HSM, many of the data variables needed for deriving the HSM calibration factors are currently unavailable in Florida's roadway characteristics inventory (RCI) database. This project attempts to identify and prioritize influential calibration variables for data collection and to determine the minimum sample sizes to estimate reliable calibration factors. For each facility type included in the HSM, this project applied the random forest technique to rank both the required and desired variables based on their importance. The variables were categorized into three groups: variables of primary importance; variables of secondary importance; and variables of lesser importance. The minimum sample sizes to estimate reliable calibration factors for different facility types were also determined. It was found that the minimum sample size of 30-50 sites with at least 100 crashes per year, as recommended by HSM, is insufficient to achieve the desired accuracy for nearly all facility types.

Compared to HSM, *SafetyAnalyst* has lesser but different data requirements. Two major efforts in applying *SafetyAnalyst* involve the conversion of local data into the strict data format required by *SafetyAnalyst* and the development of jurisdiction-specific safety performance functions (SPFs). This project developed a software program to convert the crash and roadway data for Florida's state roads to "import" files used by *SafetyAnalyst*. In addition, this project also developed the SPFs for unsignalized intersections to supplement those of other facilities which were developed under a separate project. For demonstration, *SafetyAnalyst* was applied using Florida's data to identify high crash locations. Recommendations for deploying *SafetyAnalyst* are also provided.

17. Key Word Highway Safety, Minimum Sample Size, Calibration Factors, Data Conversion, Statistical Methods, Safety Performance Functions		18. Distribution Statement	
19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 135	22. Price

Form DOT F 1700.7 (8-72)　Reproduction of completed page authorized

ACKNOWLEDGEMENTS

This research was funded by the Research Center of the Florida Department of Transportation (FDOT) under the direction of Mr. Darryll Dockstader. We are particularly grateful to our Project Manager, Mr. Joseph Santos, P.E., of the FDOT State Safety Office for his guidance and support throughout the project.

We would like to extend our earnest thanks to Dr. Kirolos Haleem, Research Associate at FIU, for his advice on random forest technique. We would also like to thank the following graduate research assistants at the FIU Lehman Center for Transportation Research (LCTR) for their assistance in data collection: Ms. Homa Fartash, Ms. Maria Guevara, Mr. Hussam Hadi, Mr. Mohammad Lavasani, Ms. Katrina Meneses, Ms. Stephanie Miranda, Ms. Anita Pourji, Mr. Asif Raihan, and Mr. Eazaz Sadeghvaziri.

A special thanks is due to Ms. Vicki Morrison of the FDOT Research Center for her review and editorial input of this report. We would also like to thank Ms. Sandra Bell, also of the FDOT Research Center, for her administrative support throughout the project.

EXECUTIVE SUMMARY

Recent research in highway safety has focused on the more advanced and statistically proven techniques of highway safety improvement. The Highway Safety Manual (HSM), *SafetyAnalyst*, and the Interactive Highway Safety Design Model (IHSDM) are the three major safety analysis tools that have the potential to define a new era in highway safety. This project focuses on the two most recent tools, the HSM and *SafetyAnalyst*.

Meeting the data requirements is considered the most challenging task in implementing these tools. In the case of HSM, many of the data variables needed for deriving the calibration factors are currently unavailable in Florida's roadway characteristics inventory (RCI) database. This project identified and prioritized influential calibration variables for data collection and determined the minimum sample sizes to estimate reliable calibration factors.

Compared to HSM, *SafetyAnalyst* has lesser but different data requirements. Two major efforts in applying *SafetyAnalyst* involve conversion of local data into the strict data format required by *SafetyAnalyst* and development of agency-specific safety performance functions (SPFs). This project developed a new conversion program to automatically generate *SafetyAnalyst* import files. It also developed SPFs for unsignalized intersections to supplement those of other facilities which were developed under a separate project. Finally, it outlined the *SafetyAnalyst* application process to be followed to identify high crash locations.

Prioritization of HSM Data Variables

Random forest technique was used to rank the required and desired variables based on their importance. Tables E-1 and E-2 give the summary of the ranking of the variables for segment and intersection subtypes, respectively.

Determination of Minimum Sample Size to Estimate HSM Calibration Factors

For the different facility types discussed in the HSM, the minimum sample sizes to estimate reliable calibration factors were recommended based on the probability that the estimated calibration factor would lie within 10% of the actual calibration factor (which is calculated from the entire data set). Tables E-3 and E-4 give the recommended minimum sample sizes required to estimate reliable calibration factors for segment and intersection subtypes, respectively. As can be observed from the tables, the HSM-recommended sample size of 30-50 sites with at least 100 crashes per year was found to be much lower than the sample sizes recommended in this study.

In addition to determining the minimum sample sizes to yield reliable calibration factors, two types of sampling procedures were researched. The calibration factors estimated from simple random sampling and stratified sampling were compared using a two-sample t-test to determine whether or not the calibration factors estimated from a stratified sample were significantly different from the calibration factors estimated from a simple random sample of the entire state data. The results indicated that it is acceptable to obtain the recommended sample size through a simple random sampling procedure from the entire state data.

Table E-1: Summary of the Ranking of Variables for Segments

Data Variable	Site Subtype[1,2]						
	R2U	R4D	SU2U	SU3T	SU4U	SU4D	SU5T
Segment length	1	1	1	1	1	1	1
Average Annual Daily Traffic (AADT)	2	2	2	2	2	2	2
Lane width	6	7	--	--	--	--	--
Shoulder type	7	NR	--	--	--	--	--
Shoulder width	4	3	--	--	--	--	--
Presence of two-way left-turn lane (TWLTL)	10	--	--	--	--	--	--
Median width	--	4	--	--	--	6	--
Presence of lighting	8	6	6	6	6	7	6
Roadside fixed object density	--	--	4	4	4	4	5
Speed limit	--	--	7	8	7	9	7
Presence of on-street parking	--	--	8	7	8	10	NR
Presence of automated speed enforcement	13	5	9	NR	NR	8	8
Presence of passing lane	9	--	--	--	--	--	--
Presence of short four-lane section	11	--	--	--	--	--	--
Presence of centerline rumble strip	12	--	--	--	--	--	--
Roadside hazard rating	5	--	--	--	--	--	--
Driveway density	3	--	--	--	--	--	--
Number of major driveways	--	--	5	5	5	5	3
Number of minor driveways	--	--	3	3	3	3	4
Horizontal curve	NR	--	--	--	--	--	--
Vertical grade	NR	--	--	--	--	--	--

1 R2U: rural two-lane undivided; R4D: rural four-lane divided; SU2U: urban and suburban two-lane undivided; SU3T: urban and suburban three-lane with TWLTL; SU4U: urban and suburban four-lane undivided; SU4D: urban and suburban four-lane divided; SU5T: urban and suburban five-lane with TWLTL.

2 NR indicates that the variable is not ranked; -- indicates that the variable is not used for that specific site subtype.

Table E-2: Summary of the Ranking of Variables for Intersections

Data Variable	Site Subtype[1,2]		
	R3ST	SU3ST	SU4SG
Major road AADT	1	1	1
Minor road AADT	2	2	2
Intersection skew angle	5	--	--
Number of approaches with left-turn lanes	4	3	6
Number of approaches with right-turn lanes	3	4	3
Presence of lighting	6	5	NR
Presence and type of left-turn signal phasing	--	--	5
Use of Right Turn On Red (RTOR) signal operation	--	--	10
Use of red-light cameras	--	--	7
Number of bus stops within 1,000 ft	--	--	4
Presence of schools within 1,000 ft	--	--	8
Number of alcohol sales establishments within 1,000 ft	--	--	9

1 R3ST: rural two-lane three-leg stop-controlled intersections; SU3ST: urban and suburban three-leg stop-controlled intersections; SU4SG: urban and suburban four-leg signalized intersections.

2 NR indicates that the variable is not ranked; -- indicates that the variable is not used for that specific site subtype.

Table E-3: Recommended Minimum Sample Sizes for Segments

Segment Site Subtype	Total Sample Length (in miles)	Sample Size	Percent of Total Length	Number of Crashes per Year	P (Estimated CF is within 10% of Actual CF)[1]
Rural Two-way Two-lane Roadway Segments	250	250	7.1%	150	90%
Rural Four-lane Divided Arterials	175	250	14.2%	270	85%
Urban and Suburban Two-lane Undivided Arterials	102	300	16.8%	180	80%
Urban and Suburban Three-lane Arterials with TWLTL	38	200	55.5%	120	94%
Urban and Suburban Four-lane Undivided Arterials	30	150	55.9%	130	90%
Urban and Suburban Four-lane Divided Arterials	140	500	10.1%	550	93%
Urban and Suburban Five-lane Arterials with TWLTL	70	275	25%	400	88%

[1] CF refers to calibration factor.

Table E-4: Recommended Minimum Sample Sizes for Intersections

Intersection Site Subtype	Sample Size	Percent of Total Intersections	Number of Crashes per Year	P (Estimated CF is within 10% of Actual CF)[1]
Rural Two-lane Three-leg Stop-controlled Intersections	150	50.3%	85	81%
Urban and Suburban Three-leg Stop-controlled Intersections	130	40.5%	200	90%
Urban and Suburban Three-leg Signalized Intersections	50	86.2%	350	87%
Urban and Suburban Four-leg Signalized Intersections	80	17.4%	1,300	87%

[1] CF refers to calibration factor.

SafetyAnalyst Data Converter

SafetyAnalyst was developed as a cooperative effort by Federal Highway Administration (FHWA) and participating state and local agencies. The software provides a suite of analytical tools to identify and manage system-wide safety improvements by incorporating all the steps in the roadway safety management process.

One of the major hurdles in deploying *SafetyAnalyst* is its stringent data requirements. *SafetyAnalyst* requires a number of import files to be generated in line with the data requirements and format recommended by the software (Harwood et al., 2010). As such, a data conversion program was developed to convert Florida data into the standard format required by *SafetyAnalyst*. The Converter uses the following input files to generate the required *SafetyAnalyst* import files for segments, intersections, ramps, and their associated crash and traffic files:

- Node list, i.e., RDWTBL25 (in .csv format)
- Linear Reference System (LRS) file, i.e., RDWTBL31 (in .csv format)
- Roadway Characteristics Inventory (RCI) database (in .accdb format)
- Crash-level (i.e., RDWTBL50) and vehicle-level (i.e., RDWTBL51) data for all the analysis years (in .txt format)

The Converter also has the capability to generate both statewide and districtwide import files. The output from the Converter can be directly inputted into the *SafetyAnalyst* Data Management Tool.

Florida-specific SPFs for *SafetyAnalyst*

To perform network screening, *SafetyAnalyst* implements the empirical Bayes (EB) method, which requires the use of SPFs. *SafetyAnalyst* is equipped with a set of national default SPFs and the software calibrates the default SPFs to represent the agency's safety performance. However, agencies are recommended to develop agency-specific SPFs whenever possible. It is believed that the agency-specific SPFs represent the agency data better than the national default SPFs calibrated to agency data. Gan et al. (2012) developed Florida-specific SPFs for segments, signalized intersections, and ramps. In this project, Florida-specific SPFs for unsignalized intersections were developed using 2011 RCI data and crash and traffic data from 2007-2010 for both total and fatal and injury (F+I) crashes. To facilitate the inclusion of as many unsignalized intersections as possible, the following Florida-specific unsignalized intersection subtypes were created: rural three-leg unsignalized intersections; rural four-leg unsignalized intersections; urban three-leg unsignalized intersections; and urban four-leg unsignalized intersections.

Application of *SafetyAnalyst*

Florida's *SafetyAnalyst* application process includes the Data Converter, the components of *SafetyAnalyst*, and the Geographic Information System (GIS) Tool. The Converter automatically generates *SafetyAnalyst* import files, and the GIS Tool spatially displays high crash locations identified by *SafetyAnalyst*. The entire *SafetyAnalyst* application process requires users to follow the steps below to identify high crash locations:

1. Input the data into the Data Converter, and run the Converter. It generates *SafetyAnalyst* import files for the roadway inventory and their associated crash and traffic data.
2. In the *SafetyAnalyst* Administration Tool, update the Florida-specific site subtypes and the Florida-specific SPFs.
3. Import the output files from the Converter into the Data Management Tool in *SafetyAnalyst.*
4. Run the import, post-process, and calibration steps in the *SafetyAnalyst* Data Management Tool.
5. Open the *SafetyAnalyst* Analytical Tool, and create the list of sites to be analyzed. For example, users can create a site list including freeways, signalized intersections, etc.
6. Run the Network Screening Module in the Analytical Tool.
7. Export the output from the Network Screening Module into .csv format.
8. Use the GIS Tool to generate high potential location maps.

Deployment of *SafetyAnalyst*

With the development of Florida-specific SPFs and the completion of a software tool to convert Florida's state road data to the data format required by *SafetyAnalyst*, Florida is now ready to deploy *SafetyAnalyst*. For the first time, Florida can have a standard system to consistently conduct safety analysis across the state. The first step in deploying *SafetyAnalyst* is to make available the *SafetyAnalyst* data sets to the district officials and their consultants. These data sets can be distributed through download from the Florida Traffic Safety Portal.

While *SafetyAnalyst* has been designed to be user-friendly, it is a relatively complex system. New users of the system would clearly benefit from a technical workshop that includes hands-on training. Accordingly, it is recommended that Florida Department of Transportation (FDOT) allocate resources to develop and conduct such a workshop at the district offices. The workshops can be made available to district safety officials, their consultants, and officials from local agencies interested in the potential use of the system. It is also recommended that FDOT provide technical support to the user community and continue to update the *SafetyAnalyst* data sets as new data become available.

In the longer run and after *SafetyAnalyst* is successfully deployed at the districts, FDOT may consider expanding its deployment to local agencies. This could be accomplished by modifying the *SafetyAnalyst* data converter from one that is based on FDOT's RCI data to one that is based on FDOT's All-Roads map.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS/ABBREVIATIONS

AADT	Average Annual Daily Traffic
AASHTO	American Association of State Highway and Transportation Officials
CAR	Crash Analysis Reporting
CF	Calibration Factor
CMF	Crash Modification Factor
CSV	Comma Separated Value
CV	Coefficient of Variation
DySeg	Dynamic Segmentation
EB	Empirical Bayes
F+I	Fatal and Injury
FDOT	Florida Department of Transportation
FHWA	Federal Highway Administration
FIU	Florida International University
GIS	Geographic Information System
HCLs	High Crash Locations
IHSDM	Interactive Highway Safety Design Model
IIA	Interchange Influence Area
LRS	Linear Referencing System
NB	Negative Binomial
NCHRP	National Cooperative Highway Research Program
NHTSA	National Highway Traffic Safety Administration
PDO	Property Damage Only
PSI	Potential for Safety Improvement
RCI	Roadway Characteristics Inventory
RDWTBL	Roadway Table
RHR	Roadside Hazard Rating
RTOR	Right Turn On Red
RTM	Regression-To-the-Mean
SA	*SafetyAnalyst*
SD	Standard Deviation
SHS	State Highway System
SPF	Safety Performance Function
TWLTL	Two-Way Left-Turn Lane
UBR	Unified Basemap Repository
VRICS	Visual Roadway Inventory Collection System

CHAPTER 1
INTRODUCTION

1.1 Background

Recent research in highway safety has focused on the more advanced and statistically proven techniques of highway safety improvement. The Highway Safety Manual (HSM), *SafetyAnalyst*, and the Interactive Highway Safety Design Model (IHSDM) are the three major safety analysis tools that have the potential to define a new era in highway safety. This research project focuses on the two most recent tools, the HSM and *SafetyAnalyst*.

1.1.1 Highway Safety Manual

The HSM marks a shift in the approach of practitioners and administrators toward more robust measures of improving highway safety by moving toward statistically proven quantitative analyses. Almost three years after its release, states and local agencies are still struggling with its implementation. Meeting the data requirements is the most challenging task in the initial stages of the HSM implementation.

The HSM provides analytical tools for quantifying the effects of potential changes at individual sites on rural two-lane roads, rural multilane highways, and urban and suburban arterials; however, the data needs are significant. Very detailed roadway characteristics and crash information is required to conduct the empirical Bayes (EB) analysis recommended by the HSM and to derive the calibration factors that are required to accurately represent the agency's safety performance. For example, to analyze intersections, data on area type, number of lanes, traffic volume, number of legs, traffic control type, intersection skew angle, number of approaches with left-turn and right-turn lanes, intersection sight distance, terrain, lighting, right-turn-on-red (RTOR), left-turn signal phasing, red-light cameras, number of bus stops, schools, and alcohol sales establishments, pedestrian activity level, and maximum lanes crossed by pedestrians are needed. Many of these variables are currently unavailable in Florida's roadway characteristics inventory (RCI) database. With these data limitations, the two main questions to be answered are which variables to collect and how much data for each variable to collect?

First, collecting and maintaining all the data variables on the entire road network for the purpose of the HSM implementation is both costly and unnecessary. Therefore, a process to streamline the data requirements without compromising the quality of analysis could result in major cost savings in meeting the data requirements. Given that not all of the variables are likely to have the same impact on safety predictions, it becomes beneficial to assess and rank the impact of each variable on safety predictions. The ranking will help prioritize the additional data to be collected such that the benefit is greatest.

Second, the manual recommends deriving calibration factors using randomly selected 30-50 roadway sites that experienced a minimum total of 100 crashes per year (AASHTO, 2010a). Given the fact that the minimum sample size is a function of sample variance, this recommendation is clearly questionable as roadway characteristics of different roadway types are likely to have different levels of homogeneity. For example, the roadway characteristics on freeways are more likely to be more homogeneous and would require a smaller sample size to achieve the same level of accuracy than local arterial streets, which are more likely to have more varying roadway characteristics. In other words, the recommended 30-50 sites could be too many for some roadway types yet insufficient for others. Similarly, the minimum total of 100 crashes per year is also questionable given the fact that the number of crashes vary widely across different roadway types. For example, intersection locations would generally experience many more crashes than mid-block segments. Similarly, arterial streets usually experience many more crashes than freeways. Research is therefore needed to determine the minimum sample sizes for different roadway types.

1.1.2 SafetyAnalyst

SafetyAnalyst was developed as a cooperative effort by Federal Highway Administration (FHWA) and participating state and local agencies. *SafetyAnalyst* is a state-of-the-art analytical tool for making system-wide safety decisions. The software incorporates all the steps in the roadway safety management process for all the three facility types: segments, intersections, and ramps (AASHTO, 2010b). Although *SafetyAnalyst* is often advertised as a companion to Part B of the HSM, the two tools are fundamentally different. Gan et al. (2012) provides a discussion on the differences between the two tools.

One of the major advantages of *SafetyAnalyst* over the existing traditional site selection methods is that *SafetyAnalyst* implements the EB method which requires the use of safety performance functions (SPFs). The software is equipped with a set of national default SPFs, and the software calibrates the default SPFs to represent the agency's safety performance. However, agencies are recommended to generate agency-specific SPFs as they represent the agency data better than the national default SPFs calibrated to agency data. As part of Project BDK80-977-07, Florida-specific SPFs were developed for segments, signalized intersections, and ramps. However, SPFs for unsignalized intersections were not developed due to the lack of required data in Florida's RCI database (Gan et al., 2012). Given the importance of intersection safety, an extra effort is needed to collect the missing data so that Florida-specific SPFs could be generated for unsignalized intersections.

One of the major hurdles in deploying *SafetyAnalyst* is its stringent data requirements. *SafetyAnalyst* requires a number of import files to be generated in line with the data requirements and format recommended by the software (Harwood et al., 2010). This process "of generating *SafetyAnalyst* import files is tedious because data may need to be retrieved and merged from multiple sources and significant amounts of data recoding may be required" (Alluri and Ogle, 2012b). Therefore, development of a program to automatically generate the *SafetyAnalyst* import files for the roadway inventory, crash, and traffic data could help state and local agencies adopt *SafetyAnalyst*.

1.2 Project Objectives

As mentioned earlier, this research project focuses on the HSM and *SafetyAnalyst*. For the HSM, the objectives are to identify and prioritize influential calibration variables for data collection, and to determine the minimum sample sizes to estimate reliable calibration factors. For *SafetyAnalyst*, the objectives are to develop a new conversion program to automatically generate *SafetyAnalyst* import files, to develop SPFs for unsignalized intersections to use with

SafetyAnalyst, and to describe the *SafetyAnalyst* application process to be followed to identify high crash locations.

1.3 Report Organization

The rest of the report is organized as follows. Chapter 2 describes in detail the data collection and preparation efforts undertaken as part of this project. Chapters 3 and 4 focus on the HSM. Chapter 3 discusses the methodology used to rank data variables based on their importance. It also provides the results on data variable prioritization. Chapter 4 focuses on determining the minimum sample size to estimate a reliable calibration factor. It discusses the methodology and results, and also provides recommendations on minimum sample sizes for different facility types.

Chapters 5 through 7 focus on *SafetyAnalyst.* Chapter 5 provides a discussion on *SafetyAnalyst*. Chapter 6 discusses in detail the data conversion program developed to convert Florida data into the standard format required by *SafetyAnalyst*. Chapter 7 focuses on the *SafetyAnalyst* application process to identify high crash locations. It includes a discussion on the Florida-specific SPFs developed for unsignalized intersections to be used with *SafetyAnalyst.* It also provides a sample list of high crash locations identified using *SafetyAnalyst*. Finally, Chapter 8 provides a summary of this project effort and the relevant findings, conclusions, and recommendations.

CHAPTER 2
DATA COLLECTION

This chapter describes the data collection and preparation efforts undertaken as part of this project. Table 2-1 lists the roadway characteristics data needed to perform the EB analysis discussed in the HSM. For each roadway facility type listed in Table 2-2, these variables were either retrieved from the RCI or collected from aerial images.

Table 2-1: Roadway Characteristics Data Requirements for Part C of the HSM (AASHTO, 2010a)

Roadway Segment Characteristics	Intersection Characteristics
• Area type • Segment length • Number of lanes • Functional classification • Average Annual Daily Traffic (AADT) • Median type and width • Lane width • Shoulder width and type • Presence of a concrete median barrier • Presence of passing lane • Presence of short-four lane section • Presence of two-way left-turn lane (TWLTL) • Horizontal curve location • Length, radius, superelevation of horizontal curve • Speed limit • Presence and type of parking • Vertical grade • Presence of centerline rumble strips • Roadside Hazard Rating (RHR) • Side slope • Driveway density • Number of roadside fixed objects • Average offset to roadside fixed objects • Presence of automated speed enforcement	• Area type • Number of lanes • Average Annual Daily Traffic (AADT) for major road • AADT for minor road • Number of legs • Traffic control type • Intersection skew angle • Number of approaches with left-turn lanes • Number of approaches with right-turn lanes • Intersection sight distance • Presence of lighting • Presence of Right Turn On Red (RTOR) • Presence and type of left-turn signal phasing • Presence of red-light cameras • Number of bus stops within 1,000 ft • Presence of schools within 1,000 ft • Number of alcohol sales establishments within 1,000 ft • Pedestrian activity level • Maximum number of lanes crossed by pedestrians

Table 2-2: Segment and Intersection Subtypes Discussed in the HSM

Rural Two-lane Roads	Rural Multilane Highways	Urban and Suburban Arterials
Segments		
• Two-way undivided	• Four-lane undivided • Four-lane divided	• Two-lane undivided • Three-lane with TWLTL • Four-lane undivided • Four-lane divided • Five-lane with TWLTL
Intersections		
• Three-leg stop-controlled • Four-leg stop-controlled • Four-leg signalized	• Three-leg stop-controlled • Four-leg stop-controlled • Four-leg signalized	• Three-leg stop-controlled • Four-leg stop-controlled • Three-leg signalized • Four-leg signalized

Note: TWLTL is two-way left-turn lane.

2.1 Segment Data

This section discusses the specific data elements for all the segment site subtypes discussed in the HSM, the availability of the required and desired data in Florida databases, and the process adopted to collect the missing data variables.

2.1.1 Segment Data Set

The RCI database is a comprehensive roadway inventory database which includes segments that are part of the state highway system (SHS), segments that are currently being constructed and yet to be added as part of the SHS, segments that are no longer maintained by the Florida Department of Transportation (FDOT), historic roads, local roads, exclusive roads (ramps, frontages roads etc.), etc. Segments that are currently not part of the SHS do not have complete roadway traffic, geometric, and crash data. Therefore, only those segments that are part of the SHS and recorded in the RCI as "active on the SHS" were extracted and included in the analysis. All other segments were excluded from further analysis. The available roadway segments were then categorized into the following three HSM-designated facility types: rural two-way two-lane roads, rural multilane highways, and urban and suburban arterials. The following four data variables in the RCI were used to determine the facility types and site subtypes (the name in the parentheses gives the description of the variable):

- FUNCLASS (functional classification),
- NOLANES (number of lanes),
- TYPEROAD (type of road), and
- RDMEDIAN (type of median).

The variable FUNCLASS was used to determine the area type (i.e., rural or urban). The variable has the following 12 codes for roadway classification (FDOT, 2012):

- 01: Rural principal arterial – interstate
- 02: Rural principal arterial – other (not interstate)
- 06: Rural minor arterial
- 07: Rural major collector
- 08: Rural minor collector
- 09: Rural local
- 11: Urban principal arterial – interstate
- 12: Urban principal arterial – other freeways and expressways
- 14: Urban principal arterial – other
- 16: Urban minor arterial
- 17: Urban collector
- 19: Urban local

Segments with functional classification coded as '02', '06', '07', and '08' were identified as rural roadway segments and those coded as '14', '16', and '17' were grouped as urban roadway segments. Note that freeways and local roads were not included in the analysis since the HSM does not discuss these facility types. The variable NOLANES was used to determine the directional number of lanes; segments with more than two lanes in each direction are not covered in the HSM, and therefore, were not considered in this study. The variable TYPEROAD was used to determine the roadway type, i.e., whether the roadway segments are divided or undivided. It has the following three codes (FDOT, 2012):

- 0: Not divided
- 2: Divided
- 4: One-way

Since the HSM does not provide guidelines for one-way streets, these sections were excluded from the analysis. The variable RDMEDIAN was used to identify sections with two-way left-turn lanes (TWLTLs).

2.1.2 Data Variables

Table 2-3 presents the data variables required and desired by the HSM for the three facilities: rural two-lane two-way roads, rural multilane highways, and urban and suburban arterials. For each data variable that is available in the RCI database, the table also includes the corresponding variable name in the RCI database.

2.1.3 Homogeneous Segments

Roadways are often segmented whenever there is a slight change in any of the data variables in the RCI database, resulting in relatively short segments. However, not all the variables in the RCI database are required for safety analysis. Therefore, longer homogeneous segments could be generated by defining segments using fewer data variables and by reducing the data sensitivity.

It is observed that greater sensitivity in data variables might not be necessary if the thresholds used during the analyses are less sensitive. In Florida, lane width is calculated by dividing the total surface width by the number of lanes, and is recorded to the nearest 0.1 ft. Likewise, shoulder width is recorded to the nearest 0.5 ft, and median width is recorded to the nearest foot. However, it is observed that these variables are mostly used in calculating crash modification factors (CMFs) to adjust for base conditions in the EB analysis (AASHTO, 2010a). These CMFs were generated based on 0.5 ft variations for lane width, 1 ft variations for shoulder width, and 10 ft variations for median width. For this study, the sensitivity of lane width, shoulder width, and median width were therefore reduced to 0.5 ft, 1 ft, and 10 ft increments, respectively. Table 2-4 lists these adjustments.

Table 2-3: Data Variables in the HSM for Segments (AASHTO, 2010a)

Data Variable	Facility Type			Data Variable in RCI (if available)
	Rural Two-way Two-lane Roads	Rural Multilane Highways	Urban and Suburban Arterials	
AADT	Required	Required	Required	SECTADT
Length and radius of horizontal curve and length of tangent *	Required			HRZCANGL, HRZDGCRV, HRZPTINT
Lane width	Required	Required		SURWIDTH, NOLANES
Shoulder type	Required	Required		SHLDTYPE
Shoulder width	Required	Required		SLDWIDTH
Presence of center TWLTL	Required		Required	
Median width		Required	Required	MEDWIDTH
Sideslope+*		Required		
Number of through traffic lanes	Required	Required	Required	NOLANES
Presence of lighting+	Desired	Required	Desired	
Presence of median			Required	RDMEDIAN
Driveway density+	Desired		Required	
Land-use+	Required	Required	Required	
Speed limit			Required	MAXSPEED
Presence of on-street parking+			Required	
Type of on-street parking+			Required	
Curb length with on-street parking+			Required	
Use of automated speed enforcement+	Desired	Desired	Desired	
Presence of passing lane+	Desired			
Presence of short four-lane section+	Desired			
Presence of centerline rumble strip+	Desired			
Presence of spiral transition for horizontal curves+*	Desired			
Superelevation variance for horizontal curves+*	Desired			
Percent grade*	Desired			GRACLASx
Roadside hazard rating+	Desired			
Roadside fixed object density+			Desired	
Offset to fixed object+			Desired	

+ Unavailable in the RCI database.
* Default values were assumed.

The Dynamic Segmentation (DySeg) application was used to generate homogeneous segments. DySeg is an in-house application that can generate segments based on different criteria such as segments having equal length, segments with uniform features, segments within a specified range of lengths, segments with desired roadway features and crash types, etc. Figure 2-1 shows the screen capture of the DySeg user interface used to specify the geometric variables to be considered for segmentation. The following were the data variables used to generate homogeneous segments:

- AADT,
- route type,
- area type,
- number of through lanes,
- median type and median width,
- shoulder type and width, and
- average lane width.

Table 2-4: Lane, Shoulder, and Median Widths: Measured and Rounded Values

Data Variable	Measured Width	Rounded Width
Lane Width (calculated as surface width/number of lanes)	≤ 9.2 ft	9 ft
	9.3 ft - 9.7 ft	9.5 ft
	9.8 ft - 10.2 ft	10 ft
	10.3 ft - 10.7 ft	10.5 ft
	10.8 ft - 11.2 ft	11 ft
	11.3 ft - 11.7 ft	11.5 ft
	≥ 11.8 ft	12 ft
Shoulder Width	≤ 0.5 ft	0 ft
	0.6 ft - 1.5 ft	1 ft
	1.6 ft - 2.5 ft	2 ft
	2.6 ft - 3.5 ft	3 ft
	3.6 ft - 4.5 ft	4 ft
	4.6 ft - 5.5 ft	5 ft
	5.6 ft - 6.5 ft	6 ft
	6.6 ft - 7.5 ft	7 ft
	≥ 7.6 ft	≥ 8 ft
Median Width	1 ft - 14 ft	10 ft
	15 ft - 24 ft	20 ft
	25 ft - 34 ft	30 ft
	35 ft - 44 ft	40 ft
	45 ft - 54 ft	50 ft
	55 ft - 64 ft	60 ft
	65 ft - 74 ft	70 ft
	75 ft - 84 ft	80 ft
	85 ft - 94 ft	90 ft
	≥ 95 ft	100 ft

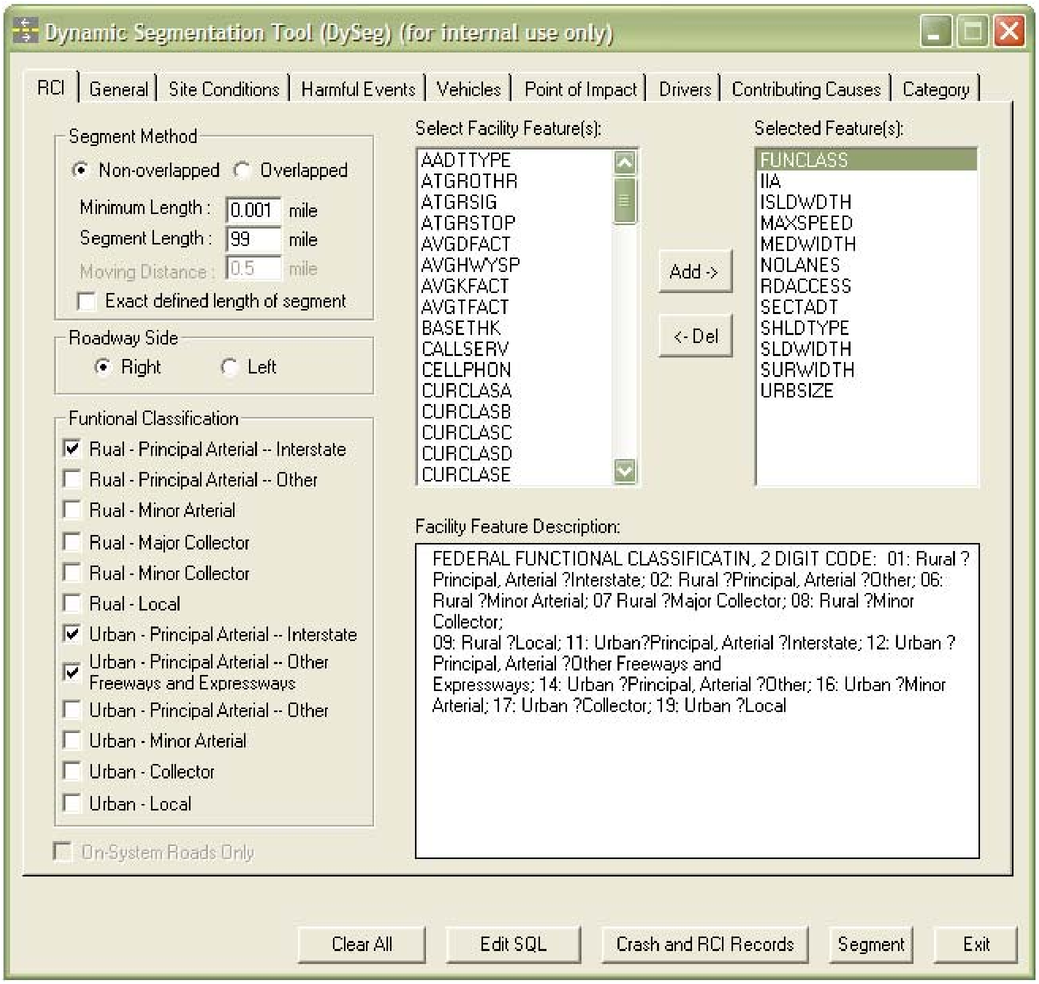


Figure 2-1: DySeg User Interface

The HSM does not recommend the minimum segment length. Based on the other states' practices on the selection of minimum segment length (Srinivasan et al., 2011 and Dixon et al., 2012), segments shorter than 0.04 miles were excluded from further analysis.

2.1.4 Data Collection

An in-house web-based data collection application, Visual Roadway Inventory Collection System (VRICS), was used to collect information on data variables that are unavailable in the RCI database. The VRICS application is developed to facilitate the process of collecting roadway data using Google Street View. Figure 2-2 shows a screen capture of the main interface of the system. The system reads a linear-referenced roadway segment/intersection, converts its coordinates to the Google Maps projection on the fly, and then displays the segment on the Street View starting from its begin milepost. Users can then let the system run the Street View through the roadway segment continuously. The system also allows users to play forward and backward at a preferred speed and can be paused to record and save observed data. After completing a segment, users can quickly have the system jump to and display the next segment to continue with data collection. Note that the map interface of the system, similar to Google Maps, provides both satellite view (i.e., aerial view) and street view to facilitate data collection.

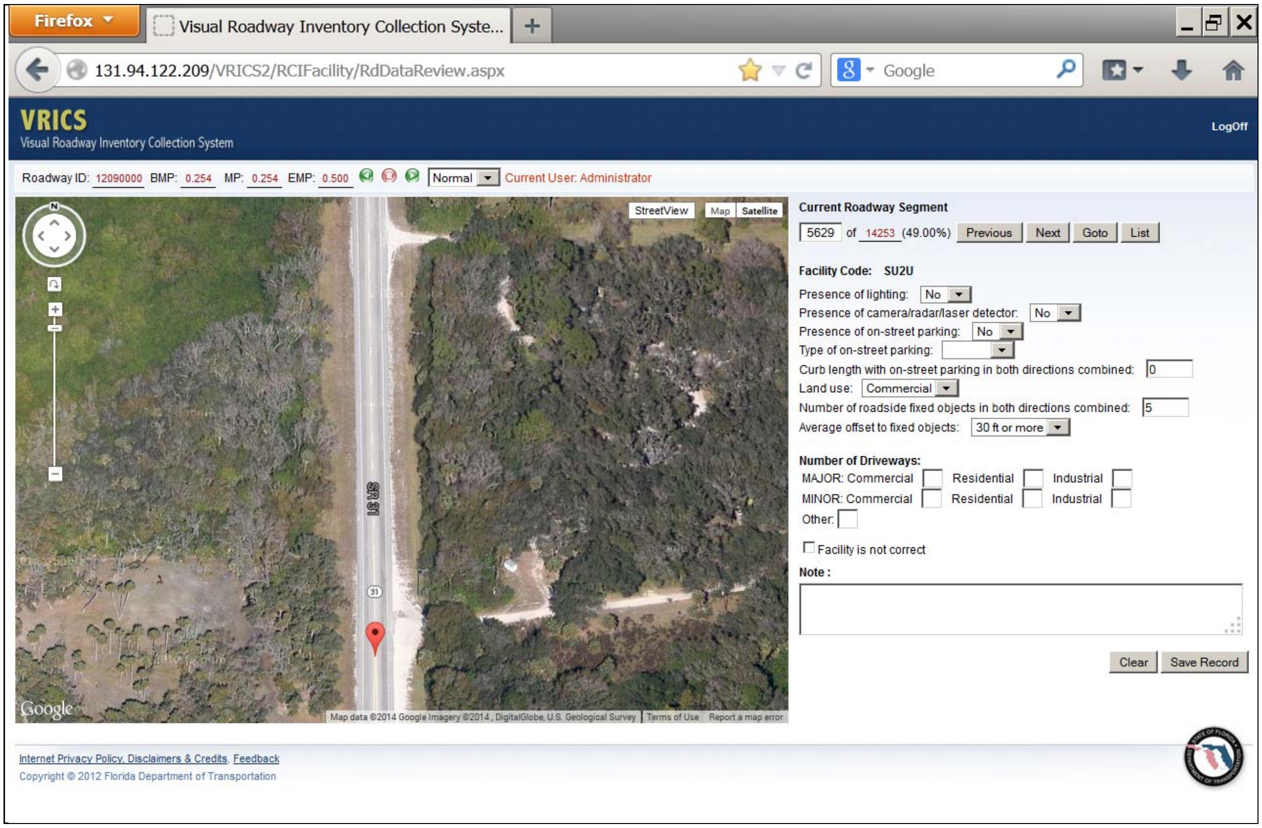


Figure 2-2: VRICS Application Customized to Collect Segment Data

Table 2-5 gives the list of variables for which the data were collected for each facility type. Table 2-6 presents the instructions provided for collecting information on different variables.

Table 2-5: Data Variables Collected for Segments Using VRICS

Rural Two-lane Roads	Rural Multilane Highways	Urban and Suburban Arterials
• Presence of passing lane • Presence of short four-lane section • Presence of centerline rumble strips • Presence of two-way left-turn lanes • Roadside hazard rating • Number of driveways • Presence of lighting • Presence of automated speed enforcement	• Presence of lighting • Presence of automated speed enforcement	• Presence of lighting • Presence of automated speed enforcement • Presence of on-street parking • Type of on-street parking • Curb length • Land use • Number of driveways by land-use type • Number of roadside fixed objects • Offset to fixed object

2.2 Intersection Data

This section discusses the specific data variables for all the intersection facility types discussed in the HSM, the availability of the required and desired data in Florida databases, and the process adopted to collect the missing data variables.

Table 2-6: Instructions for Collecting Data Variables for Segments

Facility Type	Data Variable	Instructions
Rural Two-lane Two-way Roads	Number of driveways (in both directions combined)	Count the number of driveways on both sides of the roadway and record the number.
	Presence of TWLTLs	Select "Yes" if the section has TWLTL; otherwise select "No".
	Presence of centerline rumble strips	Select "Yes" if the section has centerline rumble strips; otherwise select "No".
	Presence of lighting	Select "Yes" if there are lighting poles throughout the section; otherwise select "No".
	Presence of passing lane for each direction	Select "Yes" if the segment has passing lane in any direction; otherwise select "No".
	Presence of short four-lane section	Select "Yes" if the segment has two passing lanes side by side; otherwise select "No".
	Presence of camera/radar/laser detector	Select "Yes" if automated enforcement device is installed at any location along the segment; otherwise select "No".
	RHR in each direction	The HSM categorizes RHR into seven levels (referred to as RHR 1 to RHR 7) depending on clear zone width, sideslope, roadside surface roughness, recoverability of the roadside, and presence of roadside objects within the clear zone. Select the most appropriate RHR rating based on visual inspection.
Rural Multilane Highways	Presence of lighting	Select "Yes" if there are lighting poles throughout the section; otherwise select "No".
	Presence of camera/radar/laser detector	Select "Yes" if automated enforcement device is installed at any location along the segment; otherwise select "No".
Urban and Suburban Arterials	Presence of lighting	Select "Yes" if there are lighting poles throughout the section; otherwise select "No".
	Presence of camera/radar/laser detector	Select "Yes" if automated enforcement device is installed at any location along the segment; otherwise select "No".
	Presence of on-street parking	Select "Yes" if the section has parking spaces along the roadway on any side of the roadway; otherwise select "No".
	Type of on-street parking	Select either "angle" or "parallel" when there is on-street parking along the section.
	Curb length with on-street parking (in both directions combined)	Measure the length of on-street parking spaces on both sides of the roadway segment and record the measured length.
	Land use	Select the most appropriate land use type for the section from the following options: commercial, institutional, industrial, residential, or other.
	Number of driveways	Count the number of driveways on both sides of the roadway segment and record the number for each land use type
	Number of roadside fixed objects	Count the number of fixed objects on both sides of the roadway segment and record the number.
	Average offset to fixed objects	Record the average distance of roadside objects from the edge of the roadway through visual assessment.

2.2.1 Intersection Data Set

FDOT currently maintains a database of all nodes on the state highway system in RDWTBL 25 (i.e., Roadway Table 25). However, it was not possible to accurately extract the three-leg and four-leg signalized and unsignalized intersections. Therefore, alternative approaches to populate the intersection data set were explored. The AADT Geographic Information System (GIS) layer was used to extract all the intersections. This allows including only those intersections that have AADT data. Next, the State Roads GIS layer was overlaid on the extracted intersections layer to retrieve intersections on state roads alone. Finally, Signalized Intersections GIS layer was overlaid to identify signalized intersections. All the remaining intersections in the extracted intersections layer that were not identified as signalized intersections were recorded as unsignalized intersections. A total of 1,555 intersections were identified using this approach.

2.2.2 Data Variables

Table 2-7 presents the data variables required and desired by the HSM for the three facilities: rural two-lane two-way roads, rural multilane highways, and urban and suburban arterials. For each data variable that is available in the RCI database, the table also includes the corresponding variable name in the RCI database.

Table 2-7: Data Variables in the HSM for Intersections (AASHTO, 2010a)

Data Variable	Rural Two-lane, Two-way Roads	Rural Multilane Highways	Urban & Suburban Arterials	Data Variable in RCI (if available)
Functional classification	Required	Required	Required	FUNCLASS
Type of intersection[1]	Required	Required	Required	
Type of traffic control[1]	Required	Required	Required	
AADT for major road	Required	Required	Required	SECTADT
AADT for minor road	Required	Required	Required	SECTADT
Number of approaches with left-turn lanes[1]	Required	Required	Required	
Number of approaches with right-turn lanes[1]	Required	Required	Required	
Presence of lighting[1]	Required	Required	Required	
Intersection skew angle[1]	Desired	Desired		
Presence of left-turn signal phasing[1,2]			Required	
Type of left-turn signal phasing[1,2]			Required	
Presence of RTOR signal operation[1,2]			Required	
Presence of red-light cameras[1,2]			Required	
Number of bus stops within 1,000 ft[1,2]			Desired	
Presence of schools within 1,000 ft[1,2]			Desired	
Number of alcohol sales establishments within 1,000 ft[1,2]			Desired	
Pedestrian activity level[1,2]			Desired	
Max. number of lanes crossed by pedestrians[1,2]			Desired	

[1] Unavailable in the RCI database; [2] For signalized intersections only.

2.2.3 Data Collection

A total of 1,555 intersections were identified using the approach discussed in Section 2.2.1. Again, VRICS was customized to collect data at these 1,555 intersections. Figure 2-3 shows the screen capture of the VRICS application customized to collect intersection data. Similar to the segments, the intersections were categorized into the following three HSM-designated facility types: rural two-lane two-way roads, rural multilane highways, and urban and suburban arterials. Table 2-8 provides the list of data variables collected for each of the intersection facility types. Table 2-9 presents the instructions provided for collecting information on different variables.

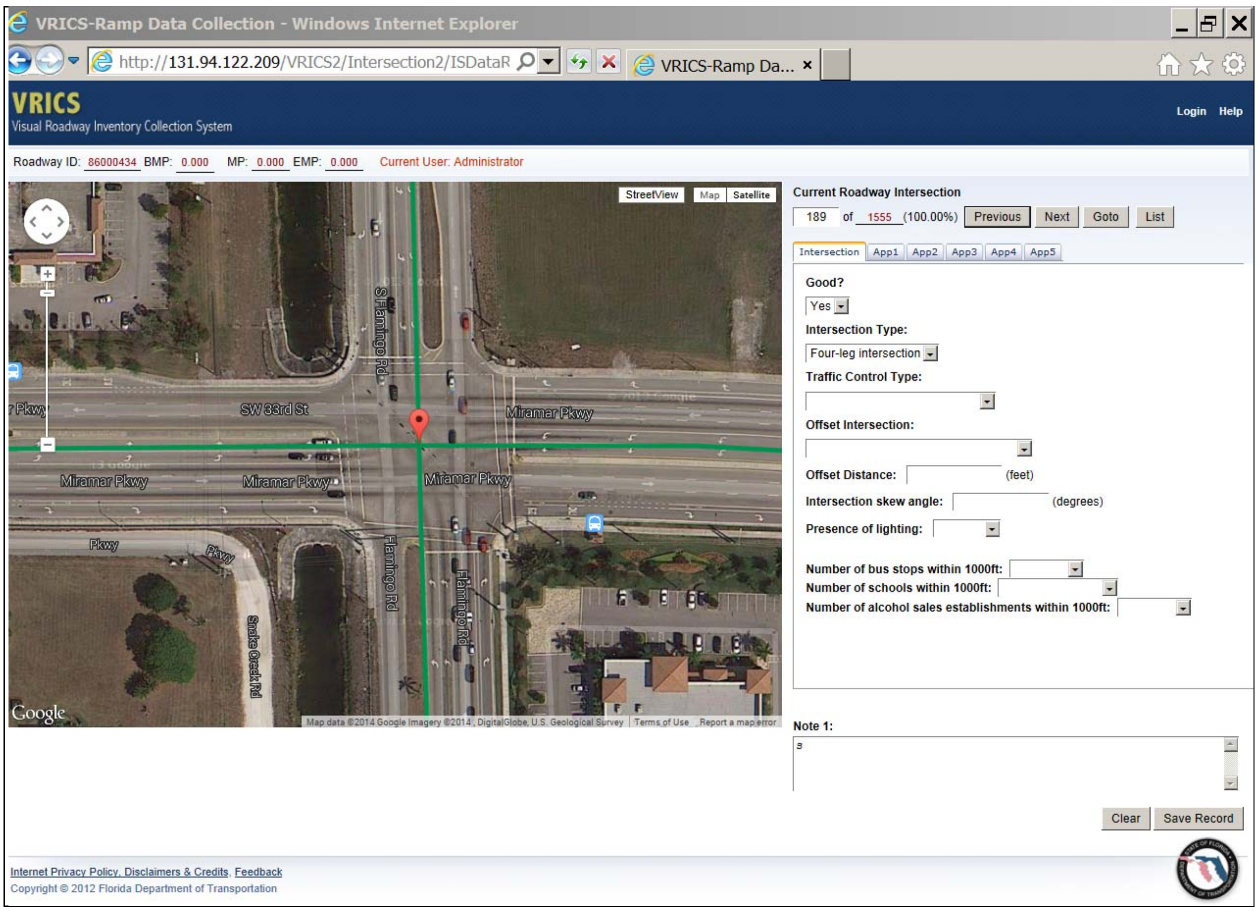


Figure 2-3: VRICS Application Customized to Collect Intersection Data

2.3 Crash Data

The HSM recommends using no more than three years of data for calibration. Therefore, three years of crash data from 2009-2011 were obtained from FDOT's crash analysis reporting (CAR) system. Note that the crashes that occurred on on-system roads alone were included in the analysis. Also, shape files of crash data for the years 2009-2011 were downloaded from FDOT Unified Basemap Repository (UBR) for on-system roads. The procedure followed to assign crashes to segments and intersections is explained in the following sections.

Table 2-8: Data Variables Collected for Intersections Using VRICS

Rural Two-lane Two-way Roads	Rural Multilane Highways	Urban and Suburban Arterials
• Type of intersection • Type of traffic control • Number of approaches with left-turn lanes • Number of approaches with right-turn lanes • Presence of lighting • Intersection skew angle	• Type of intersection • Type of traffic control • Number of approaches with left-turn lanes • Number of approaches with right-turn lanes • Presence of lighting • Intersection skew angle	• Type of intersection • Type of traffic control • Number of approaches with left-turn lanes • Number of approaches with right-turn lanes • Presence of lighting • Intersection skew angle • Presence of left-turn signal phasing[1] • Type of left-turn signal phasing[1] • Presence of RTOR signal operation[1] • Presence of red-light cameras[1] • Number of bus stops within 1,000 ft[1] • Presence of schools within 1,000 ft[1] • Number of alcohol sales establishments within 1,000 ft[1] • Pedestrian activity level[1] • Maximum number of lanes crossed by pedestrians[1]

[1] For signalized intersections only.

2.3.1 Assignment of Crashes to Segments

The crash summary records in the CAR system have crash location information, including the roadway ID and the milepost at which the crash occurred. This information was used to assign crashes to segments. In other words, crashes were assigned to segments by matching the location information. Crashes that occurred on the point between two roadway segments were consistently assigned to the beginning segment. The crash files were merged with segment characteristics files to obtain the number of different types of crashes (such as single-vehicle crashes, multi-vehicle crashes, nighttime crashes, etc.) on each segment.

2.3.2 Assignment of Crashes to Intersections

Table 2-10 lists the crash data variables used in the analysis. The CAR summary records do not have spatial attributes to spatially link crashes to intersections. Therefore, crash data shape files were used to assign crashes to intersections. These files were imported into ArcGIS 10.0. All the intersections for which the data were collected were also imported into ArcGIS. A 250 ft buffer was then created around each intersection. All the crashes that occurred within the 250 ft buffer were spatially identified. Crashes that occurred within the 250 ft buffer and those that occurred "at the intersection" and those that "are influenced by the intersection" were identified as intersection-related crashes. Crashes that fell into overlapping portion of two or more buffers were assigned to the nearest intersection by calculating linear distances from adjacent intersections. The resulting intersection-related crash files were merged with intersection characteristics files to obtain the number of different types of crashes (such as single-vehicle crashes, multi-vehicle crashes, nighttime crashes, pedestrian crashes, etc.) at each intersection.

Table 2-9: Instructions for Collecting Data Variables for Intersections

Facility Type	Data Variable	Instructions to Collect
All Intersections	Type of intersection	Record whether the intersection is Tee, Y, four-leg, or other type.
	Type of traffic control	Record whether the intersection is signalized or unsignalized.
	Number of approaches with left-turn lanes	Count the number of approaches with left-turn lanes and record the number.
	Number of approaches with right-turn lanes	Count the number of approaches with right-turn lanes and record the number.
	Presence of lighting	Select "Yes" if there are lighting poles at the intersection; otherwise select "No".
	Intersection skew angle	The skew angle for an intersection is defined as the absolute value of the deviation from an intersection angle of 90°. Estimate and record the skew angle from aerial image.
Signalized Intersections on Urban and Suburban Arterials	Presence of left-turn phasing	Determine the presence of left-turn phasing based on the number of left-turn lanes at the intersection.
	Type of left-turn phasing	Determine the type of left-turn phasing (i.e., permissive, protected/permissive, protected, or not sure) based on the number of left-turn lanes and the signal head assembly.
	Presence of RTOR signal operation	Select "Yes" if there is a sign that says "No Right Turn On Red"; otherwise select "No".
	Presence of red-light cameras	Select "Yes" if there are red-light cameras at the intersection; otherwise select "No".
	Number of bus stops within 1,000 ft	Count the number of bus stops within 1,000 ft of the intersection, and select either '0', '1-2' or '≥ 3'.
	Presence of schools within 1,000 ft	Select "Yes" if there are schools within 1,000 ft of the intersection; otherwise select "No".
	Number of alcohol sales establishments within 1,000 ft	Count the number of alcohol sales establishments (including bars, restaurants, pharmacies, and grocery stores) within 1,000 ft of the intersection, and select either '0', '1-8' or '≥ 9'.
	Pedestrian activity level	Because of limited pedestrian exposure data, pedestrian activity level was assumed to be medium.
	Maximum number of lanes crossed by pedestrians	Count the number of through lanes, left-turn lanes, and right-turn lanes in each approach and record the maximum number.

Table 2-10: Crash Data Variables Used in the Analysis

Variable Name	Variable Description
ROADWAYID	Roadway ID of crash
LOCMILEPT	Location milepost of crash
CRASHNUM	Crash number
HIGHESTINJ	Highest injury/severity level of the persons involved in the crash
SITELOCA	Type of location where the crash occurred (e.g., on ramp, at intersection, etc.)
LIGHTCOND	Lighting condition at the time of crash
CRASHEVENT1	First harmful event for the first at-fault vehicle

2.4 Traffic Data

The HSM identifies AADT as the most critical variable in predicting crashes. As such, the manual requires AADT volumes for all the segments and for both major and minor road approaches of intersections in applying the SPFs to predict segment and intersection crashes, respectively. AADT data were extracted from the RCI database based on roadway ID and milepost. Several of the segments and intersections did not have traffic data for one or more years. To include as many segments and intersections as possible in the analysis, the following assumptions were made regarding AADT data:

- If AADT data were available for only one year, that same value was assumed to apply to all the analysis years.
- If two years of AADT data were available, the AADT for the missing year was computed by either interpolation or extrapolation.
- If AADT data for a location (i.e., segment/intersection) was not available for all the three years, then the location was not included in the analysis.
- Locations with extremely high or low AADT values were considered as outliers and were excluded from the analysis.

2.5 Summary

This chapter focused on the data collection and preparation efforts undertaken in this project. The CAR summary records and crash shape files for the years 2009-2011, 2011 RCI database, 2011 RDWTBL 25 (i.e., Node list), 2011 AADT GIS layer, 2011 State Roads GIS layer, and 2011 Signalized Intersections GIS layer were used to populate the roadway segment and the intersection data sets. Several required and desired data variables are currently unavailable in the FDOT databases, and were collected from aerial images. Default values were assumed for some variables (such as side slope, pedestrian activity level, etc.) for which data could not be collected. Tables 2-11 and 2-12 give the descriptive statistics of the roadway segments and intersections that were finally included in the analysis, respectively. Appendix A provides detailed descriptive statistics of the data collected for each site subtype.

Table 2-11: Summary Statistics of Segment Subtypes

Facility Type	Segment Site Subtype	Number of Segments	Total Segment Length (miles)	Total Number of Crashes (2009-2011)	Number of Crashes per Mile per Year
Rural Two-way Two-lane Roads	Two-lane Undivided	3,546	3,549.2	6,731	0.63
Rural Multilane Highways	Four-lane Undivided	43	8.2	26	1.06
	Four-lane Divided	1,760	1,227.4	5,763	1.57
Urban and Suburban Arterials	Two-lane Undivided	1,791	616.7	3,290	1.78
	Three-lane Segment with TWLTL	359	67.0	627	3.12
	Four-lane Undivided	266	52.2	710	4.53
	Four-lane Divided	4,971	1400.9	16,503	3.93
	Five-lane Segment with TWLTL	1,095	272.5	4,758	5.82

Table 2-12: Summary Statistics of Intersection Subtypes

Facility Type	Intersection Site Subtype	Number of Intersections	Total Number of Crashes (2009-2011)	Number of Crashes per Year per Intersection
Rural Two-lane Intersections	Three-leg Stop-controlled	298	487	0.54
	Four-leg Stop-controlled	43	109	0.84
	Four-leg Signalized	21	204	3.24
Rural Multilane Intersections	Three-leg Stop-controlled	31	141	1.52
	Four-leg Signalized	27	378	4.67
Urban and Suburban Arterial Intersections	Three-leg Stop-controlled	321	1,600	1.66
	Four-leg Stop-controlled	34	222	2.18
	Three-leg Signalized	58	1,198	6.88
	Four-leg Signalized	459	23,000	16.70

CHAPTER 3
PRIORITIZATION OF DATA VARIABLES

The HSM requires very detailed roadway characteristics information to derive the calibration factors to accurately represent the agency's safety performance. This chapter focuses on ranking the data variables based on their importance. Random forest technique was used to identify and rank important variables. A review of the existing literature on data prioritization is first provided. The methodology adopted to rank the data variables is then discussed. The chapter then discusses the analysis results and finally provides a summary of the results.

3.1 Literature Review

Literature on the influence of data variables discussed in the HSM on safety predictions is limited. Only three studies, Akgüngör and Yıldız (2007), Alluri and Ogle (2012a), and Findley et al. (2012), were found to have focused on identifying and ranking influential variables. Akgüngör and Yıldız (2007) used fractional factorial method to study the sensitivity of the crash prediction model developed by Zegeer et al. (1987). The authors found that single parameters including AADT, lane width (W), paved shoulder width (PA), median roadside hazard rating (RHR), and two-parameter interactions such as AADT – W, AADT – PA, and AADT – RHR would have significant effects on crash frequency. The study ranked the variables by primary and secondary importance based on the absolute value of parameter effects at the three- and two-standard deviation thresholds, respectively. AADT was considered to be of primary importance, while all other identified parameters and parameter interactions were found to be of secondary importance.

Alluri and Ogle (2012a) conducted sensitivity analysis to assess the influence of variations of one variable on the predicted crash frequency as a function of AADT on rural two-way two-lane roads in Georgia. The authors conducted three types of sensitivity analyses: (a) effect of variation of AADT on the predicted crash frequency; (b) effect of variations of each variable on predicted crash frequency if means of all other CMFs were considered; and (c) effect of variations of each variable on predicted crash frequency if all other variables were assumed to be base conditions. Findley et al. (2012) applied the HSM predictive method to rural two-lane horizontal curves in North Carolina. The authors conducted sensitivity analysis to identify the variables that have the most significant effect on crash predictions. The predicted number of crashes calculated based on the difference between using the maximum and the minimum value and the value calculated based on the difference between using the mean value and the actual field measured value were compared. The results showed that AADT, curve length, and curve radius were the most important factors in predicting number of crashes on horizontal curve sections.

Random forest technique has been increasingly applied in fields outside of transportation to rank influential variables. Random forest technique was used in this study since it works well with highly-correlated data, data with many interactions, and data sets with small sample sizes (Shih, 2013). In this technique, a specific number of trees (usually between 500 and 1,000) were grown by randomly selecting some observations from the original data set with replacement, then searching over a randomly selected subset of variables at each split till the variable importance was ranked (Haleem et al., 2010). The variables can be ranked based on the increase in node purity (*IncNodePurity)* values, which represent the average total decrease in node impurity. Higher *IncNodePurity* value “represents higher variable importance, i.e., nodes are much 'purer'” (Jacobsson, 2012).

3.2 Methodology

The “randomForest” library in the ***R*** software package was used to conduct regression analysis using random forest technique. A total of 500 trees were grown and *p*/3 variables were used at each split where *p* is the total number of variables. About two-third of the entire data set was identified as training data, and the remaining one-third of the data, often identified as testing data, was used to estimate the variable importance. Variables were ranked based on their *IncNodePurity* values. Table 3-1 gives the descriptive statistics of the facility types analyzed.

Table 3-1: Descriptive Statistics of Segment and Intersection Subtypes

Facility Type	Site Subtype	Total Facility Length (in miles)	Total Number of Sites	Average Segment Length (in miles)
Rural Two-way Two-lane Segments	Two-lane Undivided	3,549.2	3,546	1.00
Rural Multilane Segments	Four-lane Undivided[1]	8.2	43	0.20
	Four-lane Divided	1,227.4	1,760	0.70
Urban and Suburban Arterials	Two-lane Undivided	616.7	1,791	0.34
	Three-lane with TWLTL	67.0	359	0.19
	Four-lane Undivided	52.2	266	0.20
	Four-lane Divided	1,400.9	4,971	0.28
	Five-lane with TWLTL	272.5	1,095	0.25
Rural Two-lane Intersections	Three-leg Stop-controlled	--	298	--
	Four-leg Stop-controlled[1]	--	43	--
	Four-leg Signalized[1]	--	21	--
Rural Multilane Intersections	Three-leg Stop-controlled[1]	--	31	--
	Four-leg Signalized[1]	--	27	--
Urban and Suburban Arterial Intersections	Three-leg Stop-controlled	--	321	--
	Four-leg Stop-controlled[1]	--	34	--
	Three-leg Signalized[1]	--	58	--
	Four-leg Signalized	--	459	--

[1] Site subtypes were not analyzed due to limited sample size.

AADT and segment length for segments and major and minor road AADTs for intersections were considered as the most important variables as they significantly influence crash predictions. This is evident in the results from the random forest technique as these variables have the highest *IncNodePurity* values. The remaining variables were ranked based on their *IncNodePurity* values. All the variables with *IncNodePurity* values within 15% of the highest *IncNodePurity* value of the remaining variables were identified as the variables of secondary importance. For example, Table 3-2 gives the ranking of variables for rural two-way two-lane roads along with the variables' corresponding *IncNodePurity* values. Segment length and AADT were the most important variables based on their *IncNodePurity* values. Roadside fixed object density was the next important variable with *IncNodePurity* value of 318.9. A 15% of this value is 47.8. Therefore, all the variables with *IncNodePurity* values of at least 47.8 were identified as the variables of secondary importance. The remaining variables (i.e., variables with *IncNodePurity* values < 47.8) were identified as the variables of lesser importance. Furthermore, the variables were ranked based on their corresponding *IncNodePurity* values.

3.3 Results for Segments

Random forest technique was used to rank data variables based on their importance. This section discusses the results.

3.3.1 Rural Segments

Tables 3-2 and 3-3 give the ranking of the variables for rural two-way two-lane roads and rural four-lane divided arterials, respectively. For the two site subtypes, segment length and AADT were of primary importance, and therefore, were ranked 1 and 2, respectively. For rural two-way two-lane roads, the list of variables of secondary importance included driveway density, shoulder width, RHR, lane width, and shoulder type. For rural four-lane arterials, all the variables were considered to be of secondary importance. It is to be noted that lane width on rural four-lane sections was considered to be of lesser importance as its *IncNodePurity* value is not within the predefined 15% threshold. However, lane width is still identified as the variable of secondary importance because of the extant literature that proves that lane width significantly affects crash experience.

Table 3-2: Ranking of Variables: Rural Two-way Two-lane Roads

Variables	IncNodePurity	Rank[1]	Comments[2]
Segment length	816.5	1	Variables of primary importance
AADT	735.7	2	
Driveway density	318.9	3	Variables of secondary importance
Shoulder width	151.6	4	
Roadside hazard rating	110.9	5	
Lane width	65.0	6	
Shoulder type	50.6	7	
Presence of lighting	11.2	8	Variables of lesser importance
Presence of passing lane	5.9	9	
Presence of TWLTL	4.9	10	
Presence of short four-lane section	2.3	11	
Presence of centerline rumble strip	1.7	12	
Presence of automated speed enforcement	0.2	13	

[1] Mean of squared residuals: 0.430; Percent of variance explained: 44.60%.

[2] Variables of primary importance: Variables having *IncNodePurity* value relatively greater than that of the other variables; Variables of secondary importance: Variables having *IncNodePurity* value within 15% of 318.9 (≥ 47.8); Variables of lesser importance: Variables having *IncNodePurity* value below the 15% threshold (< 47.8).

Table 3-3: Ranking of Variables: Rural Four-lane Divided Arterials

Variables	IncNodePurity	Rank[1]	Comments[2]
Segment length	1,828.6	1	Variables of primary importance
AADT	881.8	2	
Shoulder width	248.9	3	Variables of secondary importance
Median width	140.7	4	
Presence of automated speed enforcement	108.6	5	
Presence of lighting	101.4	6	
Lane width	23.9[3]	7	

[1] Mean of squared residuals: 1.070; Percent of variance explained: 58.52%.
[2] Variables of primary importance: Variables having *IncNodePurity* value relatively greater than that of the other variables; Variables of secondary importance: Variables having *IncNodePurity* value within 15% of 248.9 (≥ 37.3); Variables of lesser importance: Variables having *IncNodePurity* value below the 15% threshold (< 37.3).
[3] Although lane width has an *IncNodePurity* value that is below the 15% threshold, it is considered to be the variable of secondary importance.

3.3.2 Urban and Suburban Arterials

Tables 3-4 through 3-8 give the ranking of the required and desired variables for two-lane undivided, three-lane with TWLTL, four-lane undivided, four-lane divided, and five-lane with TWLTL arterials in urban and suburban areas, respectively. Again, as in the case of rural sections, segment length and AADT were identified as the two most influential variables, and therefore, were ranked as 1 and 2, respectively. For all the site subtypes on urban and suburban arterial sections, roadside fixed object density and the number of major and minor driveways were considered to be the most influential variables next to segment length and AADT (i.e., variables of secondary importance). Median width, whenever included in the analysis, was also found to be influential. Among the other variables, speed limit, presence of automated speed enforcement, and presence of on-street parking were found to be less important for all the site subtypes. Presence of lighting was also found to be less important for all the site subtypes except for four-lane undivided arterials.

Table 3-4: Ranking of Variables: Urban and Suburban Two-lane Undivided Arterials

Variable	IncNodePurity	Rank[1]	Comments[2]
Segment length	527.4	1	Variables of primary importance
AADT	364.2	2	
Number of minor driveways	244.7	3	Variables of secondary importance
Roadside fixed object density	209.0	4	
Number of major driveways	51.4	5	
Presence of lighting	30.2	6	Variables of lesser importance
Speed limit	10.0	7	
Presence of on-street parking	8.7	8	
Presence of automated speed enforcement	0.2	9	

[1] Mean of squared residuals: 0.571; Percent of variance explained: 38.41%.
[2] Variables of primary importance: Variables having *IncNodePurity* value relatively greater than that of the other variables; Variables of secondary importance: Variables having *IncNodePurity* value within 15% of 244.7 (≥ 36.7); Variables of lesser importance: Variables having *IncNodePurity* value below the 15% threshold (< 36.7).

Table 3-5: Ranking of Variables: Urban and Suburban Three-lane Arterials with TWLTL

Variables	IncNodePurity	Rank[1]	Comments[2]
Segment length	73.7	1	Variables of primary importance
AADT	72.9	2	
Number of minor driveways	47.6	3	Variables of secondary importance
Roadside fixed object density	39.5	4	
Number of major driveways	14.6	5	
Presence of lighting	5.1	6	Variables of lesser importance
Presence of on-street parking	5.1	7	
Speed limit	2.6	8	
Presence of automated speed enforcement[3]	0.0	--	

1 Mean of squared residuals: 0.631; Percent of variance explained: 24.78%.
2 Variables of primary importance: Variables having *IncNodePurity* value relatively greater than that of the other variables; Variables of secondary importance: Variables having *IncNodePurity* value within 15% of 47.6 (≥ 7.1); Variables of lesser importance: Variables having *IncNodePurity* value below the 15% threshold (< 7.1).
3 None of the locations have automated speed enforcement.

Table 3-6: Ranking of Variables: Urban and Suburban Four-lane Undivided Arterials

Variables	IncNodePurity	Rank[1]	Comments[2]
Segment length	92.4	1	Variable of primary importance
AADT	59.1	2	
Number of minor driveways	55.4	3	Variables of secondary importance
Roadside fixed object density	40.9	4	
Number of major driveways	30.1	5	
Presence of lighting	11.2	6	
Speed limit	4.8	7	Variables of lesser importance
Presence of on-street parking	4.4	8	
Presence of automated speed enforcement[3]	0.0	--	

1 Mean of squared residuals: 0.787; Percent of variance explained: 38.75%.
2 Variables of primary importance: Variables having *IncNodePurity* value relatively greater than that of the other variables; Variables of secondary importance: Variables having *IncNodePurity* value within 15% of 55.4 (≥ 8.3); Variables of lesser importance: Variables having *IncNodePurity* value below the 15% threshold (< 8.3).
3 None of the locations have automated speed enforcement.

Table 3-7: Ranking of Variables: Urban and Suburban Four-lane Divided Arterials

Variables	IncNodePurity	Rank[1]	Comments[2]
Segment length	3943.6	1	Variables of primary importance
AADT	2846.9	2	
Number of minor driveways	1563.8	3	Variables of secondary importance
Roadside fixed object density	1387.1	4	
Number of major driveways	1059.8	5	
Median width	505.5	6	
Presence of lighting	143.1	7	Variables of lesser importance
Presence of automated speed enforcement	56.3	8	
Speed limit	51.9	9	
Presence of on-street parking	35.5	10	

[1] Mean of squared residuals: 1.459; Percent of variance explained: 44.76%.
[2] Variables of primary importance: Variables having *IncNodePurity* value relatively greater than that of the other variables; Variables of secondary importance: Variables having *IncNodePurity* value within 15% of 1563.8 (≥ 234.6); Variables of lesser importance: Variables having *IncNodePurity* value below the 15% threshold (< 234.6).

Table 3-8: Ranking of Variables: Urban and Suburban Five-lane Arterials with TWLTL

Variables	IncNodePurity	Rank[1]	Comments[2]
Segment length	983.4	1	Variables of primary importance
AADT	908.3	2	
Number of major driveways	793.1	3	Variables of secondary importance
Number of minor driveways	561.0	4	
Roadside fixed object density	480.5	5	
Presence of lighting	35.7	6	Variables of lesser importance
Speed limit	11.8	7	
Presence of automated speed enforcement	4.3	8	
Presence of on-street parking[3]	0.0	--	

[1] Mean of squared residuals: 2.417; Percent of variance explained: 39.06%.
[2] Variables of primary importance: Variables having *IncNodePurity* value relatively greater than that of the other variables; Variables of secondary importance: Variables having *IncNodePurity* value within 15% of 793.1 (≥ 119.0); Variables of lesser importance: Variables having *IncNodePurity* value below the 15% threshold (< 119.0).
[3] None of the locations have on-street parking.

3.4 Results for Intersections

Random forest technique was used to rank data variables based on their importance. As can be observed from Table 3-1, several of the intersection site subtypes have small samples; random forest technique might not yield reliable results for these subtypes. Therefore, the following intersection site subtypes (with sample size < 100 intersections) were not analyzed:

- Rural two-lane four-leg stop-controlled intersections
- Rural two-lane four-leg signalized intersections
- Rural multilane three-leg stop-controlled intersections
- Rural multilane four-leg signalized intersections
- Urban and suburban four-leg stop-controlled intersections

- Urban and suburban three-leg signalized intersections

The random forest technique was used to rank variables for the following intersection site subtypes:

- Rural two-lane three-leg stop-controlled intersections
- Urban and suburban three-leg stop-controlled intersections
- Urban and suburban four-leg signalized intersections

3.4.1 Rural Intersections

Table 3-9 gives the ranking of the variables for rural two-lane three-leg stop-controlled intersections. Major road and minor road AADTs were considered to significantly influence the crash predictions, and therefore, were given the highest ranking in terms of variable importance. All the remaining variables were also found to be of importance.

Table 3-9: Ranking of Variables: Rural Two-lane Three-leg Stop-controlled Intersections

Variables	IncNodePurity	Rank[1]	Comments[2]
Major road AADT	49.9	1	Variables of primary importance
Minor road AADT	48.0	2	
Number of approaches with right-turn lanes	6.2	3	Variables of secondary importance
Number of approaches with left-turn lanes	5.5	4	
Presence of skewness	3.9	5	
Presence of lighting	3.8	6	

[1] Mean of squared residuals: 0.457; Percent of variance explained: 16.90%.
[2] Variables of primary importance: Variables having *IncNodePurity* value relatively greater than that of the other variables; Variables of secondary importance: Variables having *IncNodePurity* value within 15% of 6.2 (≥ 0.9); Variables of lesser importance: Variables having *IncNodePurity* value below the 15% threshold (< 0.9).

3.4.2 Urban and Suburban Intersections

Tables 3-10 and 3-11 give the ranking of the variables for urban three-leg stop-controlled and urban four-leg signalized intersections, respectively. Similar to the case of rural intersections, major road and minor road AADTs were considered to significantly influence the crash predictions, and therefore, were given the highest ranking in terms of variable importance. The variable *Number of Approaches with RTOR* was the only variable that was found to be of relatively less importance (i.e., with *IncNodePurity* value below the predefined 15% threshold).

Table 3-10: Ranking of Variables: Urban Three-leg Stop-controlled Intersections

Variables	IncNodePurity	Rank[1]	Comments[2]
Major road AADT	345.2	1	Variables of primary importance
Minor road AADT	149.1	2	
Number of approaches with left-turn lanes	62.4	3	Variables of secondary importance
Number of approaches with right-turn lanes	31.2	4	
Presence of lighting	15.7	5	

[1] Mean of squared residuals: 3.332; Percent of variance explained: 22.18%.

[2] Variables of primary importance: Variables having *IncNodePurity* value relatively greater than that of the other variables; Variables of secondary importance: Variables having *IncNodePurity* value within 15% of 62.4 (≥ 9.4); Variables of lesser importance: Variables having *IncNodePurity* value below the 15% threshold (< 9.4).

Table 3-11: Ranking of Variables: Urban Four-leg Signalized Intersections

Variables	IncNodePurity	Rank[1]	Comments[2]
Major road AADT	29,888.9	1	Variables of primary importance
Minor road AADT	20,786.1	2	
Number of approaches with right-turn lanes	5,669.6	3	Variables of secondary importance
Presence of bus stops within 1,000 ft of intersection	5,442.5	4	
Presence and type of left-turn signal phasing	4,227.1	5	
Number of approaches with left-turn lanes	2,330.1	6	
Presence of red-light camera	1,275.2	7	
Presence of schools within 1,000 ft of intersection	1,083.8	8	
Presence of alcohol sales establishments within 1,000 ft of intersection	1,032.4	9	
Number of approaches with RTOR	516.9	10	Variables of lesser importance

[1] Mean of squared residuals: 118.74; Percent of variance explained: 37.77%.

[2] Variables of primary importance: Variables having *IncNodePurity* value relatively greater than that of the other variables; Variables of secondary importance: Variables having *IncNodePurity* value within 15% of 5669.6 (≥ 850.4); Variables of lesser importance: Variables having *IncNodePurity* value below the 15% threshold (< 850.4).

3.5 Summary

Random forest technique was used to rank the required and desired variables based on their importance. Data variables for all the segment site subtypes except for rural two-lane undivided sections were ranked. For intersections, data variables were ranked only for three subtypes: rural two-lane three-leg stop-controlled intersections, urban three-leg stop-controlled intersections, and urban four-leg signalized intersections. The remaining intersection subtypes have fewer than 100 intersections, and therefore, were not included in the analysis. Tables 3-12 and 3-13 give the summary of the ranking of the variables for segments and intersections, respectively.

Table 3-12: Summary of the Ranking of Variables for Segments

Data Variable	Site Subtype[1,2]						
	R2U	R4D	SU2U	SU3T	SU4U	SU4D	SU5T
Segment length	1	1	1	1	1	1	1
AADT	2	2	2	2	2	2	2
Lane width	6	7	--	--	--	--	--
Shoulder type	7	NR	--	--	--	--	--
Shoulder width	4	3	--	--	--	--	--
Presence of TWLTL	10	--	--	--	--	--	--
Median width	--	4	--	--	--	6	--
Presence of lighting	8	6	6	6	6	7	6
Roadside fixed object density	--	--	4	4	4	4	5
Speed limit	--	--	7	8	7	9	7
Presence of on-street parking	--	--	8	7	8	10	NR
Presence of automated speed enforcement	13	5	9	NR	NR	8	8
Presence of passing lane	9	--	--	--	--	--	--
Presence of short four-lane section	11	--	--	--	--	--	--
Presence of centerline rumble strip	12	--	--	--	--	--	--
Roadside hazard rating	5	--	--	--	--	--	--
Driveway density	3	--	--	--	--	--	--
Number of major driveways	--	--	5	5	5	5	3
Number of minor driveways	--	--	3	3	3	3	4
Horizontal curve	NR	--	--	--	--	--	--
Vertical grade	NR	--	--	--	--	--	--

[1] R2U: rural two-lane undivided; R4D: rural four-lane divided; SU2U: urban and suburban two-lane undivided; SU3T: urban and suburban three-lane with TWLTL; SU4U: urban and suburban four-lane undivided; SU4D: urban and suburban four-lane divided; SU5T: urban and suburban five-lane with TWLTL.

[2] NR indicates that the variable is not ranked; -- indicates that the variable is not used for that specific site subtype.

Table 3-13: Summary of the Ranking of Variables for Intersections

Data Variable	Site Subtype[1,2]		
	R3ST	SU3ST	SU4SG
Major road AADT	1	1	1
Minor road AADT	2	2	2
Intersection skew angle	5	--	--
Number of approaches with left-turn lanes	4	3	6
Number of approaches with right-turn lanes	3	4	3
Presence of lighting	6	5	NR
Presence and type of left-turn signal phasing	--	--	5
Use of RTOR signal operation	--	--	10
Use of red-light cameras	--	--	7
Number of bus stops within 1,000 ft	--	--	4
Presence of schools within 1,000 ft	--	--	8
Number of alcohol sales establishments within 1,000 ft	--	--	9

[1] R3ST: rural two-lane three-leg stop-controlled intersections; SU3ST: urban and suburban three-leg stop-controlled intersections; SU4SG: urban and suburban four-leg signalized intersections.

[2] NR indicates that the variable is not ranked; -- indicates that the variable is not used for that specific site subtype.

CHAPTER 4
DETERMINATION OF MINIMUM SAMPLE SIZE

This chapter aims to answer the following two questions: a) what is the minimum sample size to estimate a reliable calibration factor; and b) whether there is an additional benefit in estimating calibration factors using stratified sampling procedure compared to using simple random sampling procedure. A synthesis of the existing literature on estimating calibration factors for HSM-default SPFs is first provided. The methodology adopted to determine the minimum sample sizes and to compare the calibration factors from the random sampling and the stratified sampling procedures is then described. The chapter then discusses the analysis results and finally provides recommendations.

4.1 Literature Review

The calibration factor for a particular site type is defined as the ratio of the total number of observed crashes to the total number of predicted crashes calculated using the default SPFs provided in the HSM. Thus, calibration factor (C) can be expressed as follows (AASHTO, 2010a):

$$C = \frac{\sum_{all\ sites} Observed\ Crashes}{\sum_{all\ sites} Predicted\ Crashes} \tag{4-1}$$

Calibration factors are required "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" (AASHTO, 2010a). The manual recommends deriving calibration factors using randomly selected 30-50 sites that experienced a minimum total of 100 crashes per year. Since the release of the manual in 2010, several state and local agencies have developed calibration factors as a first step in adopting the manual.

Several studies have stated that collecting the required and desired data variables to estimate calibration factors is resource intensive. For example, Sun et al. (2006) concluded that the level of effort required to obtain the data necessary to calibrate the HSM-default SPFs would be a challenge. Nonetheless, several studies have concluded that the HSM-default SPFs calibrated to local data yielded better (i.e., more reliable) crash predictions compared to the non-calibrated HSM-default SPFs. For example, Young and Park (2013) compared the safety prediction performance among jurisdiction-specific SPFs, calibrated HSM-default SPFs, and uncalibrated HSM-default SPFs based on five years of crash data (2005-2009) on intersections in the city of Regina, Saskatchewan. Among the calibrated and uncalibrated HSM-default SPFs, the calibrated HSM-default SPFs were found to yield better crash predictions compared to the HSM-default SPFs.

4.1.1 Sample Size for Calibration

State-specific calibration factors have been developed for several states including Florida (Srinivasan et al., 2011), North Carolina (Srinivasan and Carter, 2011), Utah (Brimley et al., 2012), Illinois (Williamson and Zhou, 2012), Kansas (Lubliner and Schrock, 2012), Oregon (Xie et al., 2011), Missouri (Brown et al., 2014), etc. These studies have conformed, to the extent possible, to the HSM calibration data size requirements by randomly selecting a sizeable number of sites with adequate average annual crash frequency. For site subtypes with limited crash and/or roadway network data, the authors considered all the available data for calibration.

Banihashemi (2011) concluded that some types of roadway segments require a considerably large sample size for calibration. For different site subtypes, Banihashemi conducted sensitivity analysis and recommended sample sizes so that the estimated calibration factor based on the sample size would lie within 5-10% of the actual calibration factor (i.e., calibration factor calculated based on the entire data inventory).

4.1.2 Methods to Determine Calibration Factors

Several studies have evaluated different methods for calculating calibration factors besides the method recommended in the HSM. Martinelli et al. (2009) calibrated the HSM crash prediction model using three years of crash data collected on about 938 km of rural two-lane road network in the Arezzo province of Italy. Besides the HSM recommended calibration method, additional calibration procedures were evaluated based on three ratios: the ratio between observed and predicted crashes, the ratio between densities of observed and predicted crashes, and the ratio between section length weighted averages of observed and predicted crashes. The calibration factor developed based on the product of the base SPFs and the CMFs, the HSM-recommended calibration procedure, and with the segment length weighted estimation was found to best fit the crash data. However, it was reported that a single calibration factor might not be adequate since all the estimated models resulted in overestimation of crashes at low crash locations and underestimation of crashes at high crash locations.

Lubliner and Schrock (2012) determined the calibration factor by clustering segments into 10-mile sections on rural two-lane two-way highways in Kansas. A total of 19 randomly selected 10-mile sections comprising 190 miles of highways which included 239 segments that experienced 145 crashes per year met the two major calibration criteria. The study investigated different calibration methods to select the most appropriate method to accurately predict crashes. The various calibration methods were based on geographic region (district-wise), crash distributions, prevalent crash type, etc. Based on the results of different calibration procedures, the single calibration factor determined for state-level was found most favorable in aggregate-level crash predictions.

Findley et al. (2012) applied the HSM predictive method to rural two-lane horizontal curves in North Carolina. The calibration factors were determined separately for curves and tangents, and also for composite sites combining short tangents preceding and following the curves. A paired t-test was conducted to determine whether or not the actual and predicted crash frequencies were significantly different. In other words, the t-test was intended to determine whether the calibration factor was significantly different from 1.0. The calibration factor for the curve-only-segments was found to be statistically different from 1.0, which underlined the necessity to adjust the predicted number of crashes on curves to incorporate the effect of local conditions. Since the calibration factors for tangents and composite segments were not significantly different from 1.0, agencies could use a calibration factor equal to 1.0.

Mehta and Lou (2013) introduced a new approach to determine the calibration factor, where the calibration factor was considered as a constant term in the traditional negative binomial (NB) regression model fitting crash frequency data. The regression equation was then expressed as:

$$\mu_i = exp[ln(C_r) + ln(base\ SPF\ for\ site\ i)] \tag{4-2}$$

where the natural logarithm of calibration factor is regarded as an intercept and the natural logarithm of the base SPF for site *i* is the explanatory variable whose coefficient is considered to be 1.0. This new approach of estimating calibration factor did not emerge as a better method than the HSM-recommended calibration method.

4.1.3 Summary

A review of existing literature revealed that several states have developed state-specific calibration factors for the HSM-default SPFs based on the procedure recommended in the manual. Several studies have questioned the impact of sample size and sample selection procedures on the reliability of calibration factors. However, till date, only Banihashemi (2011, 2012) evaluated the quality of the calibration factors generated from data sets of different sample sizes. However, the study was limited to the segment subtypes and did not analyze intersection subtypes.

4.2 Methodology

4.2.1 Calibration Factor

As shown in Equation 4-1, calibration factor is calculated as the ratio of the total number of observed crashes to the total number of predicted crashes. The number of predicted crashes is calculated using the following equation (AASHTO, 2010a):

$$N_{predicted} = N_{SPF} \times (CMF_{1x} \times CMF_{2x} \times CMF_{yx}) \tag{4-3}$$

where,

$N_{predicted}$ = predicted average crash frequency for a specific year for site subtype *x*,

N_{SPF} = predicted average crash frequency determined for base conditions of the SPF developed for site subtype *x*, and

CMF_{yx} = crash modification factors specific to site subtype *x* and specific geometric design and traffic control features *y*.

Tables 4-1 and 4-2 provide the list of CMFs calculated for each site subtype for segments and intersections, respectively.

Table 4-1: Crash Modification Factors for Segments

Crash Modification Factors	Rural Two-lane Two-way Roads[2]	Rural Multilane Highways[2]	Urban and Suburban Arterials[2]				
	2U	4D	2U	3T	4U	4D	5T
Lane width	Y	Y					
Shoulder type and width	Y						
Horizontal curves: length, radius, and presence of spiral transition[1]	Y						
Horizontal curves: superelevation[1]	Y						
Vertical grades[1]	Y						
Driveway density	Y						
Centerline rumble strip	Y						
Passing lanes	Y						
Two-way left-turn lanes	Y						
Roadside design (i.e., RHR)	Y						
Lighting	Y	Y	Y	Y	Y	Y	Y
Automated speed enforcement	Y	Y	Y	Y	Y	Y	Y
Right shoulder width		Y					
Median width		Y				Y	
On-street parking			Y	Y	Y	Y	Y
Roadside fixed objects			Y	Y	Y	Y	Y

[1] Default CMF of 1.0 was used.
[2] 2U: two-lane undivided; 4U: four-lane undivided; 4D: four-lane divided; 3T: three-lane with TWLTL; 5T: five-lane with TWLTL.
Y indicates CMF was applied for that particular segment site subtype.

Table 4-2: Crash Modification Factors for Intersections

Crash Modification Factors	Rural Two-lane Two-way Roads[1]			Rural Multilane Highways[1]	Urban and Suburban Arterials[1]			
	3ST	4ST	4SG	3ST	3ST	4ST	3SG	4SG
Intersection skew angle	Y	Y		Y				
Intersection left-turn lanes	Y	Y	Y	Y	Y	Y	Y	Y
Intersection right-turn lanes	Y	Y	Y	Y	Y	Y	Y	Y
Lighting	Y	Y	Y	Y	Y	Y	Y	Y
Intersection left-turn signal phasing							Y	Y
RTOR							Y	Y
Red-light cameras							Y	Y
Number of bus stops within 1,000 ft							Y	Y
Presence of schools within 1,000 ft							Y	Y
Number of alcohol sales establishments within 1,000 ft							Y	Y

[1] 3ST: three-leg stop-controlled intersections; 4ST: four-leg stop-controlled intersections; 3SG: three-leg signalized intersections; 4SG: four-leg signalized intersections.
Y indicates CMF was applied for that particular intersection site subtype.

4.2.2 Sample Size

The methodology used to determine the minimum sample size for calibration of HSM-default SPFs was based on the study conducted by Banihashemi (2011, 2012). The methodology was based on the following assumption: for each specific sample size, the calibration factors calculated using different subsets of the same data set follow normal distribution.

The calibration factor can be calculated using two approaches. The first approach is to determine yearly calibration factor for each study year and then the overall calibration factor is obtained by taking the average of the yearly calibration factors. The second approach is to divide the total number of observed crashes over the study period by the total number of predicted crashes over the study period. The calibration factors obtained from these two approaches were almost identical (vary by as little as 0.1% in some cases). In this study, the calibration factors were calculated using the second approach. Also, in determining the calibration factors, several HSM-default values were replaced with Florida-specific values. Appendix B provides the HSM-default values and their corresponding Florida-specific values.

The calibration factor calculated using the entire data set (i.e., using all the available data) was considered to be the actual calibration factor. For each site subtype, the objective was to determine the minimum sample size so that there is a high probability that the estimated calibration factor calculated from the sample size is within 10% of the actual calibration factor. For each site subtype, subsets with different sample sizes were chosen based on the available data set sizes. Simple random sampling procedure was used to extract data sets with different sample sizes. The sample sizes chosen varied by site subtype and depended on the site subtype's available data inventory. Table 4-3 provides descriptive statistics of the segment and intersection subtypes. The table includes a summary of total number of sites, total segment length and the average number of observed crashes per year for each facility type for the studied years of 2009 to 2011. It also includes the actual calibration factors calculated from the entire data inventory.

For each sample size chosen, 30 subsets of data were created. For example, for a sample size of 50 rural multilane four-leg signalized intersections, 30 subsets of data with each subset containing 50 randomly selected intersections were generated. Calibration factors were calculated for each of these 30 subsets. These 30 calibration factors were assumed to follow a normal distribution. The average calibration factor of these 30 subsets was considered as the estimated calibration factor when the sample size was 50 intersections. The estimated calibration factors were similarly calculated for the different sample sizes and for all site subtypes.

Once the estimated and the actual calibration factors were calculated for the different sample sizes, the comparisons were made based on probability. The estimated calibration factor was considered to be acceptable if it varies within 10% of the actual calibration factor. When the estimated calibration factor is found to vary within 10% of the actual calibration factor, it means that the calibration factor estimated from the sample will be between 0.90 times and 1.10 times the actual calibration factor. Based on the assumption that the estimated calibration factors follow a normal distribution, the probabilities that the estimated calibration factor is within 10% of the actual calibration factor was calculated. The difference between the probability that the estimated calibration factor will be less than 0.90 times the actual calibration factor (i.e., $P(Z < Z_{min})$) and the probability that the estimated calibration factor will be less than 1.10

times the actual calibration factor (i.e., $P(Z < Z_{max})$) (i.e., $P(Z < Z_{max}) - P(Z < Z_{min})$) gives the probability that the estimated calibration factor would lie within 10% of the actual calibration factor (see Equations 4-4 and 4-5).

$$Z_{min} = \frac{\textit{actual calibration factor} \times 0.9 - \textit{estimated calibration factor}}{\textit{standard deviation}} \tag{4-4}$$

$$Z_{max} = \frac{\textit{actual calibration factor} \times 1.1 - \textit{estimated calibration factor}}{\textit{standard deviation}} \tag{4-5}$$

Once the probabilities were calculated, the sample sizes that result in at least a 90% probability that the estimated calibration factor would lie within 10% of the actual calibration factor were identified and recommended for each site subtype.

Table 4-3: Descriptive Statistics of Roadway Segments and Intersections

Facility Type	Site Subtype	Total Facility Length (in miles)	Total Number of Sites	Average Segment Length (in miles)	Average Observed Crash Frequency[1]	Actual Calibration Factor
Rural Two-way Two-lane Segments	Two-lane Undivided	3,549.2	3,546	1.00	0.63	0.472
Rural Multilane Segments	Four-lane Undivided	8.2	43	0.20	1.06	--
	Four-lane Divided	1,227.4	1,760	0.70	1.57	0.309
Urban and Suburban Arterials	Two-lane Undivided	616.7	1,791	0.34	1.78	0.581
	Three-lane with TWLTL	67.0	359	0.19	3.12	0.474
	Four-lane Undivided	52.2	266	0.20	4.53	0.663
	Four-lane Divided	1,400.9	4,971	0.28	3.93	0.346
	Five-lane with TWLTL	272.5	1,095	0.25	5.82	0.449
Rural Two-lane Intersections	Three-leg Stop-controlled	--	298	--	0.54	0.471
	Four-leg Stop-controlled	--	43	--	0.84	0.643
	Four-leg Signalized	--	21	--	3.24	0.778
Rural Multilane Intersections	Three-leg Stop-controlled	--	31	--	1.52	1.456
	Four-leg Signalized	--	27	--	4.67	0.392
Urban and Suburban Arterial Intersections	Three-leg Stop-controlled	--	321	--	1.66	0.784
	Four-leg Stop-controlled	--	34	--	2.18	0.792
	Three-leg Signalized	--	58	--	6.89	1.721
	Four-leg Signalized	--	459	--	16.70	2.699

[1] For segment subtypes, the observed crash frequency is in crashes per year per mile; for intersection subtypes, the observed crash frequency is in crashes per year per intersection.

4.2.3 Sampling Procedure

Once the minimum sample sizes were determined, the next question was to determine whether or not there is an additional benefit in estimating calibration factors using stratified sampling procedure compared to using simple random sampling procedure. To answer this question, the entire state was divided into three regions: North Florida, Central Florida, and South Florida.

North Florida region consisted of Districts 2, and 3; Central Florida region included Districts 5 and 7; and South Florida region consisted of Districts 1, 4, and 6. For each site subtype, equal number of samples was selected from each region such that they add up to the total recommended sample size (i.e., using stratified random sampling procedure). The mean calibration factor based on this data set was calculated and compared with the calibration factor estimated from the sample randomly selected from the entire state (i.e., using random sampling procedure). A two-sample t-test with the following hypotheses was performed to compare the calibration factors from the two sampling procedures. Note that the null hypothesis is rejected if p-value is less than 0.05.

- Null hypothesis (H_0): There is no difference in means between the two data sets ($\mu_1 = \mu_2$). It means that the mean calibration factor estimated from a stratified sample is equal to the mean calibration factor estimated from a simple random sample of the entire state.

- Alternative hypothesis (H_a): There are differences in means between the two data sets ($\mu_1 \neq \mu_2$). It means that the mean calibration factor estimated from a stratified sample is not equal to the mean calibration factor estimated from a simple random sample of the entire state.

4.3 Results on Minimum Sample Size

The methodology discussed in Section 4.2.2 was applied to determine the minimum sample size for each site subtype such that there is a high probability that the estimated calibration factor (i.e., calibration factor calculated from a sample) is within 10% of the actual calibration factor (i.e., calibration factor calculated from the entire data set).

4.3.1 Roadway Segments

The analysis of the roadway segments is divided into two subsections: rural segments and urban and suburban arterials.

- *Rural Segments*: Overall, data were collected on 3,549.2 miles (3,546 segments) of rural two-way two-lane road sections. Calibration factors were calculated using data sets with the following sample sizes: 50, 75, 100, 125, 150, 175, 200, 225, 250, 275, and 300. These calibration factors were called estimated calibration factors. Table 4-4 provides the probabilities that the estimated calibration factor would fall within 10% of the actual calibration factor for the different sample sizes. When the sample size is at least 250 segments, there is a 90% probability that the estimated calibration factor is within 10% of the actual factor. Therefore, it is recommended that a sample size should include, at a minimum, 250 segments with about 150 crashes per year to produce a reliable calibration factor (i.e., when the probability that the estimated calibration factor falls within 10% of the actual calibration factor is at least 90%) for rural two-way two-lane roads.

  Table 4-5 gives the probability that the estimated calibration factor would lie within 10% of the actual calibration factor for different sample sizes for rural four-lane divided arterials. These probabilities were found to fluctuate considerably and are counterintuitive; the probabilities did not consistently increase with the increase in sample size. For example, samples with 150 sites showed 85% probability that the estimated calibration factor would lie within 10% of the actual calibration factor, whereas the probability was less than 80% when 175 and 200 sites were considered. These fluctuations, to an extent, could be attributed to the variations in the site conditions. It can be concluded that a sample size of 250 sites might yield a reliable calibration factor where there is at least 85% probability that the estimated calibration factor would lie within 10% of the actual calibration factor. The minimum sample size of 250 segments with approximately 270 crashes per year is recommended for rural four-lane divided highways.

Table 4-4: Calibration Factors by Sample Sizes for Rural Two-lane Two-way Roads

No. of Sites	Percent of Total Length	Total Crashes per Year	Actual CF[1]	Estimated CF[1]	SD[2]	Z_{min}	Z_{max}	$P(Z<Z_{min})$	$P(Z<Z_{max})$	$P(Z_{min}<Z<Z_{max})$[3]
50	1.4%	29.5	0.472	0.456	0.065	-0.480	0.972	32%	83%	52%
75	2.1%	46.3	0.472	0.469	0.057	-0.775	0.881	22%	81%	59%
100	2.8%	61.5	0.472	0.475	0.052	-0.965	0.850	17%	80%	64%
125	3.5%	76.4	0.472	0.469	0.043	-1.028	1.167	15%	88%	73%
150	4.3%	94.2	0.472	0.465	0.042	-0.957	1.290	17%	90%	73%
175	5.0%	108.3	0.472	0.466	0.036	-1.144	1.478	13%	93%	80%
200	5.6%	126.5	0.472	0.473	0.032	-1.506	1.444	7%	93%	86%
225	6.4%	140.1	0.472	0.467	0.030	-1.407	1.740	8%	96%	88%
250	7.1%	154.9	0.472	0.463	0.027	-1.415	2.081	8%	98%	90%
275	7.8%	173.8	0.472	0.474	0.027	-1.822	1.674	3%	95%	92%
300	8.5%	192.7	0.472	0.469	0.026	-1.700	1.931	4%	97%	93%

[1] CF refers to calibration factor; [2] SD refers to standard deviation; [3] probability that the estimated calibration factor is within 10% of the actual calibration factor.

Table 4-5: Calibration Factors by Sample Sizes for Rural Four-lane Divided Arterials

No. of Sites	Percent of Total Length	Total Crashes per Year	Actual CF[1]	Estimated CF[1]	SD[2]	Z_{min}	Z_{max}	$P(Z<Z_{min})$	$P(Z<Z_{max})$	$P(Z_{min}<Z<Z_{max})$[3]
50	2.8%	54.7	0.309	0.323	0.062	-0.724	0.273	23%	61%	37%
75	4.3%	78.7	0.309	0.301	0.035	-0.654	1.111	26%	87%	61%
100	5.8%	109.7	0.309	0.319	0.036	-1.136	0.581	13%	72%	59%
125	7.2%	135.2	0.309	0.311	0.033	-0.997	0.876	16%	81%	65%
150	8.5%	160.2	0.309	0.307	0.021	-1.376	1.567	8%	94%	86%
175	9.9%	191.0	0.309	0.307	0.025	-1.156	1.316	12%	91%	78%
200	11.5%	222.3	0.309	0.311	0.026	-1.265	1.112	10%	87%	76%
225	12.7%	243.7	0.309	0.311	0.018	-1.828	1.606	3%	95%	91%
250	14.2%	268.6	0.309	0.305	0.021	-1.281	1.662	10%	95%	85%
275	15.6%	296.2	0.309	0.305	0.016	-1.681	2.181	5%	99%	94%
300	16.9%	329.7	0.309	0.314	0.020	-1.795	1.295	4%	90%	87%

[1] CF refers to calibration factor; [2] SD refers to standard deviation; [3] probability that the estimated calibration factor is within 10% of the actual calibration factor.

- *Urban and Suburban Arterials:* Tables 4-6 through 4-10 give the probability that the estimated calibration factor would vary within 10% of the actual calibration factor for different sample sizes for different site subtypes in urban and suburban areas. On urban and suburban two-lane undivided arterials (see Table 4-6), when 300 sites are considered, there is an 80% probability that the estimated calibration factor would lie within 10% of the actual calibration factor. It can be concluded that a reliable calibration factor for urban and suburban two-lane undivided arterials would require a minimum of 300 roadway segments with an average frequency of 180 crashes per year.

  As shown in Table 4-7, on urban and suburban three-lane arterial sections with a center TWLTL, the estimated calibration factor was found to be within 10% of the actual calibration factor at over 90% probability when a sample size of 200 or more sites are considered. Therefore, a minimum sample size of 200 segments (constituting 35 miles) with about 120 crashes per year is recommended for urban and suburban three-lane arterial sections with a center TWLTL.

  On urban and suburban four-lane undivided arterial sections (see Table 4-8), when 150 or more sites are considered, there is a 90% probability that the estimated calibration factor would lie within 10% of the actual calibration factor. Therefore, a minimum sample size of 150 segments (constituting 30 miles) with about 130 crashes per year is recommended. Similarly, for urban and suburban four-lane divided arterial sections (see Table 4-9), a minimum sample size of 500 segments (constituting 140 miles) with about 550 crashes per year is recommended. Although 500 segments might be considered to be a large number, it is to be noted that there are over 1,400 segments within this site subtype. Therefore, 500 segments constitute to only 10.1% of the total segment length.

  As shown in Table 4-10, on urban and suburban five-lane arterial sections with a center TWLTL, samples consisting of 225 sites showed a 90% probability while samples consisting of 250 sites showed about 80% probability that the estimated calibration factor would fall within 10% of the actual factor. A sample size of 275 sites (constituting approximately 70 miles of the road network) is therefore considered to be adequate. This sample size corresponds to about 400 crashes per year.

Table 4-6: Calibration Factors by Sample Sizes for Urban Two-lane Undivided Arterials

No. of Sites	Percent of Total Length	Total Crashes per Year	Actual CF[1]	Estimated CF[1]	SD[2]	Z_{min}	Z_{max}	$P(Z<Z_{min})$	$P(Z<Z_{max})$	$P(Z_{min}<Z<Z_{max})$[3]
50	2.8%	30.5	0.616	0.616	0.104	-0.592	0.592	28%	72%	45%
75	4.2%	44.6	0.616	0.610	0.093	-0.598	0.727	27%	77%	49%
100	5.7%	61.7	0.616	0.622	0.072	-0.939	0.772	17%	78%	61%
125	7.0%	76.7	0.616	0.632	0.087	-0.892	0.524	19%	70%	51%
150	8.3%	92.4	0.616	0.628	0.068	-1.082	0.729	14%	77%	63%
175	9.6%	104.4	0.616	0.619	0.052	-1.254	1.122	10%	87%	76%
200	11.1%	121.0	0.616	0.620	0.046	-1.428	1.244	8%	89%	82%
225	12.3%	134.1	0.616	0.608	0.041	-1.315	1.718	9%	96%	86%
250	13.6%	147.8	0.616	0.605	0.050	-1.012	1.466	16%	93%	77%
275	15.1%	165.1	0.616	0.617	0.035	-1.782	1.729	4%	96%	92%
300	16.8%	181.8	0.616	0.616	0.048	-1.284	1.267	10%	90%	80%

[1] CF refers to calibration factor; [2] SD refers to standard deviation; [3] probability that the estimated calibration factor is within 10% of the actual calibration factor.

Table 4-7: Calibration Factors by Sample Sizes for Urban Three-lane Arterials with TWLTL

No. of Sites	Percent of Total Length	Total Crashes per Year	Actual CF[1]	Estimated CF[1]	SD[2]	Z_{min}	Z_{max}	$P(Z<Z_{min})$	$P(Z<Z_{max})$	$P(Z_{min}<Z<Z_{max})$[3]
50	14.2%	30.6	0.440	0.447	0.084	-0.604	0.437	27%	67%	40%
75	21.2%	44.0	0.440	0.442	0.066	-0.693	0.633	24%	74%	49%
100	28.5%	60.1	0.440	0.447	0.084	-0.604	0.437	27%	67%	40%
125	34.6%	69.3	0.440	0.427	0.040	-0.785	1.437	22%	92%	71%
150	41.9%	86.8	0.440	0.435	0.036	-1.072	1.378	14%	92%	77%
175	48.5%	102.4	0.440	0.445	0.029	-1.691	1.344	5%	91%	87%
200	55.5%	117.7	0.440	0.448	0.022	-2.374	1.675	1%	95%	94%
225	63.1%	131.8	0.440	0.440	0.024	-1.832	1.871	3%	97%	94%
250	69.3%	145.9	0.440	0.441	0.021	-2.092	2.016	2%	98%	96%
275	76.4%	159.7	0.440	0.439	0.018	-2.395	2.542	1%	99%	99%
300	83.9%	174.1	0.440	0.438	0.012	-3.436	3.824	0%	100%	100%

[1] CF refers to calibration factor; [2] SD refers to standard deviation; [3] probability that the estimated calibration factor is within 10% of the actual calibration factor.

Table 4-8: Calibration Factors by Sample Sizes for Urban Four-lane Undivided Arterials

No. of Sites	Percent of Total Length	Total Crashes per Year	Actual CF[1]	Estimated CF[1]	SD[2]	Z_{min}	Z_{max}	$P(Z<Z_{min})$	$P(Z<Z_{max})$	$P(Z_{min}<Z<Z_{max})$[3]
50	19.0%	45.0	0.346	0.356	0.058	-0.769	0.424	22%	66%	44%
75	28.2%	66.2	0.346	0.353	0.034	-1.224	0.812	11%	79%	68%
100	37.5%	88.6	0.346	0.347	0.028	-1.271	1.200	10%	88%	78%
125	46.6%	111.0	0.346	0.348	0.024	-1.525	1.358	6%	91%	85%
150	55.9%	133.8	0.346	0.346	0.021	-1.648	1.648	5%	95%	90%
175	65.9%	158.3	0.346	0.349	0.017	-2.212	1.859	1%	97%	96%
200	74.5%	177.8	0.346	0.347	0.014	-2.543	2.400	1%	99%	99%
225	83.9%	200.0	0.346	0.349	0.011	-3.418	2.873	0%	100%	100%
250	93.7%	221.8	0.346	0.346	0.007	-4.943	4.943	0%	100%	100%

[1] CF refers to calibration factor; [2] SD refers to standard deviation; [3] probability that the estimated calibration factor is within 10% of the actual calibration factor.

Table 4-9: Calibration Factors by Sample Sizes for Urban Four-lane Divided Arterials

No. of Sites	Percent of Total Length	Total Crashes per Year	Actual CF[1]	Estimated CF[1]	SD[2]	Z_{min}	Z_{max}	$P(Z<Z_{min})$	$P(Z<Z_{max})$	$P(Z_{min}<Z<Z_{max})$[3]
50	1.0%	54.3	0.642	0.645	0.121	-0.551	0.510	29%	69%	40%
75	1.5%	81.3	0.642	0.661	0.066	-1.275	0.682	10%	75%	65%
100	2.0%	112.6	0.642	0.681	0.088	-1.167	0.290	12%	61%	49%
125	2.4%	134.7	0.642	0.665	0.068	-1.277	0.600	10%	73%	62%
150	3.0%	167.6	0.642	0.671	0.067	-1.386	0.528	8%	70%	62%
175	3.5%	188.6	0.642	0.654	0.059	-1.276	0.888	10%	81%	71%
200	4.0%	220.5	0.642	0.673	0.053	-1.789	0.612	4%	73%	69%
225	4.5%	243.2	0.642	0.663	0.044	-1.949	0.990	3%	84%	81%
250	5.0%	275.2	0.642	0.663	0.044	-1.958	0.986	3%	84%	81%
275	5.6%	311.2	0.642	0.674	0.039	-2.481	0.845	1%	80%	79%
300	6.1%	336.0	0.642	0.667	0.040	-2.219	0.986	1%	84%	82%
325	6.4%	357.7	0.642	0.673	0.043	-2.236	0.774	1%	78%	77%
350	7.1%	385.2	0.642	0.657	0.041	-1.929	1.202	3%	89%	86%
375	7.5%	411.6	0.642	0.663	0.038	-2.243	1.154	1%	88%	86%
400	8.1%	445.3	0.642	0.670	0.036	-2.535	1.013	1%	84%	84%
425	8.7%	478.9	0.642	0.668	0.039	-2.330	0.991	1%	84%	83%
450	9.0%	488.6	0.642	0.666	0.038	-2.337	1.078	1%	86%	85%
475	9.7%	530.6	0.642	0.663	0.035	-2.432	1.221	1%	89%	88%
500	10.1%	554.4	0.642	0.665	0.028	-3.059	1.470	0%	93%	93%

[1] CF refers to calibration factor; [2] SD refers to standard deviation; [3] probability that the estimated calibration factor is within 10% of the actual calibration factor.

Table 4-10: Calibration Factors by Sample Sizes for Urban Five-lane Arterials with TWLTL

No. of Sites	Percent of Total Length	Total Crashes per Year	Actual CF[1]	Estimated CF[1]	SD[2]	Z_{min}	Z_{max}	$P(Z<Z_{min})$	$P(Z<Z_{max})$	$P(Z_{min}<Z<Z_{max})$[3]
50	4.6%	75.8	0.449	0.466	0.081	-0.764	0.344	22%	63%	41%
75	7.0%	116.2	0.449	0.473	0.071	-0.970	0.294	17%	62%	45%
100	9.3%	152.1	0.449	0.463	0.057	-1.033	0.542	15%	71%	56%
125	11.6%	194.7	0.449	0.475	0.041	-1.729	0.461	4%	68%	64%
150	14.0%	222.2	0.449	0.452	0.041	-1.168	1.022	12%	85%	73%
175	16.0%	253.6	0.449	0.454	0.039	-1.279	1.023	10%	85%	75%
200	18.0%	296.0	0.449	0.461	0.038	-1.497	0.866	7%	81%	74%
225	20.6%	321.2	0.449	0.443	0.027	-1.441	1.885	7%	97%	90%
250	23.1%	381.2	0.449	0.464	0.03	-1.997	0.997	2%	84%	82%
275	25.0%	401.4	0.449	0.455	0.028	-1.818	1.389	3%	92%	88%
300	27.0%	427.6	0.449	0.449	0.026	-1.727	1.727	4%	96%	92%

[1] CF refers to calibration factor; [2] SD refers to standard deviation; [3] probability that the estimated calibration factor is within 10% of the actual calibration factor.

4.3.2 Intersections

Several intersection subtypes have limited sites (i.e., < 50 sites), and therefore, the minimum sample size to calculate a reliable calibration factor could not be determined. The minimum sample size was not calculated for the following intersection subtypes:

- Rural two-lane four-leg stop-controlled intersections
- Rural two-lane four-leg signalized intersections
- Rural multilane three-leg stop-controlled intersections
- Rural multilane four-leg signalized intersections
- Urban and suburban four-leg stop-controlled intersections

The minimum sample size was determined for the following intersection subtypes:

- Rural two-lane three-leg stop-controlled intersections
- Urban and suburban three-leg stop-controlled intersections
- Urban and suburban three-leg signalized intersections
- Urban and suburban four-leg signalized intersections

The analysis of the intersections is divided into three subsections: rural unsignalized intersections, urban and suburban unsignalized intersections, and urban and suburban signalized intersections.

- *Rural Unsignalized Intersections:* Although rural two-lane three-leg intersections with minor road stop control are one of the prevalent types of intersections in Florida, crash statistics show that they experience the lowest frequency in terms of the number of crashes per year per intersection (see Table 4-3). This might result in requiring more than

100 intersections to produce a reliable calibration factor (i.e., the estimated calibration factor is within 10% of the actual factor). Table 4-11 shows that 120 intersections have a 63% probability and 150 intersections have an 81% probability that the estimated calibration factor would be within 10% of the actual factor. Therefore, a sample size of 150 intersections could produce a reliable calibration factor. These 150 intersections experienced about 85 crashes per year.

Table 4-11: Calibration Factors by Sample Sizes for Rural Two-lane Three-leg Stop-controlled Intersections

No. of Sites	Percent of Total Intersections	Crashes per Year	Actual CF[1]	Estimated CF[1]	SD[2]	Z_{min}	Z_{max}	$P(Z<Z_{min})$	$P(Z<Z_{max})$	$P(Z_{min}<Z<Z_{max})$[3]
30	10.1%	16.5	0.471	0.471	0.103	-0.457	0.457	32%	68%	35%
50	16.8%	27.9	0.471	0.489	0.083	-0.784	0.351	22%	64%	42%
60	20.1%	33.4	0.471	0.484	0.080	-0.751	0.426	23%	67%	44%
70	23.5%	38.2	0.471	0.463	0.063	-0.621	0.875	27%	81%	54%
80	26.8%	45.8	0.471	0.487	0.055	-1.147	0.565	13%	71%	59%
90	30.2%	50	0.471	0.483	0.062	-0.953	0.566	17%	71%	54%
100	33.6%	55.3	0.471	0.483	0.06	-0.985	0.585	16%	72%	56%
110	36.9%	61	0.471	0.474	0.044	-1.139	1.002	13%	84%	71%
120	40.3%	64.2	0.471	0.471	0.052	-0.906	0.906	18%	82%	63%
130	43.6%	72.2	0.471	0.475	0.038	-1.345	1.134	9%	87%	78%
140	47.0%	76.5	0.471	0.472	0.042	-1.145	1.098	13%	86%	74%
150	50.3%	82.5	0.471	0.480	0.035	-1.603	1.089	5%	86%	81%

[1] CF refers to calibration factor; [2] SD refers to standard deviation; [3] probability that the estimated calibration factor is within 10% of the actual calibration factor.

- *Urban and Suburban Unsignalized Intersections:* Table 4-12 gives the probability that the estimated calibration factor would lie within 10% of the actual calibration factor for different sample sizes for urban and suburban three-leg intersections with minor road stop control. From the table, it can be inferred that at least 100 intersections are required to obtain an acceptable calibration factor with lower standard deviation. However, when at least 130 intersections are considered, there is a 90% probability that the estimated calibration factor would lie within 10% of the actual calibration factor. These 130 intersections experienced approximately 200 crashes per year.

- *Urban and Suburban Signalized Intersections:* Tables 4-13 and 4-14 give the probability that the estimated calibration factor would lie within 10% of the actual calibration factor for different sample sizes for urban and suburban three-leg and four-leg signalized intersections, respectively. The number of urban and suburban three-leg signalized intersections is relatively low; however, this intersection subtype has the second highest number of crashes occurring per year per intersection (see Table 4-3). From Table 4-13, it can be inferred that a total of 50 intersections gives an 87% probability for the estimated calibration factor to be within 10% of the actual factor. Therefore, a minimum sample size of 50 intersections constituting approximately 350 crashes per year is recommended.

Table 4-12: Calibration Factors by Sample Sizes for Urban Three-leg Stop-controlled Intersections

No. of Sites	Percent of Total Intersections	Crashes per Year	Actual CF[1]	Estimated CF[1]	SD[2]	Z_{min}	Z_{max}	$P(Z<Z_{min})$	$P(Z<Z_{max})$	$P(Z_{min}<Z<Z_{max})$[3]
30	9.3%	50.6	0.784	0.800	0.115	-0.821	0.543	21%	71%	50%
50	15.6%	84.6	0.784	0.792	0.109	-0.793	0.646	21%	74%	53%
60	18.7%	101.1	0.784	0.826	0.105	-1.147	0.347	13%	64%	51%
70	21.8%	115.3	0.784	0.797	0.122	-0.749	0.536	23%	70%	48%
80	24.9%	131.8	0.784	0.787	0.094	-0.866	0.802	19%	79%	60%
90	28.0%	156.1	0.784	0.836	0.087	-1.499	0.303	7%	62%	55%
100	31.2%	163.5	0.784	0.782	0.069	-1.107	1.165	13%	88%	74%
110	34.3%	184.4	0.784	0.785	0.073	-1.088	1.060	14%	86%	72%
120	37.4%	198.3	0.784	0.774	0.060	-1.140	1.473	13%	93%	80%
130	40.5%	202.5	0.784	0.782	0.048	-1.592	1.675	6%	95%	90%
140	43.6%	229.4	0.784	0.777	0.047	-1.519	1.817	6%	97%	90%
150	46.7%	244.9	0.784	0.773	0.048	-1.404	1.863	8%	97%	89%

[1] CF refers to calibration factor; [2] SD refers to standard deviation; [3] probability that the estimated calibration factor is within 10% of the actual calibration factor.

Four-leg signalized intersections are the most predominant intersection subtypes in urban and suburban areas in Florida. Furthermore, these intersections are found to experience a significantly high number of crashes compared to other intersection subtypes (see Table 4-3). Because of these prevailing crash statistics, a relatively small sample size might yield a reliable calibration factor. Table 4-14 shows that as low as 40 intersections could be able to produce a calibration factor that lies within 10% of the actual factor with over 70% probability. However, it is noticeable that the standard deviation of the calibration factor is relatively higher than that for other intersection subtypes for this particular sample size. A sample of 80 intersections would be adequate to have a reliable calibration factor with lower standard deviation. The probability that the estimated calibration factor lies within 10% of actual calibration factor is 87% for a sample of 80 intersections. The average crash frequency to achieve this probability should be approximately 1,300 crashes per year. Therefore, a minimum sample size of 80 intersections constituting 1,300 crashes per year is recommended for urban and suburban four-leg signalized intersections.

Table 4-13: Calibration Factors by Sample Sizes for Urban Three-leg Signalized Intersections

No. of Sites	Percent of Total Intersections	Crashes per Year	Actual CF[1]	Estimated CF[1]	SD[2]	Z_{min}	Z_{max}	$P(Z<Z_{min})$	$P(Z<Z_{max})$	$P(Z_{min}<Z<Z_{max})$[3]
30	51.7%	206.1	1.721	1.754	0.241	-0.851	0.577	20%	72%	52%
40	69.0%	269.1	1.721	1.700	0.161	-0.939	1.199	17%	88%	71%
50	86.2%	347.9	1.721	1.725	0.115	-1.531	1.462	6%	93%	87%

[1] CF refers to calibration factor; [2] SD refers to standard deviation; [3] probability that the estimated calibration factor is within 10% of the actual calibration factor.

Table 4-14: Calibration Factors by Sample Sizes for Urban Four-leg Signalized Intersections

No. of Sites	Percent of Total Intersections	Crashes per Year	Actual CF[1]	Estimated CF[1]	SD[2]	Z_{min}	Z_{max}	$P(Z<Z_{min})$	$P(Z<Z_{max})$	$P(Z_{min}<Z<Z_{max})$[3]
30	6.5%	475.4	2.699	2.537	0.325	-0.332	1.329	37%	91%	54%
40	8.7%	674.8	2.699	2.703	0.251	-1.091	1.059	14%	86%	72%
50	10.9%	674.8	2.699	2.776	0.236	-1.470	0.817	7%	79%	72%
60	13.1%	1015.6	2.699	2.707	0.213	-1.305	1.230	10%	89%	79%
70	15.3%	1179.0	2.699	2.716	0.205	-1.400	1.234	8%	89%	81%
80	17.4%	1296.4	2.699	2.673	0.177	-1.378	1.672	8%	95%	87%
90	19.6%	1504.3	2.699	2.667	0.17	-1.399	1.776	8%	96%	88%
100	21.8%	1651.8	2.699	2.715	0.133	-2.150	1.909	2%	97%	96%

[1] CF refers to calibration factor; [2] SD refers to standard deviation; [3] probability that the estimated calibration factor is within 10% of the actual calibration factor.

4.3.3 Normality Assumption Test

As mentioned earlier, the calibration factors calculated using different subsets of the same data were assumed to follow normal distribution This assumption was verified for the recommended sample sizes using the Shapiro-Wilk test, which was appropriate for small sample sizes (< 100 samples), and which can also handle sample sizes as large as 2,000 (Gan et al., 2007). The Shapiro-Wilk test with the following hypotheses was performed:

- Null hypothesis (H_0): Calibration factors estimated from the recommended sample size are normally distributed.

- Alternative hypothesis (H_a): Calibration factors estimated from the recommended sample size are not normally distributed.

Table 4-15 gives the test results for the recommended sample size from each facility type and subtype. The null hypothesis is rejected if p-value is less than 0.05, implying that calibration factors are not normally distributed. The p-values of all the site subtypes except for urban and suburban four-lane divided arterials and urban and suburban four-leg signalized intersections are greater than 0.05. It implies that the calibration factors estimated from the recommended sample size are normally distributed.

Table 4-15: Results of Shapiro-Wilk Normality Test for Calibration Factors

Facility Type	Site Subtype	Recommended Sample Size[1]	p-value
Rural Two-lane Two-way Segments	Two-lane Undivided	250	0.085
Rural Multilane Segments	Four-lane Divided	250	0.889
Urban and Suburban Arterials	Two-lane Undivided	300	0.168
	Three-lane with TWLTL	200	0.753
	Four-lane Undivided	150	0.142
	Four-lane Divided	500	0.035
	Five-lane with TWLTL	275	0.164
Rural Two-lane Intersections	Three-leg Stop-controlled	150	0.206
Urban and Suburban Arterial Intersections	Three-leg Stop-controlled	130	0.548
	Three-leg Signalized	50	0.481
	Four-leg Signalized	80	0.046

[1] For each site subtype, a set of 30 calibration factors was tested for normality.

4.4 Results on Sampling Procedure

As mentioned in Section 4.2.3, a two-sample t-test was conducted to determine whether the calibration factor estimated from a stratified sample is statistically different from the calibration factor estimated from a simple random sample. For each site subtype, the sample size considered for stratified random sampling and simple random sampling was the sample size recommended in Section 4.5.1. Table 4-16 gives the results of this analysis.

Rural two-lane three-leg stop-controlled intersections and urban and suburban three-leg signalized intersections do not have enough samples in each region, and therefore were not analyzed. For all the remaining site subtypes except for rural two-way two-lane segments and urban and suburban four-lane divided arterials, there was no sufficient evidence to reject the null hypothesis. The results imply that the calibration factor estimated from a stratified random sample was not statistically different from the calibration factor estimated from a simple random sample of the entire state. It is, therefore, acceptable to obtain the recommended sample size through a simple random sampling procedure.

4.5 Summary

4.5.1 Sample Size

For the different facility types discussed in the HSM, the minimum sample sizes to estimate reliable calibration factors were determined. The calibration factor estimated from a sample size was considered to be reliable if there is a high probability that the estimated calibration factor would lie within 10% of the actual calibration factor (which was calculated from the entire data set). Tables 4-17 and 4-18 give probabilities that the estimated calibration factors fall within 10% of actual calibration factor for different sample sizes for different segment and intersection subtypes, respectively. Tables 4-19 and 4-20 give the recommended minimum sample sizes required to estimate reliable calibration factors for segment and intersection subtypes, respectively. For each site subtype, the tables also give the minimum sample size and the corresponding percentage of total length for segments, percentage of total intersections for intersections, and the total number of crashes per year, to yield a reliable calibration factor.

Table 4-16: Comparison of Sampling Procedures

Site Subtype	Random Sample Stratified by Region		Simple Random Sample With No Stratification		p-value	Comments
	Sample Size	Estimated Calibration Factor	Sample Size	Estimated Calibration Factor		
Rural Two-way Two-lane Roadway Segments	255	0.483	250	0.463	0.0118	Reject H_0
Rural Four-lane Divided Arterials	255	0.312	250	0.305	0.1889	Fail to Reject H_0
Urban and Suburban Two-lane Undivided Arterials	300	0.628	300	0.616	0.3157	Fail to Reject H_0
Urban and Suburban Three-lane Arterials with TWLTL	210	0.456	200	0.448	0.1082	Fail to Reject H_0
Urban and Suburban Four-lane Undivided Arterials	150	0.346	150	0.347	0.6940	Fail to Reject H_0
Urban and Suburban Four-lane Divided Arterials	510	0.646	500	0.665	0.0227	Reject H_0
Urban and Suburban Five-lane Arterials with TWLTL	285	0.446	275	0.455	0.2151	Fail to Reject H_0
Urban and Suburban Three-leg Stop-controlled Intersections	135	0.783	130	0.773	0.4710	Fail to Reject H_0
Urban and Suburban Four-leg Signalized Intersections	90	2.711	80	2.672	0.4522	Fail to Reject H_0

Table 4-17: Summary Statistics of Estimated Calibration Factors for Segments

Rural Two-way Two-lane Roadway Segments	Number of Sites	**50**	**75**	**100**	**125**	**150**	**175**	**200**	**225**	**250**	**275**	**300**
	Percentage of Total Length	1.4%	2.1%	2.8%	3.5%	4.3%	5.0%	5.6%	6.4%	7.1%	7.8%	8.5%
	Total Crashes per Year	29.5	46.3	61.5	76.4	94.2	108.3	126.5	140.1	154.9	173.8	192.7
	P (Estimated CF falls within 10% of Actual CF)	52%	59%	64%	73%	73%	80%	86%	88%	90%	92%	93%
Rural Four-lane Divided Arterials	Number of Sites	**50**	**75**	**100**	**125**	**150**	**175**	**200**	**225**	**250**	**275**	**300**
	Percentage of Total Length	2.8%	4.3%	5.8%	7.2%	8.5%	9.9%	11.5%	12.7%	14.2%	15.6%	16.9%
	Total Crashes per Year	54.7	78.7	109.7	135.2	160.2	191.0	222.3	243.7	268.6	296.2	329.7
	P (Estimated CF falls within 10% of Actual CF)	37%	61%	59%	65%	86%	78%	76%	91%	85%	94%	87%
Urban Two-lane Undivided Arterials	Number of Sites	**50**	**75**	**100**	**125**	**150**	**175**	**200**	**225**	**250**	**275**	**300**
	Percentage of Total Length	2.8%	4.2%	5.7%	7.0%	8.3%	9.6%	11.1%	12.3%	13.6%	15.1%	16.8%
	Total Crashes per Year	30.5	44.6	61.7	76.7	92.4	104.4	121.0	134.1	147.8	165.1	181.8
	P (Estimated CF falls within 10% of Actual CF)	45%	49%	61%	51%	63%	76%	82%	86%	77%	92%	80%
Urban Three-lane Arterials with TWLTL	Number of Sites	**50**	**75**	**100**	**125**	**150**	**175**	**200**	**225**	**250**	**275**	**300**
	Percentage of Total Length	14.2%	21.2%	28.5%	34.6%	41.9%	48.5%	55.5%	63.1%	69.3%	76.4%	83.9%
	Total Crashes per Year	30.6	44.0	60.1	69.3	86.8	102.4	117.7	131.8	145.9	159.7	174.1
	P (Estimated CF falls within 10% of Actual CF)	40%	49%	40%	71%	77%	87%	94%	94%	96%	99%	100%
Urban Four-lane Undivided Arterials	Number of Sites	**50**	**75**	**100**	**125**	**150**	**175**	**200**	**225**	**250**		
	Percentage of Total Length	19.0%	28.2%	37.5%	46.6%	55.9%	65.9%	74.5%	83.9%	93.7%		
	Total Crashes per Year	45.0	66.2	88.6	111.0	133.8	158.3	177.8	200.0	221.8		
	P (Estimated CF falls within 10% of Actual CF)	44%	68%	78%	85%	90%	96%	99%	100%	100%		
Urban Four-lane Divided Arterials	Number of Sites	**50**	**100**	**150**	**200**	**250**	**300**	**350**	**400**	**450**	**500**	
	Percentage of Total Length	1.0%	2.0%	3.0%	4.0%	5.0%	6.1%	7.1%	8.1%	9.0%	10.1%	
	Total Crashes per Year	54.3	112.6	167.6	220.5	275.2	336.0	385.2	445.3	488.6	554.4	
	P (Estimated CF falls within 10% of Actual CF)	40%	49%	62%	69%	81%	82%	86%	84%	85%	93%	
Urban Five-lane Arterials with TWLTL	Number of Sites	**50**	**75**	**100**	**125**	**150**	**175**	**200**	**225**	**250**	**275**	**300**
	Percentage of Total Length	4.6%	7.0%	9.3%	11.6%	14.0%	16.0%	18.0%	20.6%	23.1%	25.0%	27.0%
	Total Crashes per Year	75.8	116.2	152.1	194.7	222.2	253.6	296.0	321.2	381.2	401.4	427.6
	P (Estimated CF falls within 10% of Actual CF)	41%	45%	56%	64%	73%	75%	74%	90%	82%	88%	92%

Note: CF refers to calibration factor.

Table 4-18: Summary Statistics of Estimated Calibration Factors for Intersections

Rural Two-lane Three-leg Stop-controlled Intersections	Number of Intersections	**30**	**50**	**60**	**70**	**80**	**90**	**100**	**110**	**120**	**130**	**140**	**150**
	Percentage of Total Intersections	10.1%	16.8%	20.1%	23.5%	26.8%	30.2%	33.6%	36.9%	40.3%	43.6%	47.0%	50.3%
	Total Crashes per Year	16.5	27.9	33.4	38.2	45.8	50.0	55.3	61.0	64.2	72.2	76.5	82.5
	P(Estimated CF falls within 10% of Actual CF)[1]	35%	42%	44%	54%	59%	54%	56%	71%	63%	78%	74%	81%
Urban and Suburban Three-leg Stop-controlled Intersections	Number of Intersections	**30**	**50**	**60**	**70**	**80**	**90**	**100**	**110**	**120**	**130**	**140**	**150**
	Percentage of Total Intersections	9.3%	15.6%	18.7%	21.8%	24.9%	28.0%	31.2%	34.3%	37.4%	40.5%	43.6%	46.7%
	Total Crashes per Year	50.6	84.6	101.1	115.3	131.8	156.1	163.5	184.4	198.3	202.5	229.4	244.9
	P(Estimated CF falls within 10% of Actual CF)[1]	50%	53%	51%	48%	60%	55%	74%	72%	80%	90%	90%	89%
Urban and Suburban Three-leg Signalized Intersections	Number of Intersections	**30**	**40**	**50**									
	Percentage of Total Intersections	51.7%	69.0%	86.2%									
	Total Crashes per Year	206.1	269.1	347.9									
	P(Estimated CF falls within 10% of Actual CF)[1]	52%	71%	87%									
Urban and Suburban Four-leg Signalized Intersections	Number of Intersections	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**				
	Percentage of Total Intersections	6.5%	8.7%	10.9%	13.1%	15.3%	17.4%	19.6%	21.8%				
	Total Crashes per Year	475.4	674.8	674.8	1,015.6	1,179	1,296.4	1,504.3	1,651.8				
	P(Estimated CF falls within 10% of Actual CF)[1]	54%	72%	72%	79%	81%	87%	88%	96%				

[1] CF refers to calibration factor.

Table 4-19: Recommended Minimum Sample Sizes for Segments

Segment Site Subtype	Total Sample Length (in miles)	Sample Size	Percent of Total Length	Number of Crashes per Year	P (Estimated CF is within 10% of Actual CF)[1]
Rural Two-way Two-lane Roadway Segments	250	250	7.1%	150	90%
Rural Four-lane Divided Arterials	175	250	14.2%	270	85%
Urban and Suburban Two-lane Undivided Arterials	102	300	16.8%	180	80%
Urban and Suburban Three-lane Arterials with TWLTL	38	200	55.5%	120	94%
Urban and Suburban Four-lane Undivided Arterials	30	150	55.9%	130	90%
Urban and Suburban Four-lane Divided Arterials	140	500	10.1%	550	93%
Urban and Suburban Five-lane Arterials with TWLTL	70	275	25%	400	88%

[1] CF refers to calibration factor.

Table 4-20: Recommended Minimum Sample Sizes for Intersections

Intersection Site Subtype	Sample Size	Percent of Total Intersections	Number of Crashes per Year	P (Estimated CF is within 10% of Actual CF)[1]
Rural Two-lane Three-leg Stop-controlled Intersections	150	50.3%	85	81%
Urban and Suburban Three-leg Stop-controlled Intersections	130	40.5%	200	90%
Urban and Suburban Three-leg Signalized Intersections	50	86.2%	350	87%
Urban and Suburban Four-leg Signalized Intersections	80	17.4%	1,300	87%

[1] CF refers to calibration factor.

4.5.2 Sampling Procedure

The calibration factors estimated from two sampling procedures, simple random sampling and stratified sampling, were compared using a two-sample t-test to determine whether the calibration factors estimated from a stratified sample (by region) was significantly different from the calibration factors estimated from a simple random sample of the entire state. The calibration factor estimated from a stratified sample was not statistically different from the calibration factor estimated from a simple random sample for the following site subtypes:

- Rural four-lane divided arterials
- Urban and suburban two-lane undivided arterials
- Urban and suburban three-lane arterials with TWLTL
- Urban and suburban four-lane undivided arterials
- Urban and suburban five-lane arterials with TWLTL
- Urban and suburban three-leg stop-controlled intersections
- Urban and suburban four-leg signalized intersections

On the other hand, the calibration factor estimated from a stratified sample was statistically different from the calibration factor estimated from a simple random sample for rural two-way two-lane segments and urban and suburban four-lane divided arterials. The remaining site subtypes were not analyzed as they do not have enough samples in each region. Based on the results, it is considered to be acceptable to obtain the recommended sample size through a simple random sampling procedure from the entire state data.

CHAPTER 5
SAFETYANALYST OVERVIEW

This chapter provides an overview of *SafetyAnalyst*. It first introduces the software application and then describes the data flow structure. It further discusses the different modules and components in *SafetyAnalyst*. It also includes a detailed discussion on the data requirements. Finally, the chapter concludes with a summary.

5.1 Introduction

SafetyAnalyst was developed as a cooperative effort by FHWA and participating state and local agencies. *SafetyAnalyst* provides a suite of analytical tools to identify and manage system-wide safety improvements by incorporating all the following six steps in the roadway safety management process:

1. network screening,
2. diagnosis,
3. countermeasure selection,
4. economic appraisal,
5. priority ranking, and
6. countermeasure evaluation.

SafetyAnalyst is designed to account for the regression-to-the-mean (RTM) bias which exists in the current practice of selecting projects for safety improvements. The RTM selection bias is a direct result of the general practice of selecting high crash locations for safety improvements, which in turn results in the overestimation of crash reduction factors. To address this bias, *SafetyAnalyst* implements the EB method which requires the use of SPFs. SPFs are mathematical relationships that link crash occurrence to traffic and roadway characteristics. *SafetyAnalyst* includes a set of national default SPFs developed from multiple years of data from California, Minnesota, Ohio, and Washington. To better represent local crash experience, *SafetyAnalyst* calibrates the default SPFs to local data.

5.2 *SafetyAnalyst* Data Flow Model

SafetyAnalyst "consists of a set of multiple independent applications (or tools) that interact with a database using a two-tier, client-server architecture" (ITT, 2009). Figure 5-1 shows the data flow in *SafetyAnalyst.* As can be seen in the figure, the key data items required for performing safety analyses in *SafetyAnalyst* are the roadway inventory, traffic volume, and crash data. Another set of data (some of which are optional) include countermeasures (i.e., safety improvement projects), SPFs, operational data, etc. The software has the following four major applications (i.e., tools) that interact with each other:

1. Data Management Tool
2. Analytical Tool
3. Administration Tool
4. Implemented Countermeasures Tool

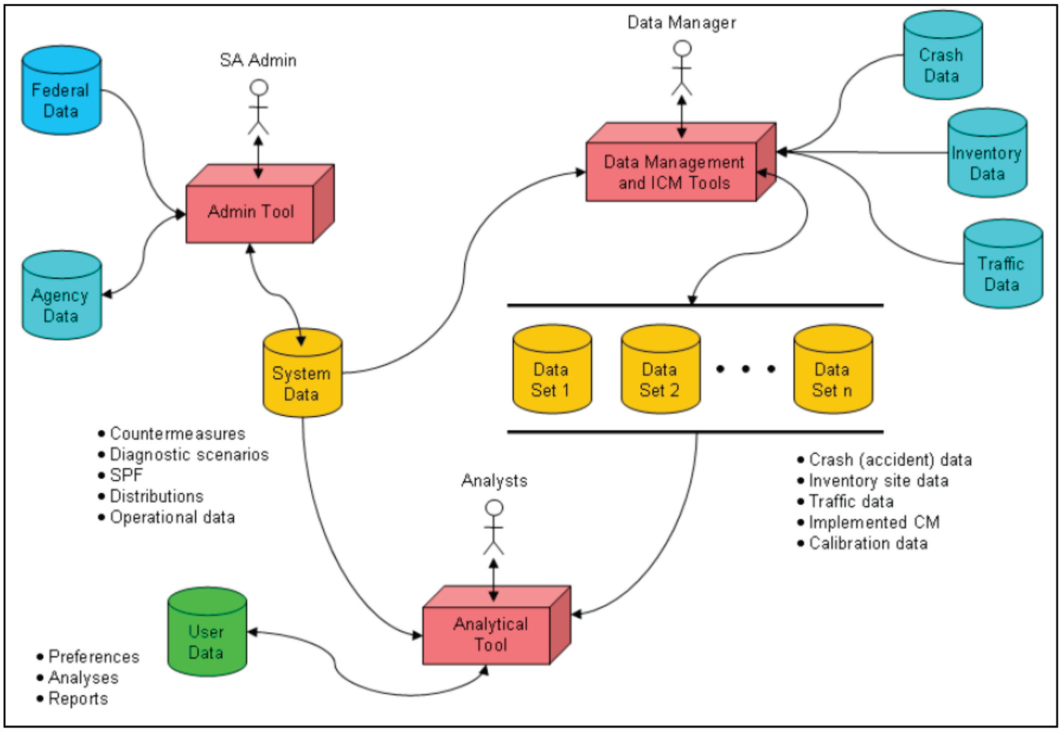


Figure 5-1: *SafetyAnalyst* Data Flow Structure (Harwood et al., 2010)

5.3 *SafetyAnalyst* Modules

SafetyAnalyst includes the following four modules and could act as a complete "safety toolbox" (AASHTO, 2010b):

1. Network Screening
2. Diagnosis and Countermeasure Selection
3. Economic Appraisal and Priority Ranking
4. Countermeasure Evaluation

These four modules are included in the *SafetyAnalyst's* Analytical Tool. A detailed discussion on these modules is provided in the following sections.

5.3.1 Network Screening Module

"The basic purpose of the Network Screening Module is to review the entire roadway network or portions of the roadway network, under the jurisdiction of a highway agency and identify and prioritize those sites that have promise as sites for potential safety improvements" (ITT, 2009). In short, it identifies and ranks sites with potential for safety improvements (PSI). The following types of screening could be performed within the Network Screening Module:

- Basic network screening with peak searching on roadway segments and coefficient of variation (CV) test
- Basic network screening with sliding window on roadway segments

- Screening for high proportion of specific crash type
- Sudden increase in mean crash frequency
- Steady increase in mean crash frequency
- Corridor screening

5.3.2 Diagnosis and Countermeasure Selection Module

This module is used to diagnose the nature of safety problems at specific sites. The countermeasure selection module assists users in selecting the countermeasures to reduce crash frequency and severity at specific sites. Within this module, the following could be conducted:

- generate collision diagrams,
- generate crash summary statistics, and
- conduct statistical tests on crash frequencies and/or proportions.

"The end result of this diagnosis process is a list of recommended countermeasures that, if implemented at the site, could serve to mitigate particular collision patterns" (ITT, 2009). The next module assists users in conducting economic evaluation of the recommended countermeasure(s).

5.3.3 Economic Appraisal and Priority Ranking Module

As the name implies, the economic appraisal module performs an economic appraisal of a specific countermeasure or several alternative countermeasures for a specific site while the priority ranking module provides a priority ranking of sites and proposed improvement projects based on the benefit and cost estimates determined by the economic appraisal tool.

5.3.4 Countermeasure Evaluation Module

This module provides the capability to estimate the safety effect of one or multiple countermeasures implemented at specific sites. The effectiveness measures are usually expressed as a percentage change in crash frequencies or in specific target crash types. Other effectiveness measures such as a shift in the proportion of specific crash types could also be evaluated. The countermeasures are evaluated using EB based before-and-after evaluations. Furthermore, the shift in the proportion of specific crash types is testes using the Wilcoxon signed rank test.

5.4 *SafetyAnalyst* Components

SafetyAnalyst includes the Data Management Tool, the Analytical Tool, the Administration Tool, and the Implemented Countermeasure Tool to perform the complete roadway safety management process. A detailed discussion on these components is provided in the following sections.

5.4.1 Data Management Tool

The Data Management Tool is used to import, post-process, and calibrate data. Once the database is set up, it is ready for data import using *SafetyAnalyst*'s Data Management Tool.

Once the data are successfully imported, post-processed, and calibrated, the data are used by the Analytical Tool to perform various safety analyses. During the entire data import process, the Data Management Tool performs the following tasks:

- connects segments to each other and to intersections,
- locates crashes on the inventory (segments, intersections, and ramps),
- assigns site subtypes to the inventory,
- merges segments into homogeneous segments (optional),
- validates traffic volume data,
- generates agency-specific crash distributions,
- calibrates data, and
- performs other data validation steps.

At each step within the Data Management Tool, the software provides a log of errors and warnings pertaining to the data. During post processing, the *SafetyAnalyst* provides an option to aggregate homogeneous segments. In other words, *SafetyAnalyst* can automatically reduce data sensitivity of certain variables and can aggregate shorter segments into longer homogeneous segments. Figure 5-2 shows the screen shot of the Homogeneous Segment Aggregation Parameters Dialog Box in the *SafetyAnalyst* Data Management Tool.

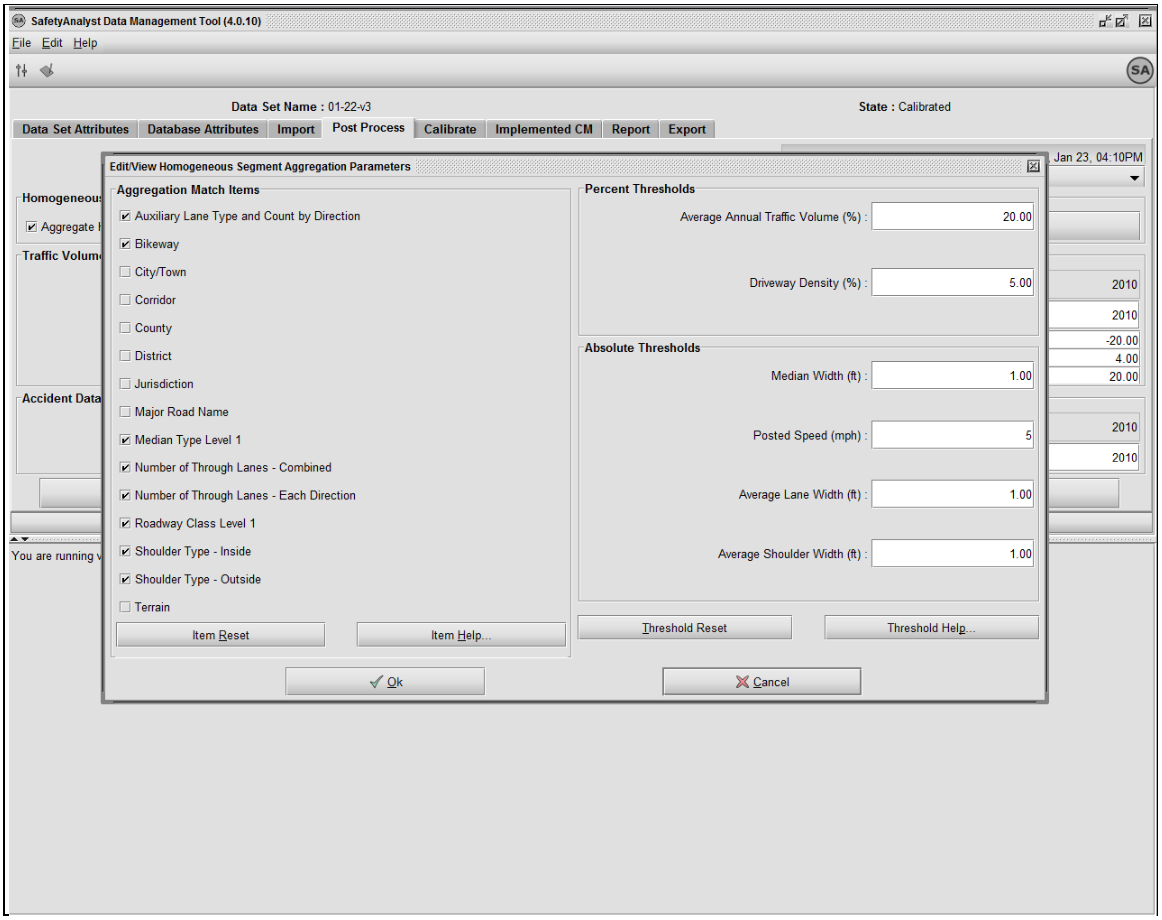


Figure 5-2: Homogeneous Segment Aggregation Parameters Dialog Box in *SafetyAnalyst* Data Management Tool

5.4.2 Analytical Tool

The Analytical Tool is used to perform analysis on the data. The four *SafetyAnalyst* modules discussed in Section 5.3 are performed within this tool. A Getting Started Wizard, as shown in Figure 5-3, helps users to get started using the Analytical Tool. All the data calibrated in the Data Management Tool will be available for analysis in the Analytical Tool. Users can create site lists by selecting one or more sites (or locations) in the inventory to conduct the analysis (i.e., run one of the four modules).

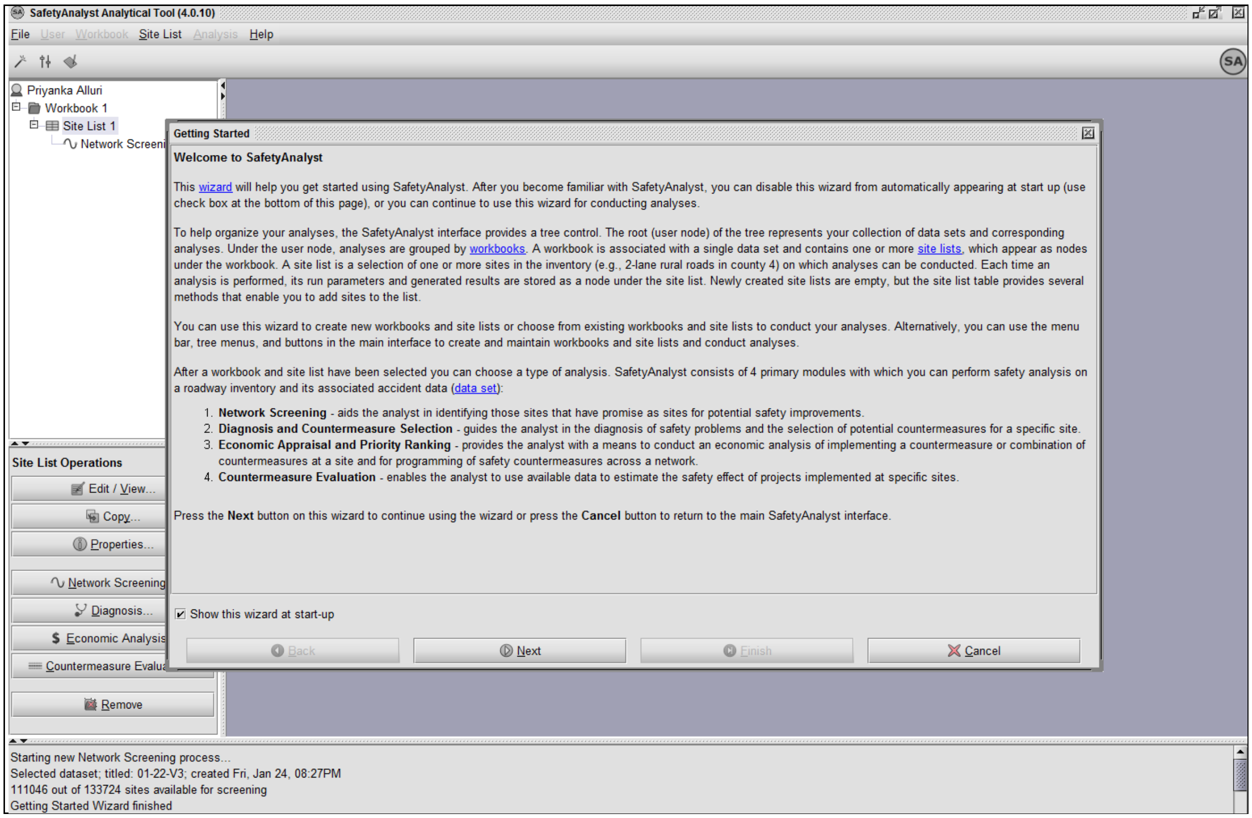


Figure 5-3: Getting Started Wizard in *SafetyAnalyst* Analytical Tool

5.4.3 Administration Tool

The Administration Tool enables users to tailor *SafetyAnalyst* for their organization and to manage the countermeasures, crash distributions, diagnostics, and SPFs used in the Analytical Tool. The Administration Tool is used to perform a variety of tasks such as adding and removing data items (with an exception of required data variables). Data recoding of various data elements' attributes could also be performed. This tool also gives access to the national default SPFs used within the software which could be replaced with agency-specific SPFs, if available. Further, diagnostic questions and countermeasures could also be edited within this tool.

In the Administration Tool, users can specifically edit the following:

- deployment-specific data attributes,

- agency-specified site subtypes,
- user permissions in the Analytical Tool,
- agency countermeasures,
- diagnostic scenarios supported by the *SafetyAnalyst* Diagnosis and Countermeasure Selection Module,
- severity-level crash distributions associated with the site subtypes supported by *SafetyAnalyst*, and
- SPFs.

The Administration Tool operates with three separate databases: Federal, Agency, and System. The Federal database is a read-only database and contains the default site subtypes, SPFs, crash distributions, etc. The Agency database is a local database that users can maintain. It contains agency-specified subtypes, agency-specific SPFs, etc. Finally, the System database merges the default Federal database and the Agency database, and it is used as the source for the *SafetyAnalyst* Analytical Tool.

5.4.4 Implemented Countermeasure Tool

The *SafetyAnalyst* Implemented Countermeasure Management Tool offers a subset of the capabilities of the *SafetyAnalyst* Data Management Tool. It can only be used to specify, modify, or remove implemented countermeasure data, and it does not allow users to modify the inventory, crash, or traffic data in a *SafetyAnalyst* data set.

5.5 *SafetyAnalyst* Data Requirements

SafetyAnalyst makes use of three groups of variables: location variables, primary data variables, and cost and function variables (ITT, 2009). Depending on the applications, some of the variables in these groups either do not apply or are optional.

5.5.1 Location Variables

SafetyAnalyst describes a location based on one of the following four reference systems:

1. Route/Milepost
2. Route/County/Milepost
3. Section/Distance
4. Route/Section/Distance

For Florida data, the Route/Milepost reference system that is used as the FDOT’s linear referencing system is based on county, section, subsection, and milepost.

5.5.2 Primary Data Variables

The latest version of the data dictionary of *SafetyAnalyst* includes a total of 16 data sets for the so-called "primary data". They include variables mainly for roadway inventory, crash, and traffic volume. These 16 data sets are (ITT, 2009):

1. *Geographic Description*: defines the attributes related to the geographic description of the site.
2. *Alternate Route Name*: defines an alternate route name associated with the geographic description.
3. *Intersection*: displays the general geometric design and traffic control data associated with the given intersection.
4. *Major Road Annual Traffic*: displays the traffic volume data for the major road associated with an intersection.
5. *Minor Road Annual Traffic*: displays the traffic volume data for the minor roads associated with an intersection.
6. *Intersection Leg*: displays geometric design and traffic control, and traffic volume data for the intersection approach.
7. *Leg Annual Traffic*: the average number of vehicles passing through this intersection from this approach in a day, for all days of the year, during a specified calendar year, expressed in vehicles per day.
8. *Leg Vehicle Movements*: indicates the average number of vehicles that travel straight through the intersection, or turn left or turn right onto a cross street, expressed as either vehicles per day or an hourly volume.
9. *Ramp*: displays the general geometric design data associated with the given ramp.
10. *Annual Traffic*: displays the traffic volume data for the ramp.
11. *Roadway Segment*: the general geometric design and traffic volume data associated with the given roadway segment inventory site.
12. *Annual Traffic*: specifies the traffic volume data associated with a roadway segment for a given year.
13. *Directional Attributes*: defines the general geometric design associated with the given roadway segment for a specific direction of travel.
14. *Auxiliary Lane*: indicates the presence of additional lanes on the roadway segment.
15. *Accident*: defines the data associated with a crash. This is a top-level data element.
16. *Vehicle-level Accident Data*: defines the vehicle data items associated with each vehicle involved in the crash.

5.5.3 Cost and Function Variables

These are variables for values such as benefit and cost calculations, etc. While *SafetyAnalyst* will come with a set of default values for many of these values, it is expected that the state would want to use state-specific values to better reflect local conditions. The current set of benefit and cost variables includes (ITT, 2009):

- *Minimum Attractive Rate of Return:* the minimum attractive rate of return.
- *Cost Fatal*: the cost for a crash with a fatality.
- *Cost Incapacitating Injury*: the cost for a crash with an incapacitating injury.
- *Cost Serious Injury:* the cost for a crash with a serious injury.
- *Cost Minor Injury:* the cost for a crash with a minor injury.
- *Cost Property Damage Only*: the cost for a crash with property damage only.
- *Weight Fatal*: the relative analysis weight for a crash with a fatality.
- *Weight Incapacitating Injury*: the relative analysis weight for a crash with an incapacitating injury.
- *Weight Serious Injury:* the relative analysis weight for a crash with a serious injury.
- *Weight Minor Injury*: the relative analysis weight for a crash with a minor injury.
- *Weight Property Damage Only*: the relative analysis weight for a crash with property damage only.
- *Total Budget for Countermeasure Construction* the budgeted number of dollars for the construction of countermeasures in the future period on which the optimization will be based.

5.6 Summary

In summary, *SafetyAnalyst* is a suite of software tools that implement the advanced EB method and automates all the steps in the roadway safety management process. Although the data requirements are intense, once the data are imported, the analysis procedures are easy requiring minimum statistical expertise. The software has the following four major applications (i.e., tools) that interact with each other:

1. Data Management Tool
2. Analytical Tool
3. Administration Tool
4. Implemented Countermeasures Tool

SafetyAnalyst includes the following four analysis modules within the Analytical Tool:

1. Network Screening Module
2. Diagnosis and Countermeasure Selection Module
3. Economic Appraisal and Priority Ranking Module
4. Countermeasure Evaluation Module

CHAPTER 6
SAFETYANALYST DATA CONVERTER

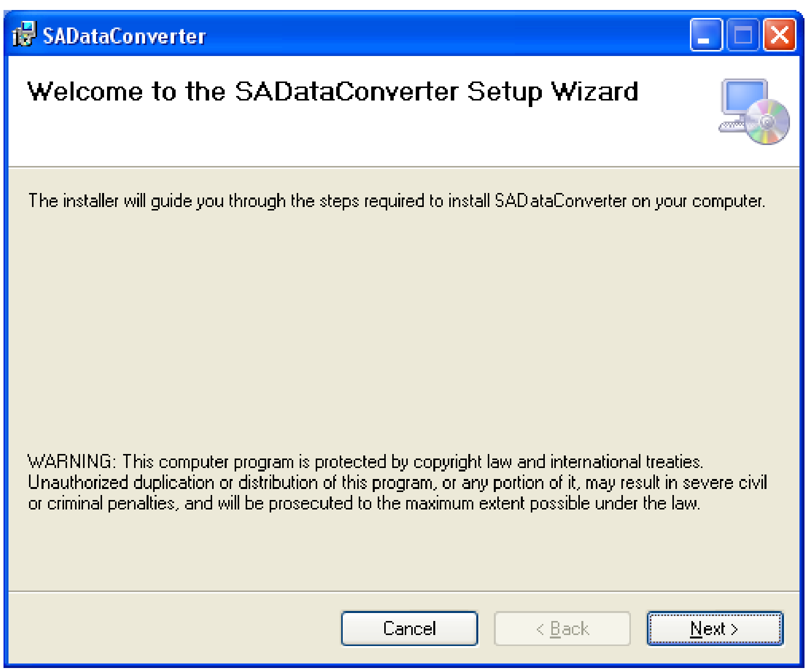
This chapter focuses on *SafetyAnalyst* Data Converter, the software application developed to convert Florida data into the standard format required by *SafetyAnalyst*. The software installation procedure is first discussed. The Converter's User Interface is then described. A detailed discussion on the input files and the data conversion is provided next. Finally, the chapter concludes with a summary.

6.1 Introduction

SafetyAnalyst Data Converter application was developed to generate import files for *SafetyAnalyst* for the entire state road network in Florida. The Converter converts the data from FDOT's CAR system, RCI, and other source databases, such as intersection node list and Linear Reference System (LRS), into the data format required by *SafetyAnalyst*. The application includes the following four conversion tools: (1) roadway segment data conversion, (2) intersection data conversion, (3) ramp data conversion, and (4) crash data conversion. The output files from these conversion tools can be directly imported into *SafetyAnalyst* Data Management Tool.

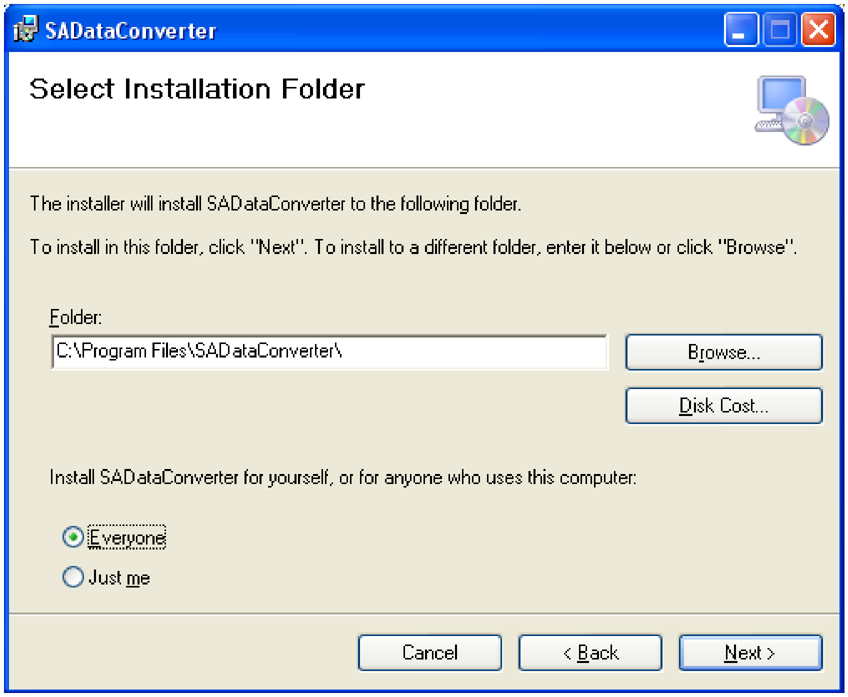
6.2 Software Installation

SafetyAnalyst Data Converter was developed using Visual C# inside the Visual Studio 2010, which is based on .NET Framework 4.0. The Converter can run on Microsoft Windows XP, Windows 7, and later versions. The installation of the Converter can be initiated by double clicking the *SafetyAnalyst* Data Converter setup file (*Setup.exe*), and by following the instructions. Figure 6-1 shows the installation steps.

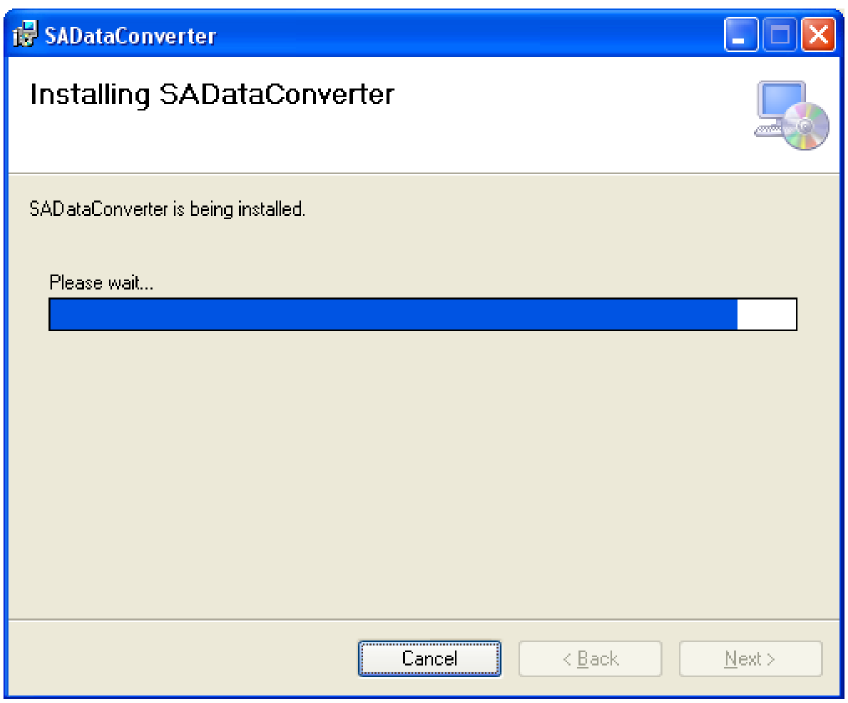


(a) Step 1

Figure 6-1: *SafetyAnalyst* Data Converter Installation Steps

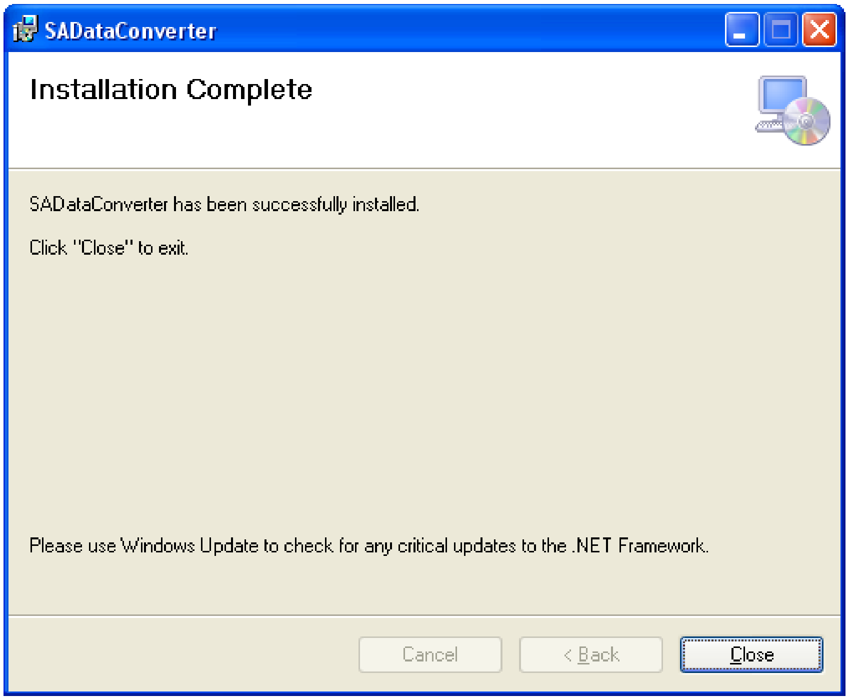


(b) Step 2



(c) Step 3

Figure 6-1: *SafetyAnalyst* Data Converter Installation Steps



(d) Step 4

Figure 6-1: *SafetyAnalyst* Data Converter Installation Steps

Once the installation has completed, a shortcut to the Converter tool will be automatically added to the desktop (See Figure 6-2).

Figure 6-2: *SafetyAnalyst* Data Converter Desktop Shortcut

6.3 User Interface

Figure 6-3 shows the main screen of the *SafetyAnalyst* Data Converter. The main screen includes the following six groups of selection boxes:

1. Jurisdiction
    a. Statewide
    b. Districtwide
2. Site Type
    a. Segment
    b. Intersection
    c. Ramp
3. Data Year
    a. From
    b. To

4. Roadway Data
    a. Node list file (RDWTBL25)
    b. Segment list file (RDWTBL31)
    c. RCI file
5. Crash Data
    a. Crash-level information files for the selected years (RDWTBL50)
    b. Vehicle-level information files for the selected years (RDWTBL51)
6. Converted *SafetyAnalyst* Files
    a. Save to folder (i.e., output file location)
    b. File name prefix (i.e., output file name prefix)

In the **Jurisdiction** group box, the *Statewide* option treats all input files as one system and the Converter generates one set of import files for the entire state. On the other hand, when the *Districtwide* option is chosen, the Converter generates different sets of import files for each district. This option is convenient when *SafetyAnalyst* has to be run separately for each district.

The **Site Type** group box allows users to select one or more facility types (i.e., segments, intersections, and ramps) for which the *SafetyAnalyst* import files have to be generated. Usually all segment, intersection, and ramp types are selected so that all the crashes can be linked to one of the three facility types. However, if users want to convert just the segment data, then users can just select *Segment* in the **Site Type** group box.

The **Data Year** group box allows users to select the period for which the import files could be generated. The Converter package has included the data year range from 2007 through 2012. In **Data Year** group box, *From* and *To* dropdown lists allow users to choose the analysis period.

The **Roadway Data** and **Crash Data** group boxes list the input data files needed for segment, intersection, ramp, and crash conversion tasks. The **Converted *SafetyAnalyst* Files** group box allows users to select the output file destination. It also allows users to include an optional prefix name to the output file names.

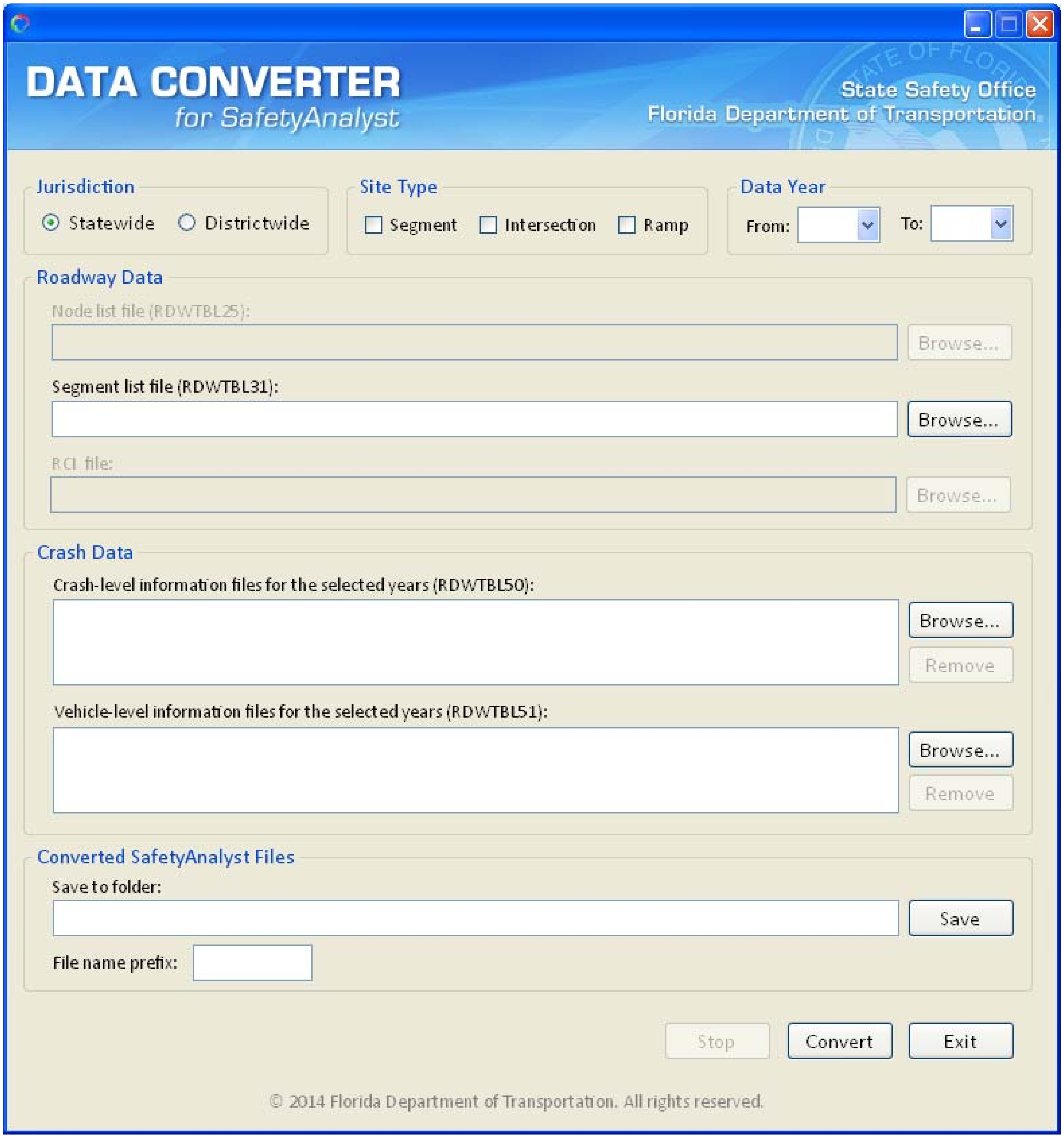


Figure 6-3: *SafetyAnalyst* Data Converter Main Screen

6.3.1 Input Data Files Required to Run the Converter

The following input files are required to run the *SafetyAnalyst* Data Converter:

- RDWTBL25 (intersection node list file)
- RDWTBL31 (segment list file, also known as LRS file)
- RCI (Roadway Characteristics Inventory file)
- RDWTBL50 (crash-level information for all the years included in the analysis)
- RDWTBL51 (vehicle-level information for all the years included in the analysis)

Table 6-1 lists all the input files required for the different conversion tasks. All input files except the RCI file are required to be either in comma separated value (CSV) format (i.e., with .csv file extension) or in text format (i.e., with .txt file extension). Also, the first row in all the input data files must include field names. The RCI file has to be inputted in Microsoft Access database format. The Converter supports both .mdb (for Microsoft Access 2003 and below versions) and .accdb (for Microsoft Access 2007 and above versions) formats.

Table 6-1: Input Data Files Required to Run the Converter

Data Source	Data Source Description	Conversion Task
RDWTBL25	Intersection node list	• Intersection
RDWTBL31	Segment list, also known as LRS file	• Intersection • Ramp • Segment
RCI Database	Roadway Characteristics Inventory data	• Intersection • Segment
RDWTBL50	Crash-level information	• Crash
RDWTBL51	Vehicle-level information	• Crash

In addition to the above identified input files, yearly AADT data are required. This data are available in the RCI file, which is updated annually. As such, for each analysis year, AADT data for the entire state road network has to be extracted from the corresponding year's RCI file. This requires users to input yearly RCI files for all the analysis years. To minimize users' efforts, the AADT data for the six most recent years (i.e., 2007-2012) were extracted from the corresponding year's RCI files and stored within the application. This allows users to select multiple years for the analysis without actually inputting the RCI data for multiple years.

SafetyAnalyst categorizes ramps into 16 subtypes based on ramp type and configuration. However, Florida's classification of ramps is different from the *SafetyAnalyst*'s classification. To analyze the maximum number of ramps, Florida-specific ramp subtypes were generated as per the classification used in Florida. Furthermore, off-ramp and on-ramp information is incomplete in the RCI database. In Project BDK80-977-07, Gan et al. (2012) collected and preprocessed this information. This file was stored within the Converter. When ramp data has to be processed, the Converter will retrieve information from this preprocessed file to generate an import file with a more complete list of ramps.

6.3.2 Output Files

Table 6-2 lists the names of all the output files generated by the Converter. The Converter provides an option to include a prefix to the output file names. Note that the output file names are predetermined and are consistent with the file names listed in the *SafetyAnalyst* User Manual (ITT, 2009). Also, the districtwide output files will include district number (i.e., D1 through D7) as an additional prefix. Appendix C provides a description of the variables in the output files.

Table 6-2: Output Data Files from the Converter

Conversion Task	Statewide Mode[1]	Districtwide Mode[1,2]
Segment	[prefix_]altSegment.csv [prefix_]altSegmentTraffic.csv	Dx_ [prefix_] altSegment.csv Dx_ [prefix_] altSegment.csv
Intersection	[prefix_]altIntersection.csv [prefix_]altMajorRoadTraffic.csv [prefix_]altMinorRoadTraffic.csv [prefix_]altLeg.csv [prefix_]altLegTraffic.csv	Dx_ [prefix_] altIntersection.csv Dx_ [prefix_] altMajorRoadTraffic.csv Dx_ [prefix_] altMinorRoadTraffic.csv Dx_ [prefix_] altLeg.csv Dx_ [prefix_] altLegTraffic.csv
Ramp	[prefix_]altRamp.csv [prefix_]altRampTraffic.csv	Dx_ [prefix_] altRamp.csv Dx_ [prefix_] altRampTraffic.csv
Crash	[prefix_]altAccident.csv	Dx_ [prefix_] altAccident.csv

[1] The Converter provides an option to add a prefix to the output file names.
[2] Dx denotes the district number (i.e., D1 through D7).

6.4 Data Conversion

As mentioned earlier, *SafetyAnalyst* import files have to be generated in line with the data requirements and format recommended by the software (Harwood et al., 2010). This requires a significant amount of data recoding and data extraction from multiple sources. The data mapping structure was originally adapted from Lu et al. (2009) and then modified. Appendix D gives the data mapping structure used within the Converter.

The Converter has two major components: roadway inventory data conversion and crash data conversion. As mentioned earlier, RDWTBL25, RDWTBL31, and RCI files are required to convert roadway inventory data, and RDWTBL50 and RDWTBL51 are required to convert crash data. It is recommended to convert both roadway inventory and crash data at the same time so that crashes can be assigned to the road network (i.e., segments, intersections, and ramps). However, roadway inventory and crash data could be also converted separately, if needed.

6.4.1 Segment and Ramp Data Conversion

Florida collects and maintains information on about 227 variables in its RCI database. With this level of detail, segmentation of road network might result in shorter segments as roadways are segmented whenever there is a slight change in any one of the 227 variables. However, not all 227 variables are required to generate import files for *SafetyAnalyst*. Therefore, longer segments can be generated when only the data variables required for generating import files for *SafetyAnalyst* are used in the process of segmentation. Within the Converter, a dynamic segment merging feature was used to merge short segments of the same roadway into longer segments. The following variables were considered in generating longer homogeneous segments:

- Route type
- Area type
- Presence of interchange influence area (IIA)
- Number of through lanes
- Median type and median width
- Operation way
- Presence and type of auxiliary lanes
- Shoulder type and width
- Average lane width
- Roadway class
- Driveway density
- Posted speed limit
- Access control
- Bikeway
- Travel direction
- Growth factor

According to the *SafetyAnalyst* User's Manual (ITT, 2009), the interchange influence area of a particular interchange covers the length of the freeway section extending approximately 0.3 miles upstream of the gore point of the first exit/entrance ramp to approximately 0.3 miles downstream of the gore point of the last entrance/exit ramp of the same interchange. Interchange influence areas are not explicitly identified within FDOT's roadway inventory database. These segments were identified through data preprocessing in ArcGIS. More details on the procedure used to identify IIAs are provided in Gan et al. (2012). The similar dynamic segment merging algorithm was also used to merge short ramp segments into longer segments.

6.4.2 Intersection Data Conversion

RDWTBL25 provides the list of nodes on the entire state road network in Florida. It provides unique node ID for each intersection along with the roadway ID, intersection mile post, and geographic coordinates. In addition to intersections, node IDs are also provided for non-intersections (for example, where the alignment of the roadway changes, etc.). Therefore, a major task was to identify intersections on state roads from the node list data file. In order to identify all intersections including the intersections where state roads cross local roads, the following steps were followed in ArcGIS:

- *Step 1:* A 25-meter buffer was created around each node location to generate a new GIS buffer layer.
- *Step 2:* The GIS buffer layer generated in Step 1 was intersected with NAVTEQ's *NAVSTREETS* street layer to retrieve all three- and four-leg intersections. In addition to the intersection-level attributes, approach-level attributes such as number of lanes and AADT, were retrieved and were used to generate import files for intersections.
- *Step 3:* As the intersections in the node list have roadway ID and mileposts of only the major roads, the mileposts of the intersecting minor roads were estimated using the linear referencing tool in ArcGIS.
- *Step 4:* FDOT's Traffic Signal GIS layer was used to identify whether the intersection is signalized or non-signalized. The final intersection list included both signalized and non-signalized intersections.

This intersection list was saved within the Converter database and is used in the intersection conversion task.

6.4.3 Crash Data Conversion

As mentioned earlier, both crash-level (RDWTBL50) and vehicle-level (RDWTBL51) information data files are required to generate *SafetyAnalyst* import files. For each analysis year, the Converter requires both RDWTBL50 and RDWTBL51 files. When users fail to import both the tables, the Converter pops up an error message window. If the segment, ramp, or intersection conversion options have not been selected (i.e., checked), the crash conversion task will automatically search all converted files with same prefix name under the folder specified inside the **Converted *SafetyAnalyst* Files** group to assign the intersection, ramp, or segment identifiers for each converted crash record.

6.5 Summary

This chapter focused on the data conversion program developed to automatically generate import files for *SafetyAnalyst*. The Converter has the capability to generate all the required import files for segments, intersections, ramps, and their associated crash and traffic files. The Converter also has the capability to generate both statewide and districtwide import files. The output from the Converter can be directly inputted into the *SafetyAnalyst* Data Management Tool.

CHAPTER 7
APPLICATION OF *SAFETYANALYST*

This chapter describes the *SafetyAnalyst* application process to be followed to identify high crash locations. It first discusses the data flow structure between *SafetyAnalyst* components and the other applications. It then provides a discussion on the SPFs used in *SafetyAnalyst.* Particularly, the SPFs developed for unsignalized intersections to be used within *SafetyAnalyst* are discussed. Finally, the process of identifying high crash locations is explained, and sample results are provided.

7.1 *SafetyAnalyst* Application Process

Besides the components of *SafetyAnalyst*, the application process includes two additional systems: the Data Converter program and the GIS Tool. As part of the current research effort, a Data Converter program was developed to automatically generate import files for *SafetyAnalyst*. Also, a GIS Tool was developed to spatially display high crash locations (Gan et al., 2012). Figure 7-1 illustrates the entire *SafetyAnalyst* application process for Florida. It further provides the interactions and the data flow between the Converter, the different components of *SafetyAnalyst*, and the GIS Tool.

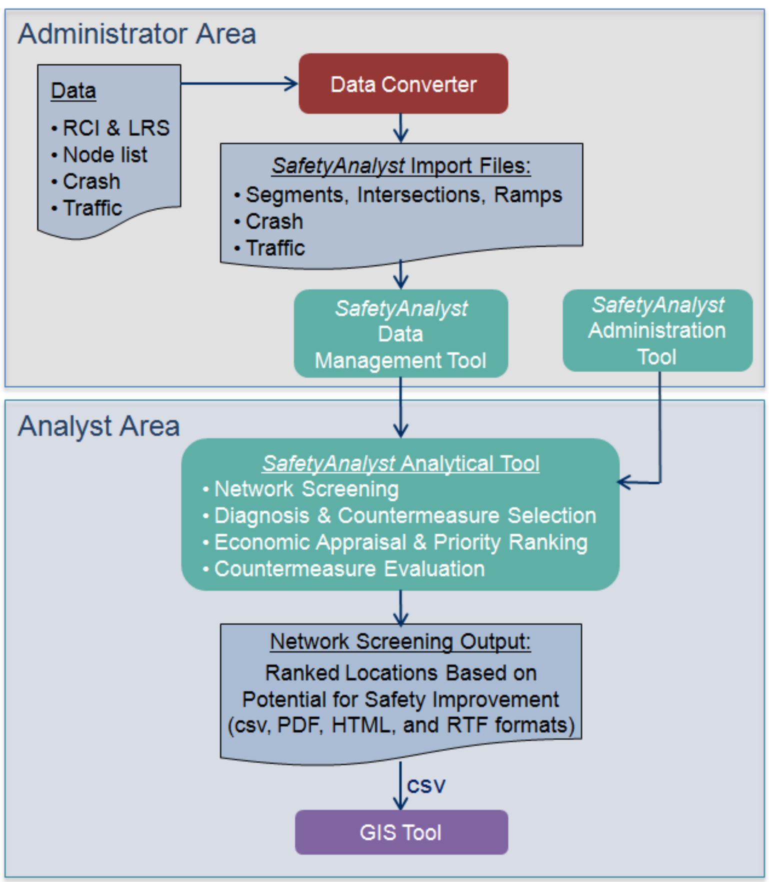


Figure 7-1: *SafetyAnalyst* Application Process

7.1.1 Data Converter

As discussed in Chapter 6, the Data Converter was developed to convert Florida data into the standard format required by *SafetyAnalyst*. The Converter has the capability to generate all the required import files for segments, intersections, ramps, and their associated crash and traffic files. The output from the Converter can be directly inputted into the *SafetyAnalyst* Data Management Tool.

7.1.2 GIS Tool

Given the spatial nature of crash data, GIS has become an essential tool for highway safety analyses. As part of a recently completed project, a new GIS system was developed specifically to work with *SafetyAnalyst* (Gan et al., 2012). Figure 7-2 shows the main screen of the GIS interface design and data layers. The new GIS system serves two purposes in the context of *SafetyAnalyst* applications:

1. provide an alternative method for selecting roadways, intersections, and ramps for analysis by *SafetyAnalyst* using a graphical display, and
2. graphically display the output from the Network Screening Module of *SafetyAnalyst*.

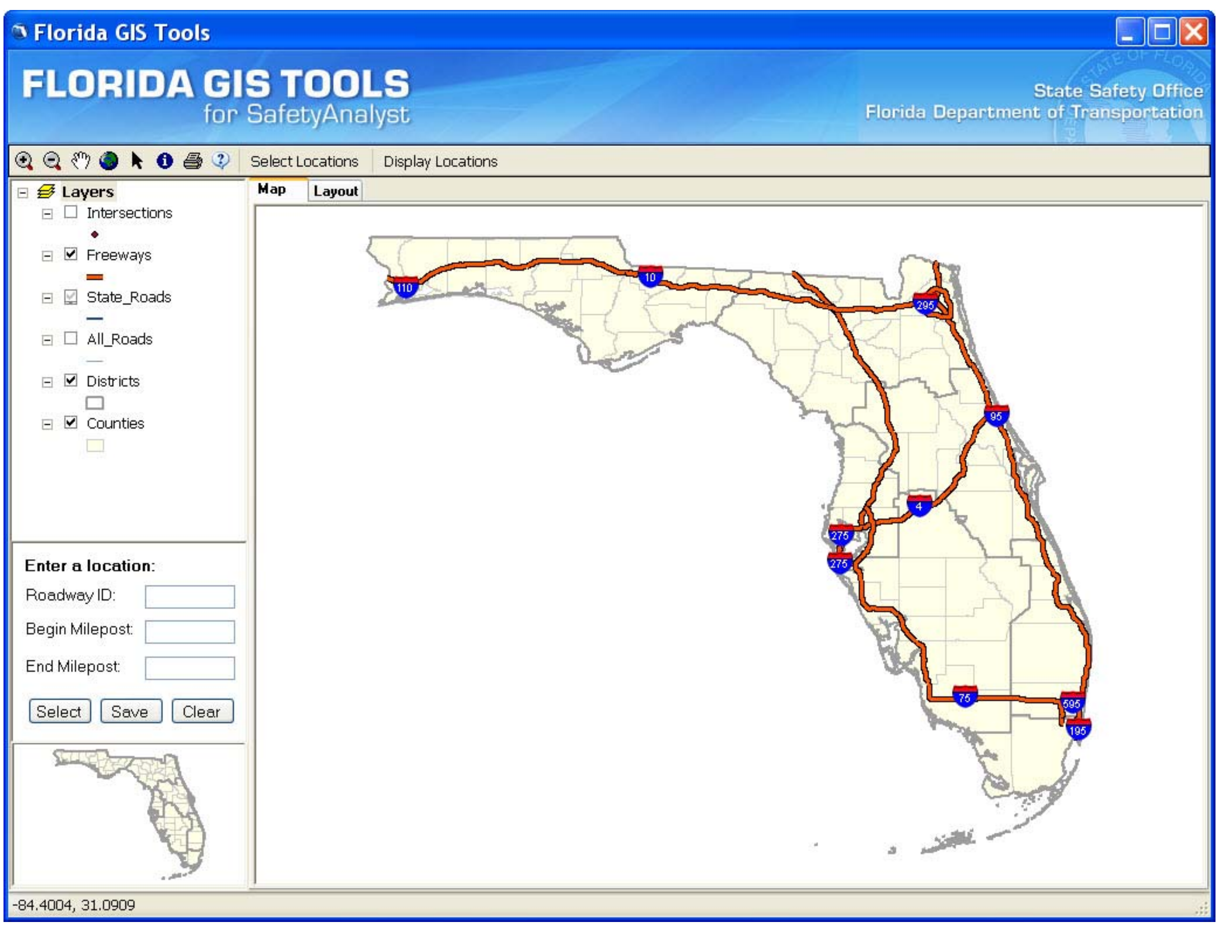


Figure 7-2: Main Screen of the GIS Interface (Gan et al., 2012)

To serve the two main functions of spatial selection and display, the system includes four major GIS tools (Ma et al., 2013):

- a basic GIS toolbox that is used to zoom in, zoom out, pan, and identify the geographic feature or place;
- a selection tool that assists users in selecting roadway locations by routes, counties, or districts to reduce the input data set to be analyzed in *SafetyAnalyst*;
- a display tool that displays specific roadway locations with potential for safety improvement and labels the major attributes of the *SafetyAnalyst* output; and
- a Google Map tool that overlays the selected roadway location on Google Map.

7.1.3 Process to Identify High Crash Locations

The *SafetyAnalyst* application process to identify high crash locations includes the following steps:

1. Input the data into the Data Converter, and run the Converter. The Converter generates *SafetyAnalyst* import files for the roadway inventory and their associated crash and traffic data.
2. In the *SafetyAnalyst* Administration Tool, update the Florida-specific site subtypes and the Florida-specific SPFs.
3. Import the output files from the Converter into the Data Management Tool in *SafetyAnalyst.*
4. Run the import, post-process, and calibration steps in the *SafetyAnalyst* Data Management Tool.
5. Open the *SafetyAnalyst* Analytical Tool, and create the list of sites to be analyzed. For example, users can create a site list including freeways, signalized intersections, etc.
6. Run the Network Screening Module in the Analytical Tool.
7. Export the output from the Network Screening Module into .csv format.
8. Use the GIS Tool to generate high potential crash location maps.

As shown in Figure 7-1, the entire application process is divided into two areas: the Administrator Area and the Analyst Area. Within this application process, the administrator is responsible for Steps 1-4, i.e., generating the *SafetyAnalyst* import files using the Converter, updating the SPFs and the site subtypes in the *SafetyAnalyst* Administration Tool, and running the *SafetyAnalyst* Data Management Tool. Once the data are successfully imported, post processed, and calibrated in the Data Management Tool, the analyst can run the Network Screening Module in the Analytical Tool, export the output (i.e., the list of locations with greater potential for safety improvement (PSI)), and use the GIS Tool to generate high potential crash location maps.

7.2 SPFs

This section focuses on the SPFs used in *SafetyAnalyst*. The functional form of the SPFs is first discussed, followed by Florida-specific SPFs.

7.2.1 Functional Form

The SPF functional form for roadway segments and ramps is as follows (Harwood et al., 2010):

$$N_{predicted} = e^{a} \times AADT^{b} \tag{7-1}$$

For fitting the NB regression models, Equation 7-1 is rewritten as:

$$N_{predicted} = \exp(a + b \times \ln(AADT)) \tag{7-2}$$

where,

$N_{predicted}$ = predicted crash frequency per mile per year,
$AADT$ = average annual daily traffic volume (vehicles per day), and
a, b = regression coefficients.

The SPF functional form for intersections is as follows (Harwood et al., 2010):

$$N_{predicted} = e^{a} \times AADT_{major}^{b} \times AADT_{minor}^{c} \tag{7-3}$$

For fitting the NB regression models, Equation 7-3 is rewritten as:

$$N_{predicted} = exp(a + b \times ln(AADT_{major}) + c \times ln(AADT_{min\,or})) \tag{7-4}$$

where,

$N_{predicted}$ = predicted crash frequency per intersection per year,
$AADT_{major}$ = average annual daily traffic volume on the major-road approaches,
$AADT_{minor}$ = average annual daily traffic volume on the minor-road approaches, and
a, b, c = regression coefficients that are estimated from the available data.

The overdispersion parameter (k), which indicates the statistical reliability of the SPF, accounts for dispersion in the data. The closer the k is to zero, the more statistically reliable the SPF is. To assess NB regression performance, the goodness-of-fit statistic, Freeman-Tukey R^2 coefficient (R^2_{FT}) is used.

Calibration of the default SPFs was performed by multiplying the default SPFs by a "calibration factor", C, which is calculated using the following equation (Harwood et al., 2010):

$$C = \frac{\sum_{All\ sites} observed\ crashes}{\sum_{All\ sites} predicted\ crashes} \tag{7-5}$$

As shown in Equation 7-5, calibration factor is calculated as the ratio of the total observed crashes to total predicted crashes obtained from the default national SPFs. Note that "*All sites*" in the equation refers to the reference sites within a specific category. The calibration factor is not needed if a local jurisdiction chooses to develop its own SPFs as the local safety trends are inherently addressed in the coefficients per Equations 7-2 and 7-4.

7.2.2 Florida-specific SPFs for Signalized Intersections, Segments, and Ramps

Tables 7-1 through 7-4 give the Florida-specific SPFs for signalized intersections, limited access facilities, non-limited access facilities, and ramps, respectively (Gan et al., 2012). The tables provide Florida-specific SPFs for both total crashes and fatal and injury (F+I) crashes. Note that the authors categorized ramps into Florida-specific site subtypes, which are different from the default site subtypes used in *SafetyAnalyst*.

Table 7-1: Florida-specific SPFs for Signalized Intersections (Gan et al., 2012)

Category	Severity	Coefficient						k	R^2_{FT}
		a		b		c			
		Estimate	p-value	Estimate	p-value	Estimate	p-value		
Rural Three-leg	Total	-8.972	<0.0001	0.728	<0.0001	0.386	0.0064	0.529	0.317
	F+I	-9.081	<0.0001	0.689	<0.0001	0.361	0.0111	0.446	0.345
Rural Four-leg	Total	-7.404	<0.0001	0.490	0.0011	0.512	0.0001	0.434	0.604
	F+I	-6.936	<0.0001	0.361	0.0247	0.514	0.0004	0.441	0.581
Urban Three-leg	Total	-9.134	<0.0001	0.664	<0.0001	0.471	<0.0001	0.430	0.400
	F+I	-8.690	<0.0001	0.624	<0.0001	0.378	<0.0001	0.364	0.341
Urban Four-leg	Total	-8.765	<0.0001	0.759	<0.0001	0.369	<0.0001	0.458	0.451
	F+I	-8.549	<0.0001	0.666	<0.0001	0.358	<0.0001	0.372	0.438

Table 7-2: Florida-specific SPFs for Freeways (Gan et al., 2012)

Category	Severity	Coefficient				k	R^2_{FT}
		a		b			
		Estimate	p-value	Estimate	p-value		
Rural Freeways with 4 Lanes							
Basic Freeway Segments	Total	-11.429	<0.0001	1.254	<0.0001	0.317	0.213
	F+I	-11.080	<0.0001	1.165	<0.0001	0.306	0.144
Segments within Interchange Influence Area	Total	-10.003	<0.0001	1.139	<0.0001	0.346	0.242
	F+I	-9.460	<0.0001	1.031	<0.0001	0.326	0.176
Rural Freeways with 6+ Lanes							
Basic Freeway Segments	Total	-10.910	<0.0001	1.182	<0.0001	0.218	0.252
	F+I	-13.283	<0.0001	1.319	<0.0001	0.150	0.207
Segments within Interchange Influence Area	Total	-10.693	0.0002	1.193	<0.0001	0.346	0.102
	F+I	-11.886	<0.0001	1.233	<0.0001	0.261	0.175
Urban Freeways with 4 Lanes							
Basic Freeway Segments	Total	-9.000	<0.0001	1.052	<0.0001	0.688	0.245
	F+I	-10.260	<0.0001	1.102	<0.0001	0.631	0.220
Segments within Interchange Influence Area	Total	-12.403	<0.0001	1.376	<0.0001	0.363	0.455
	F+I	-12.799	<0.0001	1.345	<0.0001	0.301	0.439
Urban Freeways with 6 Lanes							
Basic Freeway Segments	Total	-15.422	<0.0001	1.630	<0.0001	0.650	0.366
	F+I	-16.657	<0.0001	1.667	<0.0001	0.562	0.374
Segments within Interchange Influence Area	Total	-13.191	<0.0001	1.440	<0.0001	0.418	0.449
	F+I	-13.914	<0.0001	1.434	<0.0001	0.347	0.452
Urban Freeways with 8+ Lanes							
Basic Freeway Segments	Total	-8.355	0.0018	1.009	<0.0001	0.822	0.052
	F+I	-8.310	0.0019	0.941	<0.0001	0.705	0.060
Segments within Interchange Influence Area	Total	-8.434	<0.0001	1.041	<0.0001	0.487	0.247
	F+I	-10.576	<0.0001	1.156	<0.0001	0.427	0.295

Table 7-3: Florida-specific SPFs for Rural Roads and Urban Arterials (Gan et al., 2012)

Category	Severity	Coefficient				k	R^2_{FT}
		a		b			
		Estimate	p-value	Estimate	p-value		
Rural Two-lane Roads	Total	-6.923	<0.0001	0.874	<0.0001	0.464	0.166
	F+I	-7.660	<0.0001	0.894	<0.0001	0.444	0.118
Rural Multilane Divided Roads	Total	-5.356	<0.0001	0.689	<0.0001	0.446	0.153
	F+I	-6.016	<0.0001	0.694	<0.0001	0.413	0.118
Urban Two-lane Arterials	Total	-5.877	<0.0001	0.833	<0.0001	0.748	0.094
	F+I	-6.264	<0.0001	0.805	<0.0001	0.678	0.087
Urban Multilane Undivided Arterials	Total	-5.440	<0.0001	0.853	<0.0001	0.694	0.047
	F+I	-4.261	0.0003	0.655	<0.0001	0.571	0.052
Urban Multilane Divided Arterials	Total	-7.545	<0.0001	0.988	<0.0001	0.652	0.174
	F+I	-8.134	<0.0001	0.976	<0.0001	0.545	0.179
Urban One-way Arterials	Total	-3.144	0.0001	0.600	<0.0001	0.929	0.016
	F+I	-2.810	0.0006	0.465	<0.0001	0.819	0.021

Table 7-4: Florida-specific SPFs for Ramps (Gan et al., 2012)

Category	Severity	Coefficient				k	R^2_{FT}
		a		b			
		Estimate	p-value	Estimate	p-value		
Rural Diamond Off-ramp	Total	-4.844	<0.0001	0.776	<0.0001	0.615	0.277
	F+I	-5.317	0.0001	0.726	<0.0001	0.627	0.185
Rural Diamond On-ramp	Total	-7.783	<0.0001	1.038	<0.0001	0.522	0.369
	F+I	-9.844	<0.0001	1.214	<0.0001	0.433	0.342
Urban Diamond Off-ramp	Total	-3.335	<0.0001	0.638	<0.0001	0.883	0.061
	F+I	-3.763	<0.0001	0.598	<0.0001	0.802	0.075
Urban Diamond On-ramp	Total	-3.399	<0.0001	0.564	<0.0001	1.050	0.094
	F+I	-4.845	<0.0001	0.635	<0.0001	0.990	0.062
Urban Partial Diamond Off-ramp	Total	-3.789	0.0003	0.640	<0.0001	1.100	0.149
	F+I	-4.696	<0.0001	0.662	<0.0001	0.993	0.129
Urban Partial Diamond On-ramp	Total	-8.024	<0.0001	1.095	<0.0001	1.644	0.168
	F+I	-8.144	<0.0001	1.013	<0.0001	2.253	0.120
Urban Trumpet Off-ramp	Total	-6.428	0.0214	0.862	0.0058	1.108	0.084
	F+I	-6.918	0.0233	0.838	0.0131	0.976	0.100
Urban Trumpet On-ramp	Total	-10.228	0.0003	1.282	<0.0001	1.444	0.390
	F+I	-10.795	0.0016	1.225	0.0009	1.076	0.380
Urban Partial Cloverleaf Off-ramp	Total	-3.202	<0.0001	0.597	<0.0001	0.956	0.176
	F+I	-3.952	<0.0001	0.580	<0.0001	0.906	0.162
Urban Partial Cloverleaf On-ramp	Total	-5.722	<0.0001	0.822	<0.0001	0.660	0.225
	F+I	-6.872	<0.0001	0.869	<0.0001	0.754	0.169

7.2.3 Florida-specific SPFs for Unsignalized Intersections

Florida-specific SPFs were developed for unsignalized intersections to be used within *SafetyAnalyst*. Four-year crash data from 2008-2011 and 2011 RCI data were used to develop the SPFs. The major effort in developing the SPFs was to identify unsignalized intersections. Section 6.4.1 of Chapter 6 discusses the steps undertaken to identify unsignalized intersections. Table 7-5 provides the summary statistics of unsignalized intersections used to develop SPFs. A simple NB regression model was used to develop Florida-specific SPFs. The model form similar to the one used to generate *SafetyAnalyst* default SPFs was used (as shown in Equations 7-3 and 7-4). Table 7-6 provides the Florida-specific SPFs for unsignalized intersections. For each type of unsignalized intersection category, Florida-specific SPFs are plotted for both total and F+I crashes. Figures 7-3 through 7-6 display the plots of the predicted annual crash frequency against AADT for major road approaches. For rural unsignalized intersections, the SPFs are plotted assuming AADT for minor road approaches to be 160, 370, and 750 veh/day. Similarly, for urban unsignalized intersections, the SPFs are plotted assuming AADT for minor road approaches to be 500, 1,500, and 3,500 veh/day.

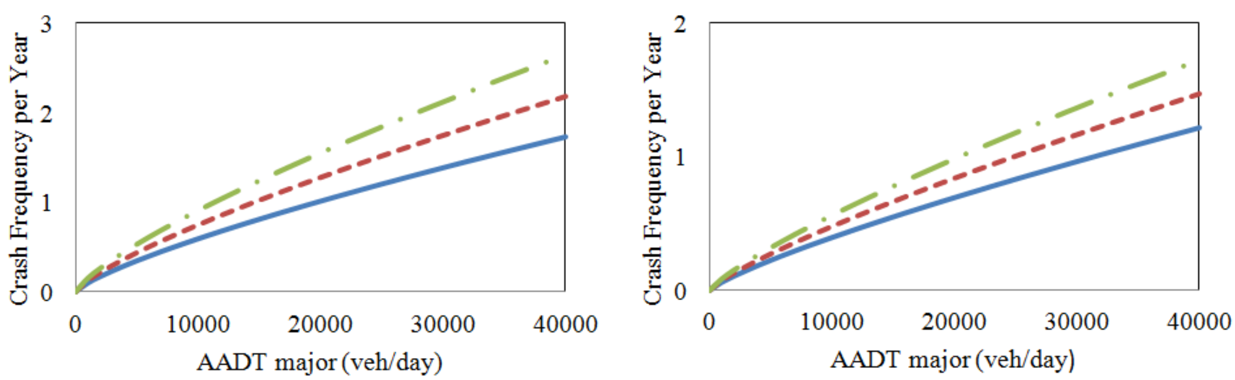
Table 7-5: Summary Statistics of Unsignalized Intersections

Area Type	Category	Number of Unsignalized Intersections	Total Number of Crashes (2008-2011)	
			Total Crashes	F+I Crashes
Rural	Three-leg	329	833	491
	Four-leg	43	143	93
Urban	Three-leg	321	2,183	1,140
	Four-leg	34	317	178

Table 7-6: Florida-Specific SPFs for Unsignalized Intersections

Category	Severity	Regression Coefficients						k	R^2_{FT}
		a		b		c			
		Estimate	p-value	Estimate	p-value	Estimate	p-value		
Rural Three-leg	Total	-9.0364	<0.0001	0.7712	<0.0001	0.2779	<0.0001	0.6379	0.298
	F+I	-9.4949	<0.0001	0.8078	<0.0001	0.2238	0.0005	0.6372	0.262
Rural Four-leg	Total	-5.8587	0.0138	0.3659	0.1438	0.4317	0.0235	0.6569	0.142
	F+I	-5.9726	0.0254	0.2875	0.3086	0.4810	0.0298	0.7813	0.120
Urban Three-leg	Total	-10.5606	<0.0001	0.9415	<0.0001	0.2176	0.0004	0.6330	0.338
	F+I	-9.7254	<0.0001	0.8674	<0.0001	0.1277	0.0524	0.5640	0.273
Urban Four-leg	Total	-7.3530	0.0409	0.7187	0.0097	0.1620	0.4019	0.5606	0.164
	F+I	-11.2344	0.0029	1.0282	0.0006	0.1986	0.2865	0.3995	0.267

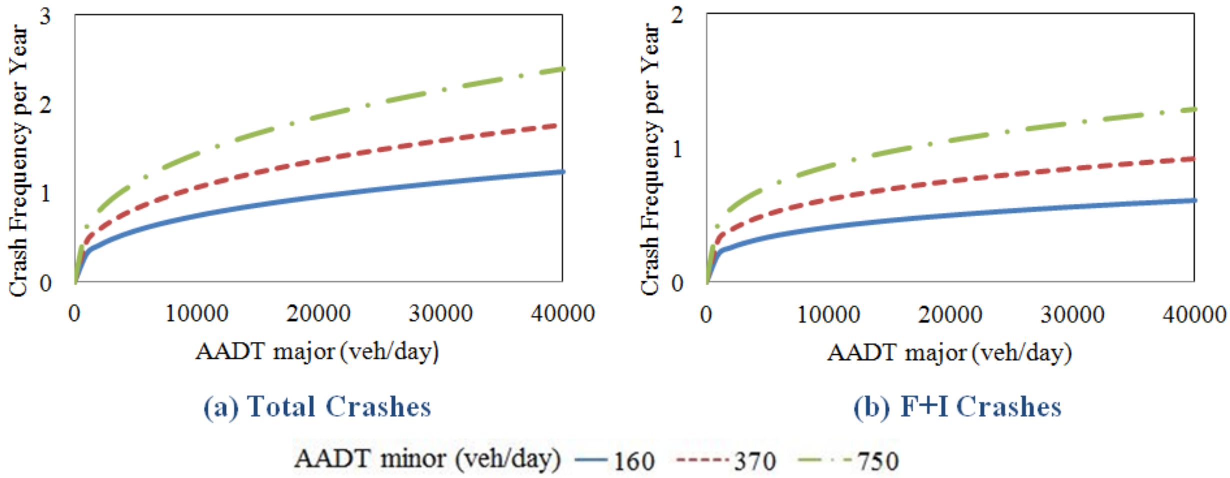
**All intersections are minor road stop control.*



(a) Total Crashes (b) F+I Crashes

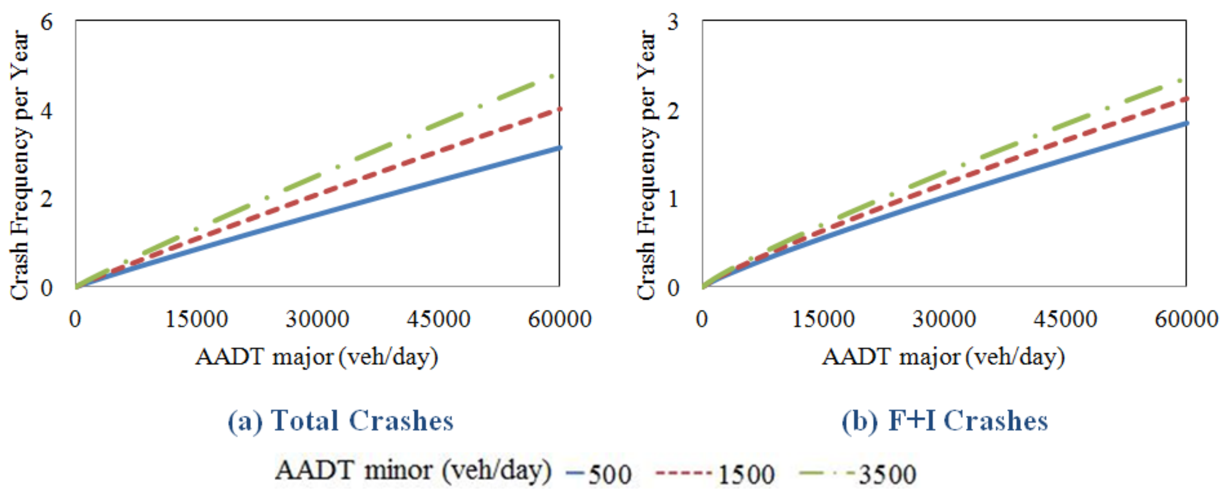
AADT minor (veh/day) — 160 ----- 370 —·- 750

Figure 7-3: SPFs for Rural Three-leg Unsignalized Intersections



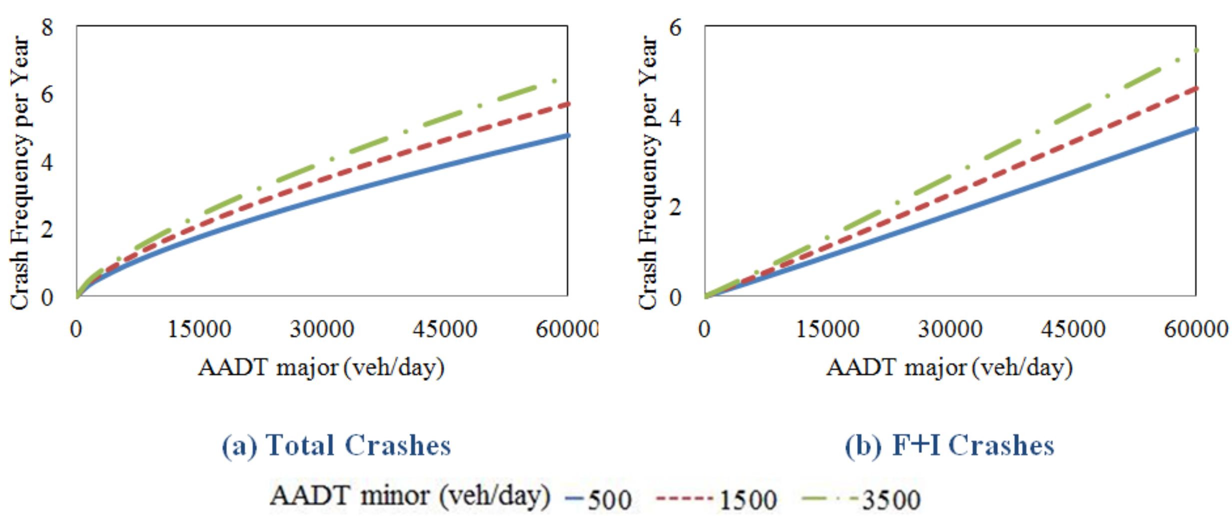
(a) Total Crashes (b) F+I Crashes

Figure 7-4: SPFs for Rural Four-leg Unsignalized Intersections



(a) Total Crashes (b) F+I Crashes

Figure 7-5: SPFs for Urban Three-leg Unsignalized Intersections



(a) Total Crashes (b) F+I Crashes

Figure 7-6: SPFs for Urban Four-leg Unsignalized Intersections

7.3 High Crash Locations

The entire process discussed in Section 7.1 was applied to identify high crash locations. As a first step, the Data Converter was used to generate *SafetyAnalyst* import files. The following input files were used:

- 2011 RDWTBL25 (in .csv format)
- 2011 RDWTBL31 (in .csv format)
- 2011 RCI database (in .accdb format)
- RDWTBL50 and RDWTBL51 for the years 2007-2010 (in .txt format)

A total of 10 files, as identified in Table 6-2, were generated by the Converter. The following are the descriptive statistics of the *SafetyAnalyst* import files.

- A total of 609,288 crashes that occurred between 2007 and 2010 were included in the *AltAccident* file. Of these, 470,482 occurred on segments, 117,951 occurred on intersections, and 20,855 occurred on ramps.
- *AltSegment* file included a total of 133,724 segments, constituting 13,130.74 miles of the state road network.
- *AltIntersection* file included a total of 21,194 intersections.
- *AltRamp* file included a total of 1,976 ramp segments, and these constitute 899.76 miles.

Florida-specific site subtypes were added in the Administration Tool. Table 7-7 gives the list of the site subtypes that were added in the Administration Tool. The national default SPFs were then replaced with Florida-specific SPFs provided in Tables 7-1 through 7-4 and Table 7-6.

The import files were inputted into the Data Management Tool in *SafetyAnalyst.* The data files were imported, post processed, and calibrated. During these steps, the software gave several warnings and errors. Some of the segments and ramps were excluded. None of the intersections were excluded because the intersection data were collected and preprocessed per the *SafetyAnalyst* requirements. Out of 1,976 ramps, a total of 75 ramps were excluded because of invalid site subtype; there were no SPFs within *SafetyAnalyst* that correspond to these ramp subtypes. Table 7-8 provides details on the reasons for exclusion of segments. As can be observed from the table, over 17,000 segments constituting over 4,400 miles were excluded because of missing or invalid site subtype. It means that *SafetyAnalyst* could not assign these segments to any one of the predefined segment subtypes, and therefore, were not included in the analysis.

Table 7-7: User-defined Florida-specific Subtypes

Facility Type	Site Subtype Code	Site Subtype Description
Intersections	261	Rural Three-leg Unsignalized Intersection
	262	Rural Four-leg Unsignalized Intersection
	263	Urban Three-leg Unsignalized Intersection
	264	Urban Four-leg Unsignalized Intersection
Ramps	411	Rural Diamond Off Ramp
	412	Rural Diamond On Ramp
	413	Rural Partial Diamond Off Ramp
	414	Rural Partial Diamond On Ramp
	415	Rural Trumpet Off Ramp
	416	Rural Trumpet On Ramp
	417	Rural Parclo Loop Off Ramp
	418	Rural Parclo Loop On Ramp
	419	Rural Direct Connection
	420	Urban Diamond Off Ramp
	421	Urban Diamond On Ramp
	422	Urban Partial Diamond Off Ramp
	423	Urban Partial Diamond On Ramp
	424	Urban Trumpet Off Ramp
	425	Urban Trumpet On Ramp
	426	Urban Parclo Loop Off Ramp
	427	Urban Parclo Loop On Ramp
	428	Urban Direct Connection
	429	Other Ramp Type

Table 7-8: Reasons for Exclusion of Segments in *SafetyAnalyst*

Reason for Exclusion	Number of Segments Excluded	Miles of Segments Excluded
Missing or invalid site subtype	17,134	4,460.63
No lane information	87	14.91
No traffic data	5,457	1,281.86
Total	22,678	5,757.40

Once the data were successfully calibrated in the Data Management Tool, users can run the analysis in the Analytical Tool. Within this Tool, users first have to create the list of sites to be analyzed. Users can use a query function to create a site list based on site subtype, geographic location such as county, roadway ID, etc. Network screening could then be performed to rank high crash locations (HCLs). Table 7-9 gives the descriptions of the various columns in the output from the Network Screening Module of *SafetyAnalyst*.

Table 7-9: Columns in the Output from the *SafetyAnalyst*'s Network Screening Module

Column in *SafetyAnalyst* Output		Description
ID		Roadway Segment/Intersection/Ramp ID
Site Type		Whether Segment/Intersection/Ramp
Site Subtype		Sub-categories in the site type
County		County where the roadway segment is located
Route		Route number of the roadway segment
Site Start Location		Start location of the roadway segment
Site End Location		End location of the roadway segment
Average Observed Accidents for Entire Site[1]		Observed crashes for the entire site
Location with Highest Potential for Safety Improvement (PSI)	Average Observed Accidents[1]	Observed crashes for the roadway sub segment
	Predicted Accident Frequency[1]	Predicted crash frequency
	Expected Accident Frequency[1]	Expected crash frequency
	Variance[2]	Variance
	Start Location	Start location of the search window
	End Location	End location of the search window
	No. of Expected Fatalities	Total number of expected fatalities at the location
	No. of Expected Injuries	Total number of expected injuries at the location
Rank		Overall Rank based on PSI
Additional Windows of Interest		Additional windows whose PSI exceeded the threshold limits, but the expected crash frequencies are between the limiting crash threshold and the highest calculated PSI for the site

[1] expressed as crashes/mile/year for segments and ramps, and as crashes/year for intersections.
[2] expressed as crashes/mile^2/year for segments and ramps, and as crashes/year for intersections.

Table 7-10 gives the statewide list of the top 25 HCLs for urban four-leg signalized intersections. Note that the following table displays only the relevant columns from the output. The following are the additional parameters considered in the analysis:

- Type of Analysis: Basic Network Screening
- Accident Severity Level: Total accidents
- Site Types: Intersections
- Screening Attribute: Accident Type and Manner of Collision
- Potential for Safety Improvement Using: Expected accident frequency
- Analysis Period: From 2007 To 2010
- Major Reconstruction: No major reconstruction occurred at any sites during the analysis period
- CV limit (intersections): 0.5
- Area Weights (Rural): 1.0
- Area Weights (Urban): 1.0
- Limiting Value (Intersections): 15.0 crashes/yr

- Number of sites in the site list: 3,269
- Number of sites evaluated: 3,269
- Number of segments evaluated: 0
- Total length of segments evaluated: 0.000
- Number of intersections evaluated: 3,269
- Number of ramps evaluated: 0
- Number of sites flagged: 221

7.4 Summary

This chapter discussed the data flow structure between the Data Converter, *SafetyAnalyst* components, and the GIS Tool. The Converter automatically generates *SafetyAnalyst* import files, and the GIS Tool spatially displays high crash locations identified by *SafetyAnalyst*. The entire *SafetyAnalyst* application process requires users to follow the steps below to identify high crash locations:

1. Input the data into the Data Converter, and run the Converter. The Converter generates *SafetyAnalyst* import files for the roadway inventory and their associated crash and traffic data.
2. In the *SafetyAnalyst* Administration Tool, update the Florida-specific site subtypes and the Florida-specific SPFs.
3. Import the output files from the Converter into the Data Management Tool in *SafetyAnalyst.*
4. Run the import, post-process, and calibration steps in the *SafetyAnalyst* Data Management Tool.
5. Open the *SafetyAnalyst* Analytical Tool, and create the list of sites to be analyzed. For example, users can create a site list including freeways, signalized intersections, etc.
6. Run the Network Screening Module in the Analytical Tool.
7. Export the output from the Network Screening Module into .csv format.
8. Use the GIS Tool to generate high potential crash location maps.

Table 7-10: Top 25 High Crash Locations - Urban Four-leg Signalized Intersections

Site Subtype	County	Route	Site Start Location	Site End Location	Average Observed Accidents for Entire Site	Location with Highest PSI						Rank
						Average Observed Accidents	Predicted Accident Frequency	Expected Accident Frequency	Variance	Start Loc.	End Loc.	
Int/Urb; 4-leg signalized	87	BR87030000	7.62	-	57.99	57.99	18.22	56.81	14.02	-	-	1
Int/Urb; 4-leg signalized	15	US15150000	25.839	-	58.23	58.23	13.02	56.42	13.42	-	-	2
Int/Urb; 4-leg signalized	87	US87120000	7.045	-	45.99	45.99	12.3	44.38	11.92	-	-	3
Int/Urb; 4-leg signalized	87	BR87030000	23.53	-	39.49	39.49	13.42	38.42	9.64	-	-	4
Int/Urb; 4-leg signalized	87	L87053000	6.089	-	41.49	41.49	5.68	38.22	9.06	-	-	5
Int/Urb; 4-leg signalized	87	L87053000	4.551	-	40.82	40.82	6.52	38.05	9.15	-	-	6
Int/Urb; 4-leg signalized	87	L87019000	2.74	-	38.37	38.37	9.41	36.72	9.04	-	-	7
Int/Urb; 4-leg signalized	87	L87066000	1.52	-	39.28	39.28	5.49	36.35	7.91	-	-	8
Int/Urb; 4-leg signalized	87	BR87030000	5.395	-	36.92	36.92	14.35	36.07	8.9	-	-	9
Int/Urb; 4-leg signalized	12	L12005000	5.283	-	36.04	36.04	12.91	35.09	8.54	-	-	10
Int/Urb; 4-leg signalized	87	BR87030000	4.744	-	34.45	34.45	15.39	33.78	8.38	-	-	11
Int/Urb; 4-leg signalized	86	L86014000	3.22	-	34.78	34.78	11.8	33.75	8.18	-	-	12
Int/Urb; 4-leg signalized	87	BR87030000	13.512	-	35.48	35.48	5.77	32.66	8.19	-	-	13
Int/Urb; 4-leg signalized	10	US10010000	15.694	-	36.82	36.82	3.42	32.32	6.82	-	-	14
Int/Urb; 4-leg signalized	87	L87066000	2.961	-	34.78	34.78	4.94	32.11	6.52	-	-	15
Int/Urb; 4-leg signalized	75	L75270000	3.063	-	32.69	32.69	9.7	31.46	7.49	-	-	16
Int/Urb; 4-leg signalized	72	L72220000	7.625	-	32.72	32.72	8.92	31.2	8.1	-	-	17
Int/Urb; 4-leg signalized	15	L15230000	3.133	-	32.95	32.95	6.9	31.04	7.18	-	-	18
Int/Urb; 4-leg signalized	86	US86100000	5.09	-	31.83	31.83	9.41	30.52	7.69	-	-	19
Int/Urb; 4-leg signalized	10	US10030000	3.522	-	32.00	32.00	7.66	30.47	6.72	-	-	20
Int/Urb; 4-leg signalized	15	US15150000	24.588	-	30.91	30.91	13.02	30.2	7.17	-	-	21
Int/Urb; 4-leg signalized	86	L86090000	6.858	-	31.25	31.25	10.6	30.13	7.88	-	-	22
Int/Urb; 4-leg signalized	87	L87001000	5.374	-	30.62	30.62	13.73	29.98	7.18	-	-	23
Int/Urb; 4-leg signalized	87	BR87030000	3.223	-	30.33	30.33	15.39	29.81	7.39	-	-	24
Int/Urb; 4-leg signalized	87	US87120000	9.056	-	30.91	30.91	9.51	29.76	6.96	-	-	25

CHAPTER 8
SUMMARY AND CONCLUSIONS

Recent research in highway safety has focused on the more advanced and statistically proven techniques of highway safety improvement. The HSM, *SafetyAnalyst*, and the Interactive Highway Safety Design Model (IHSDM) are the three major safety analysis tools that have the potential to define a new era in highway safety. This research project focuses on the two most recent tools, the HSM and *SafetyAnalyst*.

Meeting the data requirements is considered the most challenging task in implementing these tools. In the case of HSM, many of the data variables needed for deriving the calibration factors are currently unavailable in Florida's RCI database. This project identified and prioritized influential calibration variables for data collection and determined the minimum sample sizes to estimate reliable calibration factors.

Compared to HSM, *SafetyAnalyst* has lesser but different data requirements. Two major efforts in applying *SafetyAnalyst* involve conversion of local data into the strict data format required by *SafetyAnalyst* and development of agency-specific SPFs. This project aimed to develop a new conversion program to automatically generate *SafetyAnalyst* import files. It also developed SPFs for unsignalized intersections to supplement those of other facilities which were developed under a separate project. Finally, it outlined the *SafetyAnalyst* application process to be followed to identify high crash locations.

8.1 Highway Safety Manual

8.1.1 Prioritization of Data Variables

Random forest technique was used to rank the variables based on their importance. Data variables were ranked for the following segment and intersection subtypes:

- Rural two-way two-lane undivided roadway segments
- Rural multilane four-lane divided segments
- Urban and suburban two-lane undivided segments
- Urban and suburban three-lane sections with center TWLTLs
- Urban and suburban four-lane undivided segments
- Urban and suburban four-lane divided segments
- Urban and suburban five-lane sections with center TWLTLs
- Rural two-lane three-leg stop-controlled intersections
- Urban and suburban three-leg stop-controlled intersections
- Urban and suburban four-leg signalized intersections

The variables were ranked based on the increase in node purity (*IncNodePurity)* values, which represent the averaged total decrease in node impurity. Based on these values, the variables were categorized into three groups: variables of primary importance; variables of secondary importance; and variables of lesser importance. AADT and segment length for segments and major and minor road AADTs for intersections were the most important variables as they significantly influence crash predictions, and were identified as the variables of primary importance. Based on the *IncNodePurity* values, the remaining variables were either identified as variables of secondary importance or variables of lesser importance. Furthermore, the variables were ranked based on their corresponding *IncNodePurity* values. Tables 8-1 and 8-2 give the summary of the ranking of the variables for segments and intersections, respectively.

Table 8-1: Summary of the Ranking of Variables for Segments

Data Variable	Site Subtype[1,2]						
	R2U	R4D	SU2U	SU3T	SU4U	SU4D	SU5T
Segment length	1	1	1	1	1	1	1
AADT	2	2	2	2	2	2	2
Lane width	6	7	--	--	--	--	--
Shoulder type	7	NR	--	--	--	--	--
Shoulder width	4	3	--	--	--	--	--
Presence of TWLTL	10	--	--	--	--	--	--
Median width	--	4	--	--	--	6	--
Presence of lighting	8	6	6	6	6	7	6
Roadside fixed object density	--	--	4	4	4	4	5
Speed limit	--	--	7	8	7	9	7
Presence of on-street parking	--	--	8	7	8	10	NR
Presence of automated speed enforcement	13	5	9	NR	NR	8	8
Presence of passing lane	9	--	--	--	--	--	--
Presence of short four-lane section	11	--	--	--	--	--	--
Presence of centerline rumble strip	12	--	--	--	--	--	--
Roadside hazard rating	5	--	--	--	--	--	--
Driveway density	3	--	--	--	--	--	--
Number of major driveways	--	--	5	5	5	5	3
Number of minor driveways	--	--	3	3	3	3	4
Horizontal curve	NR	--	--	--	--	--	--
Vertical grade	NR	--	--	--	--	--	--

[1] R2U: rural two-lane undivided; R4D: rural four-lane divided; SU2U: urban and suburban two-lane undivided; SU3T: urban and suburban three-lane with TWLTL; SU4U: urban and suburban four-lane undivided; SU4D: urban and suburban four-lane divided; SU5T: urban and suburban five-lane with TWLTL.

[2] NR indicates that the variable is not ranked; -- indicates that the variable is not used for that specific site subtype.

8.1.2 Determination of Minimum Sample Size

For the different facility types discussed in the HSM, the minimum sample sizes to estimate reliable calibration factors were determined. The calibration factor estimated from a sample size was considered to be reliable if there is at least a 90% probability that the estimated calibration factor would lie within 10% of the actual calibration factor (which was calculated from the entire data set). Tables 8-3 and 8-4 give the recommended minimum sample sizes required to estimate reliable calibration factors for segment and intersection subtypes, respectively. As can be observed from the tables, for all the site subtypes, the HSM-recommended sample size of 30-50 sites with at least 100 crashes per year was found to be much lower than the minimum sample sizes recommended in this study.

Table 8-2: Summary of the Ranking of Variables for Intersections

Data Variable	Site Subtype[1,2]		
	R3ST	SU3ST	SU4SG
Major road AADT	1	1	1
Minor road AADT	2	2	2
Intersection skew angle	5	--	--
Number of approaches with left-turn lanes	4	3	6
Number of approaches with right-turn lanes	3	4	3
Presence of lighting	6	5	NR
Presence and type of left-turn signal phasing	--	--	5
Use of RTOR signal operation	--	--	10
Use of red-light cameras	--	--	7
Number of bus stops within 1,000 ft	--	--	4
Presence of schools within 1,000 ft	--	--	8
Number of alcohol sales establishments within 1,000 ft	--	--	9

[1] R3ST: rural two-lane three-leg stop-controlled intersections; SU3ST: urban and suburban three-leg stop-controlled intersections; SU4SG: urban and suburban four-leg signalized intersections.
[2] NR indicates that the variable is not ranked; -- indicates that the variable is not used for that specific site subtype.

Table 8-3: Recommended Minimum Sample Sizes for Segments

Segment Site Subtype	Total Sample Length (in miles)	Sample Size	Percent of Total Length	Number of Crashes per Year	P (Estimated CF is within 10% of Actual CF)[1]
Rural Two-way Two-lane Roadway Segments	250	250	7.1%	150	90%
Rural Four-lane Divided Arterials	175	250	14.2%	270	85%
Urban and Suburban Two-lane Undivided Arterials	102	300	16.8%	180	80%
Urban and Suburban Three-lane Arterials with TWLTL	38	200	55.5%	120	94%
Urban and Suburban Four-lane Undivided Arterials	30	150	55.9%	130	90%
Urban and Suburban Four-lane Divided Arterials	140	500	10.1%	550	93%
Urban and Suburban Five-lane Arterials with TWLTL	70	275	25%	400	88%

[1] CF refers to calibration factor.

Table 8-4: Recommended Minimum Sample Sizes for Intersections

Intersection Site Subtype	Sample Size	Percent of Total Intersections	Number of Crashes per Year	P (Estimated CF is within 10% of Actual CF)[1]
Rural Two-lane Three-leg Stop-controlled Intersections	150	50.3%	85	81%
Urban and Suburban Three-leg Stop-controlled Intersections	130	40.5%	200	90%
Urban and Suburban Three-leg Signalized Intersections	50	86.2%	350	87%
Urban and Suburban Four-leg Signalized Intersections	80	17.4%	1,300	87%

[1] CF refers to calibration factor.

In addition to determining the minimum sample sizes to yield reliable calibration factors, two types of sampling procedures were researched. The calibration factors estimated from simple random sampling and stratified sampling were compared using a two-sample t-test to determine whether or not the calibration factors estimated from a stratified sample were significantly different from the calibration factors estimated from a simple random sample of the entire state. Based on the results, it is considered to be acceptable to obtain the recommended sample size through a simple random sampling procedure from the entire state data.

8.2 *SafetyAnalyst*

8.2.1 SafetyAnalyst Data Converter

SafetyAnalyst was developed as a cooperative effort by FHWA and participating state and local agencies. The software provides a suite of analytical tools to identify and manage system-wide safety improvements by incorporating all the steps in the roadway safety management process.

One of the major hurdles in deploying *SafetyAnalyst* is its stringent data requirements. *SafetyAnalyst* requires a number of import files to be generated in line with the data requirements and format recommended by the software (Harwood et al., 2010). As such, *SafetyAnalyst* Data Converter was developed to automatically generate import files for *SafetyAnalyst*. The Converter uses the following input files to generate the required import files for segments, intersections, ramps, and their associated crash and traffic files:

- RDWTBL25 (in .csv format)
- RDWTBL31 (in .csv format)
- RCI database (in .accdb format)
- RDWTBL50 and RDWTBL51 for all the analysis years (in .txt format)

The Converter also has the capability to generate both statewide and districtwide import files. The output from the Converter can be directly inputted into the *SafetyAnalyst* Data Management Tool.

8.2.2 Florida-specific SPFs

To perform network screening, *SafetyAnalyst* implements the EB method, which requires the use of SPFs. *SafetyAnalyst* is equipped with a set of national default SPFs, and the software calibrates the default SPFs to represent the agency's safety performance. However, agencies are recommended to develop agency-specific SPFs whenever possible. It is believed that the agency-specific SPFs represent the agency data better than the national default SPFs calibrated to agency data. Gan et al. (2012) developed Florida-specific SPFs for segments, signalized intersections, and ramps. In this project, Florida-specific SPFs for unsignalized intersections were developed using 2011 RCI data and crash and traffic data from 2007-2010 for both total and F+I crashes. To facilitate the inclusion of as many unsignalized intersections as possible, the following Florida-specific unsignalized intersection subtypes were created:

- rural three-leg unsignalized intersections,
- rural four-leg unsignalized intersections,
- urban three-leg unsignalized intersections, and
- urban four-leg unsignalized intersections.

8.2.3 Application of SafetyAnalyst

Florida's *SafetyAnalyst* application process includes the Data Converter, the components of *SafetyAnalyst*, and the GIS Tool. The Converter automatically generates *SafetyAnalyst* import files, and the GIS Tool spatially displays high crash locations identified by *SafetyAnalyst*. The entire *SafetyAnalyst* application process requires users to follow the steps below to identify high crash locations:

1. Input the data into the Data Converter, and run the Converter. The Converter generates *SafetyAnalyst* import files for the roadway inventory and their associated crash and traffic data.
2. In the *SafetyAnalyst* Administration Tool, update the Florida-specific site subtypes and the Florida-specific SPFs.
3. Import the output files from the Converter into the Data Management Tool in *SafetyAnalyst.*
4. Run the import, post-process, and calibration steps in the *SafetyAnalyst* Data Management Tool.
5. Open the *SafetyAnalyst* Analytical Tool, and create the list of sites to be analyzed. For example, users can create a site list including freeways, signalized intersections, etc.
6. Run the Network Screening Module in the Analytical Tool.
7. Export the output from the Network Screening Module into .csv format.
8. Use the GIS Tool to generate high potential crash location maps.

8.2.4 Deployment of SafetyAnalyst

With the development of Florida-specific SPFs and the completion of a software tool to convert Florida's state road data to the data format required by *SafetyAnalyst*, Florida is now ready to deploy *SafetyAnalyst*. For the first time, Florida can have a standard system to consistently conduct safety analysis across the state. The first step in deploying *SafetyAnalyst* is to make available the *SafetyAnalyst* data sets to the district officials and their consultants. These data sets can be distributed through download from the Florida Traffic Safety Portal.

While *SafetyAnalyst* has been designed to be user-friendly, it is a relatively complex system. New users of the system would clearly benefit from a technical workshop that includes hands-on training. Accordingly, it is recommended that FDOT allocates resources to develop and conduct such a workshop at the district offices. The workshops can be made available to district safety officials, their consultants, and officials from local agencies interested in the potential use of the system. It is also recommended that FDOT provides technical support to the user community and continues to update the *SafetyAnalyst* data sets as new data become available.

In the longer run and after *SafetyAnalyst* is successfully deployed at the districts, FDOT may consider expanding its deployment to local agencies. This could be accomplished by modifying the *SafetyAnalyst* data converter from one that is based on FDOT's RCI data to one that is based on FDOT's All-Roads map.

REFERENCES

Akgüngör, A. P., & Yıldız, O. (2007). Sensitivity Analysis of an Accident Prediction Model by the Fractional Factorial Method. *Accident Analysis and Prevention*, *39*(1), 63-68.

Alluri, P., & Ogle, J. (2012a). Effects of state-specific SPFs, AADT estimations, and overdispersion parameters on crash predictions using SafetyAnalyst. In *Proceedings of the 91st Annual Meeting of the Transportation Research Board, Washington, DC*.

Alluri, P., & Ogle, J. (2012b). Road Safety Analysis in the United States: States' Current Practices and Their Future Direction. *Transportation Research Record: Journal of the Transportation Research Board, 2318*(1), 7-15.

American Association of State Highways and Transportation Officials (AASHTO). (2010a). *Highway Safety Manual* (1st Edition). Transportation Research Board of the National Academies, Washington, D.C.

American Association of State Highways and Transportation Officials (AASHTO). (2010b). *SafetyAnalyst*. Retrieved from http://www.safetyanalyst.org.

Banihashemi, M. (2011). Highway Safety Manual, New Model Parameters vs. Calibration of Crash Prediction Models. In *Proceedings of the 90th Annual Meeting of the Transportation Research Board, Washington, DC.*

Banihashemi, M. (2012). Sensitivity Analysis of Data Set Sizes for Highway Safety Manual Calibration Factors. *Transportation Research Record: Journal of the Transportation Research Board*, *2279*(1), 75-81.

Brimley, B. K., Saito, M., & Schultz, G. G. (2012). Calibration of Highway Safety Manual Safety Performance Function. *Transportation Research Record: Journal of the Transportation Research Board*, *2279*(1), 82-89.

Brown, H., Sun. C., & Edara P. (2014). Nuts and Bolts of Statewide HSM Calibration. In *Proceedings of the 93rd Annual Meeting of the Transportation Research Board, Washington, DC.*

Dixon, K., Monsere, C., Xie, F., & Gladhill, K. (2012). *Calibrating the future highway safety manual predictive methods for Oregon state highways* (No. OTREC-RR-12-02).

Findley, D., Zegeer, C., Sundstrom, C., Hummer, J., & Rasdorf, W. (2012). Applying the Highway Safety Manual to Two-lane Road Curves. *Transportation Research Record: Journal of the Transportation Research Board, 51*(3), 25-38.

Florida Department of Transportation (FDOT). (2012). *Transportation Statistics RCI Field Handbook.* Highway Data Collection/Quality Control Section, Transportation Statistics Office, Florida Department of Transportation, Tallahassee, FL.

Gan, A., Haleem, K., Alluri, P., Lu, J., Wang, T., Ma, M., & Diaz, C. (2012). *Preparing Florida for Deployment of SafetyAnalyst for All Roads*. Final Research Report, Florida Department of Transportation, Tallahassee, FL.

Gan, A., Liu, K., & Jung, R. (2007). *Vehicle Occupancy Data Collection Methods (Phase II)* (No. FDOT Project 99700-3596-119). Final Research Report, Florida Department of Transportation, Tallahassee, FL.

Haleem, K., Abdel-Aty, M., & Santos, J. (2010). Multiple applications of multivariate adaptive regression splines technique to predict rear-end crashes at unsignalized intersections. *Transportation Research Record: Journal of the Transportation Research Board*, *2165*(1), 33-41.

Harwood, D. W., Torbic, D. J., Richard, K. R., & Meyer, M. M. (2010). "SafetyAnalyst: Software Tools for Safety Management of Specific Highway Sites (No. FHWA-HRT-10-063). Midwest Research Institute, Kansas City, MO.

ITT Corp. (2009). *SafetyAnalyst User's Manual*. (Publication DTFH61-03-C-00031). Federal Highway Administration, U.S. DOT, Washington, DC.

Jacobsson, T. (2012). Towards Personalized Treatment of Cancer. Retrieved January 15, 2014, from http://www2.math.su.se/matstat/reports/serieb/2012/rep3/report.pdf.

Lu, J. (2013). *Development of Safety Performance Functions for SafetyAnalyst Applications in Florida* (Doctoral Dissertation, Florida International University, Miami, FL). Retrieved from http://digitalcommons.fiu.edu/cgi/viewcontent.cgi?article=1996&context=etd.

Lu, L., Lu, J. J., Lin, P. S., Wang, Z. & Chen, H. (2009). *Development of an Interface between FDOT's Crash Analysis Reporting System and the SafetyAnalyst* (No. BDK84 977-01). Florida Department of Transportation, Tallahassee, FL.

Lubliner, H., & Schrock, S. D. (2012). Calibration of the Highway Safety Manual Prediction Method for Rural Kansas Highways. In *Proceedings of the 91st Annual Meeting of the Transportation Research Board, Washington, DC.*

Ma, M., Alluri, P., & Gan, A. (2013). Development of a Geographic Information System for *SafetyAnalyst* for Location Selection and Output Visualization. In *Proceedings of the 92nd Annual Meeting of the Transportation Research Board, Washington, DC.*

Martinelli, F., La Torre, F., & Vadi, P. (2009). Calibration of the Highway Safety Manual's Accident Prediction Model for Italian Secondary Road Network. *Transportation Research Record: Journal of the Transportation Research Board*, *2103*(1), 1-9.

Mehta, G., & Lou, Y. (2013). Safety Performance Function Calibration and Development for the State of Alabama: Two-lane Two-way Rural Roads and Four-lane Divided Highways. In *Proceedings of the 92nd Annual Meeting of the Transportation Research Board, Washington, DC.*

Shih, S. S. (2013). Class 13a: Random Forests, for Model (and Predictor) Selection. Retrieved January 15, 2014, from http://www.stanford.edu/~stephsus/shih_randomforests.pdf.

Srinivasan, R., & Carter, D. L. (2011). *Development of safety performance functions for North Carolina* (No. FHWA/NC/2010-09). North Carolina Department of Transportation, Research and Analysis Group.

Srinivasan, S., Haas, P., Dhakar, N. S., Hormel, R., Torbic, D., & Harwood, D. (2011). *Development and Calibration of Highway Safety Manual Equations for Florida Conditions* (No. TRC-FDOT-82013-2011). Florida Department of Transportation, Tallahassee, FL.

Sun, X., Li, Y., Magri, D., & Shirazi, H. H. (2006). Application of the Highway Safety Manual Draft Chapter: Louisiana Experience. *Transportation Research Record: Journal of the Transportation Research Board, 1950*(1), 55-64.

Williamson, M., & Zhou, H. (2012). Develop calibration factors for crash prediction models for rural two-lane roadways in Illinois. *Procedia-Social and Behavioral Sciences*, *43*, 330-338.

Xie, F., Gladhill, K., Dixon, K. K., & Monsere, C. M. (2011). Calibration of Highway Safety Manual Predictive Models for Oregon State Highways. *Transportation Research Record: Journal of the Transportation Research Board*, *2241*(1), 19-28.

Young, J., & Park, P. Y. (2013). Benefits of Small Municipalities Using Jurisdiction-specific Safety Performance Functions Rather than the Highway Safety Manual's Calibrated or Uncalibrated Safety Performance Functions. *Canadian Journal of Civil Engineering*, *40*(6), 517-527.

Zegeer, C.V., Reinfurt, D., Hummer, J., Herf, L., & Hunter, W. (1987). *Safety Effect of Cross-section Design for Two-lane Roads* (No. FHWA-RD-87/008). Federal Highway Administration, Department of Transportation, USA.

APPENDIX A:
DESCRIPTIVE STATISTICS OF DATA COLLECTED FOR SEGMENTS AND INTERSECTIONS

Table A-1: Descriptive Statistics of Rural Two-lane, Two-way Roads

Data Variable	Minimum	Maximum	Mean	Standard Deviation
Segment length (miles)	0.040	2.036	1.001	0.775
AADT (veh/day)	350	25,500	5,127	3,643
Lane width (ft)	9	22	11.97	0.65
Shoulder width (ft)	0	15	10.56	2.53
Driveway density	0	127.66	13.45	15.76
Roadside hazard rating	1	6	3.48	0.99

Table A-2: Descriptive Statistics of Rural Multilane Divided Highways

Data Variable	Minimum	Maximum	Mean	Standard Deviation
Segment length (miles)	0.040	2.036	0.697	0.717
AADT (veh/day)	1,850	66,000	14,757	9,450
Shoulder width (ft)	0	8	5.103	1.88
Median width (ft)	10	100	43.44	17.03

Table A-3: Descriptive Statistics of Urban and Suburban Two-lane Undivided Arterials

Data Variable	Minimum	Maximum	Mean	Standard Deviation
Segment length (miles)	0.040	2.002	0.344	0.353
AADT (veh/day)	550	30,500	10,690	4,789
Roadside fixed object density combining both sides	0	560.98	52.25	46.50
Offset to fixed object (ft)	2	30	19.90	8.68

Table A-4: Descriptive Statistics of Urban and Suburban Three-lane Arterials with TWLTL

Data Variable	Minimum	Maximum	Mean	Standard Deviation
Segment length (miles)	0.040	1.362	0.187	0.162
AADT (veh/day)	4,000	34,500	14,566	4,825
Median width (ft)	10	20	10.33	1.80
Roadside fixed object density combining both sides	0	250.00	79.58	35.56
Offset to fixed object (ft)	2	30	9.53	5.34

Table A-5: Descriptive Statistics of Urban and Suburban Four-lane Divided Arterials

Data Variable	Minimum	Maximum	Mean	Standard Deviation
Segment length (miles)	0.040	2.034	0.282	0.336
AADT (veh/day)	1,200	89,000	26,186	10,177
Median width (ft)	10	100	31.52	14.44
Roadside fixed object density combining both sides	0	671.05	97.43	59.33
Offset to fixed object (ft)	2	30	14.87	6.53

Table A-6: Descriptive Statistics of Urban and Suburban Four-lane Undivided Arterials

Data Variable	Minimum	Maximum	Mean	Standard Deviation
Segment length (miles)	0.040	1.865	0.196	0.175
AADT (veh/day)	3,300	46,000	19,014	8,095
Roadside fixed object density combining both sides	0	291.67	74.67	40.22
Offset to fixed object (ft)	2	25	4.61	3.49

Table A-7: Descriptive Statistics of Urban and Suburban Five-lane Arterials with TWLTL

Data Variable	Minimum	Maximum	Mean	Standard Deviation
Segment length (miles)	0.040	2.000	0.249	0.227
AADT (veh/day)	3,700	55,500	25,196	8,994
Roadside fixed object density	0	250	75.47	31.38
Offset to fixed object (ft)	0	20	9.06	2.97

Table A-8: Descriptive Statistics of Rural Two-lane Three-leg Stop-controlled Intersections

Data Variable	Minimum	Maximum	Mean	Standard Deviation
AADT for major road (veh/day)	300	23,000	4,801	3,382
AADT for minor road (veh/day)	50	15,485	1,388	1,802

Number of rural two-lane three-leg stop-controlled intersections with lighting: 48
Number of rural two-lane three-leg stop-controlled intersections without lighting: 250

Table A-9: Descriptive Statistics of Rural Two-lane Four-leg Stop-controlled Intersections

Data Variable	Minimum	Maximum	Mean	Standard Deviation
AADT for major road (veh/day)	996	16,100	3,470	2,484
AADT for minor road (veh/day)	103	2,000	640	479

Number of rural two-lane four-leg stop-controlled intersections with lighting: 2
Number of rural two-lane four-leg stop-controlled intersections without lighting: 41

Table A-10: Descriptive Statistics of Rural Two-lane Four-leg Signalized Intersections

Data Variable	Minimum	Maximum	Mean	Standard Deviation
AADT for major road (veh/day)	2,100	29,500	8,702	5,747
AADT for minor road (veh/day)	350	15,600	4,138	3,431

Number of rural two-lane four-leg stop-controlled intersections with lighting: 12
Number of rural two-lane four-leg stop-controlled intersections without lighting: 9

Table A-11: Descriptive Statistics of Rural Multilane Three-leg Stop-controlled Intersections

Data Variable	Minimum	Maximum	Mean	Standard Deviation
AADT for major road (veh/day)	1,040	26,000	11,695	5,340
AADT for minor road (veh/day)	206	10,600	2,039	2,061

Number of rural multilane three-leg stop-controlled intersections with lighting: 11
Number of rural multilane three-leg stop-controlled intersections without lighting: 20

Table A-12: Descriptive Statistics of Rural Multilane Four-leg Signalized Intersections

Data Variable	Minimum	Maximum	Mean	Standard Deviation
AADT for major road (veh/day)	3,900	35,000	12,535	6,253
AADT for minor road (veh/day)	1,000	25,500	5,122	4,281

Number of rural multilane three-leg stop-controlled intersections with lighting: 22
Number of rural multilane three-leg stop-controlled intersections without lighting: 5

Table A-13: Descriptive Statistics of Urban and Suburban Three-leg Signalized Intersections

Data Variable	Minimum	Maximum	Mean	Standard Deviation
AADT for major road (veh/day)	7,800	73,500	25,585	12,914
AADT for minor road (veh/day)	450	35,500	12,448	8,125

Number of urban and suburban three-leg stop-controlled intersections with lighting: 46
Number of urban and suburban three-leg stop-controlled intersections without lighting: 12

Table A-14: Descriptive Statistics of Urban and Suburban Three-leg Stop-controlled Intersections

Data Variable	Minimum	Maximum	Mean	Standard Deviation
AADT for major road (veh/day)	250	56,000	16,248	10,244
AADT for minor road (veh/day)	80	41,500	3,542	3,287

Number of urban and suburban three-leg stop-controlled intersections with lighting: 221
Number of urban and suburban three-leg stop-controlled intersections without lighting: 100

Table A-15: Descriptive Statistics of Urban and Suburban Four-leg Signalized Intersections

Data Variable	Minimum	Maximum	Mean	Standard Deviation
AADT for major road (veh/day)	1,547	77,000	31,630	14,333
AADT for minor road (veh/day)	600	55,000	16,568	10,862

Number of urban and suburban three-leg stop-controlled intersections with lighting: 427
Number of urban and suburban three-leg stop-controlled intersections without lighting: 32

Table A-16: Descriptive Statistics of Urban and Suburban Four-leg Stop-controlled Intersections

Data Variable	Minimum	Maximum	Mean	Standard Deviation
AADT for major road (veh/day)	3,600	31,500	16,744	8,054
AADT for minor road (veh/day)	200	11,400	2,706	2,440

Number of urban and suburban four-leg stop-controlled intersections with lighting: 29
Number of urban and suburban four-leg stop-controlled intersections without lighting: 5

APPENDIX B:
FLORIDA-SPECIFIC AND HSM-DEFAULT CRASH DISTRIBUTIONS

Table B-1: Crash Proportions of Single-vehicle Run-off-road, Multiple-vehicle Head-on, and Sideswipe Crashes

Facility Type	Analysis Year		
	2009	2010	2011
Rural Two-lane, Two-way Road	0.413	0.391	0.291
Rural Four-lane Divided Highways	0.398	0.365	0.256

Table B-2: Nighttime Crash Proportions for Unlighted Roadway Segments

Facility Type	Year	Proportion of Nighttime Crashes by Severity Level				Proportion of Crashes That Occurred at Night	
		F+I		PDO		Total	
		Florida Values	HSM-Default Values	Florida Values	HSM-Default Values	Florida Values	HSM-Default Values
Rural Two-lane, Two-way Undivided Roads	2009	0.587	0.382	0.413	0.618	0.377	0.370
	2010	0.578		0.422		0.377	
	2011	0.567		0.433		0.401	
Rural Four-lane Divided Highways	2009	0.556	0.323	0.444	0.677	0.323	0.426
	2010	0.530		0.470		0.327	
	2011	0.601		0.399		0.332	
Urban and Suburban Two-lane Undivided Arterials	2009	0.612	0.424	0.388	0.576	0.247	0.316
	2010	0.645		0.355		0.231	
	2011	0.545		0.455		0.225	
Urban and Suburban Three-lane Arterials with TWLTL	2009	0.706	0.429	0.294	0.571	0.130	0.304
	2010	0.591		0.409		0.185	
	2011	0.600		0.400		0.107	
Urban and Suburban Four-lane Undivided Arterials	2009	1.000	0.517	0.000	0.483	0.027	0.365
	2010	0.750		0.250		0.091	
	2011	1.000		0.000		0.036	
Urban and Suburban Four-lane Divided Arterials	2009	0.570	0.364	0.430	0.636	0.192	0.410
	2010	0.555		0.445		0.177	
	2011	0.555		0.445		0.190	
Urban and Suburban Five-lane Arterials with TWLTL	2009	0.714	0.432	0.286	0.568	0.095	0.274
	2010	0.667		0.333		0.137	
	2011	0.636		0.364		0.080	

Table B-3: Nighttime Crash Proportions for Unlighted Intersections

Facility Type	Site Subtype	Florida-Specific Values	HSM-Default Values
Rural Two-lane Intersections	Three-leg Stop-controlled	0.346	0.260
	Four-leg Stop-controlled	0.188	0.244
	Four-leg Signalized	0.104	0.286
Rural Four-lane Intersections	Three-leg Stop-controlled	0.198	0.276
	Four-leg Signalized	0.138	0.273
Urban and Suburban Intersections	Three-leg Stop-controlled	0.155	0.238
	Four-leg Stop-controlled	0.057	0.229
	Three-leg Signalized	0.094	0.235
	Four-leg Signalized	0.091	0.235

Table B-4: Pedestrian and Bicycle Adjustment Factors for Urban and Suburban Intersections

Adjustment Factor	Site Subtype	Florida-Specific Values	HSM-Default Values
Pedestrian Adjustment Factor	Three-leg Stop-controlled	0.012	0.021
	Four-leg Stop-controlled	0.020	0.022
Bicycle Adjustment Factor	Three-leg Stop-controlled	0.020	0.016
	Four-leg Stop-controlled	0.033	0.011
	Three-leg Signalized	0.014	0.018
	Four-leg Signalized	0.012	0.015

Table B-5: Proportion of Specific Crashes on Urban and Suburban Signalized Intersections

Crash Type	Number of Intersection Legs	2009	2010	2011	HSM-Default Values
Multiple-vehicle F+I crashes represented by right-angle collisions	Three-leg	0.023	0.037	0.026	0.280
	Four-leg	0.013	0.013	0.008	0.347
Multiple-vehicle PDO crashes represented by right-angle collisions	Three-leg	0.027	0.033	0.030	0.204
	Four-leg	0.033	0.027	0.018	0.244
Multiple-vehicle F+I crashes represented by rear-end collisions	Three-leg	0.538	0.622	0.614	0.549
	Four-leg	0.512	0.505	0.541	0.450
Multiple-vehicle PDO crashes represented by rear-end collisions	Three-leg	0.423	0.489	0.533	0.546
	Four-leg	0.438	0.452	0.502	0.483

APPENDIX C:
DATA DICTIONARY FOR THE *SAFETYANALYST* IMPORT FILES

Table C-1: AltAccident File

No.	Column Name	Description
1	agencyID	Unique ID used to identify the crash
2	locSystem	Location system
3	routeType	Route type
4	routeName	Route name
5	county	County
6	locSection	Location section
7	locOffset	Location offset
8	gisID	GIS identifier
9	accidentIntersectionID	Accident intersection ID
10	accidentRampID	Accident ramp ID
11	accidentSegmentID	Accident segment ID
12	accidentDate	Accident date
13	accidentTime	Accident time
14	accidentSeverity1	Accident severity
15	numberOfFatalities	Number of fatalities
16	numberOfInjuries	Number of injuries
17	junctionRelationship	Junction relationship
18	drivewayIndicator	Driveway indicator
19	lightCondition	Lighting condition
20	weatherCondition	Weather condition
21	surfaceCondition	Surface condition
22	collisionType	Collision type
23	environmentCondition	Environment condition
24	roadCondition	Road condition
25	schoolBus	School bus-related
26	workZone	Work zone-related
27	numVehicles	Number of vehicles
28	drugInvolved	Drug involved
29	towIndicator	Tow indicator
30	runoffIndicator	Run-off-the-road indicator
31	pedestrianIndicator	Pedestrian-related
32	bikeIndicator	Bike-related
33	sideOfDividedHighway	Side of divided highway
34	v1initialTravelDirection	Initial travel direction of vehicle 1
35	v2initialTravelDirection	Initial travel direction of vehicle 2
36	v1vehicleManeuver	Vehicle maneuver of vehicle 1
37	v2vehicleManeuver	Vehicle maneuver of vehicle 2
38	v1vehicleConfiguration	Vehicle configuration of vehicle 1
39	v2vehicleConfiguration	Vehicle configuration of vehicle 2

No.	Column Name	Description
40	v1firstEvent	First harmful event of vehicle 1
41	v2firstEvent	First harmful event of vehicle 2
42	v1driverDOB	Date of birth of the driver of vehicle 1
43	v2driverDOB	Date of birth of the driver of vehicle 2

Table C-2: AltSegment File

No.	Column Name	Description
1	agencyID	Unique ID used to identify the segment
2	locSystem	Location system
3	routeType	Route type
4	routeName	Route name
5	county	County
6	startOffset	Start of segment offset
7	endOffset	End of segment offset
8	gisID	GIS identifier
9	altRouteNames	Alternate route names
10	majorRoadName	Major road name
11	segmentLength	Length of the segment
12	district	District
13	city	City/town name
14	jurisdiction	Jurisdiction
15	areaType	Area type
16	terrain	Terrain type
17	roadwayClass1	Roadway class level 1
18	d1numThruLane	Number of through lanes in direction 1
19	d2numThruLane	Number of through lanes in direction 2
20	d1auxLane1	Auxiliary lane 1 in direction 1
21	d1auxLane2	Auxiliary lane 2 in direction 1
22	d1auxLane3	Auxiliary lane 3 in direction 1
23	d2auxLane1	Auxiliary lane 1 in direction 2
24	d2auxLane2	Auxiliary lane 2 in direction 2
25	d2auxLane3	Auxiliary lane 3 in direction 2
26	d1avgLaneWidth	Average lane width in direction 1
27	d2avgLaneWidth	Average lane width in direction 2
28	medianType1	Type of median
29	medianWidth	Median width
30	d1shoulderTypeOut	Type of outside shoulder in direction 1
31	d1shoulderTypeIn	Type of inside shoulder in direction 1
32	d2shoulderTypeOut	Type of outside shoulder in direction 2
33	d2shoulderTypeIn	Type of inside shoulder in direction 2

No.	Column Name	Description
34	d1avgShoulderWidthOut	Average width of outside shoulder in direction 1
35	d1avgShoulderWidthIn	Average width of inside shoulder in direction 1
36	d2avgShoulderWidthOut	Average width of outside shoulder in direction 2
37	d2avgShoulderWidthIn	Average width of inside shoulder in direction 2
38	accessControl	Type of access control
39	drivewayDensity	Driveway density
40	growthFactor	Growth factor
41	postedSpeed	Speed limit
42	operationWay	Type of operation way (i.e., two-way vs. one-way)
43	travelDirection	Direction of travel
44	increasingMileposts	Direction of increasing mileposts
45	d1bikeway	Presence of bikeway in direction 1
46	d2bikeway	Presence of bikeway in direction 1
47	interchangeInfluence	Presence of interchange influence area on freeways
48	openedToTraffic	Date opened to traffic
49	discontinuity	Discontinuity
50	corridor	Corridor
51	comment	Comment

Table C-3: AltSegmentTraffic File

No.	Column Name	Description
1	agencyID	Associated agency segment identifier
2	calendarYear	Year
3	aadtVPD	AADT
4	comment	Comment

Table C-4: AltIntersection File

No.	Column Name	Description
1	agencyID	Unique ID used to identify the segment
2	majorRoadLocSystem	Location system of the major road
3	routeType	Route type
4	routeName	Route name
5	county	County
6	majorRoadOffset	Offset distance of the major road
7	minorRoadLocSystem	Location system of the minor road
8	minorRoadRouteType	Route type of the minor road
9	minorRoadRouteName	Route name of the minor road
10	minorRoadOffset	Offset distance of the minor road
11	gisID	GIS identifier
12	altRouteNames	Alternate route names
13	majorRoadName	Major road name
14	minorRoadName	Minor road name
15	majorRoadDirection	Major road direction
16	majBeginInfluenceZone	Beginning influence zone of major road
17	minBeginInfluenceZone	Beginning influence zone of minor road
18	majEndInfluenceZone	Ending influence zone of major road
19	minEndInfluenceZone	Ending influence zone of minor road
20	district	District
21	city	City/town
22	jurisdiction	Jurisdiction
23	areaType	Area type
24	intersectionType1	Type of intersection
25	trafficControl1	Type of traffic control
26	offsetIntersection	Offset intersection
27	offsetDistance	Offset distance
28	growthFactor	Growth factor
29	openedToTraffic	Date opened to traffic
30	corridor	Corridor
31	comment	Comment
32	agencySiteSubtype	Agency site subtype

Table C-5: AltMajorRoadTraffic File

No.	Column Name	Description
1	agencyID	Associated agency intersection identifier
2	calendarYear	Year
3	aadtVPD	AADT
4	comment	Comment

Table C-6: AltMinorRoadTraffic File

No.	Column Name	Description
1	agencyID	Associated agency intersection identifier
2	calendarYear	Year
3	aadtVPD	AADT
4	comment	Comment

Table C-7: AltLeg File

No.	Column Name	Description
1	agencyID	Associated agency intersection identifier
2	legID	Intersection leg identifier
3	legType	Type of intersection leg
4	legDirection	Direction of intersection leg
5	legNumThruLane	Number of through lanes
6	legNumLeftTurnLane	Number of left-turn lanes
7	legNumRightTurnLane	Number of right-turn lanes
8	legMedianType	Median type
9	leftTurnPhasing	Left-turn phasing
10	postedSpeed	speed limit
11	turnProhibitions	Turn prohibitions
12	operationWay	Type of operation way (i.e., two-way vs. one-way)

Table C-8: AltLegTraffic File

No.	Column Name	Description
1	agencyID	Associated agency intersection identifier
2	legID	Intersection leg identifier
3	calendarYear	Year
4	aadtVPD	AADT
5	throughVolume	Through volume
6	leftTurnVolume	Left-turn volume
7	rightTurnVolume	Right-turn volume

Table C-9: AltRamp File

No.	Column Name	Description
1	agencyID	Unique ID used to identify the ramp
2	locSystem	Location system
3	routeType	Route type
4	routeName	Route name
5	startOffset	Start of segment offset
6	endOffset	End of segment offset
7	gisID	GIS identifier
8	altRouteNames	Alternate route names
9	majorRoadName	Major road name
10	county	County
11	comment	Comment
12	district	District
13	city	City/town name
14	jurisdiction	Jurisdiction
15	areaType	Area type
16	rampType	Ramp type
17	rampConfiguration	Ramp configuration
18	rampFreewayConnection	Ramp freeway connection
19	rampCrossroadConnection	Ramp crossroad connection
20	numOfLanes	Number of lanes
21	rampLength	Length of ramp
22	growthFactor	Growth factor
23	openedToTraffic	Date opened to traffic
24	corridor	Corridor
25	agencySiteSubtype	Agency site subtype

Table C-10: AltRampTraffic File

No.	Column Name	Description
1	agencyID	Associated agency ramp identifier
2	calendarYear	Year
3	aadtVPD	AADT
4	comment	Comment

APPENDIX D:
DATA MAPPING

Table D-1: Data Mapping for AltAccident File

SafetyAnalyst (SA) Variable	SA Type	Reqd?	CAR Variable	Source Table	SA Code	Source Code	Default Value
accidentDate	Date	Y	CRASHDTE	RDWTBL50			99
accidentIntersectionID	Char	N	NA	AltIntersection			0
accidentRampID	Char	N	NA	AltRamp			n
accidentSegmentID	Char	N	NA	AltSegment			n
accidentSeverity1	Code	Y	ACCISEV	RDWTBL50	A	4	0
accidentSeverity1	Code	Y	ACCISEV	RDWTBL50	B	3	0
accidentSeverity1	Code	Y	ACCISEV	RDWTBL50	C	2	0
accidentSeverity1	Code	Y	ACCISEV	RDWTBL50	P	1	0
accidentSeverity1	Code	Y	ACCISEV	RDWTBL50	X	NA	0
accidentSeverity1	Code	Y	ACCISEV	RDWTBL50	K	5	0
accidentTime	Time	N	TIMEOFAC	RDWTBL50			0
agencyID	Char	Y	CARNUM	KEY			n
bikeIndicator	Code	N	FRST_EVNT_CAUS_CD	RDWTBL51	Y	11	X
bikeIndicator	Code	N	SCND_EVNT_CAUS_CD	RDWTBL51	Y	11	X
bikeIndicator	Code	N	FRST_EVNT_CAUS_CD	RDWTBL51	Y	12	X
bikeIndicator	Code	N	SCND_EVNT_CAUS_CD	RDWTBL51	Y	12	X
bikeIndicator	Code	N	FRST_EVNT_CAUS_CD	RDWTBL51	X	0	X
bikeIndicator	Code	N	SCND_EVNT_CAUS_CD	RDWTBL51	X	0	X
collisionType	Code	Y	FRST_EVNT_CAUS_CD	RDWTBL51	11	37	99
collisionType	Code	Y	FRST_EVNT_CAUS_CD	RDWTBL51	27	9	99
collisionType	Code	Y	FRST_EVNT_CAUS_CD	RDWTBL51	9	31	99
collisionType	Code	Y	FRST_EVNT_CAUS_CD	RDWTBL51	10	34	99
collisionType	Code	Y	FRST_EVNT_CAUS_CD	RDWTBL51	10	35	99
collisionType	Code	Y	FRST_EVNT_CAUS_CD	RDWTBL51	11	32	99
collisionType	Code	Y	FRST_EVNT_CAUS_CD	RDWTBL51	8	39	99
collisionType	Code	Y	FRST_EVNT_CAUS_CD	RDWTBL51	11	33	99
collisionType	Code	Y	FRST_EVNT_CAUS_CD	RDWTBL51	11	36	99
collisionType	Code	Y	FRST_EVNT_CAUS_CD	RDWTBL51	11	38	99
collisionType	Code	Y	FRST_EVNT_CAUS_CD	RDWTBL51	21	1	99
collisionType	Code	Y	FRST_EVNT_CAUS_CD	RDWTBL51	22	2	99
collisionType	Code	Y	FRST_EVNT_CAUS_CD	RDWTBL51	24	3	99
collisionType	Code	Y	FRST_EVNT_CAUS_CD	RDWTBL51	25	6	99
collisionType	Code	Y	FRST_EVNT_CAUS_CD	RDWTBL51	27	4	99
collisionType	Code	Y	FRST_EVNT_CAUS_CD	RDWTBL51	27	7	99
collisionType	Code	Y	FRST_EVNT_CAUS_CD	RDWTBL51	8	30	99
collisionType	Code	Y	FRST_EVNT_CAUS_CD	RDWTBL51	99	0	99
collisionType	Code	Y	FRST_EVNT_CAUS_CD	RDWTBL51	7	19	99
collisionType	Code	Y	FRST_EVNT_CAUS_CD	RDWTBL51	27	5	99
collisionType	Code	Y	FRST_EVNT_CAUS_CD	RDWTBL51	3	12	99

SafetyAnalyst (SA) Variable	SA Type	Reqd?	CAR Variable	Source Table	SA Code	Source Code	Default Value
collisionType	Code	Y	FRST_EVNT_CAUS_CD	RDWTBL51	7	22	99
collisionType	Code	Y	FRST_EVNT_CAUS_CD	RDWTBL51	8	29	99
collisionType	Code	Y	FRST_EVNT_CAUS_CD	RDWTBL51	1	8	99
collisionType	Code	Y	FRST_EVNT_CAUS_CD	RDWTBL51	3	11	99
collisionType	Code	Y	FRST_EVNT_CAUS_CD	RDWTBL51	3	13	99
collisionType	Code	Y	FRST_EVNT_CAUS_CD	RDWTBL51	4	10	99
collisionType	Code	Y	FRST_EVNT_CAUS_CD	RDWTBL51	5	15	99
collisionType	Code	Y	FRST_EVNT_CAUS_CD	RDWTBL51	6	27	99
collisionType	Code	Y	FRST_EVNT_CAUS_CD	RDWTBL51	7	16	99
collisionType	Code	Y	FRST_EVNT_CAUS_CD	RDWTBL51	7	18	99
collisionType	Code	Y	FRST_EVNT_CAUS_CD	RDWTBL51	8	28	99
collisionType	Code	Y	FRST_EVNT_CAUS_CD	RDWTBL51	7	20	99
collisionType	Code	Y	FRST_EVNT_CAUS_CD	RDWTBL51	7	23	99
collisionType	Code	Y	FRST_EVNT_CAUS_CD	RDWTBL51	7	24	99
collisionType	Code	Y	FRST_EVNT_CAUS_CD	RDWTBL51	8	26	99
collisionType	Code	Y	FRST_EVNT_CAUS_CD	RDWTBL51	7	25	99
collisionType	Code	Y	FRST_EVNT_CAUS_CD	RDWTBL51	8	21	99
collisionType	Code	Y	FRST_EVNT_CAUS_CD	RDWTBL51	7	17	99
collisionType	Code	Y	FRST_EVNT_CAUS_CD	RDWTBL51	2	14	99
county	Char	N	CONTYDOT	RDWTBL50			L
drivewayIndicator	Code	N	SITELOCA	RDWTBL50	1	11	99
drivewayIndicator	Code	N	SITELOCA	RDWTBL50	1	12	99
drivewayIndicator	Code	N	SITELOCA	RDWTBL50	1	13	99
drivewayIndicator	Code	N	SITELOCA	RDWTBL50	2	4	99
drivewayIndicator	Code	N	SITELOCA	RDWTBL50	1	10	99
drivewayIndicator	Code	N	SITELOCA	RDWTBL50	99	77	99
drivewayIndicator	Code	N	SITELOCA	RDWTBL50	1	6	99
drivewayIndicator	Code	N	SITELOCA	RDWTBL50	99	NA	99
drivewayIndicator	Code	N	SITELOCA	RDWTBL50	1	9	99
drivewayIndicator	Code	N	SITELOCA	RDWTBL50	1	1	99
drivewayIndicator	Code	N	SITELOCA	RDWTBL50	1	7	99
drivewayIndicator	Code	N	SITELOCA	RDWTBL50	1	5	99
drivewayIndicator	Code	N	SITELOCA	RDWTBL50	1	3	99
drivewayIndicator	Code	N	SITELOCA	RDWTBL50	1	2	99
drivewayIndicator	Code	N	SITELOCA	RDWTBL50	1	8	99
drugInvolved	Code	N	ALCINVCD	RDWTBL50	1	0	99
drugInvolved	Code	N	ALCINVCD	RDWTBL50	2	1	99
drugInvolved	Code	N	ALCINVCD	RDWTBL50	3	2	99
drugInvolved	Code	N	ALCINVCD	RDWTBL50	4	3	99
drugInvolved	Code	N	ALCINVCD	RDWTBL50	99	4	99

SafetyAnalyst (SA) Variable	SA Type	Reqd?	CAR Variable	Source Table	SA Code	Source Code	Default Value
environmentCondition	Code	N	NA	DEFAULT			99
gisID	Char	N	NA	NA			n
junctionRelationship	Code	Y	SITELOCA	JUNCTION	9	77	0
junctionRelationship	Code	Y	SITELOCA	JUNCTION	99	NA	0
junctionRelationship	Code	Y	SITELOCA	JUNCTION	9	13	0
junctionRelationship	Code	Y	SITELOCA	JUNCTION	9	12	0
junctionRelationship	Code	Y	SITELOCA	JUNCTION	9	11	0
junctionRelationship	Code	Y	SITELOCA	JUNCTION	9	10	0
junctionRelationship	Code	Y	SITELOCA	JUNCTION	9	09	0
junctionRelationship	Code	Y	SITELOCA	JUNCTION	2	02	0
junctionRelationship	Code	Y	SITELOCA	JUNCTION	7	05	0
junctionRelationship	Code	Y	SITELOCA	JUNCTION	5	08	0
junctionRelationship	Code	Y	SITELOCA	JUNCTION	5	07	0
junctionRelationship	Code	Y	SITELOCA	JUNCTION	4	04	0
junctionRelationship	Code	Y	SITELOCA	JUNCTION	3	03	0
junctionRelationship	Code	Y	SITELOCA	JUNCTION	1	01	0
junctionRelationship	Code	Y	SITELOCA	JUNCTION	8	06	0
lightCondition	Code	N	LGHTCOND	RDWTBL50	5	5	99
lightCondition	Code	N	LGHTCOND	RDWTBL50	2	2	99
lightCondition	Code	N	LGHTCOND	RDWTBL50	99	NA	99
lightCondition	Code	N	LGHTCOND	RDWTBL50	7	88	99
lightCondition	Code	N	LGHTCOND	RDWTBL50	3	3	99
lightCondition	Code	N	LGHTCOND	RDWTBL50	1	1	99
lightCondition	Code	N	LGHTCOND	RDWTBL50	4	4	99
locOffset	Char	Y	LOCMP	LOCMP			n
locSection	Char	N	NA	NA			n
locSystem	Code	Y	DEFAULT	DEFAULT			n
numberOfFatalities	Char	Y	TOT_OF_FATL_NUM	RDWTBL50			n
numberOfInjuries	Char	Y	TOT_OF_INJR_NUM	RDWTBL50			n
numVehicles	Char	Y	TOT_OF_VHCL_NUM	RDWTBL50			n
pedestrianIndicator	Code	N	SCND_EVNT_CAUS_CD	RDWTBL51	Y	10	99
pedestrianIndicator	Code	N	FRST_EVNT_CAUS_CD	RDWTBL51	X	0	99
pedestrianIndicator	Code	N	SCND_EVNT_CAUS_CD	RDWTBL51	X	0	99
pedestrianIndicator	Code	N	FRST_EVNT_CAUS_CD	RDWTBL51	Y	10	99
roadCondition	Code	N	FRST_RDCND_CRSH_CD	RDWTBL50	4	5	99
roadCondition	Code	N	FRST_RDCND_CRSH_CD	RDWTBL50	99	77	99
roadCondition	Code	N	FRST_RDCND_CRSH_CD	RDWTBL50	9	6	99
roadCondition	Code	N	FRST_RDCND_CRSH_CD	RDWTBL50	6	9	99
roadCondition	Code	N	FRST_RDCND_CRSH_CD	RDWTBL50	1	1	99
roadCondition	Code	N	FRST_RDCND_CRSH_CD	RDWTBL50	7	2	99

SafetyAnalyst (SA) Variable	SA Type	Reqd?	CAR Variable	Source Table	SA Code	Source Code	Default Value
routeName	Char	Y	RDWYID	RDWTBL50			A
routeType	Code	Y	USRTNO	RDWTBL50			X
runoffIndicator	Code	N	FRST_EVNT_CAUS_CD	RDWTBL51	X	0	1
runoffIndicator	Code	N	SCND_EVNT_CAUS_CD	RDWTBL51	X	0	1
runoffIndicator	Code	N	SCND_EVNT_CAUS_CD	RDWTBL51	Y	30	1
runoffIndicator	Code	N	FRST_EVNT_CAUS_CD	RDWTBL51	Y	30	1
schoolBus	Test	N	VEHUSE	RDWTBL51	1	77	0
schoolBus	Test	N	VEHUSE	RDWTBL51	2	5	0
schoolBus	Test	N	VEHUSE	RDWTBL51	2	6	0
schoolBus	Test	N	VEHUSE	RDWTBL51	99	0	0
sideOfDividedHighway	Code	N	LOCDIRCD	RDWTBL50	WB	W	X
sideOfDividedHighway	Code	N	LOCDIRCD	RDWTBL50	NB	N	X
sideOfDividedHighway	Code	N	LOCDIRCD	RDWTBL50	SB	S	X
sideOfDividedHighway	Code	N	LOCDIRCD	RDWTBL50	X	U	X
sideOfDividedHighway	Code	N	LOCDIRCD	RDWTBL50	EB	E	X
surfaceCondition	Code	N	RDSURFCD	RDWTBL50	2	2	99
surfaceCondition	Code	N	RDSURFCD	RDWTBL50	99	88	99
surfaceCondition	Code	N	RDSURFCD	RDWTBL50	10	77	99
surfaceCondition	Code	N	RDSURFCD	RDWTBL50	5	4	99
surfaceCondition	Code	N	RDSURFCD	RDWTBL50	1	1	99
surfaceCondition	Code	N	RDSURFCD	RDWTBL50	6	3	99
towIndicator	Code	N	NA	DEFAULT			N
v1driverDOB	Date	N	VHCL_OWN_BRTH_DT	RDWTBL51			99
v1firstEvent	Code	N	FRST_EVNT_CAUS_CD	RDWTBL51	38	27	99
v1firstEvent	Code	N	FRST_EVNT_CAUS_CD	RDWTBL51	20	21	99
v1firstEvent	Code	N	FRST_EVNT_CAUS_CD	RDWTBL51	17	7	99
v1firstEvent	Code	N	FRST_EVNT_CAUS_CD	RDWTBL51	17	9	99
v1firstEvent	Code	N	FRST_EVNT_CAUS_CD	RDWTBL51	17	6	99
v1firstEvent	Code	N	FRST_EVNT_CAUS_CD	RDWTBL51	18	28	99
v1firstEvent	Code	N	FRST_EVNT_CAUS_CD	RDWTBL51	99	0	99
v1firstEvent	Code	N	FRST_EVNT_CAUS_CD	RDWTBL51	19	25	99
v1firstEvent	Code	N	FRST_EVNT_CAUS_CD	RDWTBL51	17	5	99
v1firstEvent	Code	N	FRST_EVNT_CAUS_CD	RDWTBL51	25	29	99
v1firstEvent	Code	N	FRST_EVNT_CAUS_CD	RDWTBL51	27	18	99
v1firstEvent	Code	N	FRST_EVNT_CAUS_CD	RDWTBL51	29	20	99
v1firstEvent	Code	N	FRST_EVNT_CAUS_CD	RDWTBL51	30	24	99
v1firstEvent	Code	N	FRST_EVNT_CAUS_CD	RDWTBL51	31	22	99
v1firstEvent	Code	N	FRST_EVNT_CAUS_CD	RDWTBL51	32	17	99
v1firstEvent	Code	N	FRST_EVNT_CAUS_CD	RDWTBL51	36	19	99
v1firstEvent	Code	N	FRST_EVNT_CAUS_CD	RDWTBL51	39	26	99

SafetyAnalyst (SA) Variable	SA Type	Reqd?	CAR Variable	Source Table	SA Code	Source Code	Default Value
v1firstEvent	Code	N	FRST_EVNT_CAUS_CD	RDWTBL51	11	11	99
v1firstEvent	Code	N	FRST_EVNT_CAUS_CD	RDWTBL51	17	4	99
v1firstEvent	Code	N	FRST_EVNT_CAUS_CD	RDWTBL51	34	16	99
v1firstEvent	Code	N	FRST_EVNT_CAUS_CD	RDWTBL51	6	32	99
v1firstEvent	Code	N	FRST_EVNT_CAUS_CD	RDWTBL51	12	14	99
v1firstEvent	Code	N	FRST_EVNT_CAUS_CD	RDWTBL51	1	31	99
v1firstEvent	Code	N	FRST_EVNT_CAUS_CD	RDWTBL51	2	34	99
v1firstEvent	Code	N	FRST_EVNT_CAUS_CD	RDWTBL51	2	35	99
v1firstEvent	Code	N	FRST_EVNT_CAUS_CD	RDWTBL51	17	3	99
v1firstEvent	Code	N	FRST_EVNT_CAUS_CD	RDWTBL51	5	37	99
v1firstEvent	Code	N	FRST_EVNT_CAUS_CD	RDWTBL51	8	30	99
v1firstEvent	Code	N	FRST_EVNT_CAUS_CD	RDWTBL51	8	36	99
v1firstEvent	Code	N	FRST_EVNT_CAUS_CD	RDWTBL51	8	38	99
v1firstEvent	Code	N	FRST_EVNT_CAUS_CD	RDWTBL51	17	1	99
v1firstEvent	Code	N	FRST_EVNT_CAUS_CD	RDWTBL51	10	10	99
v1firstEvent	Code	N	FRST_EVNT_CAUS_CD	RDWTBL51	11	12	99
v1firstEvent	Code	N	FRST_EVNT_CAUS_CD	RDWTBL51	13	15	99
v1firstEvent	Code	N	FRST_EVNT_CAUS_CD	RDWTBL51	14	13	99
v1firstEvent	Code	N	FRST_EVNT_CAUS_CD	RDWTBL51	15	8	99
v1firstEvent	Code	N	FRST_EVNT_CAUS_CD	RDWTBL51	16	23	99
v1firstEvent	Code	N	FRST_EVNT_CAUS_CD	RDWTBL51	8	39	99
v1firstEvent	Code	N	FRST_EVNT_CAUS_CD	RDWTBL51	17	2	99
v1firstEvent	Code	N	FRST_EVNT_CAUS_CD	RDWTBL51	4	33	99
v1initialTravelDirection	Code	Y	TRAVDIR	RDWTBL51	NB	N	X
v1initialTravelDirection	Code	Y	TRAVDIR	RDWTBL51	SB	S	X
v1initialTravelDirection	Code	Y	TRAVDIR	RDWTBL51	EB	E	X
v1initialTravelDirection	Code	Y	TRAVDIR	RDWTBL51	WB	W	X
v1initialTravelDirection	Code	Y	TRAVDIR	RDWTBL51	XX	U	X
v1vehicleConfiguration	Code	N	VEHTYPE	RDWTBL51	7	4	XX
v1vehicleConfiguration	Code	N	VEHTYPE	RDWTBL51	2	3	XX
v1vehicleConfiguration	Code	N	VEHTYPE	RDWTBL51	99	0	XX
v1vehicleConfiguration	Code	N	VEHTYPE	RDWTBL51	15	8	XX
v1vehicleConfiguration	Code	N	VEHTYPE	RDWTBL51	14	9	XX
v1vehicleConfiguration	Code	N	VEHTYPE	RDWTBL51	9	6	XX
v1vehicleConfiguration	Code	N	VEHTYPE	RDWTBL51	1	1	XX
v1vehicleConfiguration	Code	N	VEHTYPE	RDWTBL51	4	11	XX
v1vehicleConfiguration	Code	N	VEHTYPE	RDWTBL51	4	12	XX
v1vehicleConfiguration	Code	N	VEHTYPE	RDWTBL51	1	2	XX
v1vehicleConfiguration	Code	N	VEHTYPE	RDWTBL51	5	7	XX
v1vehicleManeuver	Code	Y	VEHMOVE	RDWTBL51	6	3	NA

SafetyAnalyst (SA) Variable	SA Type	Reqd?	CAR Variable	Source Table	SA Code	Source Code	Default Value
v1vehicleManeuver	Code	Y	VEHMOVE	RDWTBL51	7	10	NA
v1vehicleManeuver	Code	Y	VEHMOVE	RDWTBL51	99	88	NA
v1vehicleManeuver	Code	Y	VEHMOVE	RDWTBL51	12	2	NA
v1vehicleManeuver	Code	Y	VEHMOVE	RDWTBL51	10	8	NA
v1vehicleManeuver	Code	Y	VEHMOVE	RDWTBL51	1	1	NA
v1vehicleManeuver	Code	Y	VEHMOVE	RDWTBL51	4	11	NA
v1vehicleManeuver	Code	Y	VEHMOVE	RDWTBL51	3	6	NA
v1vehicleManeuver	Code	Y	VEHMOVE	RDWTBL51	2	4	NA
v1vehicleManeuver	Code	Y	VEHMOVE	RDWTBL51	10	9	NA
v1vehicleManeuver	Code	Y	VEHMOVE	RDWTBL51	5	5	NA
v2driverDOB	Date	N	VHCL_OWN_BRTH_DT	RDWTBL51			99
v2firstEvent	Code	N	FRST_EVNT_CAUS_CD	RDWTBL51	25	29	99
v2firstEvent	Code	N	FRST_EVNT_CAUS_CD	RDWTBL51	20	21	99
v2firstEvent	Code	N	FRST_EVNT_CAUS_CD	RDWTBL51	19	25	99
v2firstEvent	Code	N	FRST_EVNT_CAUS_CD	RDWTBL51	18	28	99
v2firstEvent	Code	N	FRST_EVNT_CAUS_CD	RDWTBL51	17	9	99
v2firstEvent	Code	N	FRST_EVNT_CAUS_CD	RDWTBL51	17	7	99
v2firstEvent	Code	N	FRST_EVNT_CAUS_CD	RDWTBL51	17	5	99
v2firstEvent	Code	N	FRST_EVNT_CAUS_CD	RDWTBL51	39	26	99
v2firstEvent	Code	N	FRST_EVNT_CAUS_CD	RDWTBL51	17	6	99
v2firstEvent	Code	N	FRST_EVNT_CAUS_CD	RDWTBL51	26	77	99
v2firstEvent	Code	N	FRST_EVNT_CAUS_CD	RDWTBL51	27	18	99
v2firstEvent	Code	N	FRST_EVNT_CAUS_CD	RDWTBL51	29	20	99
v2firstEvent	Code	N	FRST_EVNT_CAUS_CD	RDWTBL51	30	24	99
v2firstEvent	Code	N	FRST_EVNT_CAUS_CD	RDWTBL51	31	22	99
v2firstEvent	Code	N	FRST_EVNT_CAUS_CD	RDWTBL51	32	17	99
v2firstEvent	Code	N	FRST_EVNT_CAUS_CD	RDWTBL51	34	16	99
v2firstEvent	Code	N	FRST_EVNT_CAUS_CD	RDWTBL51	17	4	99
v2firstEvent	Code	N	FRST_EVNT_CAUS_CD	RDWTBL51	38	27	99
v2firstEvent	Code	N	FRST_EVNT_CAUS_CD	RDWTBL51	8	39	99
v2firstEvent	Code	N	FRST_EVNT_CAUS_CD	RDWTBL51	36	19	99
v2firstEvent	Code	N	FRST_EVNT_CAUS_CD	RDWTBL51	10	10	99
v2firstEvent	Code	N	FRST_EVNT_CAUS_CD	RDWTBL51	99	0	99
v2firstEvent	Code	N	FRST_EVNT_CAUS_CD	RDWTBL51	1	31	99
v2firstEvent	Code	N	FRST_EVNT_CAUS_CD	RDWTBL51	2	34	99
v2firstEvent	Code	N	FRST_EVNT_CAUS_CD	RDWTBL51	2	35	99
v2firstEvent	Code	N	FRST_EVNT_CAUS_CD	RDWTBL51	4	33	99
v2firstEvent	Code	N	FRST_EVNT_CAUS_CD	RDWTBL51	5	37	99
v2firstEvent	Code	N	FRST_EVNT_CAUS_CD	RDWTBL51	6	32	99
v2firstEvent	Code	N	FRST_EVNT_CAUS_CD	RDWTBL51	8	30	99

SafetyAnalyst (SA) Variable	SA Type	Reqd?	CAR Variable	Source Table	SA Code	Source Code	Default Value
v2firstEvent	Code	N	FRST_EVNT_CAUS_CD	RDWTBL51	11	11	99
v2firstEvent	Code	N	FRST_EVNT_CAUS_CD	RDWTBL51	8	38	99
v2firstEvent	Code	N	FRST_EVNT_CAUS_CD	RDWTBL51	17	3	99
v2firstEvent	Code	N	FRST_EVNT_CAUS_CD	RDWTBL51	11	12	99
v2firstEvent	Code	N	FRST_EVNT_CAUS_CD	RDWTBL51	12	14	99
v2firstEvent	Code	N	FRST_EVNT_CAUS_CD	RDWTBL51	13	15	99
v2firstEvent	Code	N	FRST_EVNT_CAUS_CD	RDWTBL51	14	13	99
v2firstEvent	Code	N	FRST_EVNT_CAUS_CD	RDWTBL51	15	8	99
v2firstEvent	Code	N	FRST_EVNT_CAUS_CD	RDWTBL51	16	23	99
v2firstEvent	Code	N	FRST_EVNT_CAUS_CD	RDWTBL51	17	1	99
v2firstEvent	Code	N	FRST_EVNT_CAUS_CD	RDWTBL51	17	2	99
v2firstEvent	Code	N	FRST_EVNT_CAUS_CD	RDWTBL51	8	36	99
v2initialTravelDirection	Code	Y	TRAVDIR	RDWTBL51	NB	N	X
v2initialTravelDirection	Code	Y	TRAVDIR	RDWTBL51	XX	U	X
v2initialTravelDirection	Code	Y	TRAVDIR	RDWTBL51	WB	W	X
v2initialTravelDirection	Code	Y	TRAVDIR	RDWTBL51	SB	S	X
v2initialTravelDirection	Code	Y	TRAVDIR	RDWTBL51	EB	E	X
v2vehicleConfiguration	Code	N	VEHTYPE	RDWTBL51	7	4	99
v2vehicleConfiguration	Code	N	VEHTYPE	RDWTBL51	99	0	99
v2vehicleConfiguration	Code	N	VEHTYPE	RDWTBL51	15	8	99
v2vehicleConfiguration	Code	N	VEHTYPE	RDWTBL51	9	6	99
v2vehicleConfiguration	Code	N	VEHTYPE	RDWTBL51	5	7	99
v2vehicleConfiguration	Code	N	VEHTYPE	RDWTBL51	4	12	99
v2vehicleConfiguration	Code	N	VEHTYPE	RDWTBL51	4	11	99
v2vehicleConfiguration	Code	N	VEHTYPE	RDWTBL51	2	3	99
v2vehicleConfiguration	Code	N	VEHTYPE	RDWTBL51	1	2	99
v2vehicleConfiguration	Code	N	VEHTYPE	RDWTBL51	1	1	99
v2vehicleConfiguration	Code	N	VEHTYPE	RDWTBL51	14	9	99
v2vehicleManeuver	Code	Y	VEHMOVE	RDWTBL51	5	5	XX
v2vehicleManeuver	Code	Y	VEHMOVE	RDWTBL51	99	88	XX
v2vehicleManeuver	Code	Y	VEHMOVE	RDWTBL51	12	2	XX
v2vehicleManeuver	Code	Y	VEHMOVE	RDWTBL51	10	9	XX
v2vehicleManeuver	Code	Y	VEHMOVE	RDWTBL51	10	8	XX
v2vehicleManeuver	Code	Y	VEHMOVE	RDWTBL51	6	3	XX
v2vehicleManeuver	Code	Y	VEHMOVE	RDWTBL51	4	11	XX
v2vehicleManeuver	Code	Y	VEHMOVE	RDWTBL51	3	6	XX
v2vehicleManeuver	Code	Y	VEHMOVE	RDWTBL51	2	4	XX
v2vehicleManeuver	Code	Y	VEHMOVE	RDWTBL51	1	1	XX
v2vehicleManeuver	Code	Y	VEHMOVE	RDWTBL51	7	10	XX
weatherCondition	Code	N	WEATCOND	RDWTBL50	1	1	99

SafetyAnalyst (SA) Variable	SA Type	Reqd?	CAR Variable	Source Table	SA Code	Source Code	Default Value
weatherCondition	Code	N	WEATCOND	RDWTBL50	2	2	99
weatherCondition	Code	N	WEATCOND	RDWTBL50	3	4	99
weatherCondition	Code	N	WEATCOND	RDWTBL50	4	3	99
weatherCondition	Code	N	WEATCOND	RDWTBL50	99	NA	99
weatherCondition	Code	N	WEATCOND	RDWTBL50	10	77	99
weatherCondition	Code	N	WEATCOND	RDWTBL50	99	88	99
workZone	Code	N	FRST_RDCND_CRSH_CD	RDWTBL50	N	6	99
workZone	Code	N	FRST_RDCND_CRSH_CD	RDWTBL50	N	9	99
workZone	Code	N	FRST_RDCND_CRSH_CD	RDWTBL50	X	77	99
workZone	Code	N	FRST_RDCND_CRSH_CD	RDWTBL50	N	8	99
workZone	Code	N	FRST_RDCND_CRSH_CD	RDWTBL50	N	7	99
workZone	Code	N	FRST_RDCND_CRSH_CD	RDWTBL50	N	3	99
workZone	Code	N	FRST_RDCND_CRSH_CD	RDWTBL50	N	1	99
workZone	Code	N	FRST_RDCND_CRSH_CD	RDWTBL50	Y	2	99
workZone	Code	N	FRST_RDCND_CRSH_CD	RDWTBL50	Y	4	99
workZone	Code	N	FRST_RDCND_CRSH_CD	RDWTBL50	N	5	99

Table D-2: Data Mapping for AltSegment File

SA Variable	SA Type	Required?	RCI Variable	SA Code	RCI Code	Default Value
agencyID	Char	Y	RDWYID			n
areaType	Code	Y	URBSIZE	U	2	X
areaType	Code	Y	URBSIZE	U	3	X
areaType	Code	Y	URBSIZE	U	4	X
areaType	Code	Y	URBSIZE	U	5	X
areaType	Code	Y	URBSIZE	R	1	X
county	Code	Y				n
d1numThruLane	Num	Y	NOLANES			1
d2numThruLane	Num	Y	NOLANES			1
endOffset	Num	Y	ENDSECPT			n
interchangeInfluence	Code	Y				N
locSystem	Code	Y				A
medianType1	Code	Y	RDMEDIAN	2	14	99
medianType1	Code	Y	RDMEDIAN	9	30	99
medianType1	Code	Y	RDMEDIAN	9	29	99
medianType1	Code	Y	RDMEDIAN	0	01	99
medianType1	Code	Y	RDMEDIAN	2	4	99
medianType1	Code	Y	RDMEDIAN	9	28	99
medianType1	Code	Y	RDMEDIAN	1	27	99
medianType1	Code	Y	RDMEDIAN	1	26	99
medianType1	Code	Y	RDMEDIAN	1	25	99
medianType1	Code	Y	RDMEDIAN	3	24	99
medianType1	Code	Y	RDMEDIAN	4	22	99
medianType1	Code	Y	RDMEDIAN	1	21	99
medianType1	Code	Y	RDMEDIAN	9	20	99
medianType1	Code	Y	RDMEDIAN	3	19	99
medianType1	Code	Y	RDMEDIAN	4	2	99
medianType1	Code	Y	RDMEDIAN	1	16	99
medianType1	Code	Y	RDMEDIAN	6	9	99
medianType1	Code	Y	RDMEDIAN	0	10	99
medianType1	Code	Y	RDMEDIAN	5	8	99
medianType1	Code	Y	RDMEDIAN	3	05	99
medianType1	Code	Y	RDMEDIAN	4	17	99
medianType1	Code	Y	RDMEDIAN	2	18	99
medianType1	Code	Y	RDMEDIAN	1	13	99
medianType1	Code	Y	RDMEDIAN	3	15	99
medianType1	Code	Y	RDMEDIAN	2	12	99
medianType1	Code	Y	RDMEDIAN	4	3	99
medianType1	Code	Y	RDMEDIAN	2	23	99
medianType1	Code	Y	RDMEDIAN	2	31	99
medianType1	Code	Y	RDMEDIAN	5	11	99
medianType1	Code	Y	RDMEDIAN	1	06	99
operationWay	Code	Y	TYPEROAD	1	04	2

SA Variable	SA Type	Required?	RCI Variable	SA Code	RCI Code	Default Value
operationWay	Code	Y	TYPEROAD	2	02	2
operationWay	Code	Y	TYPEROAD	99	0	2
routeName	Char	Y	RDWYID			n
routeType	Code	Y	FUNCLASS	I	1	L
routeType	Code	Y	FUNCLASS	I	11	L
routeType	Code	Y	RTESGNCD	BR	2	L
routeType	Code	Y	RTESGNCD	BL	5	L
routeType	Code	Y	RTESGNCD	SP	4	L
routeType	Code	Y	FUNCLASS	TR	8	L
routeType	Code	Y	FUNCLASS	TR	17	L
routeType	Code	Y	FUNCLASS	L	9	L
routeType	Code	Y	FUNCLASS	L	18	L
segmentLength	Num	Y				n
startOffset	Num	Y	BEGSECPT			n
accessControl	Code	N	RDACCESS	3	3	99
accessControl	Code	N	RDACCESS	1	1	99
accessControl	Code	N	RDACCESS	2	2	99
altRouteNames	Char	N				n
city	Char	N				n
comment	Char	N				n
corridor	Char	N				n
d1auxLane1	Code	N	AUXLNTYP	7	8	99
d1auxLane1	Code	N	AUXLNTYP	3	4	99
d1auxLane1	Code	N	AUXLNTYP	3	3	99
d1auxLane1	Code	N	AUXLNTYP	2	7	99
d1auxLane1	Code	N	AUXLNTYP	2	6	99
d1auxLane1	Code	N	AUXLNTYP	5	5	99
d1auxLane2	Code	N	AUXLNTYP	7	8	99
d1auxLane2	Code	N	AUXLNTYP	2	6	99
d1auxLane2	Code	N	AUXLNTYP	2	7	99
d1auxLane2	Code	N	AUXLNTYP	3	3	99
d1auxLane2	Code	N	AUXLNTYP	3	4	99
d1auxLane2	Code	N	AUXLNTYP	5	5	99
d1auxLane3	Code	N	AUXLNTYP	3	3	99
d1auxLane3	Code	N	AUXLNTYP	5	5	99
d1auxLane3	Code	N	AUXLNTYP	7	8	99
d1auxLane3	Code	N	AUXLNTYP	3	4	99
d1auxLane3	Code	N	AUXLNTYP	2	6	99
d1auxLane3	Code	N	AUXLNTYP	2	7	99
d1avgLaneWidth	Num	N	SURWIDTH			12
d1avgShoulderWidthIn	Num	N	ISLDWDTH			n
d1avgShoulderWidthOut	Num	N	SLDWIDTH			n
d1bikeway	Code	N	BIKELNCD	1	0	99
d1bikeway	Code	N	BIKELNCD	3	1	99

SA Variable	SA Type	Required?	RCI Variable	SA Code	RCI Code	Default Value
d1shoulderTypeIn	Code	N	ISLDTYPE	5	8	99
d1shoulderTypeIn	Code	N	ISLDTYPE	5	6	99
d1shoulderTypeIn	Code	N	ISLDTYPE	5	0	99
d1shoulderTypeIn	Code	N	ISLDTYPE	1	2	99
d1shoulderTypeIn	Code	N	ISLDTYPE	1	1	99
d1shoulderTypeOut	Code	N	SHLDTYPE	5	8	99
d1shoulderTypeOut	Code	N	SHLDTYPE	1	1	99
d1shoulderTypeOut	Code	N	SHLDTYPE	2	2	99
d1shoulderTypeOut	Code	N	SHLDTYPE	2	7	99
d1shoulderTypeOut	Code	N	SHLDTYPE	3	4	99
d1shoulderTypeOut	Code	N	SHLDTYPE	4	3	99
d1shoulderTypeOut	Code	N	SHLDTYPE	5	6	99
d1shoulderTypeOut	Code	N	SHLDTYPE	5	0	99
d2auxLane1	Code	N	AUXLNTYP	5	5	99
d2auxLane1	Code	N	AUXLNTYP	7	8	99
d2auxLane1	Code	N	AUXLNTYP	2	6	99
d2auxLane1	Code	N	AUXLNTYP	2	7	99
d2auxLane1	Code	N	AUXLNTYP	3	3	99
d2auxLane1	Code	N	AUXLNTYP	3	4	99
d2auxLane2	Code	N	AUXLNTYP	7	8	99
d2auxLane2	Code	N	AUXLNTYP	2	7	99
d2auxLane2	Code	N	AUXLNTYP	3	3	99
d2auxLane2	Code	N	AUXLNTYP	5	5	99
d2auxLane2	Code	N	AUXLNTYP	3	4	99
d2auxLane2	Code	N	AUXLNTYP	2	6	99
d2auxLane3	Code	N	AUXLNTYP	2	7	99
d2auxLane3	Code	N	AUXLNTYP	2	6	99
d2auxLane3	Code	N	AUXLNTYP	3	3	99
d2auxLane3	Code	N	AUXLNTYP	3	4	99
d2auxLane3	Code	N	AUXLNTYP	5	5	99
d2auxLane3	Code	N	AUXLNTYP	7	8	99
d2avgLaneWidth	Num	N	SURWIDTH			12
d2avgShoulderWidthIn	Num	N	ISLDWDTH			n
d2avgShoulderWidthOut	Num	N	SLDWIDTH			n
d2bikeway	Code	N	BIKELNCD	1	0	99
d2bikeway	Code	N	BIKELNCD	3	1	99
d2shoulderTypeIn	Code	N	ISLDTYPE	5	8	99
d2shoulderTypeIn	Code	N	ISLDTYPE	1	1	99
d2shoulderTypeIn	Code	N	ISLDTYPE	5	6	99
d2shoulderTypeIn	Code	N	ISLDTYPE	5	0	99
d2shoulderTypeIn	Code	N	ISLDTYPE	1	2	99
d2shoulderTypeOut	Code	N	SHLDTYPE	2	7	99
d2shoulderTypeOut	Code	N	SHLDTYPE	3	4	99
d2shoulderTypeOut	Code	N	SHLDTYPE	1	1	99

SA Variable	SA Type	Required?	RCI Variable	SA Code	RCI Code	Default Value
d2shoulderTypeOut	Code	N	SHLDTYPE	5	6	99
d2shoulderTypeOut	Code	N	SHLDTYPE	4	3	99
d2shoulderTypeOut	Code	N	SHLDTYPE	1	2	99
d2shoulderTypeOut	Code	N	SHLDTYPE	5	8	99
d2shoulderTypeOut	Code	N	SHLDTYPE	5	0	99
discontinuity	Code	N				X
district	Code	N				n
drivewayDensity	Num	N	ACMANCLS	342	6	990
drivewayDensity	Num	N	ACMANCLS	342	5	990
drivewayDensity	Num	N	ACMANCLS	990	2	990
drivewayDensity	Num	N	ACMANCLS	550	3	990
drivewayDensity	Num	N	ACMANCLS	550	4	990
drivewayDensity	Num	N	ACMANCLS	125	7	990
gisID	Char	N				n
growthFactor	Num	N				n
increasingMilesposts	Code	N				X
jurisdiction	Code	N				99
majorRoadName	Char	N				n
medianWidth	Num	N	MEDWIDTH			n
openedToTraffic	Char	N				n
postedSpeed	Num	N	MAXSPEED			n
roadwayClass1	Code	N	FUNCLASS	6	08	99
roadwayClass1	Code	N	FUNCLASS	5	07	99
roadwayClass1	Code	N	FUNCLASS	4	16	99
roadwayClass1	Code	N	FUNCLASS	4	06	99
roadwayClass1	Code	N	FUNCLASS	3	14	99
roadwayClass1	Code	N	FUNCLASS	3	02	99
roadwayClass1	Code	N	FUNCLASS	2	12	99
roadwayClass1	Code	N	FUNCLASS	1	11	99
roadwayClass1	Code	N	FUNCLASS	7	09	99
roadwayClass1	Code	N	FUNCLASS	7	18	99
roadwayClass1	Code	N	FUNCLASS	1	01	99
roadwayClass1	Code	N	FUNCLASS	0	17	99
terrain	Code	N	TERRAIN	L	1	X
terrain	Code	N	TERRAIN	R	2	X
travelDirection	Code	N				X

Table D-3: Data Mapping for AltIntersection File

SA Variable	SA Type	Required?	RCI Variable	SA Code	RCI Code	Default Value
agencyID	Char	Y	Node Number			n
agencySiteSubtype	Code	Y				
areaType	Code	Y	URBSIZE			X
county	Code	Y	RDWYID			FIRST PART
intersectionType1	Code	Y				99
majorRoadLocSystem	Code	Y	RDWYID			
majorRoadLocSystem	Code	Y	BEGSECPT			
majorRoadLocSystem	Code	Y	ENDSECPT			
majorRoadLocSystem	Code	Y				A
majorRoadOffset	Num	Y	BEGSECPT			0
minorRoadLocSystem	Code	Y	RDWYID			
minorRoadLocSystem	Code	Y	BEGSECPT			
minorRoadLocSystem	Code	Y	ENDSECPT			
minorRoadLocSystem	Code	Y				A
minorRoadOffset	Num	Y	BEGSECPT			0
minorRoadRouteName	Char	Y	RDWYID			n
minorRoadRouteType	Code	Y	USROUTE	US		X
minorRoadRouteType	Code	Y	STROADNO	SR		X
minorRoadRouteType	Code	Y	STROADNO	CR		X
minorRoadRouteType	Code	Y		O		X
minorRoadRouteType	Code	Y		X		X
minorRoadRouteType	Code	Y	FUNCLASS	I	01	X
minorRoadRouteType	Code	Y	FUNCLASS	TR	08	X
minorRoadRouteType	Code	Y	FUNCLASS	L	09	X
minorRoadRouteType	Code	Y	FUNCLASS	I	11	X
minorRoadRouteType	Code	Y	FUNCLASS	TR	17	X
minorRoadRouteType	Code	Y	FUNCLASS	L	18	X
minorRoadRouteType	Code	Y	RTESGNCD	BR	2	X
minorRoadRouteType	Code	Y	RTESGNCD	SP	4	X
minorRoadRouteType	Code	Y	RTESGNCD	BL	5	X
routeName	Char	Y	RDWYID			n
routeType	Code	Y	USROUTE,US_RTE_NUM_ID	US		L
routeType	Code	Y	STROADNO	SR		L
routeType	Code	Y	STROADNO,ST_RD_NUM_ID	CR		L
routeType	Code	Y		O		L
routeType	Code	Y		X		L
routeType	Code	Y	FUNCLASS	I	01	L
routeType	Code	Y	FUNCLASS	TR	08	L

SA Variable	SA Type	Required?	RCI Variable	SA Code	RCI Code	Default Value
routeType	Code	Y	FUNCLASS	L	09	L
routeType	Code	Y	FUNCLASS	I	11	L
routeType	Code	Y	FUNCLASS	TR	17	L
routeType	Code	Y	FUNCLASS	L	18	L
routeType	Code	Y	RTESGNCD	BR	2	L
routeType	Code	Y	RTESGNCD	SP	4	L
routeType	Code	Y	RTESGNCD	BL	5	L
trafficControl1	Code	Y				99
altRouteNames	Char	N				n
city	Char	N				n
comment	Char	N				n
corridor	Char	N				n
district	Code	N				n
gisID	Char	N				n
growthFactor	Num	N				n
jurisdiction	Code	N				99
majBeginInfluenceZone	Num	N				0.0473485
majEndInfluenceZone	Num	N				0.0473485
majorRoadDirection	Code	N	RCI Section Direction	NS	1	RDWTBL 22
majorRoadDirection	Code	N	RCI Section Direction	X	2	RDWTBL 22
majorRoadDirection	Code	N	RCI Section Direction	EW	3	RDWTBL 22
majorRoadDirection	Code	N	RCI Section Direction	X	4	RDWTBL 22
majorRoadDirection	Code	N	RCI Section Direction	NS	5	RDWTBL 22
majorRoadDirection	Code	N	RCI Section Direction	X	6	RDWTBL 22
majorRoadDirection	Code	N	RCI Section Direction	EW	7	RDWTBL 22
majorRoadDirection	Code	N	RCI Section Direction	X	8	RDWTBL 22
majorRoadDirection	Code	N				X
majorRoadName	Char	N	LOCALNAM			n
minBeginInfluenceZone	Num	N				0.0473485
minEndInfluenceZone	Num	N				0.0473485
minorRoadName	Char	N	LOCALNAM			n
offsetDistance	Num	N				n
offsetIntersection	Code	N				X
openedToTraffic	Char	N				n

Table D-4: Data Mapping for AltRamp File

SA Variable	SA Type	Required?	RCI Variable	SA Code	RCI Code	Default Value
agencyID	Char	Y	RDWYID			n
agencySiteSubtype	Code	Y	INTERCHG			429
altRouteNames	Char	N				n
areaType	Code	Y	URBSIZE	R	1	R
areaType	Code	Y	URBSIZE	U	2	R
areaType	Code	Y	URBSIZE	U	3	R
areaType	Code	Y	URBSIZE	U	4	R
areaType	Code	Y	URBSIZE	U	5	R
city	Char	N				n
comment	Char	N				n
corridor	Char	N				n
county	Code	Y	RDWYID			n
district	Code	N				n
endOffset	Num	Y	ENDSECPT			0
gisID	Char	N				n
growthFactor	Num	N				n
jurisdiction	Code	N				99
locSystem	Code	Y				A
majorRoadName	Char	N				n
numOfLanes	Num	Y	NOLANES			1
openedToTraffic	Char	N				n
rampConfiguration	Code	Y	INTERCHG	1	01	99
rampConfiguration	Code	Y	INTERCHG	1	02	99
rampConfiguration	Code	Y	INTERCHG	5	03	99
rampConfiguration	Code	Y	INTERCHG	2	05	99
rampConfiguration	Code	Y	INTERCHG	0	09	99
rampCrossroadConnection	Code	N				99
rampFreewayConnection	Code	N				99
rampLength	Num	N				n
rampType	Code	Y	LOCALNAM			99
RouteType	Code	Y	RTESGNCD	BR	02	X
RouteType	Code	Y	RTESGNCD	BL	05	X
RouteType	Code	Y	FUNCLASS	I	1	X
RouteType	Code	Y	FUNCLASS	I	11	X
RouteType	Code	Y	FUNCLASS	TR	17	X
RouteType	Code	Y	FUNCLASS	L	18	X
RouteType	Code	Y	RTESGNCD	SP	4	X
RouteType	Code	Y	FUNCLASS	TR	8	X
RouteType	Code	Y	FUNCLASS	L	9	X
routeName	Char	Y	RDWYID			n
startOffset	Num	Y	BEGSECPT			0

Table D-5: Data Mapping for AltLeg File

SA Variable	SA Type	Required?	RCI Variable	SA Code	RCI Code	Default Value
agencyID	Char	Y				n
leftTurnPhasing	Code	N		98		99
leftTurnPhasing	Code	N		99		99
leftTurnPhasing	Code	N	TURNLANL	1	1	99
leftTurnPhasing	Code	N	TURNLANL	2	3	99
leftTurnPhasing	Code	N	TURNLANL	3	4	99
leftTurnPhasing	Code	N	TURNLANL	4	5	99
legDirection	Code	Y	RCI Section Direction	NB	1	n
legDirection	Code	Y	RCI Section Direction	X	2	n
legDirection	Code	Y	RCI Section Direction	EB	3	n
legDirection	Code	Y	RCI Section Direction	X	4	n
legDirection	Code	Y	RCI Section Direction	SB	5	n
legDirection	Code	Y	RCI Section Direction	X	6	n
legDirection	Code	Y	RCI Section Direction	WB	7	n
legDirection	Code	Y	RCI Section Direction	X	8	n
legID	Code	N		1		n
legID	Code	N		2		n
legID	Code	N		3		n
legID	Code	N		4		n
legID	Code	N		5		n
legID	Code	N		6		n
legMedianType	Code	N	RDMEDIAN	5		99
legMedianType	Code	N	RDMEDIAN	0		99
legMedianType	Code	N	RDMEDIAN	99		99
legMedianType	Code	N	RDMEDIAN	4	01	99
legMedianType	Code	N	RDMEDIAN	1	02	99
legMedianType	Code	N	RDMEDIAN	1	03	99
legMedianType	Code	N	RDMEDIAN	1	04	99
legMedianType	Code	N	RDMEDIAN	1	06	99
legMedianType	Code	N	RDMEDIAN	2	08	99
legMedianType	Code	N	RDMEDIAN	3	09	99
legMedianType	Code	N	RDMEDIAN	3	10	99
legMedianType	Code	N	RDMEDIAN	1	12	99
legMedianType	Code	N	RDMEDIAN	1	13	99
legMedianType	Code	N	RDMEDIAN	1	16	99
legMedianType	Code	N	RDMEDIAN	1	17	99
legMedianType	Code	N	RDMEDIAN	1	23	99
legMedianType	Code	N	RDMEDIAN	1	31	99
legMedianType	Code		RDMEDIAN			99

SA Variable	SA Type	Required?	RCI Variable	SA Code	RCI Code	Default Value
legNumLeftTurnLane	Int		AUXLNTYP			3
legNumRightTurnLane	Int		AUXLNTYP			4
legNumThruLane	Int		NOLANES			0
legType	Code	Y		1		n
legType	Code	Y		2		n
legType	Code	Y		3		n
legType	Code	Y		4		n
legType	Code	Y		98		n
legType	Code	Y		99		n
operationWay	Code		TYPEROAD			2
postedSpeed	Num		MAXSPEED			0
turnProhibitions	Code	N		1		99
turnProhibitions	Code	N		3		99
turnProhibitions	Code	N	TURNLANL	2	5	99
turnProhibitions	Code	N	TURNLANR	4	5	99